

# Trade and Divergence in Education Systems:

## Technical Appendix

Pao-Li Chang and Fali Huang\*

Singapore Management University, Singapore

August 9, 2013

### 1. The Public Choice of Education System under Trade

The main results of the paper are obtained where the education policy is chosen by the government maximizing the aggregate welfare of all individuals in a country. This seems to be a reasonable approximation to reality, where education policies are often determined by government bureaucracies rather than by a referendum or other voting process. Nonetheless, it is still possible that they are indirectly affected by voter preferences, for example, through the choice of the ruling party or the education minister. We develop a full-blown political economy model in this section based on both the majority voting rule and the probabilistic voting approach, and show that our main result (i.e., educational choices become more divergent after trade) still holds under a variety of reasonable political scenarios.

It is natural to use an overlapping generation framework to study the voting process. Suppose there are overlapping generations where each generation  $n = 1, 2, \dots$  lives for two sub-periods labeled as childhood and adulthood. All children go to school, become adults at graduation and

---

\*Please address correspondence to: Pao-Li Chang, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903. Phone: +65-68280830. Fax: +65-68280833. E-mail: [plchang@smu.edu.sg](mailto:plchang@smu.edu.sg). Huang gratefully acknowledges the financial support of SMU Research Grant C244/MSS6E019.

enter the workforce either in the automobile ( $A$ ) or software ( $S$ ) industry. At the beginning of each period  $n$ , the education system characterized by  $\delta_{n+1}$  is to be chosen collectively by all adults and then to be implemented on children who will become adults of generation  $n + 1$  in the next period. Suppose there is a positive correlation of talent across generations such that the talent of a child from family  $i$  is equal to his parent's initial talent  $t_{i0}$  with probability  $q$ , and is a random draw from the distribution  $G(\cdot)$  with probability  $1 - q$ , where  $q \in (0, 1)$ . Note that if there is no correlation of talent between parent and child so that  $q = 0$ , all individuals' problems would be identical with the social planner's problem. We focus on the steady state results where  $\delta$  is constant across generations.

An individual  $i$ 's optimal choice  $\delta^i$  maximizes his child's expected utility

$$\mathcal{V}_i(\delta) \equiv q\nu_i(\delta) + (1 - q)U(\delta),$$

where  $\nu_i(\delta) \equiv \xi p^{-(1-\beta)} w(t_i) - k\delta$  is the indirect utility function of a child if inherited with initial ability  $t_{i0}$ , net of the utility cost, given an education system  $\delta$ , and  $\xi \equiv \beta^\beta (1 - \beta)^{1-\beta}$  as derived in the paper. Recall that  $U(\delta) = \sum \nu_i(\delta)$ , which is the expected net welfare of a child under the same education system if talent is a random draw.

In the following analysis, we study the case where individuals take prices as exogenously given when making decisions on the optimal education system  $\delta$ , while at the aggregate level, the price is determined endogenously by the chosen  $\delta$  under autarky in each country, or by  $\delta_J$  and  $\delta_U$  together under trade. In the equilibrium, it can be shown that there exists a unique combination of  $\delta$  and  $p$  under autarky (or  $\delta_J$ ,  $\delta_U$ , and  $p$  under trade) that are consistent with each other, the proof of which is similar to that in the paper and thus omitted.

Note that in the scenario where talent is not inherited but a random draw, the FOC for the most preferred education system by an individual  $i$  is:

$$\begin{aligned} \frac{dU(\delta)}{d\delta} &= \sum_i \frac{d\nu_i(\delta)}{d\delta} = \sum_i \xi p^{-(1-\beta)} \left( \frac{\partial w(t_i)}{\partial \delta} + \frac{dw(t_i)}{dp} \frac{\partial p}{\partial \delta} - (1 - \beta) \frac{w(t_i)}{p} \frac{\partial p}{\partial \delta} \right) - \sum_i k \\ &= \xi p^{-(1-\beta)} \left( \frac{\partial Y_A}{\partial \delta} + p \frac{\partial Y_S}{\partial \delta} \right) - k, \end{aligned}$$

where the last equality is obtained by the fact that the price effects  $\frac{\partial p}{\partial \delta}$  are not taken into account by individual voters and note that  $\frac{\partial Y_A}{\partial \delta} = \sum_{i \in A} \frac{\partial w(t_i)}{\partial \delta}$  and  $p \frac{\partial Y_S}{\partial \delta} = \sum_{i \in S} \frac{\partial w(t_i)}{\partial \delta}$ . As expected, when the endogenous price effects are not taken into account, the FOC above is identical to that for the socially optimal outcome, an observation also made in the paper. In the socially optimal outcome, the price (or terms of trade) effects are canceled out for the whole population (or the whole world); thus, it is equivalent to considering the direct effects of  $\delta$  on outputs alone as assumed for the individual voters here.

### 1.1. Conflict of Interests in Educational Choice

For an individual  $i$ , the marginal gain to his child from increasing  $\delta$  is

$$\begin{aligned} \frac{d\mathcal{V}_i(\delta)}{d\delta} &= q \frac{d\nu_i(\delta)}{d\delta} + (1-q) \frac{dU(\delta)}{d\delta} \\ &= q\xi p^{-(1-\beta)} \frac{\partial w(t_i)}{\partial \delta} - qk + (1-q) \frac{dU(\delta)}{d\delta}. \end{aligned}$$

The conflict of interests in choosing  $\delta$  across individuals is reflected by the direct effect of  $\delta$  on wage  $\frac{\partial w(t_i)}{\partial \delta}$ , which differs not only across industries but also among individuals in the same industry.

Before we start, we delineate the constraints on the mean effect parameter  $\gamma$ . The parameter  $\bar{\gamma}$  as introduced below plays a similar role as  $\tilde{\gamma}$  in the paper that sets an upper bound on the mean effect of education on talent. For  $\gamma < \bar{\gamma}$ , an increase in  $\delta$  decreases the software output for each pair of workers in the sector. The aggregate software output as a result will also decrease in  $\delta$ .

**Definition 1**  $\bar{\gamma} \equiv \min_{\{\delta, t_{i0}\}} \gamma(\delta, t_{i0}) > 1$ , where  $\gamma(\delta, t_{i0})$  defines the critical value for  $\gamma$  such that  $\frac{\partial Y_{S,i}}{\partial \delta}$  is equal to zero and  $Y_{S,i} \equiv F^S[t_i, m(t_i)]$  is the software output of a pair of workers with talents  $t_i$  and  $m(t_i)$ .

**Explanation.** Define  $Y_{S,i}^b \equiv F^S[t_{ib}, m(t_{ib})]$  as the software output of a pair of workers excluding the mean effect. Recall that  $t_i \equiv \gamma^\delta t_{ib}$  and  $t_{ib} \equiv (1-\delta)t_{i0} + \delta \bar{t}_0$ . It follows that  $Y_{S,i} = \gamma^\delta Y_{S,i}^b$  and  $\frac{\partial Y_{S,i}}{\partial \delta} = (\gamma^\delta \ln \gamma) Y_{S,i}^b + \gamma^\delta \frac{\partial Y_{S,i}^b}{\partial \delta}$ . Note that  $\frac{\partial Y_{S,i}^b}{\partial \delta} = (F_1^S - F_2^S) \left( \frac{\bar{t}_b - t_{ib}}{1-\delta} \right) < 0$  and  $\frac{\partial Y_{S,i}}{\partial \delta}$  increases in  $\gamma$ . Thus, there exists a critical value  $\gamma(\delta, t_{i0}) = \exp\left(-\frac{1}{Y_{S,i}^b} \frac{\partial Y_{S,i}^b}{\partial \delta}\right) > 1$  such that  $\frac{\partial Y_{S,i}}{\partial \delta} = 0$ . ■

**Lemma 1** For  $1 < \gamma < \min\{e, \bar{\gamma}\}$ , a majority of workers have  $\frac{\partial w(t_i)}{\partial \delta} > 0$ ; they include all software workers with abilities  $t_{i0} \in [t_{i0}, \hat{t}_0]$  and auto workers with  $t_{i0} \in [\tilde{t}_0, \min\{m(\hat{t}_0), \tilde{t}_0\}]$ , where  $\tilde{t}_0 \equiv \frac{1+\delta \ln \gamma}{1+\delta \ln \gamma - \ln \gamma} \bar{t}_0 > \bar{t}_0$ .

**Proof.** (1) Auto workers: If a voter works in the auto industry, his child's income will be  $w(t_i) = \frac{\lambda_A t_i}{2}$  with probability  $q$ , where  $t_i = \gamma^\delta [(1 - \delta)t_{i0} + \delta \bar{t}_0]$ . It is easy to see that

$$\begin{aligned} \frac{\partial w(t_i)}{\partial \delta} &= \frac{\lambda_A}{2} \frac{\partial t_i}{\partial \delta} = \frac{\lambda_A}{2} \gamma^\delta \{[(1 - \delta)t_{i0} + \delta \bar{t}_0] \ln \gamma + (\bar{t}_0 - t_{i0})\} \\ &= \frac{\lambda_A}{2} \gamma^\delta \{(1 - \delta) \ln \gamma - 1\} t_{i0} + (\delta \ln \gamma + 1) \bar{t}_0. \end{aligned}$$

Note that  $\frac{\partial w(t_i)}{\partial \delta} > 0$  holds for any  $t_{i0} \leq \bar{t}_0$  given  $\gamma > 1$ . In addition,

$$\frac{\partial^2 w(t_i)}{\partial \delta \partial t_{i0}} = \frac{\lambda_A}{2} \gamma^\delta [(1 - \delta) \ln \gamma - 1] < 0$$

holds given  $\gamma < e$ . That is, the marginal wage gain from a higher  $\delta$  is decreasing in the initial ability  $t_{i0}$ . Given that  $\frac{\partial w(t_i)}{\partial \delta} > 0$  for  $t_{i0} = \bar{t}_0$ , we know that there must exist a unique threshold  $\tilde{t}_0 > \bar{t}_0$  such that  $\frac{\partial w(t_i)}{\partial \delta} = 0$ . When  $\tilde{t}_0 < m(\hat{t}_0)$  holds, there are conflicting interests among auto workers, where those with lower abilities ( $t_{i0} < \tilde{t}_0$ ) have  $\frac{\partial w(t_i)}{\partial \delta} > 0$  while others ( $t_{i0} > \tilde{t}_0$ ) have the opposite result  $\frac{\partial w(t_i)}{\partial \delta} < 0$ . When  $\tilde{t}_0 \geq m(\hat{t}_0)$ , all auto workers have  $\frac{\partial w(t_i)}{\partial \delta} > 0$ .

(2) Software workers: A voter with low ability  $t_i \in [t_l, \hat{t}]$  works in the software industry, and his child's income will be

$$w(t_i) = \frac{\lambda_A \hat{t}}{2} - \int_{t_i}^{\hat{t}} pF_1^S[s, m(s)] ds$$

with probability  $q$ , where

$$\frac{\partial w(t_i)}{\partial \delta} = pF_1^S[t_i, m(t_i)] \frac{\partial t_i}{\partial \delta} + \left\{ \frac{\lambda_A}{2} - pF_1^S[\hat{t}, m(\hat{t})] \right\} \frac{\partial \hat{t}}{\partial \delta} > 0$$

is true, since  $\frac{\partial t_i}{\partial \delta} > 0$  for  $t_i < \hat{t}$ ,  $\frac{\partial \hat{t}}{\partial \delta} = \gamma^\delta (\ln \gamma) \hat{t}_b > 0$ , and  $\frac{\lambda_A}{2} > pF_1^S[\hat{t}, m(\hat{t})]$ . To see the last result, note that the wage profile of the software sector is convex in  $t_i$  and has a smaller slope than the linear wage schedule of the auto sector at  $\hat{t}$  (Grossman and Maggi, 2000, Figure 4).

Note that  $pY_{S,i} = w(t_i) + w(m(t_i))$  holds given perfect competition. Since  $p \frac{\partial Y_{S,i}}{\partial \delta} < 0$  holds for each pair of software workers given  $\gamma < \bar{\gamma}$ , the fact  $\frac{\partial w(t_i)}{\partial \delta} > 0$  for the low-ability software workers

implies  $\frac{\partial w(m(t_i))}{\partial \delta} < 0$  for the high-ability software workers. In conclusion, a majority of workers, with initial abilities below  $\min\{m(\hat{t}_0), \tilde{t}_0\}$ , have  $\frac{\partial w(t_i)}{\partial \delta} > 0$ . ■

## 1.2. Divergent Educational Choices after Trade in Majority Voting Model

Since the main focus of our paper is the effect of trade on education policies, we assume that countries have the same political institutions. In the following, we consider the choice of education system based on majority voting, where recall that the majority consists of all auto workers with initial abilities lower than  $\min\{m(\hat{t}_0), \tilde{t}_0\}$ , and all low-ability software workers.

**Proposition 1** *If the median voter is an auto worker of low abilities with  $t_{i0} < \min\{m(\hat{t}_0), \tilde{t}_0\}$ , the educational choices become more divergent across countries after trade for any  $q$ . The same result holds regardless of the identity of the median voter if  $q$  is sufficiently small.*

**Proof.** When an auto worker with  $t_{i0} < \min\{m(\hat{t}_0), \tilde{t}_0\}$  is the median voter, his preferred choice will be the voting result. In the case of autarky, the FOC for his optimal choice  $\delta_j^{a*}$  in country  $j$  is

$$\frac{d\mathcal{V}_i(\delta_j^{a*})}{d\delta} \equiv q \left[ \xi(p_j^{a*})^{-(1-\beta)} \frac{\partial w(t_i)}{\partial \delta} - k_j \right] + (1-q) \frac{dU(\delta_j^{a*})}{d\delta} = 0.$$

With no difference in political institutions and with the second order condition  $\frac{d^2 \mathcal{V}_i(\delta_j^{a*})}{d\delta^2} < 0$ , the difference in the utility cost of centralization  $k_J < k_U$  implies that the optimal choice of curriculum centralization will be higher in Japan than in the US under autarky, i.e.,  $\delta_J^{a*} > \delta_U^{a*}$ . Hence, Japan will have a comparative advantage in cars, as  $\frac{\partial Y_A}{\partial \delta} > 0 > \frac{\partial Y_S}{\partial \delta}$  given  $1 < \gamma < \bar{\gamma}$ .

When the countries open to trade, with individuals taking trade price  $p$  as given, the FOC for the median voter's optimal choice  $\delta_j^{o*}$  in country  $j$  becomes

$$\frac{\partial \mathcal{V}_i(\delta_j^{o*}; p)}{\partial \delta} \equiv q \left[ \xi p^{-(1-\beta)} \frac{\partial w(t_i)}{\partial \delta} - k_j \right] + (1-q) \frac{\partial U(\delta_j^{o*}; p)}{\partial \delta} = 0.$$

Note that

$$\frac{\partial \mathcal{V}_i(\delta; p)}{\partial p} \equiv q \xi p^{-(1-\beta)} \left\{ -(1-\beta)p^{-1}w(t_i) + \frac{dw(t_i)}{dp} \right\} + (1-q) \frac{\partial U(\delta; p)}{\partial p}.$$

Since  $\frac{dw(t_i)}{dp} = 0$  for an auto worker, it follows that

$$(1) \quad \frac{\partial^2 \mathcal{V}_i(\delta; p)}{\partial \delta \partial p} = q\xi p^{-(1-\beta)} \left\{ -(1-\beta)p^{-1} \frac{\partial w(t_i)}{\partial \delta} \right\} + (1-q) \frac{\partial^2 U(\delta; p)}{\partial \delta \partial p} < 0,$$

where the inequality follows by  $\frac{\partial w(t_i)}{\partial \delta} > 0$  for an auto worker with  $t_{i0} < \min\{m(\hat{t}_0), \tilde{t}_0\}$ , and by  $\frac{\partial^2 U(\delta; p)}{\partial \delta \partial p} < 0$  as shown in the proof of Proposition 1 of the paper. This implies that

$$\frac{d\delta_j^{o*}}{dp} = \frac{\partial^2 \mathcal{V}_i(\delta_j^{o*}; p)}{\partial \delta \partial p} / \left( -\frac{\partial^2 \mathcal{V}_i(\delta_j^{o*}; p)}{\partial \delta^2} \right) < 0,$$

where  $\frac{\partial^2 \mathcal{V}_i(\delta_j^{o*}; p)}{\partial \delta^2} < 0$  holds by the SOC for  $\delta_j^{o*}$ .

Note that the FOCs under autarky and trade are identical if the trade price were the same as the autarky price. After trade, the equilibrium trade price  $p^f$  in fact falls in between the two countries' autarky prices,  $p_J^{a*} > p^f > p_U^{a*}$ . Thus, the median voter in the auto exporting country  $J$  facing a lower relative price for software will choose a higher  $\delta$  after trade. The opposite occurs in the software exporting country  $U$ . The desired result therefore follows.

If, instead, a low-ability software worker is the median voter, the term  $\frac{dw(t_i)}{dp}$  is not zero. Based on similar analysis as in Grossman and Maggi (2000, Section IV) but with a change in the numeraire good, it can be shown that  $\frac{dw(t_i)}{dp} = p^{-1} \left\{ w(t_i) + \frac{\lambda_A}{2} (m(\hat{t}) - \hat{t}) \right\} > 0$  for  $t_i < \hat{t}$ . Thus, for a low-ability software worker:

$$(2) \quad \begin{aligned} \frac{\partial^2 \mathcal{V}_i(\delta; p)}{\partial \delta \partial p} &= q\xi p^{-(1-\beta)} \left\{ -(1-\beta)p^{-1} \frac{\partial w(t_i)}{\partial \delta} + \frac{\partial^2 w(t_i)}{\partial \delta \partial p} \right\} + (1-q) \frac{\partial^2 U(\delta; p)}{\partial \delta \partial p}, \\ &= q\xi p^{-(2-\beta)} \left\{ \beta \frac{\partial w(t_i)}{\partial \delta} + \lambda_A (\ln \gamma) (\bar{t} - \hat{t}) \right\} + (1-q) \frac{\partial^2 U(\delta; p)}{\partial \delta \partial p}, \end{aligned}$$

where  $\frac{\partial^2 w(t_i)}{\partial \delta \partial p} = p^{-1} \left\{ \frac{\partial w(t_i)}{\partial \delta} + \lambda_A (\ln \gamma) (\bar{t} - \hat{t}) \right\} > 0$ , which more than offsets the first term. Thus, there remains a positive individual-specific effect ( $\frac{\partial^2 \mathcal{V}_i(\delta; p)}{\partial \delta \partial p} > 0$ ) in case talent is inherited, against a negative aggregate welfare consideration ( $\frac{\partial^2 U(\delta; p)}{\partial \delta \partial p} < 0$ ) in case talent is a random draw. Since  $\frac{\partial^2 \mathcal{V}_i(\delta; p)}{\partial \delta \partial p}$  in (2) increases in  $q$  and is negative when  $q = 0$ , there exists a critical threshold  $\tilde{q}_i > 0$  such that  $\frac{\partial^2 \mathcal{V}_i(\delta; p)}{\partial \delta \partial p}$  is equal to zero. Hence, the desired result holds if the probability  $q$  (that talent is inherited) is sufficiently small such that  $q < \tilde{q} \equiv \min_i \tilde{q}_i$ . ■

The above findings are not dependent on the choice of the numeraire good, as can be verified easily. In essence, among the majority of voters, each auto worker' interest is aligned with the social

planner's in terms of the tradeoff between  $\delta$  and  $p$ : the marginal gain in the individual-specific welfare given an increase in  $\delta$  is decreasing with the relative price of software ( $\frac{\partial^2 \nu_i(\delta; p)}{\partial \delta \partial p} < 0$ ), just as the aggregate welfare is ( $\frac{\partial^2 U(\delta; p)}{\partial \delta \partial p} < 0$ ). In contrast, the low-ability software worker's individual-specific interest works in the opposite direction of the average interest. Thus, the main result derived in the paper will continue to hold under majority voting for any  $q$  if the median voter is an auto worker. Alternatively, if the probability  $q$  that talent is inherited is sufficiently small, the result will hold regardless of the median voter's sector affiliation.

### 1.3. Divergent Educational Choices after Trade in Probabilistic Voting Model

As an extension, one may also consider the probabilistic voting model. In this approach, voters' preferences could be affected by non-economic factors. For example, voters may have ideological beliefs about the ideal education policy and such beliefs are randomly distributed in the population; alternatively, individuals may form special interest groups to engage in lobbying. It can be shown that the equilibrium choice under these circumstances is equivalent to maximizing a weighted sum of the indirect utilities of individuals, where the individual-specific weights reflect the effects of ideological beliefs or lobbying (Acemoglu and Robinson, 2006).

Specifically, let  $h_i$  denote the weight for individual  $i$ , which does not necessarily depend on an individual's initial ability  $t_{i0}$ . Then a generic objective function in a probabilistic voting model is

$$\begin{aligned} \max_{\delta} \sum_i h_i \mathcal{V}_i(\delta) &= \sum_i q h_i \nu_i(\delta) + (1 - q) U(\delta) \\ &= \sum_i q h_i \nu_i(\delta) + (1 - q) \sum_i \nu_i(\delta) \\ &= \sum_i (q h_i + 1 - q) \nu_i(\delta) \equiv \sum_i \eta_i \nu_i(\delta), \end{aligned}$$

where  $\eta_i \equiv q h_i + 1 - q$  is used to simplify the notation. Note that the median voter model would be a special case where  $h_i = 1$  for the median voter while  $h_i = 0$  for all others. In contrast, when  $h_i = 1$  holds for all individuals, it is exactly the social planner's objective function. We continue to assume the same political institution across countries, so that  $h_i$ 's are identical across countries for individuals with the same initial talent  $t_{i0}$ . The probabilistic voting approach allows for potentially

a very rich set of political scenarios, as the decision makers could be any (weighted) subset of the population. We consider some likely political scenarios in the following proposition.

**Proposition 2** *Under the probabilistic voting model, if the auto sector and/or the software sector is represented as a whole in the political process, i.e.,  $h_i = h_A \geq 0$  for all  $i \in A$  and  $h_i = h_S \geq 0$  for all  $i \in S$ , the educational choices will become more divergent across countries after trade. Alternatively if the individuals are not organized by sector, the result will still hold if any subset of the low-ability auto workers with  $t_{i0} \in [\hat{t}_0, \min\{m(\hat{t}_0), \tilde{t}_0\}]$  and the high-ability software workers with  $t_{i0} \in [m(\hat{t}_0), t_{h0}]$  are the dominant decision makers.*

**Proof.** The interest of a low-ability auto or software worker has been shown in Proposition 1. We now look at the interest of a high-ability auto or software worker.

For a high-ability auto worker with  $t_{i0} > \min\{m(\hat{t}_0), \tilde{t}_0\}$ , he has  $\frac{\partial w(t_i)}{\partial \delta} < 0$ ; thus, in (1), there is a conflict of individual-specific interest ( $\frac{\partial^2 \nu_i(\delta; p)}{\partial \delta \partial p} > 0$ ) with aggregate welfare consideration ( $\frac{\partial^2 U(\delta; p)}{\partial \delta \partial p} < 0$ ), just as in the case of low-ability software workers.

For a high-ability software worker, to determine his wage, note that for any pair of software workers,  $pF^S[t_i, m(t_i)] = w(t_i) + w(m(t_i))$  for  $t_i < \hat{t}$ . Taking the derivative of the equation with respect to  $p$ , we have  $F^S[t_i, m(t_i)] = \frac{dw(t_i)}{dp} + \frac{dw(m(t_i))}{dp}$ . Taking the derivative of the above equation further with respect to  $\delta$ , we have  $\frac{\partial F_i^S}{\partial \delta} = \frac{\partial^2 w(t_i)}{\partial \delta \partial p} + \frac{\partial^2 w(m(t_i))}{\partial \delta \partial p}$ . Since  $\frac{\partial F_i^S}{\partial \delta} < 0$  given  $\gamma < \bar{\gamma}$  and  $\frac{\partial^2 w(t_i)}{\partial \delta \partial p} > 0$  as shown in Proposition 1, it follows that  $\frac{\partial^2 w(m(t_i))}{\partial \delta \partial p} < 0$ . Recall also  $p \frac{\partial F_i^S}{\partial \delta} = \frac{\partial w(t_i)}{\partial \delta} + \frac{\partial w(m(t_i))}{\partial \delta} < 0$  and  $\frac{\partial w(t_i)}{\partial \delta} > 0$  as shown in Lemma 1; thus,  $\frac{\partial w(m(t_i))}{\partial \delta} < 0$ .

Thus, for a high-ability software worker,

$$\begin{aligned} (3) \quad \frac{\partial^2 \nu_i(\delta; p)}{\partial \delta \partial p} &= q\xi p^{-(1-\beta)} \left\{ -(1-\beta)p^{-1} \frac{\partial w(m(t_i))}{\partial \delta} + \frac{\partial^2 w(m(t_i))}{\partial \delta \partial p} \right\} + (1-q) \frac{\partial^2 U(\delta; p)}{\partial \delta \partial p}, \\ &= q\xi p^{-(2-\beta)} \left\{ \beta \frac{\partial w(m(t_i))}{\partial \delta} - \lambda_A (\ln \gamma) (\bar{t} - \hat{t}) \right\} + (1-q) \frac{\partial^2 U(\delta; p)}{\partial \delta \partial p} < 0, \end{aligned}$$

where the second equality follows by the facts that  $p \frac{\partial F_i^S}{\partial \delta} = \frac{\partial w(t_i)}{\partial \delta} + \frac{\partial w(m(t_i))}{\partial \delta}$  and  $\frac{\partial F_i^S}{\partial \delta} = \frac{\partial^2 w(t_i)}{\partial \delta \partial p} + \frac{\partial^2 w(m(t_i))}{\partial \delta \partial p}$ , and the results in (2). Thus, for a high-ability software worker, his individual-specific consideration is also aligned with the social planner's in terms of the tradeoff between  $\delta$  and  $p$ .



Thus, if any subset of the low-ability auto workers and the high-ability software workers are the dominant decision makers, the result of increased divergence in educational choices across countries after trade will continue to hold.

Alternatively, if either or both sectors as a whole are represented, e.g. by an interest group, the result will also hold. This is because the high-ability software workers' interests will dominate those of the low-ability software workers; similarly, the low-ability auto workers' interests will dominate in its sector. For example, by summing up (2) and (3) for all workers in the software sector, it follows that,

$$\frac{\partial^2 \mathcal{V}_S(\delta; p)}{\partial \delta \partial p} = q\xi p^{-(1-\beta)} \left\{ \beta \frac{\partial Y^S}{\partial \delta} \right\} + (1-q) \frac{\partial^2 U(\delta; p)}{\partial \delta \partial p} < 0;$$

similarly, for the auto sector as a whole,

$$\frac{\partial^2 \mathcal{V}_A(\delta; p)}{\partial \delta \partial p} = q\xi p^{-(1-\beta)} \left\{ -(1-\beta)p^{-1} \frac{\partial Y^A}{\partial \delta} \right\} + (1-q) \frac{\partial^2 U(\delta; p)}{\partial \delta \partial p} < 0,$$

given  $1 < \gamma < \bar{\gamma}$ . The result therefore follows. ■

In summary, when the education policy is made in a political economy model where individuals have conflict of interests, our main result still goes through in various scenarios. In particular, we show that when political coalition is organized at the industry level, for example, when individuals in each sector are represented collectively by a lobby group, then regardless of how the final decision is determined (whether it is dominated by the auto or the software sector or by a weighted combination of them), the education policies across countries will become more divergent after trade than under autarky. The key insight is that even though the two sectors have a conflict of interests in setting the absolute level of  $\delta$ , they are in agreement with each other in the direction of change in  $\delta$  after trade, that is, they all prefer a higher  $\delta$  when the relative price of software  $p$  is lower. Thus, the political decision makers in the auto exporting country facing a lower  $p$  after trade will want to adjust  $\delta$  upward, while the political decision makers in the software exporting country facing a higher  $p$  after trade will want to adjust  $\delta$  downward. The implications are thus the same as in the paper with a social planner.

## 2. Alternative Model Setups without Utility Cost of Centralization

### 2.1. Initial Difference as a Result of Multiple Optimal Choices in Autarky

Our paper's mechanism will go through so long as there is some initial difference in autarky education systems. Different utility (psychic) costs of centralization are one feasible rationale for such a difference across countries with identical economic structures.

Alternatively, two countries may be identical in all aspects but still choose different autarky education systems if there are multiple optimal education systems. This is possible as the outer contour of the PPFs as a function of  $\delta$  may not be concave (see Figure 1). Under certain combinations of preferences and production structures, it is possible for the national indifference curve to be tangent to two (or more) PPFs; thus, there could exist multiple optimal  $\delta$ 's. This scenario is illustrated in Figure 1, where  $\delta_1$  and  $\delta_3$  attain the same optimal welfare level under autarky;  $\delta_2$  and all the remaining possible  $\delta$ 's (not shown explicitly) correspond to PPFs that attain lower welfare.

If the two countries  $J$  and  $U$  happen to choose different systems under autarky ( $\delta_3$  by  $J$  and  $\delta_1$  by  $U$  as illustrated), then with trade, all the propositions in the paper will go through by exactly the same mechanism: Japan will face a lower relative price of software after trade than under autarky, which will prompt it to increase its  $\delta$  further, while the exactly opposite occurs to the US; the terms-of-trade consideration will dampen the incentive to diverge but will not eliminate it completely. With a world social planner, the same incentive exists to enlarge the cross-country difference in education systems.

The shortcoming with this framework is that the existence of multiple optimal education systems only implies the 'possibility' of increased divergence but does not predict it. It is equally likely that the two countries will choose the same system under autarky and remain in that state. In contrast, our paper's framework (with different psychic costs of centralization) predicts increased divergence as an inevitable outcome and not as a mere possibility.

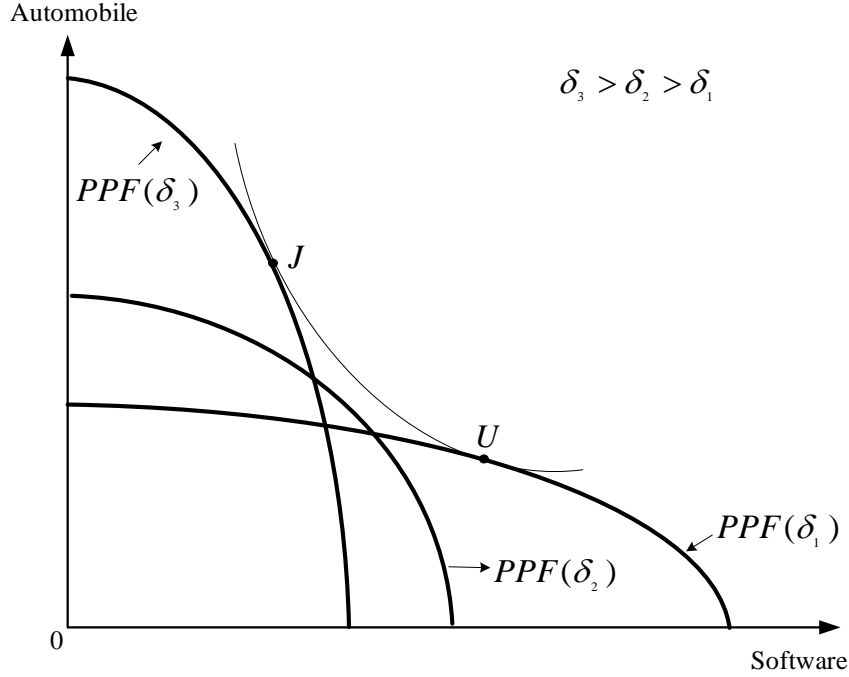


Figure 1: Multiple Optimal Education Systems in Autarky

## 2.2. Unstability of Symmetric Equilibria without Initial Differences

The discussions above focus on the scenario where there is an initial difference in autarky education systems across countries, as a result of either an exogenous difference in utility costs of centralization or multiple optimal education systems. The mechanism that drives the countries apart in education systems after trade rests upon the gain from specialization and trade.

If countries start with exactly the same education system, it is still possible to have diverging education systems after trade. In the cooperative setup where the world social planner chooses the optimal systems, increased divergence is optimal after trade if the surface of the world welfare function at the symmetric autarky equilibrium is concave in the direction of the  $45^\circ$  line where  $d\delta_J = d\delta_U$  (so that the chosen autarky system is optimal) but convex in the directions where  $(d\delta_J)(d\delta_U) \leq 0$  (so that divergence is preferable). In the noncooperative setup with each country choosing its optimal education system, increased divergence is possible if the symmetric equilibrium is unstable; in other words, the best response function of each country has a slope greater than

one (in absolute value) at the symmetric autarky equilibrium. We derive the general conditions for these scenarios below.

**World Optimal Choice.** Recall that the world welfare is

$$(4) \quad U_w(\delta_J, \delta_U) = \beta^\beta (1 - \beta)^{1-\beta} p^{-(1-\beta)} (Y_{AJ} + pY_{SJ} + Y_{AU} + pY_{SU}),$$

where we have assumed away the utility cost of centralization:  $k_J = k_U = 0$ , to focus on the scenario where countries are symmetric. It is straightforward to show that

$$(5) \quad d^2U_w = \frac{\partial^2 U_w}{\partial \delta_J^2} (d\delta_J)^2 + 2 \frac{\partial^2 U_w}{\partial \delta_J \partial \delta_U} (d\delta_J)(d\delta_U) + \frac{\partial^2 U_w}{\partial \delta_U^2} (d\delta_U)^2.$$

As discussed above, starting with a symmetric autarky equilibrium ( $\delta_J^a = \delta_U^a$ ), increased divergence is optimal after trade if at the symmetric autarky equilibrium, the surface of the world welfare function is:

- (i) concave in the direction of the 45° line where  $d\delta_J = d\delta_U$ : Along the 45° line, the two national education systems remain the same, so the two countries remain in autarky. In this case,  $U_w$  coincides with  $2U_J (= 2U_U)$ , which has a relative maximum at  $\delta_J^a (= \delta_U^a)$ . Thus, the world welfare function must be locally concave at  $(\delta_J^a, \delta_U^a)$  along the 45° line. Given (5), this implies that when  $d\delta_J = d\delta_U$ ,

$$(6) \quad \begin{aligned} d^2U_w|_{(\delta_J^a = \delta_U^a)} &= 2 \frac{\partial^2 U_w}{\partial \delta_J^2} (d\delta_J)^2 + 2 \frac{\partial^2 U_w}{\partial \delta_J \partial \delta_U} (d\delta_J)^2 < 0 \\ &\Rightarrow \frac{\partial^2 U_w}{\partial \delta_J^2} < -\frac{\partial^2 U_w}{\partial \delta_J \partial \delta_U}, \quad \text{at } \delta_J^a = \delta_U^a, \end{aligned}$$

where the first equality follows, because  $\frac{\partial^2 U_w}{\partial \delta_J^2} = \frac{\partial^2 U_w}{\partial \delta_U^2}$  holds at  $\delta_J^a = \delta_U^a$  since  $U_w$  is symmetric in  $(\delta_J, \delta_U)$ . Note that  $\frac{\partial^2 U_w}{\partial \delta_J \partial \delta_U} < 0$ .

- (ii) convex along opposite directions of change where  $(d\delta_J)(d\delta_U) \leq 0$ : This condition implies that divergence in  $(\delta_J, \delta_U)$  from the symmetric autarky equilibrium  $\delta_J^a = \delta_U^a$  increases the world welfare. Thus, the world optimal education systems will diverge across countries after trade.

Use (5) again. This condition implies that when  $(d\delta_J)(d\delta_U) \leq 0$  but not both zero,

$$\begin{aligned}
d^2U_w|_{(\delta_J^a=\delta_U^a)} &= \frac{\partial^2 U_w}{\partial \delta_J^2} (d\delta_J)^2 + 2\frac{\partial^2 U_w}{\partial \delta_J \partial \delta_U} (d\delta_J)(d\delta_U) + \frac{\partial^2 U_w}{\partial \delta_U^2} (d\delta_U)^2 > 0 \\
(7) \quad &\Rightarrow \frac{\partial^2 U_w}{\partial \delta_J^2} > 0, \text{ at } \delta_J^a = \delta_U^a.
\end{aligned}$$

To see this, note that there could only be four possible changes in the systems (on both sides of the 45° line) such that  $(d\delta_J)(d\delta_U) \leq 0$  but not both zero: (1)  $d\delta_J = 0$  and  $d\delta_U < 0$ : in this case, the condition implies that  $\frac{\partial^2 U_w}{\partial \delta_U^2} > 0$ ; (2)  $d\delta_J < 0$  and  $d\delta_U = 0$ : in this case, the condition implies that  $\frac{\partial^2 U_w}{\partial \delta_J^2} > 0$ ; (3)  $d\delta_J > 0$  and  $d\delta_U < 0$ : in this case, because  $\frac{\partial^2 U_w}{\partial \delta_J \partial \delta_U} < 0$ , the condition on the second derivatives is weaker (they could be negative); (4)  $d\delta_J < 0$  and  $d\delta_U > 0$ : this case is symmetric to the previous one with the same implication. Recall that  $\frac{\partial^2 U_w}{\partial \delta_J^2} = \frac{\partial^2 U_w}{\partial \delta_U^2}$  at  $\delta_J^a = \delta_U^a$ . Together, the condition that  $U_w$  be convex at  $(\delta_J^a, \delta_U^a)$  for any arbitrary changes such that  $(d\delta_J)(d\delta_U) \leq 0$  implies the result in (7).

In sum, starting with a symmetric autarky equilibrium ( $\delta_J^a = \delta_U^a$ ), increased divergence is socially optimal after trade for the world if

$$(8) \quad 0 < \frac{\partial^2 U_w}{\partial \delta_J^2} \left( = \frac{\partial^2 U_w}{\partial \delta_U^2} \right) < -\frac{\partial^2 U_w}{\partial \delta_J \partial \delta_U}, \text{ at } \delta_J^a = \delta_U^a.$$

Using (4), we can translate (8) into conditions on the production functions. Note that

$$(9) \quad \frac{\partial U_w}{\partial \delta_J} = \beta^\beta (1-\beta)^{1-\beta} (p^f)^{-(1-\beta)} \left( \frac{\partial Y_{AJ}}{\partial \delta_J} + p^f \frac{\partial Y_{SJ}}{\partial \delta_J} \right),$$

based on similar derivations as in the paper, except that  $k_J = 0$  here. Recall that the world welfare  $U_w$  coincides with the autarky welfare along the 45° line and has a relative maximum at  $(\delta_J^a, \delta_U^a)$ .

This implies that when  $d\delta_J = d\delta_U$ ,  $dU_w|_{(\delta_J^a=\delta_U^a)} = 0$  holds. This further implies that

$$\begin{aligned}
dU_w|_{(\delta_J^a=\delta_U^a)} &= \frac{\partial U_w}{\partial \delta_J} d\delta_J + \frac{\partial U_w}{\partial \delta_U} d\delta_U \\
(10) \quad &= 2\frac{\partial U_w}{\partial \delta_J} d\delta_J = 0
\end{aligned}$$

where the first equality is by definition, and the second equality follows because  $U_w$  is symmetric in  $(\delta_J, \delta_U)$  and we are evaluating  $dU_w$  along the direction where  $d\delta_J = d\delta_U$ .

Use the above FOC condition that  $\frac{\partial U_w}{\partial \delta_J} = 0$  at  $(\delta_J^a, \delta_U^a)$  and the competitive profit condition  $p = MRT$  (which implies that  $\frac{\partial Y_{AJ}}{\partial p^f} + p^f \frac{\partial Y_{SJ}}{\partial p^f} = 0$ ). We obtain

$$(11) \quad \frac{\partial^2 U_w}{\partial \delta_J^2} = \beta^\beta (1 - \beta)^{1-\beta} (p^f)^{-(1-\beta)} \left( \frac{\partial^2 Y_{AJ}}{\partial \delta_J^2} + p^f \frac{\partial^2 Y_{SJ}}{\partial \delta_J^2} + \frac{\partial Y_{SJ}}{\partial \delta_J} \frac{\partial p^f}{\partial \delta_J} \right), \text{ at } \delta_J^a = \delta_U^a.$$

Similarly, we have

$$(12) \quad \frac{\partial^2 U_w}{\partial \delta_J \partial \delta_U} = \beta^\beta (1 - \beta)^{1-\beta} (p^f)^{-(1-\beta)} \left( \frac{\partial Y_{SJ}}{\partial \delta_J} \frac{\partial p^f}{\partial \delta_U} \right), \text{ at } \delta_J^a = \delta_U^a.$$

Note  $\frac{\partial p^f}{\partial \delta_J} = \frac{\partial p^f}{\partial \delta_U}$  at  $\delta_J^a = \delta_U^a$ , where the two countries are symmetric. Together, given (11) and (12), the condition in (8) is equivalent to the following

$$(13) \quad -\frac{\partial Y_{SJ}}{\partial \delta_J} \frac{\partial p^f}{\partial \delta_J} < \frac{\partial^2 Y_{AJ}}{\partial \delta_J^2} + p^f \frac{\partial^2 Y_{SJ}}{\partial \delta_J^2} < -2 \frac{\partial Y_{SJ}}{\partial \delta_J} \frac{\partial p^f}{\partial \delta_J}, \text{ at } \delta_J^a = \delta_U^a.$$

Note that  $-\frac{\partial Y_{SJ}}{\partial \delta_J} \frac{\partial p^f}{\partial \delta_J} > 0$ . Thus, loosely speaking, starting with a symmetric autarky equilibrium ( $\delta_J^a = \delta_U^a$ ), increased divergence is socially optimal after trade for the world if the direct effect of education systems on national income through production is sufficiently convex, but not too convex. The findings in this section is summarized in the following.

**Summary 2 (World Socially Optimal Symmetry Breaking)** *Starting with a symmetric autarky equilibrium ( $\delta_J^a = \delta_U^a$ ), increased divergence is socially optimal after trade for the world if*

$$0 < \frac{\partial^2 U_w}{\partial \delta_J^2} \left( = \frac{\partial^2 U_w}{\partial \delta_U^2} \right) < -\frac{\partial^2 U_w}{\partial \delta_J \partial \delta_U}, \text{ at } \delta_J^a = \delta_U^a.$$

*Given no utility cost of centralization, this is equivalent to*

$$-\frac{\partial Y_{SJ}}{\partial \delta_J} \frac{\partial p^f}{\partial \delta_J} < \frac{\partial^2 Y_{AJ}}{\partial \delta_J^2} + p^f \frac{\partial^2 Y_{SJ}}{\partial \delta_J^2} < -2 \frac{\partial Y_{SJ}}{\partial \delta_J} \frac{\partial p^f}{\partial \delta_J}, \text{ at } \delta_J^a = \delta_U^a.$$

**Nash Equilibrium Choice.** Starting with a symmetric autarky equilibrium, divergence in education systems will arise after trade in a noncooperative setting, if the symmetric autarky equilibrium is unstable, or in other words, the best response function of each country has a slope greater than one (in absolute value) at the symmetric autarky equilibrium.

Since the two countries are symmetric, it suffices to study the condition for either one of the two countries, say, Japan. Recall that the national welfare for Japan, without utility cost of centralization, is

$$(14) \quad U_J(\delta_J; \delta_U) = \beta^\beta (1 - \beta)^{1-\beta} p^{-(1-\beta)} (Y_{AJ} + pY_{SJ}).$$

The best response function for Japan  $\delta_J^n(\delta_U)$  satisfies the condition that  $\frac{\partial U_J}{\partial \delta_J} = 0$ , from which we derive the slope of the best response function as

$$(15) \quad \frac{d\delta_J^n}{d\delta_U} = -\frac{\frac{\partial^2 U_J}{\partial \delta_J^n \partial \delta_U}}{\left(\frac{\partial^2 U_J}{(\partial \delta_J^n)^2}\right)}.$$

Since both the numerator and the denominator in (15) are negative (by the gains-from-trade argument and the second order condition for  $\delta_J^n$ , respectively), it follows that  $\left|\frac{d\delta_J^n}{d\delta_U}\right| > 1$  if and only if  $\frac{\partial^2 U_J}{(\partial \delta_J^n)^2} > \frac{\partial^2 U_J}{\partial \delta_J^n \partial \delta_U}$ . We translate this condition to conditions on the production functions below.

Recall from the paper, without utility cost of centralization,

$$(16) \quad \frac{\partial U_J(\delta_J; \delta_U)}{\partial \delta_J} = \beta^\beta (1 - \beta)^{1-\beta} (p^f)^{-(1-\beta)} \left( \frac{\partial Y_{AJ}}{\partial \delta_J} + p^f \frac{\partial Y_{SJ}}{\partial \delta_J} + T \frac{\partial p^f}{\partial \delta_J} \right).$$

where  $T \equiv -(1 - \beta)(p^f)^{-1} Y_{AJ} + \beta Y_{SJ} = \beta \frac{Y_{SJ} Y_{AU} - Y_{SU} Y_{AJ}}{Y_{AJ} + Y_{AU}} < 0$  is the negative terms-of-trade effect of raising  $\delta_J$ . To begin, note that  $(\delta_J^a, \delta_U^a)$  is a Nash equilibrium. There is no trade at this point, thus  $p^f = p^a(\delta_J^a)$  and  $T = 0$ ; since  $\delta_J^a$  satisfies the FOC for autarky, it also satisfies the FOC (16) for the best response  $\delta_J^n$ . Next, note that given (16), we can obtain

$$(17) \quad \begin{aligned} \frac{\partial^2 U_J}{\partial \delta_J^2} &= \beta^\beta (1 - \beta)^{1-\beta} (p^f)^{-(1-\beta)} \left( \frac{\partial^2 Y_{AJ}}{\partial \delta_J^2} + p^f \frac{\partial^2 Y_{SJ}}{\partial \delta_J^2} \right. \\ &\quad \left. + \frac{\partial Y_{SJ}}{\partial \delta_J} \frac{\partial p^f}{\partial \delta_J} + \frac{\partial T}{\partial \delta_J} \frac{\partial p^f}{\partial \delta_J} + \frac{\partial T}{\partial p^f} \left( \frac{\partial p^f}{\partial \delta_J} \right)^2 + T \frac{\partial^2 p^f}{\partial \delta_J^2} \right), \end{aligned}$$

where to obtain the result, we have used the competitive profit condition  $p = MRT$  and the FOC  $\frac{\partial U_J}{\partial \delta_J} = 0$ . Similarly, we can show that

$$(18) \quad \frac{\partial^2 U_J}{\partial \delta_J \partial \delta_U} = \beta^\beta (1 - \beta)^{1-\beta} (p^f)^{-(1-\beta)} \left( \frac{\partial Y_{SJ}}{\partial \delta_J} \frac{\partial p^f}{\partial \delta_U} + \frac{\partial T}{\partial p^f} \frac{\partial p^f}{\partial \delta_U} \frac{\partial p^f}{\partial \delta_J} + T \frac{\partial^2 p^f}{\partial \delta_J \partial \delta_U} \right).$$

Note that  $\frac{\partial p^f}{\partial \delta_J} = \frac{\partial p^f}{\partial \delta_U}$  and  $T = 0$  when evaluated at  $\delta_J^a = \delta_U^a$ , where countries are symmetric. Thus, we have

$$(19) \quad \left. \frac{\partial^2 U_J}{(\partial \delta_J^n)^2} - \frac{\partial^2 U_J}{\partial \delta_J^n \partial \delta_U} \right|_{(\delta_J^a = \delta_U^a)} = \beta^\beta (1 - \beta)^{1-\beta} (p^f)^{-(1-\beta)} \left( \frac{\partial^2 Y_{AJ}}{\partial \delta_J^2} + p^f \frac{\partial^2 Y_{SJ}}{\partial \delta_J^2} + \frac{\partial T}{\partial \delta_J} \frac{\partial p^f}{\partial \delta_J} \right).$$

Given the definition of  $T$ , we have  $\frac{\partial T}{\partial \delta_J} = -(1-\beta)(p^f)^{-1} \frac{\partial Y_{AJ}}{\partial \delta_J} + \beta \frac{\partial Y_{SJ}}{\partial \delta_J} < 0$  given that  $\frac{\partial Y_A}{\partial \delta} > 0 > \frac{\partial Y_S}{\partial \delta}$ .

Thus, the condition for the autarky symmetric equilibrium to be unstable is equivalent to

$$(20) \quad -\frac{\partial T}{\partial \delta_J} \frac{\partial p^f}{\partial \delta_J} < \frac{\partial^2 Y_{AJ}}{\partial \delta_J^2} + p^f \frac{\partial^2 Y_{SJ}}{\partial \delta_J^2}, \text{ at } \delta_J^a = \delta_U^a.$$

Compare (13) and (20); their lower bounds turn out to be the same. To see this, note that  $-\frac{\partial T}{\partial \delta_J} - (-\frac{\partial Y_{SJ}}{\partial \delta_J}) = (1-\beta)(p^f)^{-1} \left( \frac{\partial Y_{AJ}}{\partial \delta_J} + p^f \frac{\partial Y_{SJ}}{\partial \delta_J} \right) = 0$ , where the last equality follows by (9) and (10). Alternatively, it also follows from the FOC for  $\delta_J^n$  in (16) and the fact that  $T = 0$  at  $\delta_J^a = \delta_U^a$ .

Finally, given (17), the second order condition for  $\delta_J^n$  requires that  $\frac{\partial^2 U_J}{\partial \delta_J^2} < 0$  at  $\delta_J^n$ . As argued above,  $\frac{\partial T}{\partial \delta_J} = \frac{\partial Y_{SJ}}{\partial \delta_J}$  and  $T = 0$  at  $\delta_J^a = \delta_U^a$ . Note also that  $\frac{\partial T}{\partial p^f} = -(1-\beta)(p^f)^{-1} \frac{\partial Y_{AJ}}{\partial p^f} + \beta \frac{\partial Y_{SJ}}{\partial p^f} + (1-\beta)(p^f)^{-2} Y_{AJ} > 0$  given that  $\frac{\partial Y_A}{\partial p} < 0 < \frac{\partial Y_S}{\partial p}$ . Together, these suggest an upper bound smaller than in (13):

$$(21) \quad \frac{\partial^2 Y_{AJ}}{\partial \delta_J^2} + p^f \frac{\partial^2 Y_{SJ}}{\partial \delta_J^2} < -2 \frac{\partial T}{\partial \delta_J} \frac{\partial p^f}{\partial \delta_J} - \frac{\partial T}{\partial p^f} \left( \frac{\partial p^f}{\partial \delta_J} \right)^2, \text{ at } \delta_J^a = \delta_U^a.$$

**Summary 3 (Nash Equilibrium Symmetry Breaking)** *Starting with a symmetric autarky equilibrium ( $\delta_J^a = \delta_U^a$ ), increased divergence is a stable Nash equilibrium outcome after trade if*

$$-\frac{\partial Y_{SJ}}{\partial \delta_J} \frac{\partial p^f}{\partial \delta_J} < \frac{\partial^2 Y_{AJ}}{\partial \delta_J^2} + p^f \frac{\partial^2 Y_{SJ}}{\partial \delta_J^2} < -2 \frac{\partial Y_{SJ}}{\partial \delta_J} \frac{\partial p^f}{\partial \delta_J} - \frac{\partial T}{\partial p^f} \left( \frac{\partial p^f}{\partial \delta_J} \right)^2, \text{ at } \delta_J^a = \delta_U^a.$$

The final question is whether the lower bound is lower than the upper bound, so the range is not an empty set. We show that this is the case below. Recall the free trade equilibrium condition  $p^f = \frac{1-\beta}{\beta} \frac{Y_{AJ} + Y_{AU}}{Y_{SJ} + Y_{SU}}$ . Define  $V \equiv -(1-\beta)(p^f)^{-1}(Y_{AJ} + Y_{AU}) + \beta(Y_{SJ} + Y_{SU})$ . Given any change in  $\delta_J$ , this condition always holds, which implies that  $\frac{\partial V}{\partial \delta_J} + \frac{\partial V}{\partial p^f} \frac{\partial p^f}{\partial \delta_J} = 0$ . Recall that  $T \equiv -(1-\beta)(p^f)^{-1} Y_{AJ} + \beta Y_{SJ}$ . It is straightforward to show that

$$(22) \quad \begin{aligned} \frac{\partial T}{\partial \delta_J} + \frac{\partial T}{\partial p^f} \frac{\partial p^f}{\partial \delta_J} &= -(1-\beta)(p^f)^{-1} \frac{\partial Y_{AJ}}{\partial \delta_J} + \beta \frac{\partial Y_{SJ}}{\partial \delta_J} \\ &+ \left( -(1-\beta)(p^f)^{-1} \frac{\partial Y_{AJ}}{\partial p^f} + \beta \frac{\partial Y_{SJ}}{\partial p^f} + (1-\beta)(p^f)^{-2} Y_{AJ} \right) \frac{\partial p^f}{\partial \delta_J} \\ &= \frac{\partial V}{\partial \delta_J} + \frac{\partial V}{\partial p^f} \frac{\partial p^f}{\partial \delta_J} \\ &+ \left( (1-\beta)(p^f)^{-1} \frac{\partial Y_{AU}}{\partial p^f} - \beta \frac{\partial Y_{SU}}{\partial p^f} - (1-\beta)(p^f)^{-2} Y_{AU} \right) \frac{\partial p^f}{\partial \delta_J} < 0, \end{aligned}$$



where the inequality holds since  $\frac{\partial Y_A}{\partial p} < 0 < \frac{\partial Y_S}{\partial p}$  and  $\frac{\partial p^f}{\partial \delta_J} > 0$ . Note that in Summary 3, the difference of the upper and the lower bound is  $-\left(\frac{\partial Y_{SJ}}{\partial \delta_J} + \frac{\partial T}{\partial p^f} \frac{\partial p^f}{\partial \delta_J}\right) \frac{\partial p^f}{\partial \delta_J} = -\left(\frac{\partial T}{\partial \delta_J} + \frac{\partial T}{\partial p^f} \frac{\partial p^f}{\partial \delta_J}\right) \frac{\partial p^f}{\partial \delta_J} > 0$ , given (22). This completes the analysis.

To sum up this section, increased divergence in education systems after trade can still arise without any exogenous difference in country characteristics. In particular, it can arise in the scenario where initial autarky education systems differ as a result of multiple optimal systems; the same mechanism as in the paper of gains from trade is at work here in driving the result. Alternatively, in the scenario where countries start with exactly the same initial education system, increased divergence in education systems after trade is socially optimal in a cooperative setting or is a Nash equilibrium in a non-cooperative setting, if the direct effect of education systems on national income through production is sufficiently convex, but not too convex. It is worthwhile noting that the mechanism emphasized in the second scenario, if present, will reinforce the mechanism of gains from trade highlighted in the paper and drive further divergence in education systems than predicted in the paper.

### 3. The Case of Perfect Specialization

Given the paper's setup, there is a possibility of perfect specialization under trade because the slope of the PPF is bounded away from zero and infinity. Below, we analyze the scenario with perfect specialization. In short, our paper's main propositions are not affected by the possibility of perfect specialization, except that we need to modify the free-trade equilibrium condition and the FOCs of policy choice for the fact that a country may produce only one good, and that we need to allow weak inequalities in the main propositions in some cases.

#### 3.1. The Possibility of Perfect Specialization

Recall from the paper that the slope of the PPF is  $MRT = \frac{\lambda_A \bar{t}}{F^S(\hat{t}, 2\bar{t} - \hat{t})}$  for  $\hat{t} \in [t_l, \bar{t}]$ . If  $\hat{t} \rightarrow \bar{t}$ , all workers are engaged in the software sector and  $Y_A \rightarrow 0$ . The slope of the PPF in this scenario (at

the intercept of the PPF with the software axis) is:

$$(23) \quad MRT = \frac{\lambda_A \bar{t}}{F^S(\bar{t}, \bar{t})} = \frac{\lambda_A \bar{t}}{\lambda_S \bar{t}} = \frac{\lambda_A}{\lambda_S} \equiv p^h, \quad \text{where } \lambda_S \equiv F^S(1, 1).$$

Alternatively, if  $\hat{t} \rightarrow t_l$ , all workers are engaged in the auto sector and  $Y_S \rightarrow 0$ . The slope of the PPF in this scenario (at the intercept of the PPF with the auto axis) is:

$$(24) \quad MRT = \frac{\lambda_A \bar{t}}{F^S(t_l, t_h)} \equiv p^l(\delta) \rightarrow \begin{cases} p^h, & \text{as } \delta \rightarrow 1; \\ \frac{\lambda_A \bar{t}_0}{F^S(t_{l0}, t_{h0})} < p^h, & \text{as } \delta \rightarrow 0. \end{cases}$$

In (23) and (24), we have used the fact that the production technology of the software sector  $F^S(\cdot, \cdot)$  is of constant returns to talent (so that  $F^S(\bar{t}, \bar{t}) = \lambda_S \bar{t}$ ) and submodular (so that  $F^S(t_{l0}, t_{h0}) > F^S(\bar{t}_0, \bar{t}_0)$ ).

Thus, there is incomplete specialization for  $p \in (p^l(\delta), p^h)$ . If  $p \geq p^h$ , a country completely specializes in software. On the other hand, if  $p \leq p^l(\delta)$ , a country completely specializes in auto. In the extreme case where  $\delta = 1$ , all workers have the same talent level and the PPF becomes a straight line with a slope equal to  $p^h$ ; in this case, there is incomplete specialization only if  $p = p^h$ .

### 3.2. The Results with Perfect Specialization

Let's fix the scenario as in the paper where under autarky, Japan chooses a higher  $\delta$  than the US ( $\delta_J^a > \delta_U^a$ ). In autarky, perfect specialization is not possible given that preferences are Cobb-Douglas. Thus, autarky prices must be such that  $p^l(\delta) < p^a < p^h$  for  $\delta < 1$  and  $p^a = p^h$  for  $\delta = 1$ . With trade, three scenarios of perfect specialization are possible:

- (i) the US completely specializes in  $S$  and Japan completely specializes in  $A$ : the former implies that the trade price must be such that  $p \geq p^h$  and the latter implies that  $p \leq p^l(\delta_J) \leq p^h$ . Together, this implies that  $p = p^h$  and  $\delta_J = 1$  (If  $\delta_J \neq 1$ , the PPF is not a straight line. Then given  $p = p^h$ , Japan will also completely specialize in  $S$  contradicting the scenario).
- (ii) the US completely specializes in  $S$  and Japan produces both goods: the former implies that  $p \geq p^h$  and the latter implies that  $p \leq p^h$ . Together, this implies that  $p = p^h$  and  $\delta_J = 1$ .

(iii) the US produces both goods and Japan completely specializes in  $A$ : the former implies that  $p^l(\delta_U) < p < p^h$  and the latter implies that  $p \leq p^l(\delta_J)$ . Together, this implies that  $p^l(\delta_U) < p \leq p^l(\delta_J)$ .

We can then check how the FOCs and propositions derived in the paper under small open economy, world optimal choice, and Nash equilibrium are affected by the possibility of perfect specialization.

**Small Open Economy.** In scenarios (i) and (ii),  $p = p^h$  and  $\delta_J = 1$  under trade. Given  $\delta_J = 1$ , the PPF of Japan is a straight line with  $MRT = p^h$ . It follows that the competitive profit condition  $p = MRT$  still holds for Japan and so does Proposition 1 (with ‘=’ in (26) of the paper replaced by ‘ $\geq$ ’ to accommodate a corner solution in  $\delta_J$ ). On the other hand, the US completely specializes in  $S$ , its FOC taking the trade price as given becomes

$$(25) \quad \frac{\partial U(\delta; p)}{\partial \delta} = \beta^\beta (1 - \beta)^{1-\beta} p^{-(1-\beta)} \left( p \frac{\partial Y_S}{\partial \delta} \right) - k < 0,$$

since  $\frac{\partial Y_S}{\partial \delta} < 0$ . This implies that, with trade, the US will choose  $\delta_U = 0$  regardless of  $p$ . Thus, Proposition 1 holds (weakly) for the US.

In scenario (iii), the proof for Proposition 1 is still valid for the US, as it produces both goods. The FOC for Japan becomes

$$(26) \quad \frac{\partial U(\delta; p)}{\partial \delta} = \beta^\beta (1 - \beta)^{1-\beta} p^{-(1-\beta)} \left( \frac{\partial Y_A}{\partial \delta} \right) - k = 0,$$

and

$$\begin{aligned} \frac{\partial U(\delta; p)}{\partial p} &= \beta^\beta (1 - \beta)^{1-\beta} p^{-(1-\beta)} \left[ \frac{\partial Y_A}{\partial p} - (1 - \beta) p^{-1} Y_A \right] \\ &= \beta^\beta (1 - \beta)^{1-\beta} p^{-(1-\beta)} [-(1 - \beta) p^{-1} Y_A], \end{aligned}$$

where the second equality obtains because  $\frac{\partial Y_A}{\partial p} = 0$  since Japan has already completely specialized in  $A$ . Based on the above condition, we get

$$\frac{\partial^2 U(\delta; p)}{\partial \delta \partial p} = \beta^\beta (1 - \beta)^{1-\beta} p^{-(1-\beta)} \left[ -(1 - \beta) p^{-1} \frac{\partial Y_A}{\partial \delta} \right] < 0,$$

since  $\frac{\partial Y_A}{\partial \delta} > 0$ . This implies  $\frac{d\delta^o}{dp} = \frac{\partial^2 U(\delta^o; p)}{\partial \delta \partial p} / \left( -\frac{\partial^2 U(\delta^o; p)}{\partial \delta^2} \right) < 0$ . Thus, Proposition 1 continues to hold.

**World Optimal Choice.** In scenarios (i) and (ii), the US completely specializes in  $S$ . The world social planner's FOC for  $\delta_U$  becomes:

$$\begin{aligned}
\frac{\partial U_w(\delta_J, \delta_U)}{\partial \delta_U} &= \beta^\beta (1-\beta)^{1-\beta} (p^f)^{-(1-\beta)} \left[ p^f \frac{\partial Y_{SU}}{\partial \delta_U} \right. \\
&\quad + \left( \frac{\partial Y_{AJ}}{\partial p^f} + p^f \frac{\partial Y_{SJ}}{\partial p^f} + Y_{SJ} + p^f \frac{\partial Y_{SU}}{\partial p^f} + Y_{SU} \right) \frac{\partial p^f}{\partial \delta_U} \\
&\quad \left. - (1-\beta)(p^f)^{-1} (Y_{AJ} + p^f Y_{SJ} + p^f Y_{SU}) \frac{\partial p^f}{\partial \delta_U} \right] - k_U \\
(27) \qquad \qquad \qquad &= \beta^\beta (1-\beta)^{1-\beta} (p^f)^{-(1-\beta)} \left( p^f \frac{\partial Y_{SU}}{\partial \delta_U} \right) - k_U < 0,
\end{aligned}$$

where the last equality follows by the facts that: (1)  $\frac{\partial Y_{AJ}}{\partial p^f} + p^f \frac{\partial Y_{SJ}}{\partial p^f} = 0$  because  $p = p^h = MRT$  in Japan, (2)  $\frac{\partial Y_{SU}}{\partial p^f} = 0$  because the US has already completely specialized in  $S$ , and (3) the free-trade equilibrium condition  $p^f = \frac{1-\beta}{\beta} \frac{Y_{AJ}}{Y_{SJ}+Y_{SU}}$  holds. Thus, it follows that  $\delta_U^w = 0$ . Recall that  $\delta_J = 1$  in these two scenarios (with trade). Since under autarky,  $\delta_J^a > \delta_U^a$ , while with trade,  $\delta_J = 1$  and  $\delta_U = 0$ , the socially optimal education systems in the two countries must have weakly diverged after trade compared to autarky, so Proposition 2 remains valid (possibly weakly).

In scenario (iii), the US produces both goods and Japan completely specializes in  $A$ . Note that the appropriate conditions in this case are: (1)  $\frac{\partial Y_{AU}}{\partial p^f} + p^f \frac{\partial Y_{SU}}{\partial p^f} = 0$  by the competitive profit condition  $p = MRT$ , which holds because the US produces both goods, (2)  $\frac{\partial Y_{AJ}}{\partial p^f} = 0$  because Japan has already completely specialized in  $A$ , and (3) the free-trade equilibrium condition  $p^f = \frac{1-\beta}{\beta} \frac{Y_{AJ}+Y_{AU}}{Y_{SU}}$ . The world social planner's FOC for  $\delta_U$  turns out to be the same as in the paper, basically since the US still produces both goods. The world social planner's FOC for  $\delta_J$  is:

$$(28) \qquad \qquad \qquad \frac{\partial U_w(\delta_J, \delta_U)}{\partial \delta_J} = \beta^\beta (1-\beta)^{1-\beta} (p^f)^{-(1-\beta)} \left( \frac{\partial Y_{AJ}}{\partial \delta_J} \right) - k_J = 0,$$

from which it follows that

$$(29) \qquad \qquad \qquad \frac{\partial^2 U_w}{\partial \delta_J \partial \delta_U} = \beta^\beta (1-\beta)^{1-\beta} (p^f)^{-(1-\beta)} \left( -(1-\beta)(p^f)^{-1} \frac{\partial Y_{AJ}}{\partial \delta_J} \frac{\partial p^f}{\partial \delta_U} \right) < 0.$$

Thus, it still holds that  $\frac{d\delta_J^w(\delta_U)}{d\delta_U} < 0$  and  $\frac{d\delta_U^w(\delta_J)}{d\delta_J} < 0$  in this case of perfect specialization, so Proposition 2 remains valid.

**Nash Equilibrium Choice.** Recall again in scenarios (i) and (ii),  $p = p^h$  and  $\delta_J = 1$ ; the US completely specializes in  $S$  and Japan may or may not specialize completely in  $A$ . The FOC for a best response  $\delta_U^n(\delta_J)$  given  $\delta_J$  becomes

$$(30) \quad \frac{\partial U_U(\delta_U; \delta_J)}{\partial \delta_U} = \beta^\beta (1 - \beta)^{1-\beta} (p^f)^{-(1-\beta)} \left[ p^f \frac{\partial Y_{SU}}{\partial \delta_U} + \beta Y_{SU} \frac{\partial p^f}{\partial \delta_U} \right] - k_U.$$

Compare (27) with (30). Note that  $\frac{\partial U_U(\delta_U; \delta_J)}{\partial \delta_U} > \frac{\partial U_w(\delta_J, \delta_U)}{\partial \delta_U}$ , which implies that  $\delta_U^n(\delta_J) > \delta_U^w(\delta_J)$ . Similarly, it can be easily verified that Japan's FOCs for the world optimal choice and the Nash equilibrium choice differ by a negative terms-of-trade component:  $-\beta Y_{SU} \frac{\partial p^f}{\partial \delta_J}$ , whether Japan completely specializes in  $A$  or not. Thus, it still follows that  $\delta_J^n(\delta_U) < \delta_J^w(\delta_U)$  if the solution is interior. If, in the Nash equilibrium,  $\delta_J$  has hit the upper bound as stipulated in this scenario, it implies that  $\delta_J^w = 1$  as well. Thus, Proposition 3 still holds but weakly for Japan:  $1 = \delta_J^w = \delta_J^n \geq \delta_J^a > \delta_U^a > \delta_U^n > \delta_U^w$ .

In scenario (iii), recall that Japan completely specializes in  $A$  and the US produces both goods. This again implies that: (1)  $\frac{\partial Y_{AU}}{\partial p^f} + p^f \frac{\partial Y_{SU}}{\partial p^f} = 0$  (2)  $\frac{\partial Y_{AJ}}{\partial p^f} = 0$  and (3)  $p^f = \frac{1-\beta}{\beta} \frac{Y_{AJ} + Y_{AU}}{Y_{SU}}$ , based on which we can derive the FOC for Japan's best response as:

$$(31) \quad \frac{\partial U_J(\delta_J; \delta_U)}{\partial \delta_J} = \beta^\beta (1 - \beta)^{1-\beta} (p^f)^{-(1-\beta)} \left( \frac{\partial Y_{AJ}}{\partial \delta_J} - \beta \frac{Y_{SU} Y_{AJ}}{Y_{AJ} + Y_{AU}} \frac{\partial p^f}{\partial \delta_J} \right) - k_J = 0,$$

which has an extra negative terms-of-trade component compared to (28). Similarly, the FOC for the US's best response differs from that of the world optimal choice by a positive terms-of-trade component  $\beta \frac{Y_{SU} Y_{AJ}}{Y_{AJ} + Y_{AU}} \frac{\partial p^f}{\partial \delta_U}$ . Thus, Proposition 3 remains valid.

## 4. The n-Country Case

Our theory predicts a wider divergence in education policies in a world with two countries, if they open to trade with each other. Below, we verify that it is possible to extend this general prediction to a setting with many countries.

With only two goods, the trade pattern is not unique if there are more than two countries. For example, suppose there are three countries and let  $M_{Ai}$  be the net imports of auto by country  $i$  for  $i = 1, 2, 3$ . Similarly, let  $M_{Si}$  be the net imports of software by country  $i$ . A negative value of  $M$

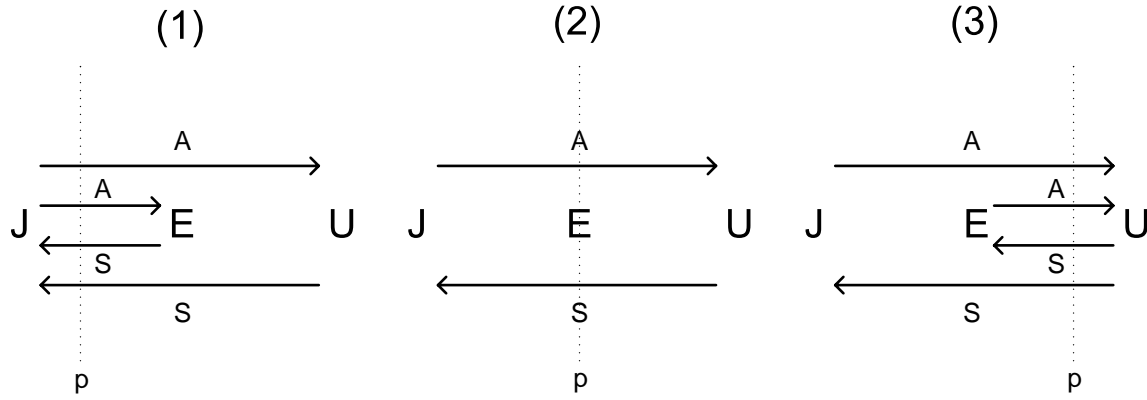


Figure 2: Trade pattern with 3 countries

corresponds to positive exports. The market clearing conditions require that:

$$(32) \quad M_{A1} + M_{A2} + M_{A3} = 0$$

$$(33) \quad M_{S1} + M_{S2} + M_{S3} = 0.$$

The trade-balance conditions require that  $M_{A_i} + pM_{S_i} = 0$  for  $i = 1, 2, 3$ , where  $p$  as in the paper is the relative price of software. Thus, if there are  $n$  countries and 2 goods, there are  $n + 2$  conditions but  $2n$  trade-flow variables, so the trade pattern is not unique for  $n > 2$ .

However, if we introduce some trade friction such that a country does not re-export what it imports, it is possible to pin down the trade pattern among a set of  $n$  countries to some extent.

For example, suppose there are 3 countries, Japan, Europe, and the US, with autarky education systems such that  $\delta_J > \delta_E > \delta_U$  (which implies that  $p_J^a > p_E^a > p_U^a$ ). Given trade friction, a country will either import or export a good (but not simultaneously), and must export one good in exchange for imports of the other good (of the same value). Consider Japan. Since it has the strongest comparative advantage in auto, it will export auto and import software. By the same token, the US will export software and import auto. The EU will then either: (1) export software and import auto with Japan; (2) remain in autarky; or (3) export auto and import software with the US. It will not simultaneously do (1) and (3) because of trade cost. The potential trade patterns are illustrated in Figure 2.

Thus, the middle country turns out to trade (at most) with only one partner country, and the overall multilateral trade flow can be decomposed into two bilateral trade flows, where the ranking of autarky education systems between two countries in any pair and the change in the relative price after trade relative to autarky satisfy our propositions' setting. For example, if scenario (1) occurs, it implies that the trade price lies in between the autarky prices of Japan and the EU. Considering this pair of countries alone, it holds that  $\delta_J^a > \delta_E^a$  and  $p_J^a > p > p_E^a$ . Thus, the conditions for the mechanisms to work in our propositions are met. Similarly, for the bilateral trade flows between Japan and the US, it is also true that  $\delta_J^a > \delta_U^a$  and  $p_J^a > p > p_U^a$ ; thus, similar mechanisms will also work. In other words, for each pair of trading countries, the gains from specialization and trade will tend to amplify the autarky difference in their education systems.

Similar results can be obtained for a setting with more than 3 countries, by using the same conditions as above (that each country will either import or export a good, and trade is balanced for each country). We can rule out the possibility where two countries with  $\delta_i^a > \delta_j^a$  trade with each other when  $p > p_i^a > p_j^a$  or  $p_i^a > p_j^a > p$ . Thus, the multilateral trade flows can be decomposed into a series of bilateral trade flows, where in each bilateral relationship  $ij$  with positive trade,  $p_i^a > p > p_j^a$  holds if  $\delta_i^a > \delta_j^a$  for  $i \neq j \in \{1, 2, \dots, n\}$ . Thus, the mechanisms underlying the paper's propositions would still work to enlarge bilateral divergence in education systems.

## References

- ACEMOGLU, D. AND J. A. ROBINSON, *Economic Origins of Dictatorship and Democracy* (New York: Cambridge University Press, 2006).
- GROSSMAN, G. M. AND G. MAGGI, "Diversity and Trade," *American Economic Review* 90 (2000), 1255–1275.