Submission Number: PED11-11-00050

Stereotypes and inequality: a `signaling' theory of identity choice

Young chul Kim
Korea Development Institute

Glenn C. Loury
Brown University

Abstract

We develop an identity choice model based on the stereotyping and signaling framework. Inequality of collective reputation between exogenous groups in equilibrium is due to feedback between group reputation and individual human capital investment activities(Arrow, 1973; Coate and Loury, 1993). But it entails no positive selection into or out of the groups and human capital cost distributions among groups' members are equal. When group membership is endogenous and if the groups' reputations differ in equilibrium, the group with a higher reputation not only engages in more human capital investment activities, but the group itself also consists disproportionately of members with low human-capital-investment cost, who have more to gain from joining the favored group. This causes human capital cost distributions between groups to endogenously diverge, reinforcing incentive-feedbacks. We examine the existence and stability of stereotyping equilibria with endogenous group membership. We show that inequality deriving from stereotyping of endogenously constructed social groups is at least as great as the inequality that can emerge between exogenously given groups.

Stereotypes and Inequality: A 'Signaling' Theory of Identity Choice

Young Chul Kim[*]   Glenn C. Loury[†]

April 27, 2011

Abstract

We develop an identity choice model based on the stereotyping and signaling framework. Inequality of collective reputation between exogenous groups in equilibrium is due to feedback between group reputation and individual human capital investment activities (Arrow, 1973; Coate and Loury, 1993). But it entails no positive selection into or out of the groups and human capital cost distributions among groups' members are equal. When group membership is endogenous and if the groups' reputations differ in equilibrium, the group with a higher reputation not only engages in more human capital investment activities, but the group itself also consists disproportionately of members with low human-capital-investment cost, who have more to gain from joining the favored group. This causes human capital cost distributions between groups to endogenously diverge, reinforcing incentive-feedbacks. We examine the existence and stability of stereotyping equilibria with endogenous group membership. We show that inequality deriving from stereotyping of endogenously constructed social groups is at least as great as the inequality that can emerge between exogenously given groups.

KEYWORDS: Stereotypes, Statistical Discrimination, Identity Choice, Signaling.

1 Introduction

We develop an identity choice model based on the stereotyping and signaling framework. If a worker's true productivity is not perfectly observable, employers have an incentive to use the collective reputation of the job applicants in the screening process. The individuals who belong to a group with a better collective reputation have a greater incentive to invest in skills, (and vice versa). With their greater (smaller) skill investment rate, the group maintains a better (worse) collective reputation. Therefore, there are multiple self-confirming equilibria of collective reputation (Arrow, 1973; Coate and Loury, 1993). In previous work related to such

[*]Korea Development Institute, Dongdaemoon-gu, Seoul 130-868, South Korea. Email: yckim@kdi.re.kr.

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

statistical discrimination, group identity is immutable and each group member is affected by the collective reputation of his own group only. We handle the dynamics between the collective reputation and the identity choice problem by relaxing the immutability assumption.

For instance, "Passing" is an apparent identity choice behavior. Talented young members in the group with the worse collective reputation may consider passing for the group with the better collective reputation when the return for passing, such as better treatment in the labor market, outweighs the cost of passing, such as the disconnection from their own ties. For example, a meaningful number of the black population consistently passes for White or some other race (Sweet, 2005).[1] The Korean descendants in Japan, most of whom are descendants of forced laborers in mines and factories who were brought to Japan from the Korean peninsula during the period of Japanese imperialism, have passed for native Japanese, changing both surnames and given names around the time when they sought formal employment and marriage. This way they tried to escape negative stereotypes and prejudices.[2]

When passing for a member of the advantaged group with high reputation is not possible due to immutability, the most talented of the stereotyped group are more likely to seek styles of self-presentation that aim to communicate "I'm not one of THEM; I'm one of YOU!" because they are the ones who gain most by separating themselves from the mass (Loury, 2002). Taking the example of the Black population in the US, methods that are known to be used for "partial passing" are: affections of speech, dressing up rather than wearing casual clothes, spending more on conspicuous consumption, and migration to affluent residential areas. There is evidence that the more educated (or talented) blacks tend to speak standard American English rather than African American English (Grogger, 2008). That is, the most talented of the stereotyped group "pass for" the slightly better-off subgroup that maintains a higher reputation than the stereotyped population as a whole by adopting the cultural traits of the subgroup.

This selective out-migration to the better-off subgroup may undermine solidarity in the disadvantaged population and cause conflict among them, such as the accusation of "Acting White" against the ones who practice the "partial passing" methods (Fryer and Torelli, 2010).

---

[1]The National Longitudinal Survey conducted by the Department of Labor of the US shows that 1.87 percent of those who had originally answered "Black" to the interviewer's race question in 1979 switched to either "White," "I don't know, " or "other," by 1998.

[2]Every year about 10,000 Koreans living in Japan, out of around 600,000 Korean descendants holding Korean nationality, choose to be naturalized, giving up their names and original nationality (Fukuoka et al., 1998).

However, there might be a social gain through this practice: at least some cultural subgroups of the stereotyped population might be able to recover their reputation when the talented young members gather around certain cultural traits. The usage of the observable cultural traits in the screening process can to some extent cure the social inefficiency caused by imperfect information about the true characteristics of workers.

The emergence of an elite social group out of a population can also be explained through identity choice behavior. The usage of a cultural instrument that is intrinsically irrelevant for productivity to form an elite group is well discussed in Fang (2001) as an explanation for the complexity of elite etiquette in European (or Confucian) societies and the respect for "Oxford Accent." Skilled and unskilled workers have different incentives to join a group with unique cultural traits that are *expensive* to obtain. Thus, the cultural group is treated preferentially by employers due to the higher fraction of skilled workers, even though the cultural traits of the group are not relevant for productivity. We may see an autonomously growing elite subgroup with differentiated cultural traits whose members are considered as distinguished from their peers.

The identity choice model in this paper starts with a standard statistical discrimination framework (Coate and Loury, 1993). We identify multiple self-confirming prior beliefs, which we call Phenotypic Stereotyping Equilibria (PSE). Inequality of collective reputation between exogenous groups in equilibrium is due to feedback between group reputation and individual investment activities. But it entails no positive selection into or out of the groups. Therefore, human capital cost distributions among groups' members are equal.

However, when group membership is endogenous, and if the groups' reputations differ in equilibrium, the favored group not only faces great human capital investment incentives, but it also consists disproportionately of low human capital investment cost types, who gain more from joining a favored group, thereby causing human capital cost distributions between groups to endogenously diverge, reinforcing incentive-feedbacks. We call the multiple equilibria with positive selection non-trivial Affective Stereotyping Equilibria (ASE).[3] For the development of a theoretical model, we introduce two affects, A and B. The cost to choose affect A rather than B varies across the population. Agents choose affect A if and only if the anticipated return exceeds the agent's cost of choosing affect A. The cost distribution for the affects is irrelevant for the cost distribution of human capital investment or skill achievement. In equilibrium, we

---

[3]Note that PSE automatically generate trivial ASE which does not entail positive selection.

show that the more talented members, that is, lower human capital investment cost types, tend to choose affect A when the collective reputation of the affect A group is better than that of the affect B group.

We prove that inequality deriving from stereotyping of *endogenously constructed* social groups is at least as great as the inequality that can emerge between *exogenously given* groups. While the inequality between exogenous groups involves no positive selection, low human capital cost types are disproportionately drawn to the favored group when groups are endogenous, causing the skill disparity between the groups to endogenously diverge.

We further prove that there always exist multiple non-trivial Affective Stereotyping Equilibria whenever multiple Phenotypic Stereotyping Equilibria can be constructed. Even more, in the overlapping generation framework, those non-trivial ASE are the only stable equilibria when the society has a critical fraction of newborns whose identity choice cost is sufficiently low. That is, the skill composition of the society converges to a non-trivial ASE in the long run. In addition, we show that non-trivial ASE can exist even under the unique PSE. Even when phenotypic discrimination cannot generate inequality between any identity groups, which could happen due to the uniqueness of the PSE, affective discrimination may bring about inequality between affective groups that are endogenously being constructed in a society.[4]

The paper is organized into the following sections. Section 2 summarizes the related literature on stereotypes, sorting and matching. Section 3 develops the model with the identify choice and skill investments. Section 4 defines Phenotypic and Affective Stereotyping Equilibria. Section 5 and Section 6 each identify Affective Stereotyping Equilibria with multiple PSE and with unique PSE. Section 7 presents further discussions and Section 8 concludes.

2 Related Literature

1. Sorting-matching: Becker (marriage); Sattinger (job assignment); Benabou (location with HC spillovers); Fernandez/Rogerson (location with local public goods); Sethi/Somanathan (location with race/income differences). In all cases, key role played by some 'single crossing' property. Also, Rosen ('superstars' - audience sorting); Costrell-Loury (job

---

[4]The example of Fang (2001) is a special case in which non-trivial ASE exists given the uniqueness of the PSE that is zero.

sorting)

2. Stereotypes

   (a) Economics: Arrow (1970); Coate-Loury (1993); Fang (2001)

   (b) Sociology: Goffman ('Presentation of Self'); Loury ('Anatomy of Racial Inequality'); Anderson ('streetwise'); Sampson ('perceptions of disorder)

   (c) Social Psychology: Fiske (prejudice); Banaji (implicit bias); Steele ('stereotype threat')

3   Model with the Identity Choice

Workers' Affective/Expressive Behavior: Agents choose affect $i \in \{A, B\}$. The cost to choose the affect A is $k \in R$. k can be negative: the affect A can generate benefits for some agents. CDF of the affective behavior cost is denoted by $H(k)$. We assume the affective symmetry: $H(k) = 1 - H(-k)$. Agents choose the affect A if and only if the anticipated return exceeds the agent's cost k. Otherwise, they choose the affect B. WLOG, it is natural to assume that PDF of the cost k, $h(k)$, has one peak at $k = 0$: $h'(k) > 0$ for any $k \in (-\infty, 0)$ and $h'(k) < 0$ for any $k \in (0, \infty)$.[5]

Workers' Skill Acquisition Behavior: Agents choose whether to be skilled or not: $e \in \{0, 1\}$. The cost to be skilled is c, which is non-negative. CDF of the skill acquisition cost is $G(c)$, in which $G(0) \geq 0$ and $G(\infty) = 1$.[6] WLOG, it is natural to assume that PDF of the cost c, $g(c)$, has one peak at $\hat{c}$: $g'(c) > 0$ for any $c \in (\hat{c}, 0)$ and $g'(c) < 0$ for any $c \in (\hat{c}, \infty)$. An agent chooses ($e = 1$) if the return from doing so exceeds that agent's cost for the skill acquisition (c). We impose that c and k are independently distributed.

Employers' Wage-setting Behavior: Skill e is not fully identified. Employers observe group identity and noisy signal $t \in R^+$ distributed conditional on e. PDF of the signal conditional on e is $f_e(t)$ and its CDF is $F_e(t)$. Let us define the function $f(\pi, t)$ as $f(\pi, t) \equiv \pi f_1(t) + (1 - \pi) f_0(t)$, which indicates the distribution of signal t of agents belonging to a group of which the skill level is believed to be π. WLOG, we assume that $f_1(t)/f_0(t)$ increases with respect to t, which is denoted by MLRP: Monotonic Likelihood Ratio Property. The employers' belief

---

[5]This is not a critical assumption in the model.
[6]With $G(c) \geq 0$, we allow that a fraction of workers always invest for skills.

that an agent with signal t is skilled is $\rho(\pi,t)(\equiv Pr[e=1|\pi,t]) = \frac{\pi f_1(t)}{f(\pi,t)}$. Under MLRP, $\rho(\pi,t)$ increases with both π and t. The productivity of a skilled worker is w and that of an unskilled worker is zero. We assume that the wage is proportional to the expected skill level:

$$
\begin{aligned}
W(\pi,t) &= w \cdot \rho(\pi,t), \text{ for some } w > 0 \\
&= w \cdot \frac{\pi f_1(t)}{\pi f_1(t) + (1-\pi)f_0(t)}.
\end{aligned}
\tag{1}
$$

Workers' Payoffs:

The expected wage from acquiring skill level e is denoted by $V_e(\pi)$:

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

in which $V'_e(\pi)$ is positive for any $e \in \{0,1\}$. Workers' expected return acquiring human capital $(R(\pi))$ is defined as

$$
R(\pi) \equiv V_1(\pi) - V_0(\pi).
\tag{3}
$$

$R(\pi)$ is expressed as

$$
\begin{aligned}
R(\pi) &= \int_0^1 (f_1(t) - f_0(t))W(\pi,t)\,\mathrm{d}t \\
&= w\pi \int_0^1 \frac{(f_1(t) - f_0(t))f_1(t)}{f(\pi,t)}\,\mathrm{d}t.
\end{aligned}
\tag{4}
$$

The followings can be easily seen

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

$$
R''(\pi) = -2w \int_0^1 \frac{(f_1(t) - f_0(t))^2 f_1(t)f_0(t)}{f(\pi,t)^3}\,\mathrm{d}t \quad (< 0)
\tag{6}
$$

Thus, $R(\pi)$ is concave and $R(0) = R(1) = 0$, which implies that $\lim_{\pi \to 0} R'(\pi) > 0$ and $\lim_{\pi \to 1} R'(\pi) < 0$. Let us denote $argmax\{R(\pi)\}$ by $\bar{\pi}$: $R'(\bar{\pi}) = 0$.

The first derivatives of $V_0(\pi)$ and $V_1(\pi)$ are

$$
V'_0(\pi) = \int_0^1 w f_1(t)f_0(t)^2 f(\pi,t)^{-2}\,\mathrm{d}t,
\tag{7}
$$

$$
V'_1(\pi) = \int_0^1 w f_1(t)^2 f_0(t) f(\pi,t)^{-2}\,\mathrm{d}t.
\tag{8}
$$

Note that $V_0'(0) = w$ and $V_1'(1) = w$. Since we know $R'(0) > 0$ and $R'(1) < 0$, we have $V_1'(1) > w$ and $V_0'(0) > w$. It is more likely that $V_0'(\pi)$ tends to increase as π increases and $V_0'(\pi)$ tends to decrease as π increases. WLOG, we impose that relative marginal benefits $(V_1'(\pi)/V_0'(\pi))$ declines over π. Let us call the property the Marginal Benefits Ratio Property (MBRP): $\frac{V_1'(\pi)}{V_0'(\pi)} > \frac{V_1'(\pi+\delta)}{V_0'(\pi+\delta)}$ for any $\delta > 0$.[7]

Thus, a worker with cost c, in a group believed to be investing at rate π has the payoff:

$$U(\pi, c) = \max\{V_1(\pi) - c;\ V_0(\pi)\}, \tag{9}$$

in which the function $U(\pi, c)$ is increasing in π ($\because V_e'(\pi) > 0, \forall e \in \{0, 1\}$) and non-increasing in c.

4 Equilibrium

4.1 Equilibrium with No Affective Stereotyping

Given the employers' prior belief (π) about human capital investment rate in a population, the fraction of workers who choose ($e = 1$) is $G(R(\pi))$. Let us denote an equilibrium belief/investment rate by $\hat{\pi} \in [0, 1]$: $\hat{\pi} = G(R(\hat{\pi}))$. The set of all such equilibria is denoted by Ω_{CL} (Coate and Loury 1993). Let us call them Phenotypic Stereotyping Equilibria (PSE).

Absent affective discrimination, workers choose 'affect' based on their "natural" orientation: $i = B$ if $k > 0$ and $i = A$ if $k < 0$. This implies that the human capital cost distribution, namely $G(c)$, is the same for both affective groups.

It is most likely that there exists either one or three equilibria in the economy, because $G(c)$ is $S-$shaped as displayed in Figure 1. Multiple equilibria $\hat{\pi} \in \Omega_{CL}$ create possibility of Phenotypic Stereotyping (PS) wherein groups are exogenously and visibly distinct, though equally well endowed. Nevertheless, they fare unequally in the equilibrium.

The socially optimal level of human capital investment is $G(w)$. However, human capital investment is socially inadequate in any PSE $\hat{\pi}$: $G(R(\hat{\pi})) < G(w)$ for any $\hat{\pi} \in \Omega_{CL}$. Note that

---

[7]Consider a simple example that $f_1(t) = 1 - P_1$ for any $t \in (0, 1)$ and $f_1(t) = P_1$ for any $t \in (1, 2)$ together with $f_0(t) = 1 - P_0$ for any $t \in (0, 1)$ and $f_1(t) = P_0$ for any $t \in (1, 2)$. Define $Pr(e = 1|0 < t < 1, \pi) = W_N(\pi)$ and $Pr(e = 1|1 < t < 2, \pi) = W_P(\pi)$. It is easily seen that $W_N'' > 0$ and $W_P'' < 0$. We have $V_1(\pi) = (1 - P_1)W_N(\pi) + P_1 W_P(\pi)$ and $V_0(\pi) = (1 - P_0)W_N(\pi) + P_0 W_P(\pi)$. Using these results, we can confirm that the following MBRP property is true for this example: $\frac{\partial [V_1'(\pi)/V_0'(\pi)]}{\partial \pi} < 0$.

$w - R(\pi) > 0$ because $w \int_0^1 f_1(t) \, \mathrm{d}t - R(\pi) = w \int_0^1 f_1(t) f_0(t) f(\pi, t)^{-1} \, \mathrm{d}t > 0$.

4.2   Affective Stereotyping Equilibria

Let π_i be employer belief about human capital investment rate in affective group i. Consider two affective groups A and B. Let us define a function $\Delta U(\pi_A, \pi_B; c)$ as the payoff difference between a A-type worker and a B-type worker given their skill acquisition cost level c: $\Delta U(\pi_A, \pi_B; c) \equiv U(\pi_A, c) - U(\pi_B, c)$. Given $\pi_A > \pi_B$, $\Delta U(\pi_A, \pi_B; c)$ is positive because $\partial U(\pi, c)/\partial \pi > 0$. Note that $\Delta U(\pi_A, \pi_B; c) = -\Delta U(\pi_B, \pi_A; c)$ and $\Delta U(\pi, \pi; c) = 0$.

An agent with the cost set (c, k) chooses affective behavior $i = A$ if and only if $\Delta U(\pi_A, \pi_B; c) \leq k$. Otherwise, he chooses affective behavior $i = B$. Given that c and k are independent, the fraction of agents choosing $(i = A)$ is given by

$$\Sigma^A \equiv \int_0^\infty H(\Delta U(\pi_A, \pi_B; c)) \, \mathrm{d}G(c). \tag{10}$$

The fraction of workers choosing $(i = A)$ and $(e = 1)$ is given by

$$\sigma^A \equiv \int_0^{R(\pi_A)} H(\Delta U(\pi_A, \pi_B; c)) \, \mathrm{d}G(c). \tag{11}$$

Then, the fraction of agents choosing $(i = B)$ is obtained using $\Sigma^B = 1 - \Sigma^A$ and $\Delta U(\pi_A, \pi_B; c) = -\Delta U(\pi_B, \pi_A; c)$:

$$\Sigma^B \equiv \int_0^\infty H(\Delta U(\pi_B, \pi_A; c)) \, \mathrm{d}G(c). \tag{12}$$

Consequently, the fraction of workers choosing $(i = B)$ and $(e = 1)$ is given by

$$\sigma^B \equiv \int_0^{R(\pi_B)} H(\Delta U(\pi_B, \pi_A; c)) \, \mathrm{d}G(c). \tag{13}$$

Given the employer belief about human capital investment rates (π_A, π_B), the actual investment rates for the affective groups denoted by $\phi(\pi_A, \pi_B)$ and $\phi(\pi_B, \pi_A)$ are

$$\begin{cases} Pr\{e = 1 | i = A, \pi_A, \pi_B\} (\equiv \phi(\pi_A, \pi_B)) = \sigma^A / \Sigma^A, \\ Pr\{e = 1 | i = B, \pi_B, \pi_A\} (\equiv \phi(\pi_B, \pi_A)) = \sigma^B / \Sigma^B. \end{cases} \tag{14}$$

It is noteworthy that when employers' belief is the same for both affective groups $(\pi_A = \pi_B)$,

$\Delta U(\pi_B, \pi_A; c)$ is zero and we have $R(\pi_A) = R(\pi_B)$. This implies that the affective behavior does not affect the human capital investment activities: $\phi(\pi_A, \pi_B) = \phi(\pi_B, \pi_A)(= G(R(\pi_A)))$.

An equilibrium with affective stereotyping (ASE) is defined as a pair of investment rates for the affective groups $(\pi_A^*, \pi_B^*) \in [0,1]^2$ such that $\pi_A^* = \phi(\pi_A^*, \pi_B^*)$ and $\pi_B^* = \phi(\pi_B^*, \pi_A^*)$. The set of all such equilibria is denoted by Ω_F. Note that every PSE corresponds to trivial ASE where differences in affect are uninformative: $(\hat{x}, \hat{x}) \in \Omega_F$ if $\hat{x} \in \Omega_{CL}$ because $\phi(\hat{x}, \hat{x}) = G(R(\hat{\pi})) = \hat{x}$. Affective stereotyping discrimination occurs if and only if $\pi_A^* \neq \pi_B^*$.

For notation simplicity, we use a and b instead of π_A and π_B. $\Delta U(a, b; c)$ can be expressed by

$$\Delta U(a, b; c) = \max\{R(a) - c; 0\} + V_0(a) - \max\{R(b) - c; 0\} - V_0(b). \tag{15}$$

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

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

$$\Delta U(a, b; c) = \begin{cases} V_1(a) - V_0(b) - c & \text{if } R(a) \geq R(b), \\ V_0(a) - V_1(b) + c & \text{if } R(a) < R(b). \end{cases} \tag{16}$$

The above lemma is summarized in Figure 2. Panel 1 of the figure displays the case with $a > b$ and panel 2 does the case with $a < b$. It is easily seen that $\Delta U(a, b; c) > 0$ for any c if and only if $a > b$. Therefore, we have the following result.

Proposition 1. *When employers have different beliefs about two affective groups($\pi_A \neq \pi_B$), the number 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: $\Sigma^i > \Sigma^j$ if $\pi_i > \pi_j$ for any $i, j \in \{A, B\}$.*

That is, in the current setting with symmetric cost distribution, more than half workers adopt the 'affect' that corresponds to the more favorable employers' belief: $\Sigma^i > .5$ and $\Sigma^j < .5$ if $\pi_i > \pi_j$. The Lemma 1 implies that $\Delta U(a, b; c)$ is non-increasing with respect to c whenever $R(a) > R(b)$, and non-decreasing whenever $R(b) > R(a)$. It leads to the following useful lemma.

Lemma 2. *Whenever $R(a) > R(b)$, the following holds: $\phi(a,b) > \phi(a,a)$ and $\phi(b,a) < \phi(b,b)$. In a symmetric way, whenever $R(a) < R(b)$, the following holds: $\phi(a,b) < \phi(a,a)$ and $\phi(b,a) > \phi(b,b)$. When $R(a) = R(b)$ and $a \neq b$, the following holds: $\phi(a,b) = \phi(b,a) = \phi(a,a) = \phi(b,b)$.*

The above lemma implies the following proposition.

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

For any b except for $\bar{\pi}$, we can find b' such that $R(b) = R(b')$. The following should hold for the combination (b,b'): $\phi(b,b) = \phi(b',b) = G(R(b))$. The overall shape of $\phi(a,b)$ is displayed in Panel A of Figure 3 for three different levels of b below $\bar{\pi}$, $b_1 < b_2 < b_3 < \bar{\pi}$, together with the the shape of $\phi(a,a)(= G(R(a)))$, in which $\bar{\pi} > \pi_h$. Also, Panel B of the figure displays the shape of $\phi(a,a)$ for the case with $\bar{\pi} < \pi_h$ and the overall shape of $\phi(a,b)$ for three different levels of b below $\bar{\pi}$, $b_4 < b_5 < b_6 < \bar{\pi}$. Note that the $\phi(a,b)$ curve intercepts the $\phi(a,a)(= G(R(a)))$ curve at $a = b$ and $a = b'$. We have the following lemma for the relative positions of $\phi(a,b)$s.

Lemma 3. *For any b_1 and b_2 such that $b_1 < b_2 < \bar{\pi}$, $\phi(a,b_1)$ is placed above $\phi(a,b_2)$: $\phi(a,b_1) > \phi(a,b_2)$, $\forall a \in (0,1)$. Also, for any b_1 and b_2 such that $\bar{\pi} < b_1 < b_2$, $\phi(a,b_2)$ is placed above $\phi(a,b_1)$: $\phi(a,b_2) > \phi(a,b_1)$, $\forall a \in (0,1)$.*

Proof. Let us prove the first part. First, consider an arbitrary level of b such that $b < a < \bar{\pi}$. For very small δ_1 and δ_2, the following approximation holds: $h(\Delta U(a,b;c)) \approx h(\Delta U(a,b-\delta_1;c)) \approx h(\Delta U(a,b-\delta_1-\delta_2;c))$, which is denoted by $\tilde{h}(a,b,c)$. The small incremental decrease of b as much as δ leads $V_0'(b) \cdot \delta$ increase of ΔU for any $c \in (R(a),\infty)$ and $V_1'(b) \cdot \delta$ increase of ΔU for any $c \in (0, R(b-\delta))$. Therefore, the incremental decrease as much as δ_1 and subsequent decrease as much as δ_2 generate the different levels of ΔU as shown in Appendix Figure 1.

Let us impose that $V_0'(b-\delta_1) \cdot \delta_2 = V_0'(b) \cdot \delta_1$. Then, the incremental impact of decreased b on the overall human capital investment rate depends on the relative size of the skilled population of area P and that of area Q. As far as the skilled population of area P is greater

than the skilled population of area Q, it is assured that the incremental decrease of b leads to the increase of $\phi(a, b)$: $\partial\phi(a, b)/\partial b < 0$ for any $b < a < \bar{\pi}$.

Let the skilled population in area P and area Q be denoted by $\sigma^A[P]$ and $\sigma^B[Q]$:

$$
\begin{aligned}
\sigma^A[P] &\approx [V_1'(b - \delta_1) - V_0'(b - \delta_1)]\delta_2 \cdot \tilde{h}(a, b, 0) \cdot G(R(b - \delta_1)) \\
&\quad - [V_1'(b - \delta_1) - V_0'(b - \delta_1)]^2\delta_2^2 \cdot \tilde{h}(a, b, 0) \cdot g(R(b - \delta_1)) \\
&\approx \left[\frac{V_1'(b - \delta_1)}{V_0'(b - \delta_1)} - 1\right]\delta_1 V_0'(b) \cdot h(a, b, 0) \cdot G(R(b))
\end{aligned}
\tag{17}
$$

$$
\begin{aligned}
\sigma^B[Q] &\approx [V_1'(b) - V_0'(b)]\delta_1 \cdot \tilde{h}(a, b, 0) \cdot G(R(b)) \\
&\quad - [V_1'(b) - V_0'(b)]^2\delta_2^2 \cdot \tilde{h}(a, b, 0) \cdot g(R(b)) \\
&\approx \left[\frac{V_1'(b)}{V_0'(b)} - 1\right]\delta_1 V_0'(b) \cdot h(a, b, 0) \cdot G(R(b))
\end{aligned}
\tag{18}
$$

Using the declining marginal benefits ratio property (MBRP), $\frac{V_1'(\pi)}{V_0'(\pi)} > \frac{V_1'(\pi+\delta)}{V_0'(\pi+\delta)}$ for any $\delta > 0$, we confirm that $\sigma^A[P] > \sigma^B[Q]$. Therefore, given $b_1 < b_2 < a < \bar{\pi}$, $\phi(a, b_1)$ is placed above $\phi(a, b_2)$: $\phi(a, b_1) > \phi(a, b_2)$. In the identical way, we can show the same results for other levels of a given $b_1 < b_2 < \bar{\pi}$. Also, we can prove the second part of the lemma (concerning the cases under $\bar{\pi} < b_1 < b_2$) using the similar methodology. (The proof needs to be improved further.) ∎

The following lemma helps us understand how the $\phi(a, b)$ curve cross over the $\phi(a, a)$ curve:

Lemma 4. *The slope of the $\phi(a, b)$ curve at the point where it crosses over the $\phi(a, a)$ curve is*

$$
\left.\frac{\partial\phi(a, b)}{\partial a}\right|_{a=b} \approx g(R(b))R'(b) + 2H'(0)R'(b)G(R(b))(1 - G(R(b))).
\tag{19}
$$

Proof. Consider a very small $\delta > 0$ such that $a = b+\delta$. Define $\Delta(\delta)$ as $\Delta(\delta) \equiv R(b+\delta) - R(b)$: $\Delta' = R'(b + \delta)$. We have $H'(k) \approx H'(0)$ for small enough k. Using Lemma 1 and Panel A of Figure 2, we can calculate $\sigma^A(\delta)$ and $\Sigma^A(\delta)$, and consequently $\sigma^{A\prime}(\delta)$ and $\Sigma^{A\prime}(\delta)$:

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

(20)

(21)

(Note that the last terms, $-.5H'(0)R'(b)\Delta^2$ and $-H'(0)g(R(b))\Delta R'(b+\delta)$, are added only when $R'(b) > 0$).

$$
\begin{aligned}
\Sigma^A(\delta) &\approx .5 + H'(0)(V_0(b+\delta) - V_0(b)) + G(R(b) + \Delta)H'(0)\Delta - .5H'(0)g(R(b))\Delta^2, \quad (22)\\
\Sigma^{A\prime}(\delta) &\approx H'(0)V_0'(b+\delta) + G(R(b) + \Delta)H'(0)R'(b+\delta) + g(R(b) + \Delta)R'(b+\delta)H'(0)\Delta\\
&\quad -H'(0)g(R(b))\Delta R'(b+\delta). \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (23)
\end{aligned}
$$

The slope of the $\phi(a,b)$ curve given a=b can be expressed as follows:

$$
\begin{aligned}
\left.\frac{\partial \phi(a,b)}{\partial a}\right|_{a=b} &= \lim_{\delta \to 0} \frac{\phi(b+\delta, b) - \phi(b,b)}{\delta}\\
&= \lim_{\delta \to 0} \frac{\sigma^A(\delta)/\Sigma^A(\delta) - \sigma^A(0)/\Sigma^A(0)}{\delta}\\
&= \lim_{\delta \to 0} \left[\frac{[\sigma^A(\delta) - \sigma^A(0)]\Sigma^A(0)}{\delta} - \frac{[\Sigma^A(\delta) - \Sigma^A(0)]\sigma^A(0)}{\delta} \right] \cdot \frac{1}{\Sigma^A(\delta)\Sigma^A(0)}\\
&\approx \frac{\sigma^{A\prime}(0)\Sigma^A(0) - \sigma^A(0)\Sigma^{A\prime}(0)}{\Sigma^A(0)^2} \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (24)
\end{aligned}
$$

We can achieve the following results:

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

Consequently, we have $\left.\frac{\partial \phi(a,b)}{\partial a}\right|_{a=b} \approx g(R(b))R'(b) + 2H'(0)R'(b)G(R(b))[1 - G(R(b))]. \quad \blacksquare$

The above lemma implies that the slope of $\phi(a,b)$ at the crossing point is positive (negative) whenever $R'(b)$ is positive (negative). Also, the slope of $\phi(a,b)$ at the crossing point is greater (smaller) than the slope of $\phi(a,a)(= g(R(b))R'(b))$ whenever $R'(b)$ is positive (negative).

5 Affective Stereotyping Equilibria with Multiple PSE

Let us define a correspondence $\Gamma(y)$:

$$\Gamma(y) = \{x : x = \phi(x, y)\}. \tag{26}$$

Note that any $\hat{\pi} \in \Omega_{CL}$ satisfies $\hat{\pi} \in \Gamma(\hat{\pi})$ and any $\hat{\pi} \in \Gamma(\hat{\pi})$ satisfies $\hat{\pi} \in \Omega_{CL}$. Thus, the set of PSE is represented as follows using the correspondence: $\Omega_{CL} = \{x : x \in \Gamma(x)\}$. The set of affective stereotyping equilibria can be expressed as $\Omega_F = \{(x, y) : x \in \Gamma(y) \text{ and } y \in \Gamma(x)\}$.

Consider the case with multiple PSE. WLOG, we assume that there are three: π_h, π_m and π_l. We will examine the case with a unique PSE in the next section.

5.1 Existence of Affective Stereotyping Equilibria

It is most likely that there exist either one or three values in $\Gamma(y)$. Let us denote the three values by $\Gamma(y)^h$, $\Gamma(y)^m$ and $\Gamma(y)^l$ as displayed in Figures 4 and 5. Panels A, B and C of Figure 4 describe the case with $\bar{\pi} > \pi_h$ and Panel A of Figure 5 the case with $\bar{\pi} < \pi_h$. If there exists a unique value for some range of y, $\Gamma(y)$ with its unique value is denoted by $\Gamma(y)^i$ as it is connected to nearby $\Gamma(y)^i$ for $i \in \{h, m, l\}$, which is an element of $\Gamma(y)$ with multiple values. We can infer the following result using Lemma 3.

Lemma 5. *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 $\bar{\pi}$. Also, we have $\pi_h < \Gamma(0)^h < 1$ and $\pi_h < \Gamma(1)^h < 1$.*

This lemma also implies that $\min \Gamma(y)^l = \Gamma(\bar{\pi})^l$ and $\arg \min \Gamma(y)^l = \bar{\pi}$. $\Gamma(a)$ and $\Gamma(b)$ are overlapped in Figure 4. Using the local linearization described in Appendix Figure 2, we can calculate the slope of correspondence curve at each trivial ASE, $\Gamma'(\hat{\pi})$.

Lemma 6. *The slope of correspondence curve at a trivial ASE $(\hat{x}, \hat{x})$, which is denoted by $\Gamma'(\hat{x})$, is approximated by*

$$\Gamma'(\hat{x}) \approx \frac{2H'(0)R'(\hat{x})\hat{x}(1 - \hat{x})}{g(R(\hat{x}))R'(\hat{x}) - 1 + 2H'(0)R'(\hat{x})\hat{x}(1 - \hat{x})}. \tag{27}$$

Proof. Given the slope of $\phi(x, y)$ at $(\hat{x}, \hat{x})$ denoted by $\frac{\partial \phi(x,y)}{\partial x}\big|_{x=y=\hat{x}}$ and the slope of $\phi(x, x)$ at the same point, $g(R(\hat{x}))R'(\hat{x})$, we can find a correspondence value x' such that $x' = \phi(x', \hat{x}+\Delta)$

13

using the following equation:

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

Therefore, we have $\Gamma'(\hat{\pi})$, which is approximately equal to $\frac{x'-\hat{x}}{\Delta}$:

$$\Gamma'(\hat{x}) \approx \left[g(R(\hat{x}))R'(\hat{x})) - \frac{\partial \phi(x,y)}{\partial x}\Big|_{x=y=\hat{x}} \right] \bigg/ \left[1 - \frac{\partial \phi(x,y)}{\partial x}\Big|_{x=y=\hat{x}} \right]. \tag{29}$$

From Lemma 4 and $G(R(\hat{x})) = \hat{x}$, we have $\frac{\partial \phi(x,y)}{\partial x}\big|_{x=y=\hat{x}} = g(R(\hat{x}))R'(\hat{x})+2H'(0)R'(\hat{x})\hat{x}(1-\hat{x})$. Then, we have the given result for $\Gamma'(\hat{x})$. ∎

Using the above lemma, we can describe the correspondence curves more accurately. First, note that the slope of $\Gamma(\pi)$ at trivial ASE (π_m, π_m) always satisfies $0 < \Gamma'(\pi_m) < 1$, because the slope of the $\phi(a,a)$ curve at $a = \pi_m$ is greater than one: $g(R(\pi_m))R'(\pi_m) > 1$. Secondly, only when $\bar{\pi} < \pi_h$ as shown in Panel B of Figure 3, we have $R'(\pi_h) < 0$. Then, we have $0 < \Gamma'(\pi_h) < 1$ out of Lemma 6, as displayed in Panel A of Figure 5. (However, note that even when $\bar{\pi} < \pi_h$, we have $R'(\pi_l) > 0$.) Thirdly, as far as $\bar{\pi} > \pi_h$, any PSE $\hat{\pi}$ satisfies $R'(\hat{\pi}) > 0$. We have the following summary for all of the above cases.

Lemma 7. *The slope of correspondence at trivial ASE (π_m, π_m) always satisfies $0 < \Gamma'(\pi_m) < 1$. Given $R'(\pi_h) < 0$, the slope of correspondence at trivial ASE (π_h, π_h) is $0 < \Gamma'(\pi_h) < 1$. Given $R'(\hat{x}) > 0$ for $\hat{x} \in \{\pi_h, \pi_l\}$, the slope of correspondence at a trivial ASE $(\hat{x}, \hat{x})$ depends on the the density of identity cost k around zero, $H'(0)$:*

$$\begin{cases} -1 < \Gamma'(\hat{x}) < 0 & if \quad H'(0) < \frac{1-g(R(\hat{x}))R'(\hat{x})}{4R'(\hat{x})\hat{x}(1-\hat{x})} \\ \Gamma'(\hat{x}) < -1 & if \quad \frac{1-g(R(\hat{x}))R'(\hat{x})}{4R'(\hat{x})\hat{x}(1-\hat{x})} < H'(0) < \frac{1-g(R(\hat{x}))R'(\hat{x})}{2R'(\hat{x})\hat{x}(1-\hat{x})}, \; \forall \hat{x} \in \{\pi_h, \pi_l\}. \\ \Gamma'(\hat{x}) > 1 & if \quad H'(0) > \frac{1-g(R(\hat{x}))R'(\hat{x})}{2R'(\hat{x})\hat{x}(1-\hat{x})} \end{cases} \tag{30}$$

Proof. Given $R'(\hat{x}) > 0$ for $\hat{x} \in \{\pi_h, \pi_l\}$, we have $0 < g(R(\hat{x}))R'(\hat{x}) < 1$. Under this condition, Lemma 6 derives the given result. ∎

The lemma implies that given $R'(\hat{x}) > 0$, when the sensitivity of identity choice represented by $H'(0)$ is above a certain level, $\frac{1-g(R(\hat{x}))R'(\hat{x})}{4R'(\hat{x})\hat{x}(1-\hat{x})}$, the absolute value of the slope of correspondence curve $|\Gamma'(\hat{x})|$ at the trivial ASE $(\hat{x}, \hat{x})$ is greater than one.

Theorem 1. *Given multiple PSE, there always exist at least two non-trivial ASE.*

14

Proof. First, consider the case with $\bar{\pi} > \pi_h$. $\Gamma(b)$ passes through the points (π_h, π_h) and a-intercept $(a, b) = (0, \Gamma(0)^h)$, in which $\pi_h < \Gamma(0)^h < 1$. $\Gamma(a)$ passes through (π_l, π_l) and b-intercept $(b, a) = (1, \Gamma(1)^h)$, in which $\pi_h < \Gamma(1)^h < 1$. Thus, there should be at least one ASE which satisfies $\pi_A^* > \pi_B^*$. In the same way, we can find at least one ASE which satisfies $\pi_B^* > \pi_A^*$. Secondly, consider the case with $\bar{\pi} < \pi_h$. Using Lemma 3, we can find that $\pi_h < \Gamma(\pi_l)^h < b_4'$, in which $\phi(b_4', \pi_l) = \pi_l$ as shown in Panel A of Figure 5. We can draw the shape of $\phi(a, \Gamma(\pi_l)^h)$, which pass through the $\phi(a, a)$ curve both at some $a > \pi_l$ and at some $a < \pi_l$, as well as the point $(\Gamma(\pi_l)^h, \phi(\Gamma(\pi_l)^h, \Gamma(\pi_l)^h))$ on the curve. This implies that $\Gamma(\Gamma(\pi_l)^h)^l < \pi_l$ and $\Gamma(\Gamma(\pi_l)^h)^m > \pi_l$. From this, we can infer that there exist at least four non-trivial ASE. (Note that even when the number of PSE is two instead of three (for example, $\phi(a, a)$ is tangent to the 45 degree line at $a = \pi_l$), the proof goes in the same way.) ∎

At least two non-trivial ASE exist as far as multiple PSE exist. Whether there are more than two ASE or not depends on the curvature of $\Gamma(a)$ and $\Gamma(b)$ around trivial ASE $(\hat{x}, \hat{x})$. The slope of correspondence at a trivial ASE is important to examine the exact number of non-trival ASE. WLOG, the condition $|\Gamma'(\hat{x})| < 1$ for $\hat{x} \in \{\pi_h, \pi_l\}$ generates two more non-trivial ASE near to a trivial ASE $(\hat{x}, \hat{x})$, while the condition $|\Gamma'(\hat{x})| > 1$ for $\hat{x} \in \{\pi_h, \pi_l\}$ does not generate such non-trivial ASE around a trivial ASE $(\hat{x}, \hat{x})$. Refer to Panel A of Figure 4 and Panel A of Figure 5 for the case with the condition $|\Gamma'(\hat{x})| < 1$ for $\hat{x} \in \{\pi_h, \pi_l\}$, and Panels B and C of Figure 4 for the case with the condition $|\Gamma'(\hat{x})| > 1$ for $\hat{x} \in \{\pi_h, \pi_l\}$.

Proposition 3. *WLOG, it is most likely that the number of non-trivial ASE is six when both* $|\Gamma'(\pi_h)| < 1$ *and* $|\Gamma'(\pi_l)| < 1$ *and it is only two when both* $|\Gamma'(\pi_h)| > 1$ *and* $|\Gamma'(\pi_l)| > 1$.

Panel A of Figure 4 and Panel A of Figure 5 display six non-trivial ASE given $|\Gamma'(\pi_h)| < 1$ and $|\Gamma'(\pi_l)| < 1$, and Panels B and C of Figure 4 display two non-trivial ASE given $|\Gamma'(\pi_h)| > 1$ and $|\Gamma'(\pi_l)| > 1$. Let us call the two non-trivial ASE that exist regardless of the curvatures of the correspondences $\Gamma(a)$ and $\Gamma(b)$ "Persistent ASE," and denote them (π_H^{**}, π_L^{**}) and (π_L^{**}, π_H^{**}).

Proposition 4. *The two "Persistent ASE",* (π_H^{**}, π_L^{**}) *and* (π_L^{**}, π_H^{**}), *that consistently exist given multiple PSE (regardless of* $|\Gamma'(\pi_h)|$ *and* $|\Gamma'(\pi_l)|$*) satisfy*

$$\pi_L^{**} < \min\{\Omega_{CL}\} < \max\{\Omega_{CL}\} < \pi_H^{**}. \tag{31}$$

Proof. Given multiple PSE, using Lemma 3, we can find that $\pi_h < \Gamma(\pi_l)^h < \tilde{b}$, in which $\phi(\tilde{b}, \pi_l) = \pi_l$ as shown in Figure 3. We can draw the shape of $\phi(a, \Gamma(\pi_l)^h)$, which pass through the $\phi(a, a)$ curve both at some $a > \pi_l$ and at some $a < \pi_l$. This implies that $\Gamma(\Gamma(\pi_l)^h)^l < \pi_l$. Since $\Gamma(b)^h$ decreases over $b \in (0, \bar{\pi})$, there must be an intercept of $\Gamma(b)$ and $\Gamma(a)$, (π_L^{**}, π_H^{**}), which satisfies $\pi_L^{**} < \pi_l$ and and $\pi_H^{**} > \pi_h$. Out of the symmetricity, there must be another ASE (π_H^{**}, π_L^{**}). ∎

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

5.2   Stability of Affective Stereotyping Equilibria

Consider an intergenerational population structure. Every period, the randomly chosen α fraction of the workers die and the same number of agents are newly born. The newborn agents incur the cost c of skill achievement and the cost k to choose the affect A: k can be negative. Each newborn agent with his cost set (c, k) decides whether to invest for skills or not and which 'affect' to choose among A and B in the early days of his life. After those days of education and affect adaption, newborns join the labor market and receive wage set by employers. We assume that employers set the newborns' lifetime wage $W(\pi, t)$ proportional to the estimated skill level $\rho(\pi, t)$: $W(\pi_j, t) = w \cdot \rho(\pi_j, t)$ for the entering newborns with group identity $j \in \{A, B\}$ and the noisy signal t, given $\rho(\pi_j, t) = \pi_j f_1(t) / f(\pi_j, t)$. Employers use the skill composition of the current workers belonging to identity group j to estimate π_j. Therefore, we have the following dynamics:

$$\dot{\pi}_A > (<)0 \quad \Leftrightarrow \quad \phi(\pi_A, \pi_B) > (<)\pi_A, \tag{32}$$

$$\dot{\pi}_B > (<)0 \quad \Leftrightarrow \quad \phi(\pi_B, \pi_A) > (<)\pi_B. \tag{33}$$

The direction arrows in Panel A of Figure 3 describe the law of motions of π_A given π_B fixed as b_1: $\dot{\pi}_A > 0$ for any $\pi_A \in (0, \Gamma(b_1)^l)$ and any $\pi_A \in (\Gamma(b_1)^m, \Gamma(b_1)^h)$, and $\dot{\pi}_A < 0$ for any $\pi_A \in (\Gamma(b_1)^l, \Gamma(b_1)^m)$ and any $\pi_A \in (\Gamma(b_1)^h, 1)$. Therefore, direction arrows of $\dot{a}$ are upward between $\Gamma(b)^h$ and $\Gamma(b)^m$ and below $\Gamma(b)^l$ in the (b, a) plain, and downward between $\Gamma(b)^m$ and $\Gamma(b)^l$ and above $\Gamma(b)^h$. The direction arrows of $\dot{b}$ are rightward between $\Gamma(a)^h$ and $\Gamma(a)^m$ and at the lefthand side of $\Gamma(a)^l$ in the (b, a) plain, and leftward between $\Gamma(a)^m$ and $\Gamma(a)^l$ and

16

at the righthand side of $\Gamma(a)^h$, as displayed in Figures 4 and 5. From the described direction arrows, we can infer the following theorem.

Theorem 2. *Given Multiple PSE, two "Persistent ASE", (π_H^{**}, π_L^{**}) and (π_L^{**}, π_H^{**}), are stable and all other non-trivial ASE are unstable.*

The theorem together with Proposition 4 implies that inequality between endogenous groups in non-trivial ASE should be greater than inequality between exogenous groups in any PSE in the long run, because stable non-trivial ASE must be "Persistent ASE."

Proposition 5. *The middle trivial ASE (π_m, π_m) is always unstable. Other trivial ASEs, (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 the direction arrows, we can easily confirm the above proposition as well. Therefore, given $\bar{\pi} < \pi_h$, the trivial ASE (π_h, π_h) is stable because of $0 < \Gamma'(\pi_h) < 1$ (Lemma 7). Using Lemma 7 and the above proposition, we have the following result.

Theorem 3. *Given $R'(\hat{\pi}) > 0, \forall \hat{x} \in \Omega_{CL}(= \{\pi_l, \pi_m, \pi_h\})$, the trivial ASE (π_m, π_m) is unstable and other trivial ASE, (π_h, π_h) or (π_l, π_l), is stable if and only if $H'(0) \leq \frac{1-g(R(\hat{x}))R'(\hat{x})}{4R'(\hat{x})\hat{x}(1-\hat{x})}$, for $\hat{x} \in \{\pi_h, \pi_l\}$.*

The theorem implies the following interesting result:

Corollary 1. *Given $R'(\hat{\pi}) > 0, \forall \hat{x} \in \Omega_{CL}(= \{\pi_l, \pi_m, \pi_h\})$, the stable ASE are "Persistent ASE", (π_H^{**}, π_L^{**}) and (π_L^{**}, π_H^{**}), and all other ASE are unstable if and only if $H'(0) > \frac{1-g(R(\hat{x}))R'(\hat{x})}{4R'(\hat{x})\hat{x}(1-\hat{x})}, \forall \hat{x} \in \{\pi_h, \pi_l\}$.*

Therefore, when the society has enough fraction of newborns whose identity choice cost k is very low (i.e. $H'(0)$ is sufficiently big), balanced skill rates between two identity groups, (π_h, π_h) or (π_l, π_l), are not sustainable due to the incentives for the talented members to choose the "affect" associated with the slightly better collective reputation. The skill composition of the society converges to a non-trivial ASE in the long run, in which inequality between endogenous identity groups is greater than that of exogenous groups in any PSE: $|\pi_H^{**} - \pi_L^{**}| > |\pi_i - \pi_j|, \forall i, j \in \{l, m, h\}$.

Now imagine that the society is trapped by the low skill investment rates: the society is placed in a stable ASE (π_l, π_l). As far as two identity groups are feasible and the identity

17

choice is available for a fraction of workers, the social coordinator such as a government can mobilize the society to move out of the low investment trap by treating one of the identity groups favorably. The favorable treatment will lead more talented newborns to join the selected identity group. The skill level of the group can improve quickly with the higher skill investment activities of the newborns and by joining disproportionately more talented newborns to the group. However, the skill level of the other group which is not supported by the social coordinator may continue to be left behind in the low skill investment trap. For example, as shown in Panel A of Figure 4, the governmental intervention to relocate the skill composition from (π_l, π_l) to the point Q in the basin of attraction to (π_L^{**}, π_H^{**}) can mobilize the society to carry the much enhanced skill investment activities and, consequently, to arrive at a "Persistent ASE" (π_L^{**}, π_H^{**}) in which overall skill rate of the economy is much greater than the original skill rate π_l.

Proposition 6. *When the society is in low skill investment trap (π_l, π_l), the affective stereotyping may improve the social efficiency as the skill composition of the society can move to a "Persistent ASE" with a little push for an identity group to advance.*

6 Affective Stereotyping Equilibria with Unique PSE

In this section, we consider the case with unique PSE. Let us denote it by π_u: $G(R(\pi_u)) = \pi_u$. We show that non-trivial ASE can exist even under the unique PSE. It is surprising that even when phenotypic discrimination cannot generate the inequality between any groups, affective discrimination may bring about the inequality between affective groups forming endogenously in a society.

6.1 Existence of Affective Stereotyping Equilibria

Every PSE corresponds to trivial ASE: a trivial ASE (π_u, π_u) exists which satisfies $\phi(\pi_u, \pi_u) = \pi_u$. $\Gamma(b)$ passes through the points (π_u, π_u) and a-intercept $(b, a) = (0, \Gamma(0)^h)$, in which $\pi_u < \Gamma(0)^h < 1$. $\Gamma(a)$ passes through (π_u, π_u) and b-intercept $(a, b) = (1, \Gamma(1)^h)$, in which $\pi_u < \Gamma(1)^h < 1$. Therefore, as far as $|\Gamma'(\pi_u)| > 1$, there should be at least one non-trivial ASE which satisfies $\pi_A^* > \pi_B^*$ and at least one non-trivial ASE which satisfies $\pi_B^* > \pi_A^*$. (An example is described in Panel B of Figure 6 given $\Gamma'(\pi_u) < -1$.) Because $\Gamma(b)^h$ is decreasing

when $b < \bar{\pi}$, WLOG, there are two non-trivial ASE given $|\Gamma'(\pi_u)| > 1$.

Proposition 7. *Given unique PSE (π_u) and $|\Gamma'(\pi_u)| > 1$, WLOG, there exist two non-trivial ASE.*

However, the existence of non-trivial ASE is not guaranteed when $|\Gamma'(\pi_u)| < 1$. Panel A of Figure 6 and Panel A of Figure 7 show cases with existing non-trivial ASE while Panel B of Figure 7 shows a case without existing non-trivial ASE. Given $|\Gamma'(\pi_u)| < 1$, the curvature of $\Gamma(\pi)$ is critical for the determination of non-trivial ASE's existence: the closer the $\phi(x, y)$ curve is to the 45 degree line, the more likely that non-trivial ASE exist. If any non-trivial ASE exists, WLOG, it is most likely that there are four non-trival ASE given $|\Gamma'(\pi_u)| < 1$.

Corollary 2. *Given unique PSE (π_u) and $|\Gamma'(\pi_u)| < 1$, the existence of non-trivial ASE depends on the curvature of $\Gamma(\pi)$. Once they exist, WLOG, there are four non-trivial ASE.*

With the careful examination of the relative position of $\Gamma(a)$ and $\Gamma(b)$, we can confirm the following result:

Proposition 8. *Given unique PSE $(\pi_u \in (0, 1))$, any pair of non-trivial ASE, (π_H^*, π_L^*) and (π_L^*, π_H^*), satisfies the following condition:*

$$\pi_L^* < \pi_u < \pi_H^*. \tag{34}$$

At any non-trivial ASE, the collective reputation of an affective group is better than the PSE level π_u and that of the other affective group is worse than the level π_u.

6.2   Stability of Affective Stereotyping Equilibria

Using the direction arrows in phase diagrams in Figures 6 and 7, we can confirm the following results:

Proposition 9. *When two non-trivial ASE exist, both of them are stable. When four non-trivial ASE exist, two of them closer to the 45 degree line are unstable and the other two near the corners are stable.*

Lemma 8. *The trivial ASE (π_u, π_u) is stable if $|\Gamma'(\pi_u) \leq 1|$ and unstable if $|\Gamma'(\pi_u) > 1|$.*

Note that, given $R'(\pi_u) > 0$, we have $|\Gamma'(\pi_u) > 1|$ if and only if $H'(0) > \frac{1-g(R(\pi_u))R'(\pi_u)}{4R'(\pi_u)\pi_u(1-\pi_u)}$ (Lemma 7). Therefore, we achieve the following interesting result:

Proposition 10. *Given $R'(\pi_u) < 0$, the trivial ASE (π_u, π_u) is stable because $0 < \Gamma'(\pi_u) < 1$. Given $R'(\pi_u) > 0$, it is stable if and only if $H'(0) \leq \frac{1-g(R(\pi_u))R'(\pi_u)}{4R'(\pi_u)\pi_u(1-\pi_u)}$.*

Theorem 4. *Given $R'(\pi_u) > 0$ and $H'(0) > \frac{1-g(R(\pi_u))R'(\pi_u)}{4R'(\pi_u)\pi_u(1-\pi_u)}$, the only stable ASE are non-trivial ones while the trivial ASE (π_u, π_u) is unstable.*

Therefore, when the society has enough newborns whose 'affect' choice cost k is low, the equal society cannot be stable due to the emerging affective stereotyping. The society must converge to a non-trivial ASE, in which one group's skill level is greater than π_u and another group's skill level is less than π_u.

7 Discussions on Fang(2000)

The example of Fang(2000) is a special case of the given model that there exists a unique PSE which is zero: $\pi_u = 0$. Fang(2000)'s Proposition 2 proves that there exists at least one non-trivial ASE if and only if $\phi(a, 0) > a$ for some $a \in (0, 1)$. (Refer to Panel A of Appendix Figure 3.)

Using Lemma 6, we know $\Gamma'(0) = 0$. We also confirm $\Gamma(b)^l = 0, \forall b \in [0, 1]$ from the $\phi(a, b)$ curves in Panel A of the figure. Given $\phi(a, 0) > a$ for some $a \in (0, 1)$, we have both $\Gamma(0)^h > 0$ and $\Gamma(0)^m > 0$. Existence of non-trivial ASE is easily confirmed from the $\Gamma(a)$ and $\Gamma(b)$ curves in Panel B of the figure: $\Gamma(0)^j \in \Gamma(0)$ and $0 \in \Gamma(\Gamma(0)^j), \forall j \in \{m, h\}$. Corollary 2 shows that, WLOG, there are four non-trivial ASE once any non-trivial ASE exists. Those four non-trivial ASE are denoted in Panel B of the figure. According to Proposition 9, two of them closer to the 45 degree line are unstable and the other two near the corners are stable, as displayed in the panel.

8 Conclusion

In this paper, we develop an identity choice model that can explain social activities such as passing and selective out-migration from a stereotyped group, loosening the assumption of group identity immutability in standard statistical discrimination models. More talented

members with low human capital investment cost have a greater incentive to identify themselves with a group that has a better collective reputation. The positive selection into a favored group plays a critical role in causing human capital cost distribution between groups to endogenously diverge. This model can be applied to many other social settings such as code switching (Goffman, 1959) and generating certificates to fight negative stereotypes.

In the given model, agents are myopic in the sense that they do not account for long-term expectations of groups' reputations. In Kim and Loury (2008), we discuss the stability of multiple equilibria in a dynamic setting. We identify the balanced dynamic paths to high and low stable reputation equilibria, and the 'overlap' range in which expectations about the future determine the final economic outcomes. The model in this paper can be extended to such a dynamic setting to generate further implications for identity choice behavior.

Reference

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

Arrow, K.J. (1971). "Political and Economic Evaluation of Social Effects and Externalities," in Michael Intriligator (eds.), *Frontiers of Quantitative Economics.* Amsterdam: North-Holland: 841-877.

Arrow, K.J. (1973). "The Theory of Discrimination," in Ashenfelter and Rees (eds.), *Discrimination in Labor Markets.*

Chaudhuri, S. and R. Sethi (2008). "Statistical Discrimination with Peer Effects: Can Integration Eliminate Negative Stereotypes?" *Review of Economic Studies* 78: 579-596.

Charles, Kerwin K., Erik Hurst and Nikolai Roussanov (2007). "Conspicuous Consumption and Race," Unpublished Manuscript.

Coate, S. and G. C. Loury (1993). "Will Affirmative-Action Policies Eliminate Negative Stereotype?" *American Economic Review* 83: 1220-1240.

Fang, Haming (2001). "Social Culture and Economic Performance," *American Economic Review* September: 924-937.

Fukuoka, Yasunori, Yukiko Tsujiyama and John Russell (1998). "MINTOHREN: Young Koreans Against Ethnic Discrimination in Japan," *The Bulletin of Chiba College of Health Science* 10.

Fryer, Roland and Paul Torelli (2010). "An Empirical Analysis of 'Acting White'," *Journal of Public Economics* 94: 380-396.

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

Grogger, Jeffrey (2008). "Speech Patterns and Racial Wage Inequality," Unpublished Manuscript.

Loury, Glenn C. (1995). "Economic Discrimination: Getting to the Core of the Problem," in *One By One From the Inside Out: Essays and Reviews on Race and Responsibility in America.* New York: The Free Press.

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

Myrdal, Gunnar. (1962), *An American Dilemma: The Negro Problem and Modern Democracy.* New York: Pantheon Books.

O'Flaherty, D. and R. Sethi (2008). "Racial Stereotypes and Robbery," *Journal of Economic Behavior & Organization* 68: 511-524.

Phelps, E. (1972). "The Statistical Theory of Racism and Sexism," *American Economic Review* 62: 659-661.

Schwab, S. (1986). "Is Statistical Discrimination Efficient?" *American Economic Review* 76: 229-234.

Spence, Michael A. (1974). *Market Signaling: Information Transfer in Hiring and Related Screening Processes*, Cambridge, MA: Harvard University Press.

Sweet, Frank W. (2005). "The Rate of Black-White "Passing"," in *Legal History of the Color Line: The Rise and Triumph of the One-Drop Rule*, Backintyme.

Figure 1. Phenotypic Stereotyping Equilibria

Panel A. Unique PSE
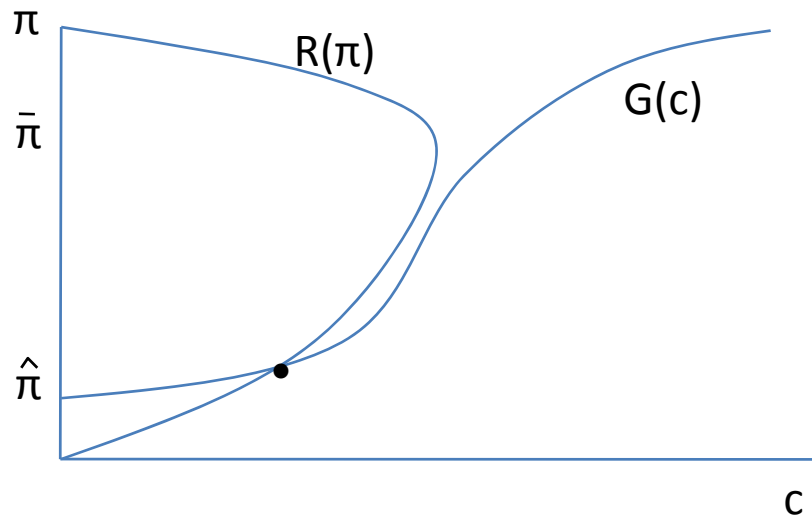
Panel B. Multiple PSE
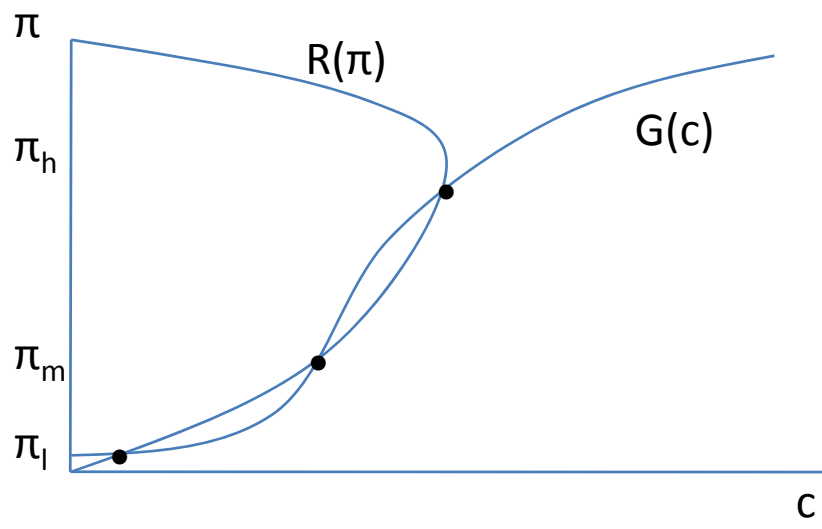
Figure 2. Human Capital Investment and Affective Behavior

Panel A. Case with a>b
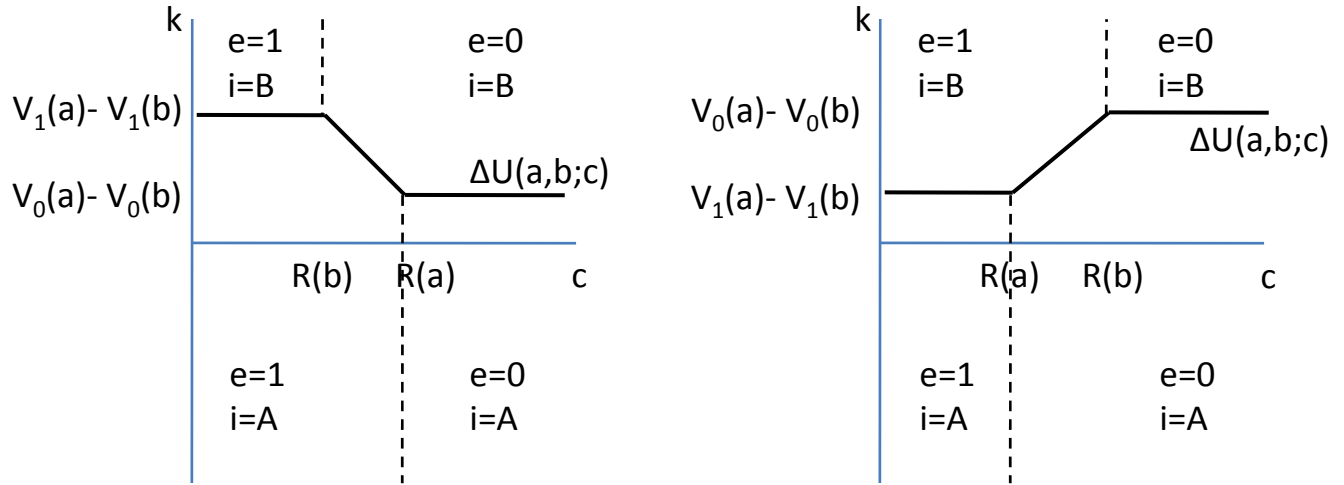
Panel B. Case with a<b
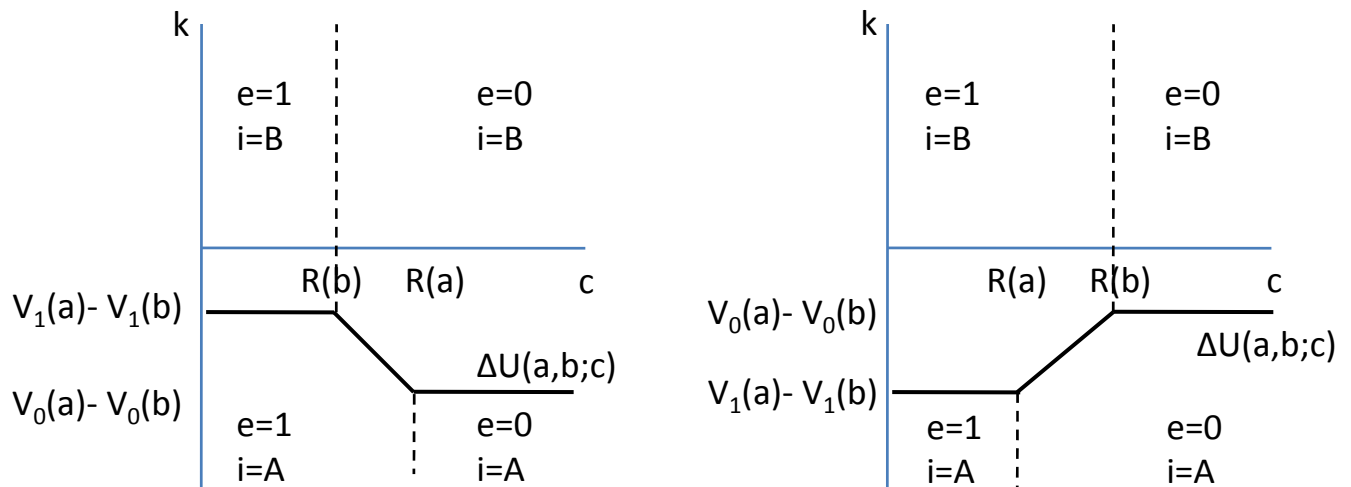
Figure 3. Human Capital Investment Rate

Panel A. Case with $\bar{\pi} > \pi_h$
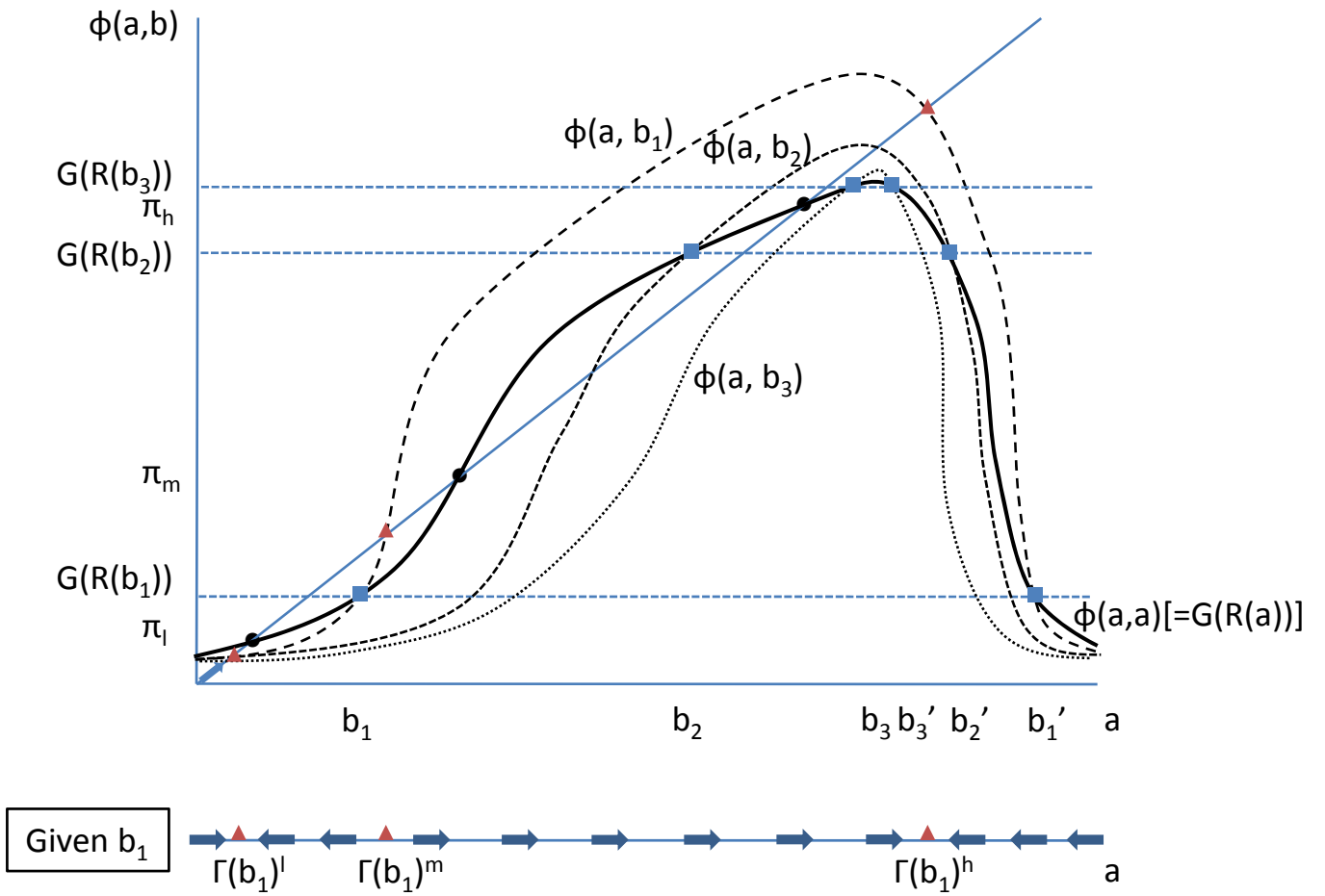
Panel B. Case with $\bar{\pi} < \pi_h$
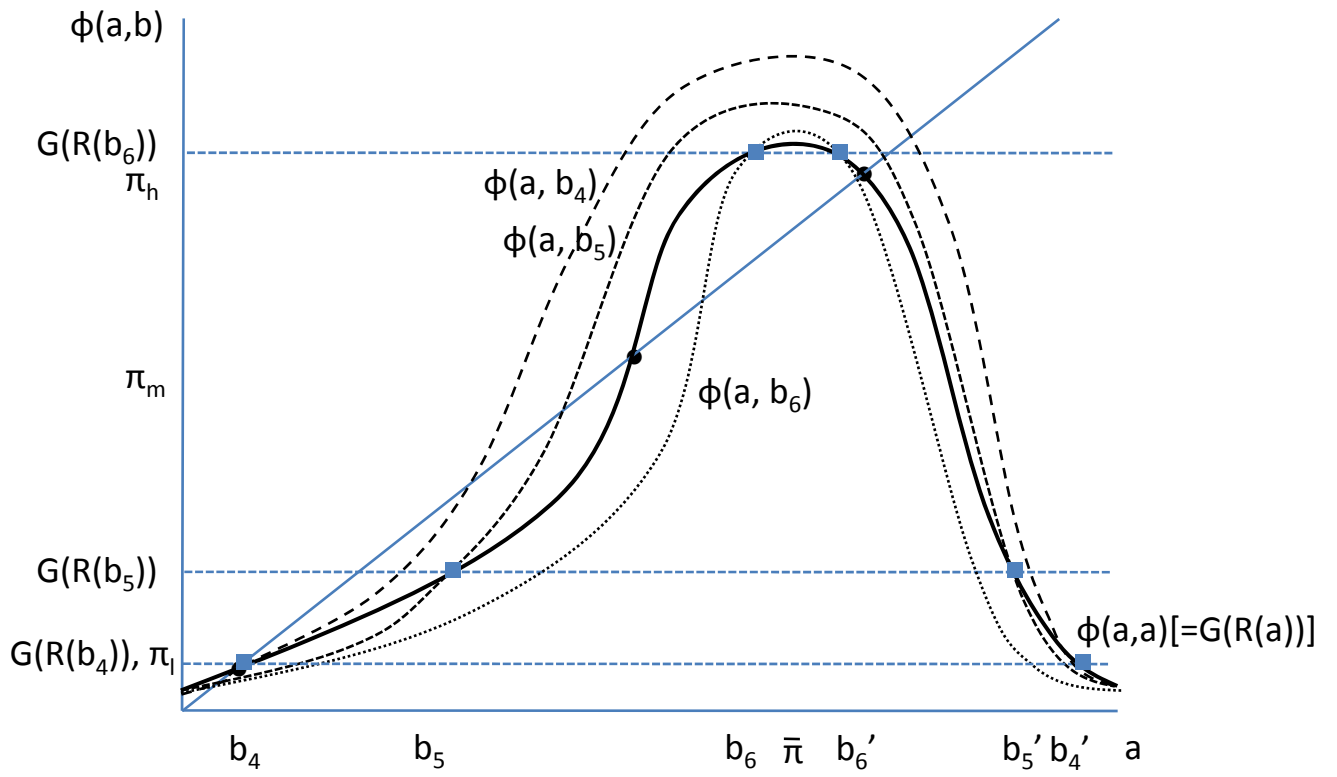
Figure 4. ASE given Multiple PSE: Case with $\bar{\pi} > \pi_h$
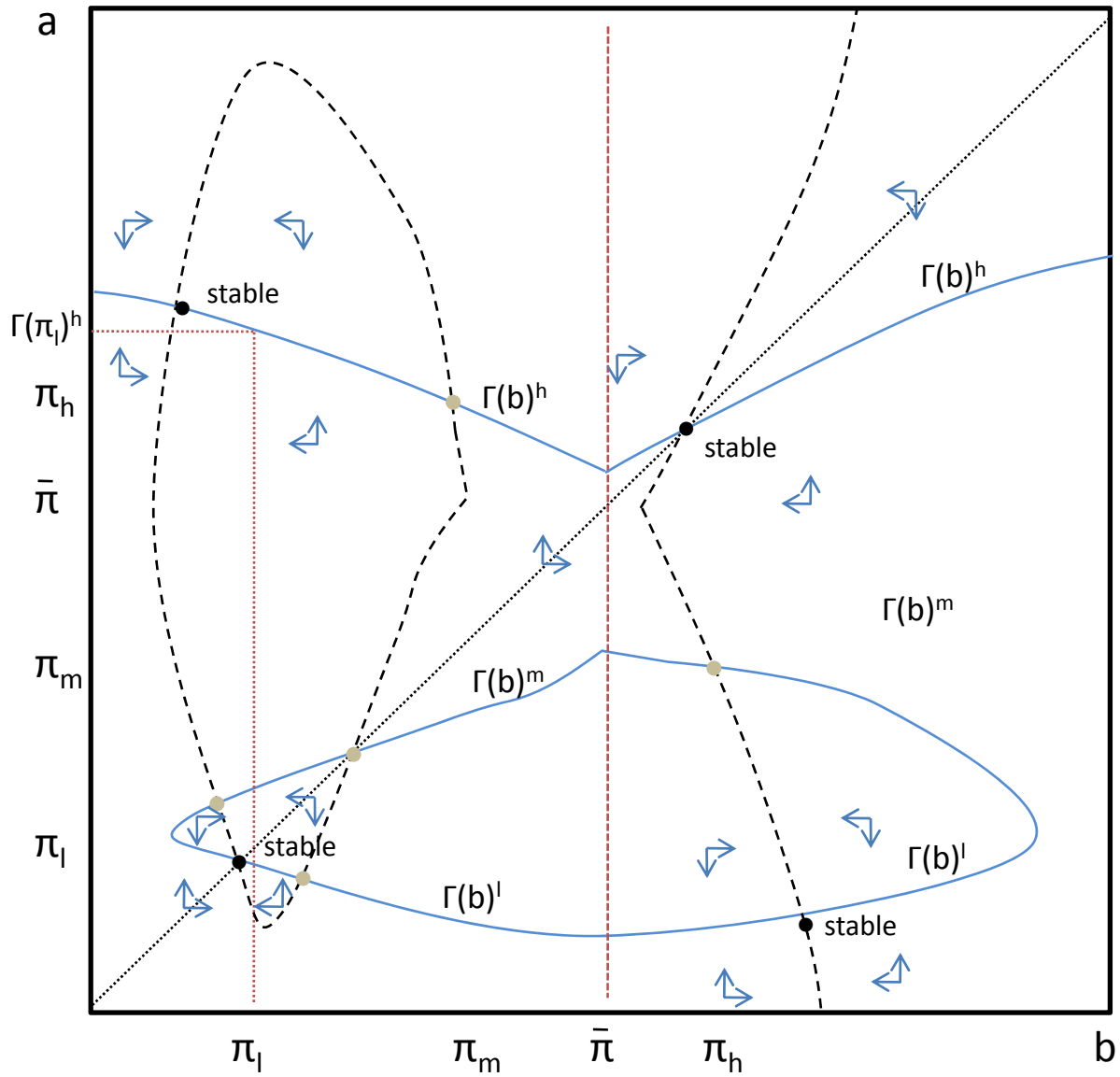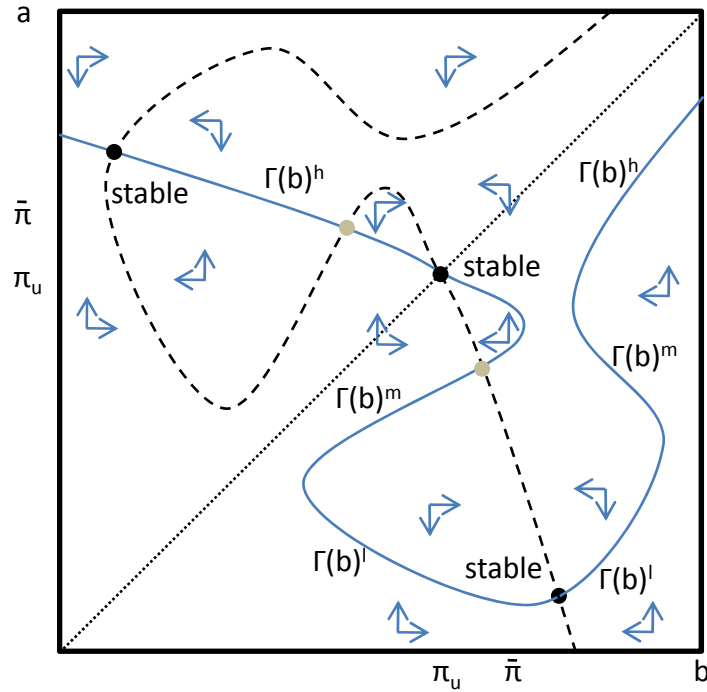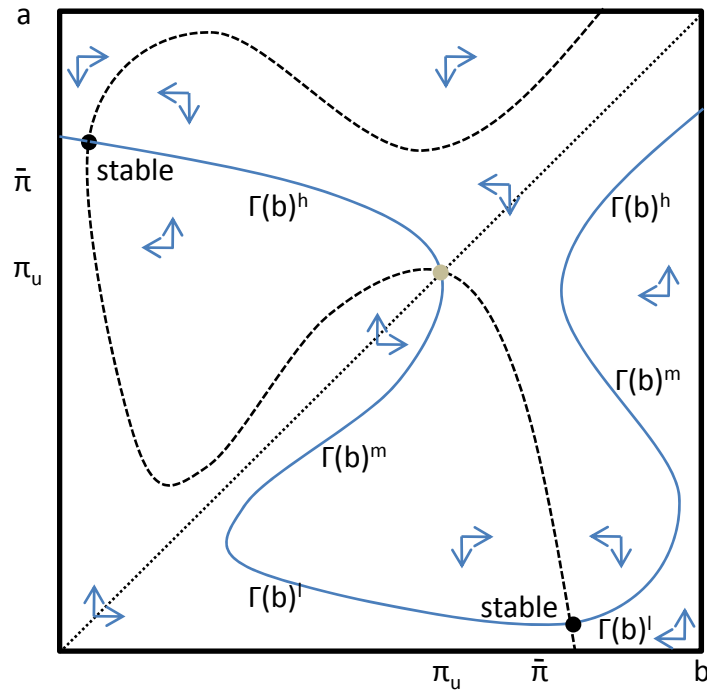
Panel A. Given both -1<Γ'(π_h) and -1<Γ'(π_l)<0
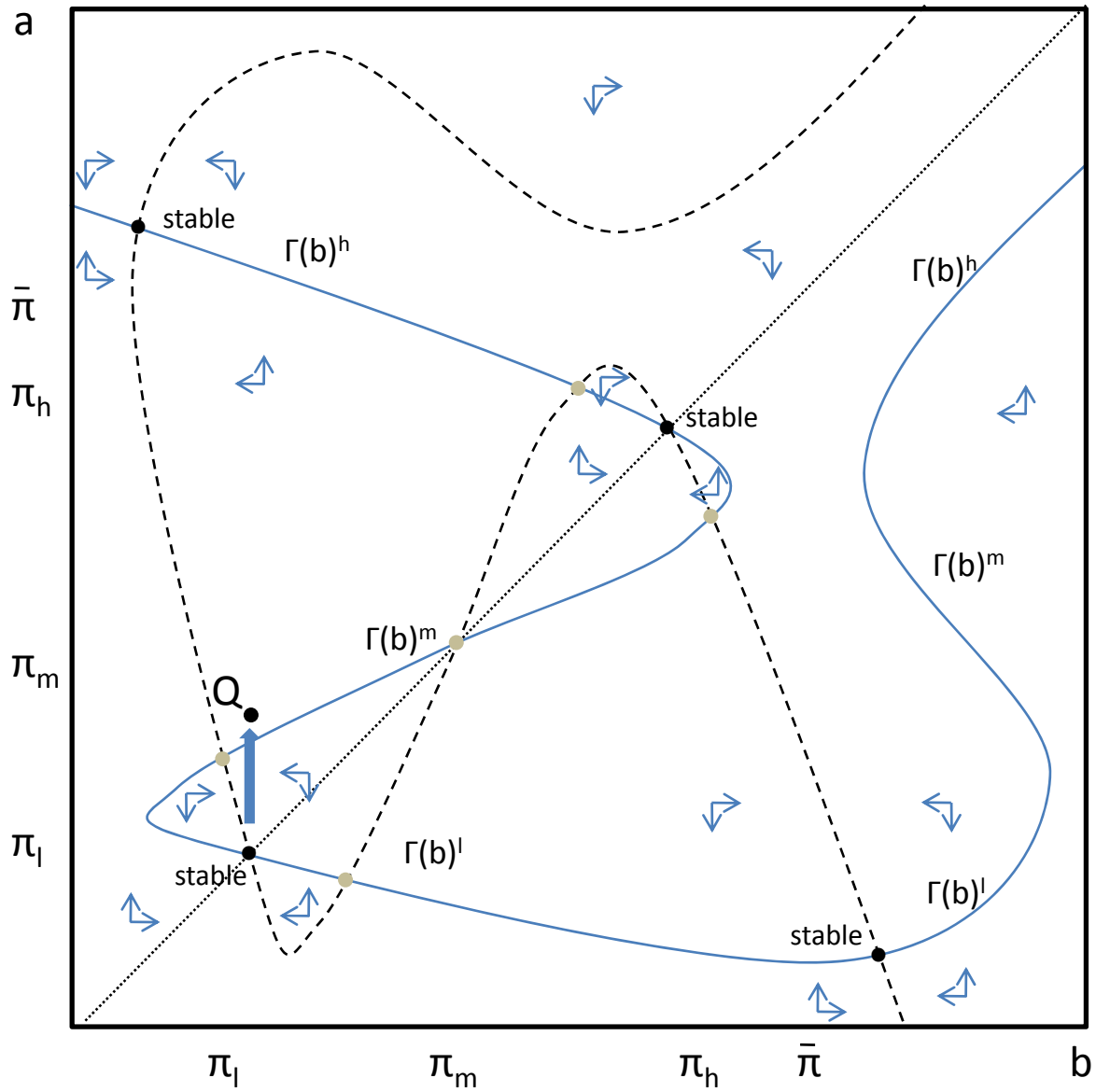
Panel B. Given both $\Gamma'(\pi_h)<-1$ and $\Gamma'(\pi_l)<-1$
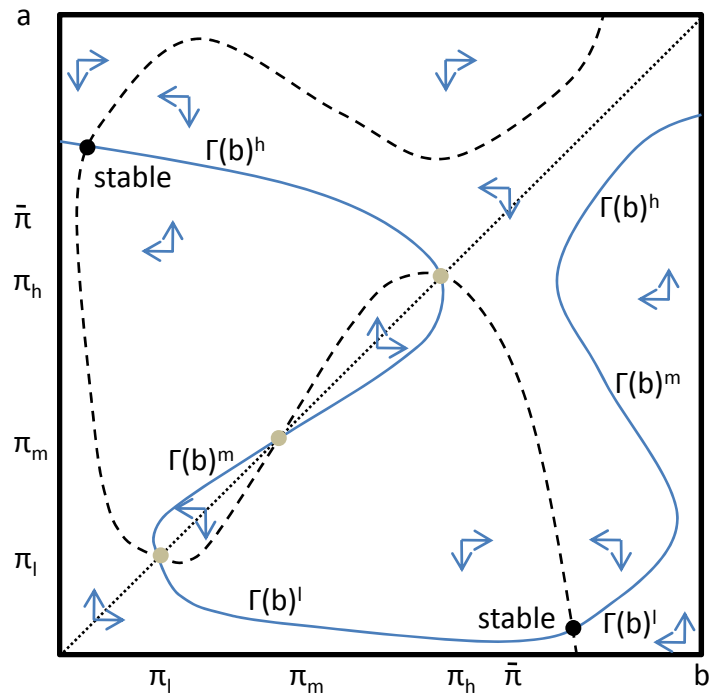
Panel C. Given both $\Gamma'(\pi_h)>1$ and $\Gamma'(\pi_l)>1$
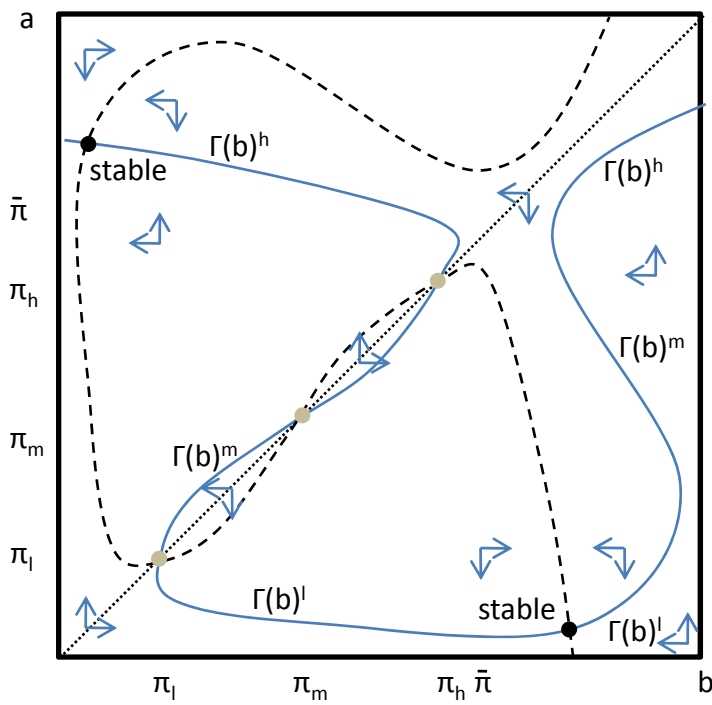
Figure 5. ASE given Multiple PSE: Case with $\bar{\pi}<\pi_h$

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

Figure 6. ASE given Unique PSE: Case with $\bar{\pi} > \pi_u$

Panel A. Given -1< Γ'(π_u)<0

Panel B. Given Γ'(π_u)<-1

Figure 7. ASE given Unique PSE: Case with $\bar{\pi}<\pi_u$

Panel A. Multiple ASE given 0< $\Gamma'(\pi_u)$<1
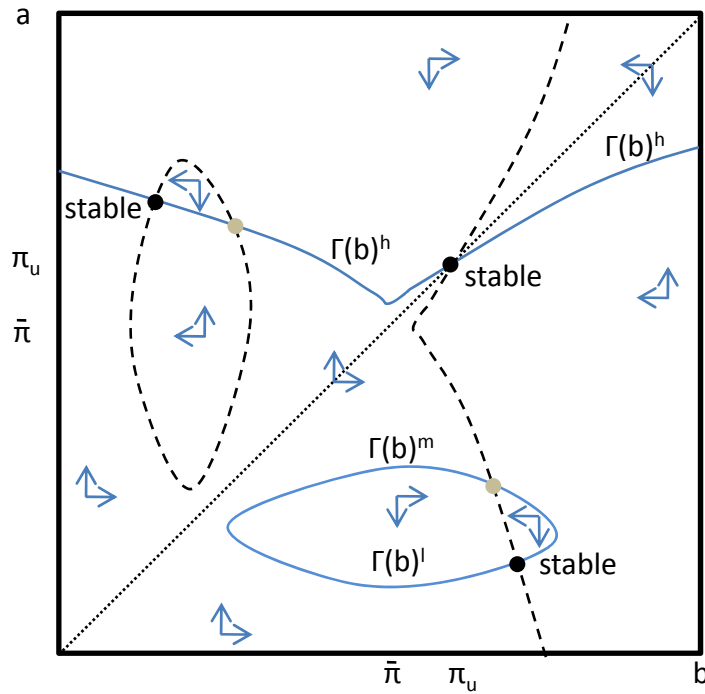
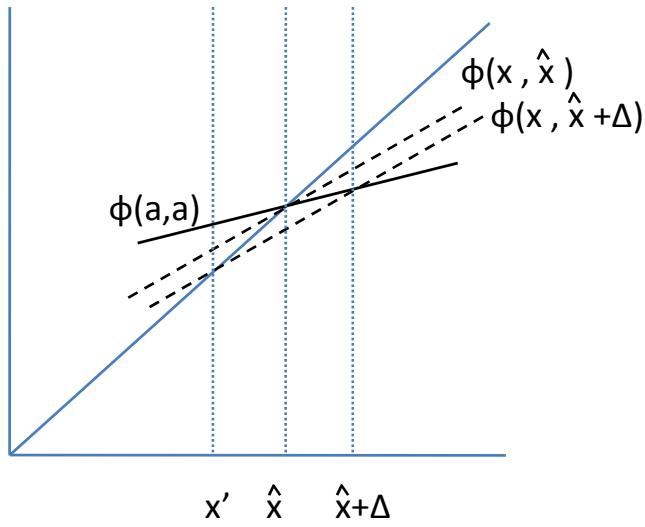## Panel B. Unique ASE given 0< $\Gamma'(\pi_u)$<1
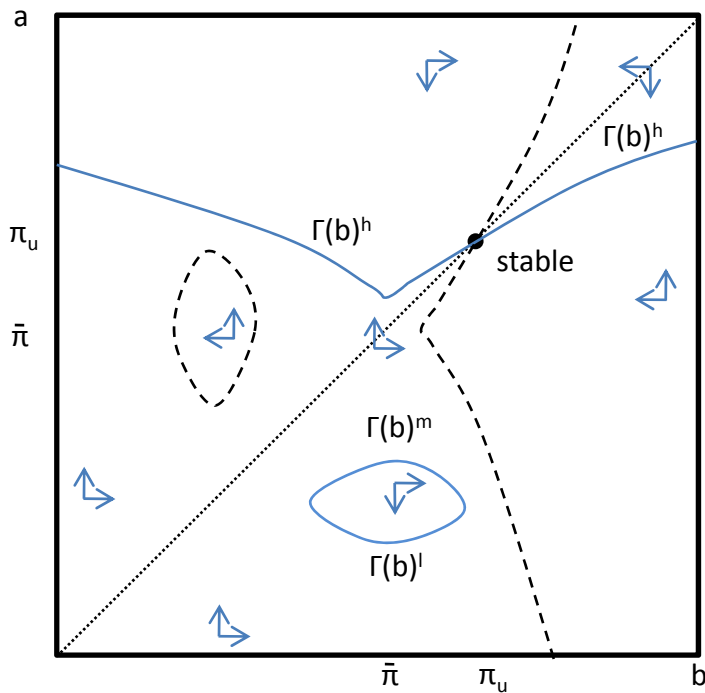
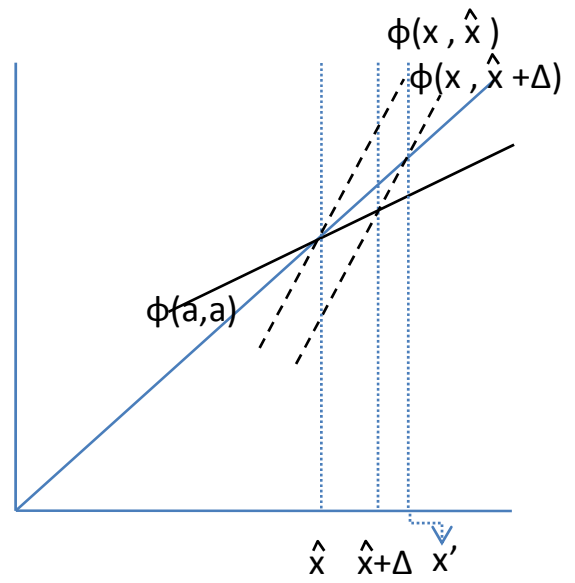# Appendix Figure 1. Proof of Lemma 3

Panel A. Case with $b<a<\bar{\pi}$
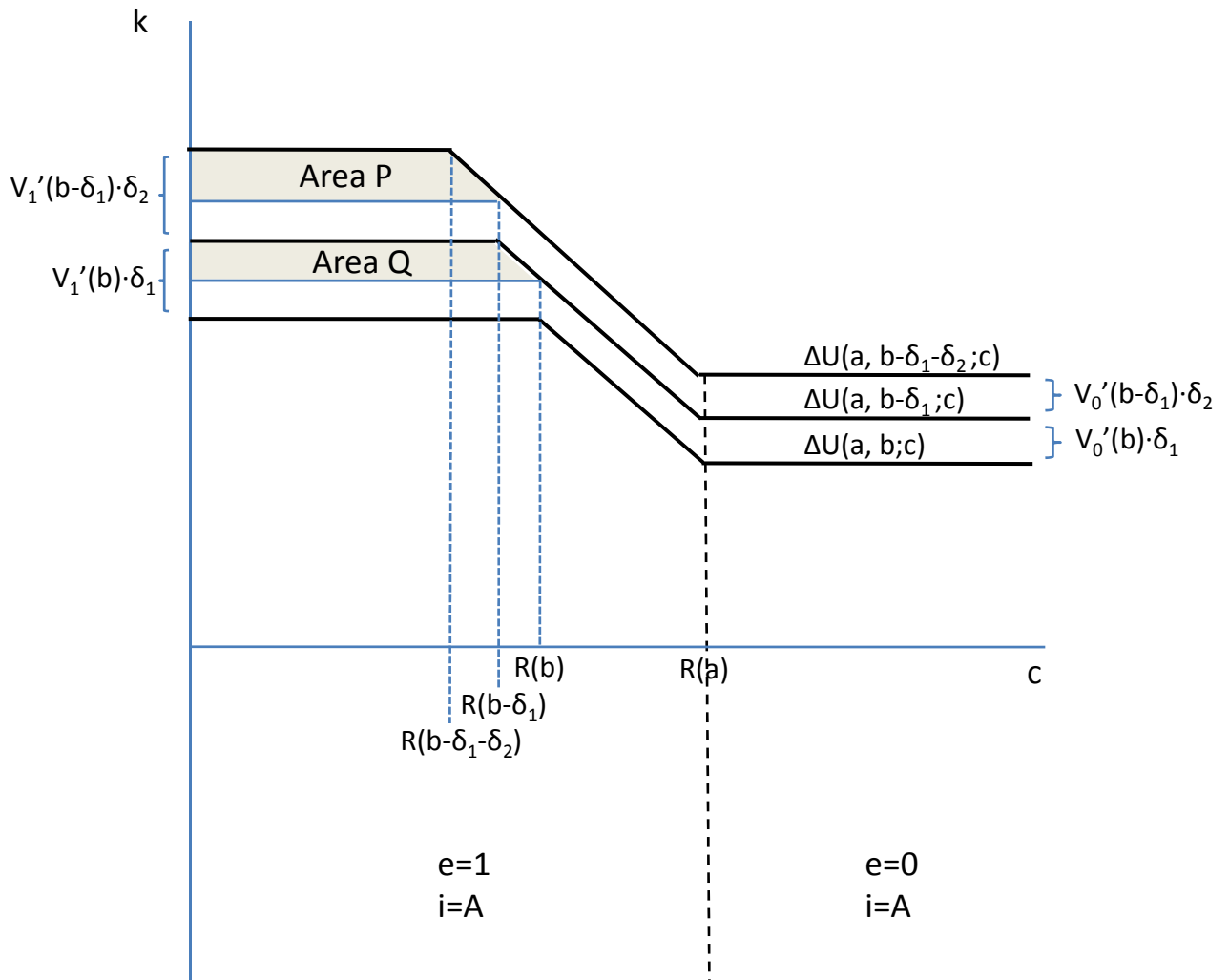
Appendix Figure 2. Slope of Correspondence at trivial ASE
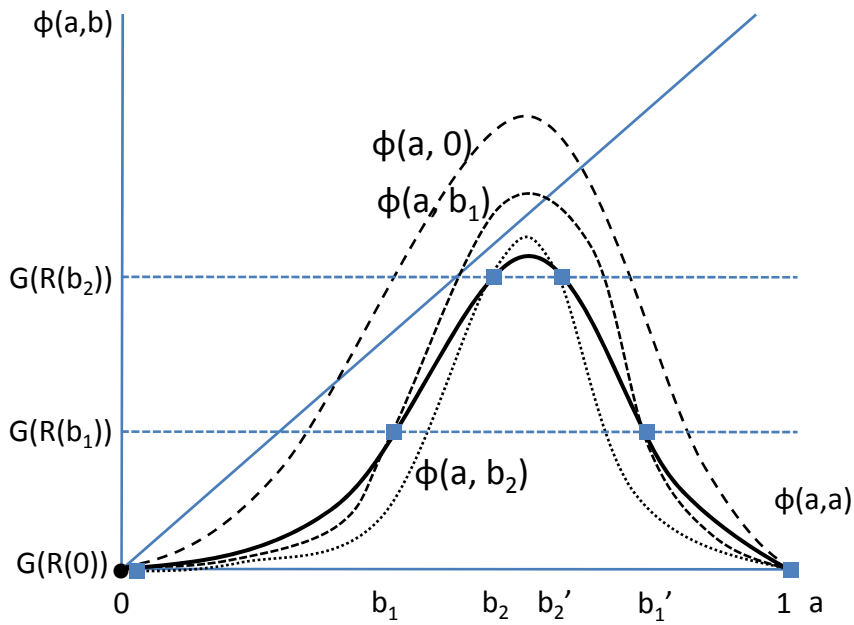
Panel A. Example for $\Gamma'(\hat{x})<1$

Panel B. Example for $\Gamma'(\hat{x})>1$

$\phi(x, \hat{x})$

$\phi(x, \hat{x}+\Delta)$

$\phi(a,a)$

x' $\hat{x}$ $\hat{x}+\Delta$

$\phi(x, \hat{x})$

$\phi(x, \hat{x}+\Delta)$

$\phi(a,a)$

$\hat{x}$ $\hat{x}+\Delta$ x'

Appendix Figure 3. An example of Fang (2000)

Panel A. Given Unique PSE: $\pi_u=0$

Panel B. Multiple ASE (When $\phi(a, 0)>a$ for some a)