International Trade, Factor Mobility and the Persistence of Cultural-Institutional Diversity[§]

Marianna Belloc[*] and Samuel Bowles[**]

26 April, 2010

Abstract

We present a model in which specialization and trade occur not as a result of exogenous differences in factor endowments or technologies, but because of endogenous differences in culture (preferences including social norms) and institutions (contracts). Goods differ in the kinds of contracts that are appropriate for their production, and so strategic complementarities between contracts and the nature of social norms may result in a multiplicity of cultural-institutional equilibria that provide the basis for comparative advantage and specialization. In our evolutionary model of endogenous preferences and institutions under autarchy, trade and factor mobility, transitions among multiple asymptotically stable cultural-institutional conventions may occur as a result of decentralized and un-coordinated contractual or behavioral innovations by employers or employees. We show that: *i*) specialization and trade may arise and enhance welfare even when the countries are identical other than their cultural-institutional conventions; *ii*) trade liberalization does not lead to convergence, it reinforces the cultural-institutional differences upon which comparative advantage is based and may thus impede even Pareto-improving cultural-institutional transitions; and *iii*) by contrast, greater mobility of factors of production favors decentralized transitions to a superior cultural-institutional convention by reducing the minimum number of cultural or institutional innovators necessary to induce a transition as well as the cost of innovating.

JEL CODES: C73, D23, F15, F16

KEYWORDS: institutions, partnership, endogenous preferences, evolutionary game theory, culture, trade integration, factor mobility, globalization

---

[§] We would like to thank [to be completed] for helpful comments. Thanks also to the Behavioral Science Program of the Santa Fe Institute and the U.S. National Science Foundation for support of this project.

[*] Sapienza University of Rome (marianna.belloc@uniroma1.it).

[**] Santa Fe Institute and University of Siena (samuel.bowles@gmail.com).

1. Introduction

Among history's great puzzles are the many instances of centuries-long persistence of institutional and cultural differences between populations, often enduring long after their initial causes have disappeared. Institutions and elite cultures that owed their origin to the 16th century exploitation of slaves and coerced Native American labor persisted long after sugar and gold had lost their central role in the Latin American economies (Sokoloff and Engerman, 2000). Current levels of distrust in distinct African populations vary inversely with the exposure to the slave trade that ended two centuries ago (Nunn and Wantchekon, 2010). Differing levels of cooperation and civic values among Italian urban areas appear to be the legacy of autonomous city-state institutions or their absence half a millennium earlier (Guiso, Sapienza, and Zingales, 2009). The effects of the differing tax and land tenure systems imposed by the British Raj in the 18th and 19th century persisted in post-independence India (Banerjee and Iyer, 2005).

In epochs and social orders marked by limited contact and restricted competition among geographically separated areas, persistent cultural and institutional differences are hardly surprising. But this is not the case in a globally integrated world economy. In this paper we explain how the decentralized updating of both preferences and contractual choices can support durable cultural and institutional differences that may provide a basis for specialization, comparative advantage, and hence trade, which in turn stabilizes the cultural and institutional differences. Our explanation hinges on the endogenous codetermination of institutions, cultures, and economic specialization, a nexus long-studied by economists with a historical bent (Gerschenkron, 1944; Greif and Tabellini, 2010; Kindleberger, 1962; Sokoloff and Engerman, 2000), but not heretofore formally modeled.

We refer to differences across economies in the distribution of employment contracts as institutional differences, while between-economy variations in the distribution of preferences (including social norms) are termed cultural differences. We thus develop a two-country/two-contract/two-preference/two-good model in which countries may differ in their institutions and cultures. Production and distribution are governed by employers' choice between two contracts, either joint residual claimancy under share contracts (partnerships) or forcing contracts with the employer as residual claimant. The relevant preference differences are captured by assuming that employees are either reciprocal or self-regarding. Finally, goods differ in the extent to which their production depends on qualitative labor, namely that which is prohibitively costly to verify and hence cannot be cost-effectively secured by a forcing contract requiring an explicit labor input. Where non-verifiable aspects of work are important to production, social norms such as reciprocity or a positive work ethic may be required for high levels of productivity.

The main novelty of our approach is that, rather than treating institutions and preferences as exogenous or determined by a national-level constitutional bargain, we use evolutionary game theory to model the interacting dynamics of both as the result of decentralized non-cooperative interactions among economic agents. Like Greif (1994), Guiso, Sapienza and Zingales (2009), Tabellini (2008) and Spolaore and Wacziarg (2009), we study the economic importance of cultural differences. Unlike all above papers but in common with Bisin and Verdier (2001), Bowles (1998), Fershtman and Bar-Gill (2005), Galor and Maov (2002), and Doepke and Zilibotti (2008), we model cultural evolution.

In our model, the choice of contract that maximizes employers' profits depends on the preferences which prevail in a given country, so firms face a problem of matching contracts and preferences as in Prendergast (2008). Partnership contracts, for example, are more profitable where social preferences like the work ethic or reciprocity are common. The distribution of preferences in turn is based on a cultural updating process in which the payoffs associated with different preferences (and the behaviors they support) depend on the distribution of contracts in the economy. It is this mutual dependence of preferences and contracts and the differences among goods in the importance of non-verifiable qualitative labor that supports the multiplicity of equilibria and provides the basis for national specialization in our model, thus playing a role analogous to technology-based economies of scale in Paul Krugman's (1987) model of trade among countries with identical factor endowments and technologies. Transitions may occur among these cultural-institutional conventions when sufficiently many innovators deviate from the status quo convention (adopting non-best response preferences or contracts) due to individual experimentation and other forms of idiosyncratic play. We derive three key results.

First, for historical reasons two otherwise identical countries may experience different cultural-institutional conventions, and these cross-country differences in the institutional and cultural environment, like differences in technologies in the Ricardian approach or factor endowments in the standard Heckscher-Ohlin model, are an independent source of comparative advantage.

Second, economic integration reinforces rather than destabilizes institutional and cultural diversity and may impede transitions, even to Pareto-improving conventions. This result contradicts the view, popular among critics of trade liberalization since John Maynard Keynes (1933), that trade will lead to institutional and cultural convergence and thus defeat attempts by nations that, as he put it (p. 762), would prefer to "have a try at working out our own salvation". This is especially thought to be true when one nation's cultural-institutional equilibrium confers absolute advantage in both products. But since trade allows countries to specialize in the goods that are relatively more advantaged given their institutions and preferences, it increases the joint surplus in the cultural-institutional status quo even in the absolutely disadvantaged country. These gains from trade increase the joint surplus available to employers and employees and, hence, the cost of a mismatch that is likely to occur as the result of deviations from the prevalent preferences and contracts. By making experimentation more costly the gains from trade thus increase the impediments to cultural-institutional transitions. Trade may also increase the number of preference or contractual innovators required to induce a transition to the superior convention. Thus, in an open-economy setting a nation's cultural-institutional convention may persist over very long periods, even when a Pareto-superior convention exists and when the status quo convention confers absolute disadvantage with respect to other countries in all goods. The source of persistent inefficiency in this model is the coordination failure arising from the decentralized nature of preference formation and contractual choice.

Third, in contrast to trade, factor market integration facilitates convergence to superior culture and institutions. The reason is that factor mobility provides a kind of "innovation insurance" as it lowers the expected costs of deviating from the status quo; it also reduces the minimum number of innovators necessary to induce Pareto-improving cultural-institutional transitions. Factor market integration thus reduces both the size and (loosely speaking) the depth of the basin of attraction of the inferior equilibrium.

We begin with the basic assumptions of our model and the empirical evidence motivating them (section 2). We then develop a model of endogenous preferences and contractual choice, extending the standard 2x2 model of international exchange to a 2x2x2x2 model, and illustrate cultural-institutional comparative advantage (section 3). In section 4 we introduce the model's dynamics and show that multiple asymptotically stable cultural-institutional equilibria may exist. We then explore the persistence of cultural and institutional differences following trade integration (section 5), and factor mobility (section 6). Section 7 discusses related literature and concludes.

2. Goods, preferences and contracts

An economy is populated by employers and employees. Employers hire employees to produce one of two goods, the employment relationship being a random employee-employer match for a single interaction in which the employer offers a contract under which the employee works. Labor is perfectly mobile across industries but (until section 6) immobile across countries. Our model is based on four distinctive assumptions that we believe are of broad empirical relevance.

First, there are two aspects of labor. Qua*N*titative labor (denoted by the subscript N) includes time at work, compliance with directions, simple effort readily measured either by input or output, and other aspects of work that are readily observable, either directly or that may be inferred from the associated outputs. Because it is readily observable, quantitative labor is cost-effectively verifiable and can be enforced by contracts. By contrast, qua*L*itative work (denoted by the subscript L) consists of care, creativity, problem solving and other non-routine aspects of work that are more difficult to verify, and hence not cost-effectively subject to explicit contracts conditional on individual performance. Production of all goods requires quantitative labor and is also enhanced by qualitative labor (though, as we will see, in differing degree). Each employee may provide either quantitative labor alone or both quantitative and qualitative labor.

Second, there are two goods. One is intensive in quantitative labor and termed transparent (the *t*-good) because the labor activities that are readily observed are relatively more important in its production. The production of the opaque good (*o*-good), by contrast, depends more intensively on qualitative aspects of work. Examples of the latter are knowledge-intensive goods (and services), complex and quality-variable manufactured goods, personal services ranging from legal advice to preparing meals, and care-sensitive agricultural products (such as tobacco, many vegetables and fruits) and wine. For these goods the necessary labor inputs cannot be verified because they are not directly observable and cannot be indirectly inferred from the resulting output. Transparent goods include standardized manufactured goods (exemplified by most good produced on an assembly line and any good the production of which is cost effectively compensated by piece rates), most grains and sugar.

Hence, denoting by Q_L^i the quantity of good i ($i = o,t$) obtained using one unit of both qualitative and quantitative labor, and by Q_N^i the output obtained with a single unit of quantitative labor, we have:

$$\frac{Q_L^o}{Q_N^o} > \frac{Q_L^t}{Q_N^t}, \qquad (1)$$

that is, the increase in production obtained employing both quantitative and qualitative labor rather than quantitative labor only is relatively greater in the opaque than in the transparent sector.

Our third assumption is that some employees have preferences over the form of the contract under which they work *per se*, that is, in addition to the material payoffs. For some individuals, close supervision and threats of sanctions for non-compliance signal distrust or otherwise offend reciprocal or other social preferences essential to mutually beneficial exchange. This is found in a large number of natural environments (Bewley, 1999) and experimental studies (Fehr, Klein, and Schmidt, 2007; Falk and Kosfeld, 2006; surveyed in Bowles, 2008, and Bowles and Polania, 2009). We simplify by assuming just two kinds of employees. Those who we term Reciprocators (denoted by the superscript *R*) who care about the form of the contract *per se*: in a dyadic interaction their utility is increasing in their own payoffs and may be either increasing or decreasing in the payoffs of the employer depending on the individual's belief about the type of the other, in the spirit of Rabin (1993), Levine (1998) and Fehr and Falk (2002).Thus, the utility of employee *h* who is matched with employer *k* depends on his own material payoff (π_h), including the disutility of labor, and on the payoff of employer *k* (π_k):

$$U_{hk} = \pi_h + \alpha_h \gamma_{hk} \pi_k, \tag{2}$$

where α_h (>0 for Reciprocator and =0 otherwise) is the strength of *h*'s reciprocity preferences and γ_{hk} (= −1, 1) is *h*'s belief about *k*'s type, the latter depending on the form of contract that *k* offers *h*. In the model below a Partnership (denoted by the superscript *P*), in which the employer and the employee are joint residual claimants on the firm's output and the employee is free to choose whether to supply both qualitative and quantitative labor, or only quantitative labor, signals the good will and trust of the employer, leading to γ_{hk}=1; while the employer's close surveillance and the threat of termination under a Forcing contract (superscript *F*), signals distrust with γ_{hk}= −1 as a result. Consequently, there may be a mismatch between the firm's contractual structure and the employees' preferences.

Other employees, who we will term Homo economicus (superscript *E*), care only about their own material payoffs ($\alpha_h = 0$) so that $U_{hk} = \pi_h$ for any *k*. We refer to preferences of this kind as self-regarding. As we will see in section 3, from this it follows that social preferences such as a strong work ethic, truth telling and intrinsic motivation may be essential to the production of opaque goods, because Forcing contracts appealing to conventional self-regarding motives cannot elicit qualitative labor due to the lack of verifiability of this input.

The final assumption is that while both culture and institutions are endogenous, neither is the result of instantaneous individual maximization or collective choice. Rather both are durable characteristics of individuals and organizations that evolve in a decentralized environment under the influence of long-run society-wide payoff differences. Institutions and preferences are acquired and abandoned by a trial and error process often taken place at critical times, the birth of a firm, for example, for contractual forms, or early childhood or adolescence for preference formation. Because childhood socialization and the other processes by which preferences are acquired take place under the influence of religious instruction, schooling and other effects operating at the national level, we represent this process of cultural evolution by a society-wide dynamics operating prior to economic matching for production. Thus individuals do not condition the updating of their preferences on the kind of contract (Partnership, Forcing) they are offered in any period; rather they periodically update by best responding to the distribution of contracts in the past. Similarly firms do not condition their contractual offers on the type of the employee (Reciprocator, Homo economicus) with whom they are paired in a given period; rather they occasionally update by best responding to the past distribution of employee preferences.

The correspondence between preferences, contracts and specialization implied by these four assumptions is widely observed. Eric Nilsson (1994) studied the effects on comparative advantage and specialization resulting from the emancipation of slaves at the time of the U.S. Civil War. Cotton, according to Nilsson, was a “slave commodity” for which kinds of labor beyond that which could be coerced from the worker were of little importance. For other commodities – manufactures and tobacco in Nilsson’s empirical study – variations in the labor quality were more important, and impossible to secure by coercion. Nilsson exploited the natural experiment provided by the end of slavery to study the effect of an exogenous institutional shock on production specialization in 169 counties in the Confederacy. He found that the end of slavery brought about a significant shift away from the “slave commodity” (cotton) and towards manufactures and tobacco. Stefano Fenoaltea’s (1984) study of slave and non-slave production makes a similar distinction between “care intensive” and “effort intensive” productive activities, the former being opaque in our terminology and the latter transparent. A similar distinction between sugar and tobacco was made by Fernando Ortiz (1963) who contrasted the coerced labor and hierarchical and authoritarian culture of the sugar plantation regions of Cuba with the self-motivated labor and liberal culture of the tobacco family-farming areas.

Norms and preferences influencing economic behavior differ significantly among societies (Inglehart, 1977; Henrich, Boyd, Bowles et al., 2005). In particular, reciprocal social preferences appear to be more prevalent in the higher income countries. Among subjects in 15 countries, the level of cooperation sustained in a public goods experiment in which the altruistic punishment of free riders was possible was much higher in wealthier nations (Herrmann, Thoni, and Gaechter, 2008). For these reasons we represent an economy whose cultural-institutional equilibrium is characterized by partnerships and extensive social preferences such as trust and the positive work ethic as having a “good” cultural-institutional environment and, as a result, enjoying absolute advantage with respect to other countries in which forcing contracts and high levels of monitoring may elicit quantitative (but not qualitative) labor services from entirely self-regarding economic agents. This view is consistent with the observation that opaque goods make up a substantial fraction of the output of the more advanced economies (production and distribution of information-intensive goods and many services ranging from health care to entertainment and other recreational services), whereas poorer nations produce large shares of agricultural and manufactured goods that are closer to the transparent pole of the opaque-transparent continuum.

3. Cultural-institutional equilibrium under autarchy

Employers maximize profits, while employees maximize utility. Agents consume a composite bundle (indicated by c) of the two goods produced. For simplicity, we assume that the composite good is made up of one unit of the transparent (t) and one unit of the opaque (o) good; thus prices have no effect on consumption proportions. Denoting by p^o and p^t the price of the o-good and the price of the t-good, we define $\rho^o = p^o/(p^o + p^t)$ and $\rho^t = p^t/(p^o + p^t)$ respectively the value of the opaque good in terms of the composite good (how many units of the c-good one can purchase with one unit of the o-good) and the value of the transparent good in terms of the composite good (how many units of the c-good one can purchase with one unit of the t-good). Payoffs (profits and utility respectively) are measured in the number of units of the composite good commanded. Markets are competitive in the sense that employers take the price of the good as exogenously given.

The (risk-neutral) utility function of employees is additive in consumption of the composite good, the subjective utility associated with the contract (for the reciprocal agents) and the disutility associated with the type of labor provided in production. Supplying quantitative labor incurs a cost η (>0), while supplying both quantitative and qualitative labor costs $\delta\,(>\eta)$.

The key difference between a Forcing (F) contract and a Partnership (P) is that in the former the motivation to work is provided by the fear of being fired (as in many secondary labor market jobs), while in the latter the primary motivation is gain-sharing with the employer based on joint residual claimancy (as in many legal practices, financial consulting, and software design). Under the Forcing contract the employee is offered a fixed compensation (w>0) set by the employer to satisfy the participation constraint of the worker, is closely monitored and required to provide quantitative labor as a condition of continued employment. Under the Partnership the “employee” is offered half of the revenue of the Partnership and selects the type of labor (quantitative alone or both quantitative and qualitative) without supervision.

In the *F*-contract, providing quantitative work is sufficient for the worker to remain employed and paid, so the Homo economicus (*E*-type) employees offer quantitative labor, incurring the associated disutility η. If offered a *P*-contract, the *E*-worker also provides quantitative labor only as we assume that the worker's share of increased output associated with qualitative labor is less than the greater disutility required, i.e. $\rho^i Q_N^i / 2 - \eta > \rho^i Q_L^i / 2 - \delta$ (with *i*=*o*,*t*). By contrast, as we have seen, reciprocal (*R*-type) employees have preferences over the kind of contract that is offered by the employer *per se*. Under a Forcing contract the *R*-worker values the payoff of the employer negatively ($\gamma = -1$; the subscript *hk* for the individuals is hereafter omitted with no loss of clarity), subtracting $\alpha(\rho^i Q_N^i - w - \mu_R)$ from his utility. As a result the *R*-worker (like the *E*-worker) provides quantitative labor only (also at a cost η). Under the Partnership, however, the *R*-worker's positive valuation of the payoff to the partner (γ=1) is sufficient to offset the greater disutility of labor, i.e. $(1+\alpha)\rho^i Q_L^i / 2 - \delta > \rho^i Q_N^i / 2 - \eta$ (with *i*=*o*,*t*), and so the reciprocal type employee provides, in addition, qualitative aspects of work contributing to production (at a greater cost δ).

Both kinds of employees receive a rent under the *F*-contract: $w - \eta$ for the *E*-employee and $w - \eta - \alpha(\rho^i Q_N^i - w - \mu_R)$ for the *R*-employee, where μ_R is the employer's cost of monitoring the reciprocal worker. The level of monitoring cost sufficient so that supplying quantitative labor is a best response by the employee is greater for the (dissatisfied) reciprocal worker so $\mu_R \geq \mu_E$, where μ_E is the cost of monitoring the self-regarding worker (the following results are not affected if the monitoring costs are the same).

We now determine the conditions under which each of the four contract-preference pairs ({*F*,*E*}, {*F*,*R*}, {*P*,*E*} and {*P*,*R*}) may be Nash equilibria in the absence of trade. This will depend on relative prices of the goods which, because of the differing relative importance of qualitative labor in the production of the two goods, will in turn depend on whether (quantitative and) qualitative as well as quantitative labor is a best response of the employees. In autarchic equilibrium the only relative price, p^o / p^t, such that both goods are produced in the given country will be equal to the domestic marginal rate of transformation, namely Q_L^t / Q_L^o, for pairs in which qualitative in addition to quantitative labor is a best response, or Q_N^t / Q_N^o, where only quantitative labor is a best response. Using the subscript 1 and 2 to denote contract-preference pairs in which both quantitative and qualitative or just quantitative labor, respectively, are provided, we define $p_1^o / p_1^t = Q_L^t / Q_L^o$ and $p_2^o / p_2^t = Q_N^t / Q_N^o$. Accordingly, the relative price of the opaque (transparent) good in terms of the composite good respectively in the two situations will be $\rho_1^o = Q_L^t /(Q_L^t + Q_L^o)$ ($\rho_1^t = Q_L^o /(Q_L^t + Q_L^o)$) and $\rho_2^o = Q_N^t /(Q_N^t + Q_N^o)$ ($\rho_2^t = Q_N^o /(Q_N^t + Q_N^o)$).

Table 1 reports the matrix of payoffs measured in units of composite good. Because by construction autarchic prices make producers indifferent to the choice of which product to produce, we know that $\rho_1^o Q_L^o = \rho_1^t Q_L^t$ and $\rho_2^o Q_N^o = \rho_2^t Q_N^t$. Thus the entries in Table 1 are invariant across sectors. To find the Nash equilibria note that from the above description of the production process and prices we know that $(1+\alpha)\rho_1^i Q_L^i / 2 - \delta > \rho_2^i Q_N^i / 2 - \eta$, for any *i*=*o*,*t* (because, as shown in appendix A.1, $\rho_1^i Q_L^i > \rho_2^i Q_N^i$). To exclude uninteresting cases where cultural-institutional differences could not occur in equilibrium, we further assume that $\rho_2^i Q_N^i > 2(w + \mu_E)$, for any *i*=*o*,*t*.

Two Nash equilibria in pure strategies exist, namely {*P*,*R*}, that is the Partnership contract matched with the reciprocal employee, and {*F*,*E*}, that is the Forcing contract matched with the Homo economicus (see appendix A.1). We term these stable outcomes cultural-institutional conventions, meaning that conforming to them is a mutual best response as long as virtually all members of each sub-population (employers and employees) expect virtually all members of the other to conform to it. We denote the two conventions respectively by subscript 1 and 2. As we are interested in the effect of trade and factor market liberalization on the quality of cultural-institutional conventions, we assume that output with both qualitative and quantitative labor is sufficiently productive so that $Q_L^i / 2 > Q_N^i$ allowing an unambiguous ranking of the two outcomes by guaranteeing that the {*P*,*R*} Nash equilibrium Pareto-dominates the {*F*,*E*} equilibrium. But this does not guarantee that {*P,R*} will be observed in practice in a dynamic setting because the second "inferior" convention is also asymptotically stable. (A third unstable Nash equilibrium in mixed strategies exists; it will play an essential role in the dynamics of convention-switching studied in sections 4 and 5.)

	Employee/Preferences	
Employer/Contract	Reciprocator	Homo economicus
Partnership	$\boldsymbol{\rho_1^i Q_L^i/2,\ (1+\alpha)\rho_1^i Q_L^i/2-\delta}$	$\rho_2^i Q_N^i / 2, \rho_2^i Q_N^i / 2 - \eta$
Forcing contract	$\rho_2^i Q_N^i - w - \mu_R,\ w - \eta - \alpha(\rho_2^i Q_N^i - w - \mu_R)$	$\boldsymbol{\rho_2^i Q_N^i - w - \mu_E,\ w - \eta}$

***Table 1*: Matrix of payoffs.** (NOTE: Payoffs in bold type indicate pure stable Nash equilibria)

Assume now that the world economy comprises two countries, 1 and 2, identical in all relevant respects (same technology, same demand function, no difference in worker skills or in the preferences they may adopt), except for different cultural-institutional conventions. Let us suppose that country 1 is near equilibrium 1 ($\{P,R\}$), so that all pairs are reciprocal types working under Partnership contracts, whereas country 2 is near equilibrium 2 ($\{F,E\}$), so that all pairs are self-regarding employees working under Forcing contracts. Because the two countries are identical other than their cultural-institutional equilibria, hereafter the subscript 1 (2) denotes country 1 (country 2) and also equilibrium 1 (equilibrium 2).

In Figure 1 we represent the production possibility frontiers of the two countries, the slope of the dashed lines indicating the international terms of trade lying strictly between the two countries' marginal rates of substitution. Because $Q_L^o > Q_N^o$ and $Q_L^t > Q_N^t$, country 1 enjoys an absolute advantage in the production of both goods. However, the cultural and institutional differences across countries (like differences in endowments or technologies in the standard model) result in differences in the ratios of marginal costs of goods in autarchy and, as a result, confer different comparative advantages to the two countries considered. Country 1, where the established cultural-institutional equilibrium is able to elicit qualitative (in addition to quantitative) labor in all the employment relations, is superior in the production of both commodities, but has a relatively greater advantage in the production of the *o*-good where qualitative aspects of work are relatively more important. By contrast, country 2 has a culture and institutions for which employees are willing to provide quantitative labor only; this country, as a consequence, has a comparative advantage in the production of the *t*-good that is relatively less intensive in non-verifiable labor services.

Since in autarchic equilibrium the relative prices of the two countries are equal to the domestic marginal rates of transformation, and given inequality (1), it follows:

$$\frac{p_1^o}{p_1^t} = \frac{Q_L^t}{Q_L^o} < \frac{Q_N^t}{Q_N^o} = \frac{p_2^o}{p_2^t}, \qquad (3)$$

or, equivalently, $\rho_1^o < \rho_2^o$ ($\rho_1^t > \rho_2^t$). Providing that the international terms of trade, p_T^o / p_T^t (the subscript "*T*" refers to trade), falls strictly between the autarchic relative prices of the two countries, specialization and trade will take place. Given the linearity of the two production possibility frontiers, country 1 will specialize entirely in the production of (and will export) the opaque good, while country 2 will specialize in the production of (and will export) the transparent good.

Unless the two economies happen to be of the "right" size, given the fixed proportions in the composite consumption good there will either be excess supply of one of the two goods under complete specialization following trade integration. To retain the valuable simplifications due to both complete specialization and fixed proportions in consumption we could (artificially, but harmlessly) assume that under trade integration the "smaller" nation specializes and that firms in the other country produce a joint product of the two goods in the proportions necessary to satisfy global demands for the two goods. We opt for the simpler assumption that the countries are of a size to equilibrate world commodity markets, thereby avoiding notational clutter associated with joint production in one country.

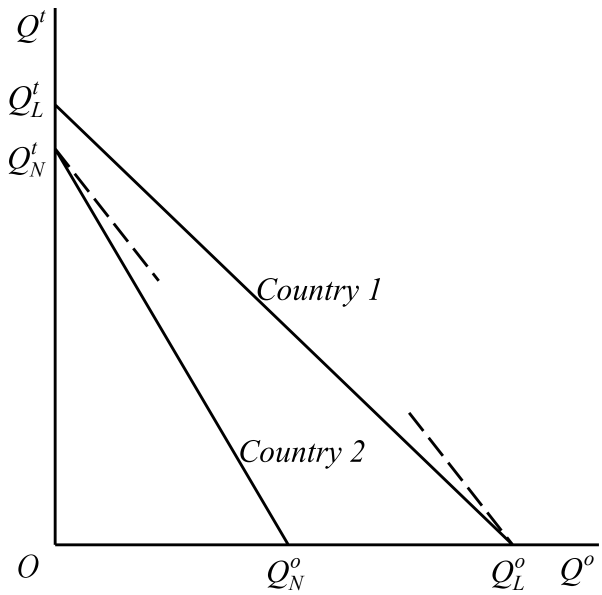

***Figure 1*: Production possibility frontiers in the two countries.**
(NOTE: Each country has a normalized labor endowment of 1)

Compared to autarchy, specialization and trade benefit both classes of individuals in country 1 and employers in country 2. When cross-country barriers to trade are removed and in absence of transportation costs, the relative price of the opaque (transparent) good increases in country 1 (country 2), whereas the relative price of the transparent (opaque) good decreases. It follows that $\rho_T^o > \rho_1^o$ and $\rho_T^t > \rho_2^t$: in both countries the good in which the country specializes becomes relatively more valuable in terms of the *c*-good (with one unit of the *o*-good (*t*-good) in country 1 (country 2) one can purchase a greater number of units of the *c*-good under trade than in autarchy). Thus, as expected, $\rho_T^o Q_L^o > \rho_1^o Q_L^o$ and $\rho_T^t Q_N^t > \rho_2^t Q_N^t$: the *c*-good value of output in the two countries increases as a result of specialization. All the other terms (δ, η, w, μ and γ) in the payoff matrix (Table 1) are measured in units of the composite goods and so remain unaltered.

4. Dynamics

To provide a framework for understanding the process of transitions from one convention to the other, we now study the asymptotic stability properties of the two conventions. We express the expected payoffs of employers and employees as a function of the distribution of contracts and worker types in each country, given the prevailing prices. For each economy there are two sets of prices to consider: autarchic prices (denoted by subscript 1 and 2, as above) and the prices common to both countries following trade (denoted by subscript "*T*", as above). Employers and employees are matched after having updated their contracts and preferences based on the distribution of play in the past. Writing the fraction of the employees who were Reciprocators in the previous period as ω and using the payoffs in Table 1 with nationally specific equilibrium prices, the expected payoffs to employers offering the *P*- and *F*-contracts are

$$\begin{aligned} v_P(\omega_j) &= \omega_j \frac{\rho_j^i Q_L^i}{2} + (1-\omega_j)\frac{\rho_j^i Q_N^i}{2}, \\ v_F(\omega_j) &= \omega_j[\rho_j^i Q_N^i - (w+\mu_R)] + (1-\omega_j)[\rho_j^i Q_N^i - (w+\mu_E)], \end{aligned} \tag{4}$$

where i=o,t and j=1,2. Similarly, writing the fraction of the employers offering Partnerships in the previous period as ϕ, the expected payoffs to the *R*- and *E*-employees are respectively:

$$\begin{aligned} v_R(\phi_j) &= \phi_j\left[(1+\alpha)\frac{\rho_j^i Q_L^i}{2} - \delta\right] + (1-\phi_j)[w - \eta - \alpha(\rho_j^i Q_N^i - w - \mu_R)], \\ v_E(\phi_j) &= \phi_j\left(\frac{\rho_j^i Q_N^i}{2} - \eta\right) + (1-\phi_j)(w-\eta). \end{aligned} \tag{5}$$

where again i=o,t and j=1,2. These expected payoff functions are illustrated in Figure 2.

To model the mutual dependence of preferences and contracts, suppose that both employers and employees periodically update the contracts they offer and their preferences (respectively) by best responding to the distribution of play in the other class in the previous period.

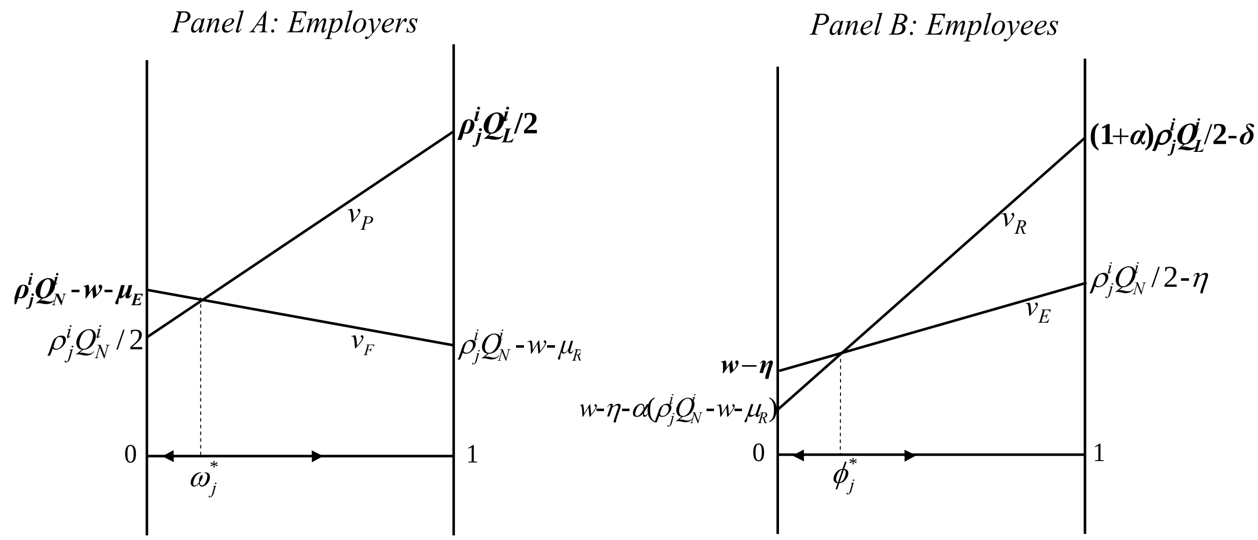

***Figure 2*: Expected payoffs under autarchy to *P*- and *F*-employers (*panel A*) and to *R*- and *E*-employees (*panel B*).** (NOTE: ϕ_j is the fraction of the employers offering Partnerships and ω_j the fraction of the employees being Reciprocators in the previous period and in country *j*. The vertical intercepts are from Table 1 using nationally specific equilibrium prices in autarchy (j=1,2 and i=o,t); payoffs in bold type refer to the stable pure Nash equilibria)

The updating process works as follows (Bowles, 2004). At the beginning of each period, individuals are exposed to a cultural or institutional model randomly selected from their sub-population: for instance, an employer, named A, has the opportunity to observe the contract offered by another employer, named B, and to know her payoff. If employer B is the same type as employer A, A does not update. But if B is a different type, A compares the two payoffs and, if B has the greater payoff, switches to B's type with a probability equal to β (>0) times the payoff difference, retaining her own type otherwise. It is easily shown that this process gives the replicator equations:

$$\frac{\mathrm{d}\phi_j}{\mathrm{d}\tau} = \phi_j(1-\phi_j)\beta[v_P(\omega_j) - v_F(\omega_j)],$$
$$\frac{\mathrm{d}\omega_j}{\mathrm{d}\tau} = \omega_j(1-\omega_j)\beta[v_R(\phi_j) - v_E(\phi_j)], \tag{6}$$

where j=1,2 and τ stands for time. We are interested in the stationary states, such that $\mathrm{d}\phi_j / \mathrm{d}\tau = 0$ and $\mathrm{d}\omega_j / \mathrm{d}\tau = 0$. It is easy to see that:

$$
\begin{aligned}
&\frac{\mathrm{d}\phi_j}{\mathrm{d}\tau}=0\,\text{for}\,\phi_j=0,\ \phi_j=1\,\text{and}\,\omega_j^*=\frac{\dfrac{\rho_j^i Q_N^i}{2}-(w+\mu_E)}{\dfrac{\rho_j^i Q_L^i}{2}-\dfrac{\rho_j^i Q_N^i}{2}+(\mu_R-\mu_E)},\\
&\frac{\mathrm{d}\omega_j}{\mathrm{d}\tau}=0\,\text{for}\,\omega_j=0,\ \omega_j=1\,\text{and}\,\phi_j^*=\frac{\alpha[\rho_j^i Q_N^i-(w+\mu_R)]}{\left[(1+\alpha)\dfrac{\rho_j^i Q_L^i}{2}-\delta\right]-\left(\dfrac{\rho_j^i Q_N^i}{2}-\eta\right)+\alpha[\rho_j^i Q_N^i-(w+\mu_R)]},
\end{aligned}
\tag{7}
$$

where $i=o,t$ and $j=1,2$. The resulting dynamical system is illustrated by the vector field in Figure 3 where the arrows indicate the out-of-equilibrium adjustment given by the replicator dynamic (equations 6) and subscripts j are omitted with no loss of clarity. The states where $\mathrm{d}\phi/\mathrm{d}\tau=0$ and $\mathrm{d}\omega/\mathrm{d}\tau=0$ are cultural-institutional equilibria. The state (ϕ^*,ω^*) is stationary, but it is a saddle: small movements away from ϕ^* or ω^* are not self-correcting. (Two additional unstable stationary states, namely $(\phi=1,\omega=0)$ and $(\phi=0,\omega=1)$ are of no interest.) The asymptotically stable states are (1,1) (corresponding to convention 1, i.e. {*P*,*R*}, in Table 1) and (0,0) (corresponding to convention 2, i.e. {*F*,*E*}, in Table 1).

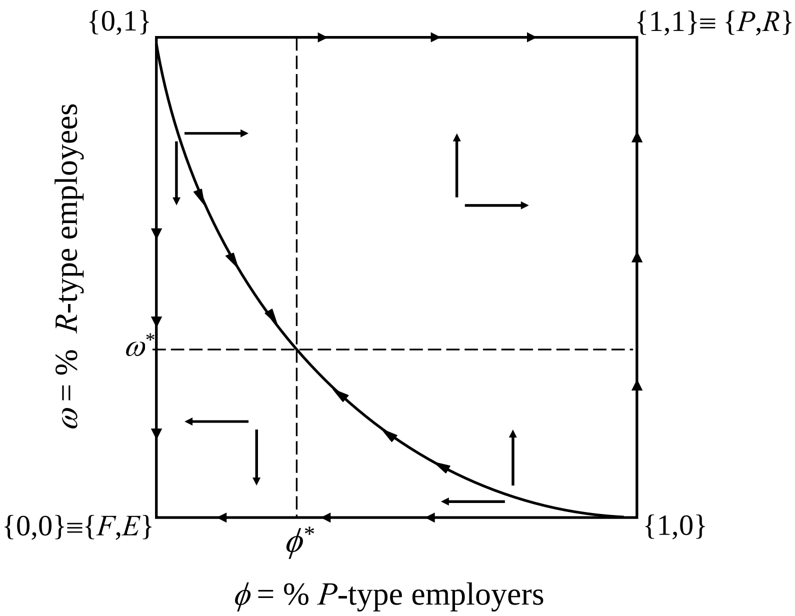

***Figure 3*: Co-evolution of preferences and institutions, and persistence of two cultural-institutional equilibria in a given country.**

In this deterministic setting, the initial state determines which of these two asymptotically stable states occurs. Of course institutions (and, in some cases, even cultures) may be altered by a joint decision of hypothetical representatives of one or both classes (Acemoglu and Robinson, 2006).

But non-cooperative (that is decentralized, bottom-up) transitions are also possible. To study such a process we assume that occasional idiosyncratic (non-best response) updating of both preferences and contractual offers occurs (Kandori, Mailath, and Rob, 1993; Young, 1993, 1998). Suppose that with probability $1-\varepsilon$ the myopic best response updating process described above occurs, but with a small probability ε the employee chooses randomly from the two preferences and the employer likewise randomizes her contractual offer. The preference or contractual innovations represented by idiosyncratic play may be due to deliberate experimentation, error, or any other reason for non-best response play. We assume throughout that the rate of idiosyncratic play is sufficiently small that the equilibrium conventions described above are persistent, defined as having an expected duration of more than one period (i.e. $\varepsilon <$ critical number that would induce a transition to the other convention), so that in equilibrium 1 $\varepsilon < 1-\omega^*$ and $\varepsilon < 1-\phi^*$, while in equilibrium 2 $\varepsilon < \omega^*$ and $\varepsilon < \phi^*$. Jointly these persistence conditions imply $\varepsilon < 1/2$.

In the resulting perturbed Markov process over the long run both $\{P,R\}$ and $\{F,E\}$ will occur, with infrequent transitions between the basins of attraction of these two equilibria (Young, 1998). In the absence of system-level exogenous shocks, for even moderately large populations and plausible rates of idiosyncratic play cultural-institutional equilibria will persist over very long periods and the system will spend more time at the convention with the larger basin of attraction. Thus equilibrium 1 will be more persistent in this sense if $\phi^*\omega^* < (1-\phi^*)(1-\omega^*)$ that is, if $\{P,R\}$ is the risk-dominant equilibrium, and conversely for equilibrium 2.

5. Trade integration and the persistence of inefficient equilibria

Our finding that the culture and institutions prevailing in each country are a source of comparative advantage, and that opening up to trade enables the two otherwise identical countries to enjoy welfare gains, would be of little interest if trade were to erode the differences upon which cultural-institutional comparative advantage depends. Because both culture and institutions are endogenous in our model, we can determine if the two asymptotically stable cultural-institutional equilibria persist after the two countries open up to international exchange, or equivalently, the two critical values, ϕ_j^* and ω_j^*, remain in the unit interval following trade.

We ask whether in a stochastic environment trade favors cultural and institutional convergence. As we are interested in convergence to superior cultural-institutional conventions, we consider the effect of trade (and, subsequently, factor market integration) on the stability of country 2's inferior {*F*,*E*} convention (technically, we ask: what is the effect of trade on the expected waiting time for a transition from equilibrium 2 to equilibrium 1.) Because cultural-institutional transitions occur as a result of deviations from the status quo convention, the effect of trade on convergence can be explored in two ways: by looking at either the minimum number of innovators (deviants) required to induce a transition (termed the "resistance" to a transition) or the expected cost that an innovator incurs, the latter measuring the incentives against innovating and the selection pressures operating against those who do. Though we do model the innovation process formally here, in a more complete model with state dependent rates of idiosyncratic play (Bergin and Lipman, 1996) the increased cost of innovating plausibly would reduce the rate of innovation, thereby prolonging the expected duration of the convention.

Figure 4 shows how the expected payoffs for each group of individuals change as a result of trade (expected payoff lines after trade are drawn in dashed type). Payoffs received by the individuals in the autarchic equilibrium are in bold fonts in the relevant panel.

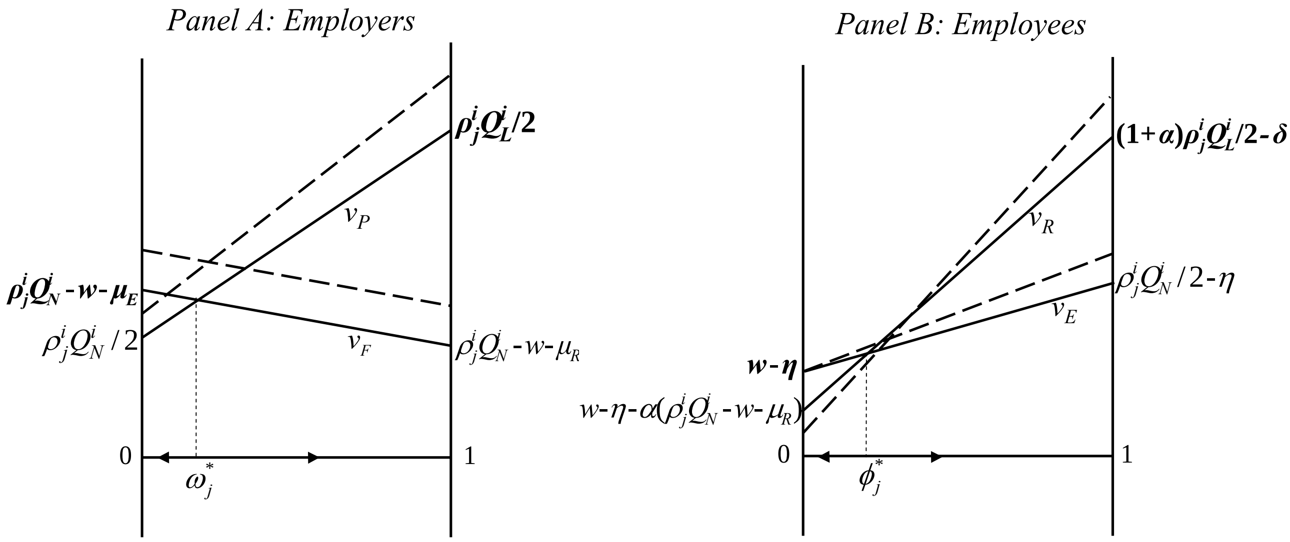

***Fig. 4*: Payoff changes to *P*- and *F*-employers (*panel A*) and *R*- and *E*-employees (*panel B*) after trade openness.** (NOTE: ϕ_j is the fraction of the employers offering Partnerships and ω_j the fraction of the employees being Reciprocators in the previous period and in country *j*. The vertical intercepts are from Table 1 using nationally specific equilibrium prices (*j*=1,2 and *i*=*o*,*t*); payoffs in bold type refer to the stable pure Nash equilibria. Dashed lines represent expected payoff lines after trade integration. The post-trade corresponding vertical intercepts differ from the autarchic case because of the increase in ρ_j)

Trade increases the amount of the composite good that may be purchased with one unit of the good in which each country specializes ($\rho^o_T > \rho^o_1$ and $\rho^t_T > \rho^t_2$), giving the dashed lines in the figure. It is readily confirmed (from inspection of their definition in equations (7)) that after trade, the critical values of ϕ_j and ω_j remain within the unit interval in both countries, implying that trade integration does not destroy the cultural-institutional differences upon which specialization is based.

Inspection of the figure also confirms that trade increases the cost of deviating from the status quo cultural-institutional convention for both groups in both countries, implying that non-coordinated convergence from one equilibrium to the other is less likely under trade integration than under autarchy. This can be seen from equations (6), along with the fact that trade increases both $[v_P(\omega_1)-v_F(\omega_1)]$ and $[v_R(\phi_1)-v_E(\phi_1)]$ when $\omega_1=1=\phi_1$ (equilibrium 1) and increases both $[v_F(\omega_2)-v_P(\omega_2)]$ and $[v_E(\phi_2)-v_R(\phi_2)]$ when $\omega_2=0=\phi_2$ (equilibrium 2) (see appendix A.2.1). The reason is that deviating from the convention almost always entails a mismatch, the result being forgoing some of or all the surplus, the value of which is higher after trade integration.

In addition to increasing the incentive not to innovate and the selection pressures operating against those who do, trade may even increase the number of innovators necessary to induce a transition from the inferior equilibrium 2 to equilibrium 1. To see this we study the effect of trade (increase in ρ_j^i) on ϕ_j^* and ω_j^*. In the case of ω_j^* the result is unambiguous: trade increases the critical fraction of reciprocal workers necessary to induce the *F*-type employers to best respond by adopting *P*-contracts. Indeed (see appendix A.2.2):

$$\frac{\mathrm{d}\omega_j^*}{\mathrm{d}\rho_j^i}=\frac{\frac{Q_N^i}{2}(\mu_R-\mu_E)+\left(\frac{Q_L^i}{2}-\frac{Q_N^i}{2}\right)(w+\mu_E)}{\left[\frac{\rho_j^iQ_L^i}{2}-\frac{\rho_j^iQ_N^i}{2}+(\mu_R-\mu_E)\right]^2}>0 \tag{8}$$

where *j*=1,2 and *i*=*o*,*t*. The reason can be seen by noting that the critical values ϕ_j^* and ω_j^* are simply given by the cost (for respectively employees and employers) of deviating from the {*F*,*E*} convention divided by the sum of this cost and the cost of deviating from the {*P*,*R*} convention. While the costs of deviating from both equilibria increase for the employers, trade increases the cost of deviating from the {*F*,*E*} equilibrium of country 2 proportionally more.

The effect of trade on ϕ_j^* cannot be signed in general, but (under plausible conditions) it too may increase following trade integration. We have (see appendix A.2.2)

$$\frac{\mathrm{d}\phi^*}{\mathrm{d}\rho_j^i}=\frac{Q_N^i(-\delta+\eta)+\left[(1+\alpha)\frac{Q_L^i}{2}-\frac{Q_N^i}{2}\right](w+\mu_R)}{\left\{(1+\alpha)\frac{\rho_j^i Q_L^i}{2}-\delta-\left(\frac{\rho_j^i Q_N^i}{2}-\eta\right)+\alpha(\rho_j^i Q_N^i-w-\mu_R)\right\}^2}>0 \qquad (9)$$

where j=1,2 and i=o,t, if and only if $[(1+\alpha)Q_L^i-Q_N^i](w+\mu_R)/2-Q_N^i(\delta-\eta)>0$. This will be the case if the degree of reciprocity and the relative productiveness of qualitative labor are sufficiently great (or if the excess disutility of providing qualitative labor is sufficiently small).

Thus removing impediments to international exchange need not destabilize and, indeed, may fortify the preexisting cultural and institutional differences upon which specialization and trade are based even if there exists an alternative cultural-institutional equilibrium that confers absolute advantage and to which a transition would be Pareto-improving. Trade impedes cultural-institutional convergence because it raises the costs of deliberate or accidental experimentation with uncommon preferences and contracts. Under plausible conditions it also increases the number of cultural or institutional innovators necessary to induce a decentralized transition from the low to the high productivity equilibrium.

While the waiting time for a transition from the inferior to the superior cultural-institutional convention may be increased by trade, a transition to the superior culture and institutions can be induced by a one-time tariff even in the absence of idiosyncratic play. It is readily shown that there exists a tariff protecting the (imported) opaque good in country 2 such that a best response-induced cultural-institutional transition will occur, country 2 adopting the {P,R} cultural-institutional nexus. Assuming that the international price ratio is not affected by the tariff, let θ_ω^* and θ_ϕ^* be the ad-valorem tariff rates on the opaque (imported) good which will implement an (after-tax) domestic price ratio in country 2 such that, respectively, $\omega_2^*=0$ and $\phi_2^*=0$. The transition-inducing tariff is given by $\theta^*=\min[\theta_\omega^*,\theta_\phi^*]$. Using equations (7) it can be shown (see appendix A.2.3) that:

$$(1+\theta_\omega^*)=\left[\frac{Q_N^t}{2(w+\mu_E)}-1\right]\frac{p_T^t}{p_T^o} \text{ and } (1+\theta_\phi^*)=\left(\frac{Q_N^t}{w+\mu_R}-1\right)\frac{p_T^t}{p_T^o}. \qquad (10)$$

It is readily seen that $\theta_\omega^*<\theta_\phi^*$ as long as $w+2\mu_E>\mu_R$.

The logic of the transition-inducing tariff is exactly the opposite of the mechanism underlying the fact that trade liberalization is transition-impeding. The tariff makes the transparent good less valuable in terms of the units of the composite good it can command and hence reduces the joint surplus available to the employer and the employee. So the tariff reduces the cost of deviation from the $\{F,E\}$ convention, and a sufficiently large tariff will eliminate the deviation cost entirely. The level that eliminates the cost of deviation for either of the two classes is the transition inducing tariff θ^*. If $\theta^* = \theta^*_\omega$, under the minimal transition inducing tariff it would be the employers who induce the transition because the real cost (in terms of *t* goods) of wages and monitoring has risen to such an extent that they do no better by offering Forcing contracts than by offering Partnerships. Any tariff greater than this makes the Partnership a strict best response for the employers. If, on the contrary, $\theta^* = \theta^*_\phi$, the tariff would reduce profits under the Forcing contract to zero and would make employees indifferent to being reciprocal or self-regarding (if the employer is making zero profits the reciprocal employee is not offended by a Forcing contract).

6. Factor market integration and transitions to efficient equilibria

Many of the effects of international economic integration – like factor price equalization in Paul Samuelson's theorem (Samuelson, 1948) – are independent of whether integration is accomplished through the elimination of barriers to trade in commodities or through the mobility of factors of production. Where comparative advantage is based on country differences in culture and institutions, as in our model, however, this is not the case.

As we are interested in convergence to superior cultural-institutional conventions, we model the effect of factor market integration on the stability of country 2's inferior $\{F,E\}$ convention. In contrast to trade integration, factor market integration facilitates a Pareto-improving cultural-institutional transition in country 2. It does this by having the opposite of the two effects of trade integration: in the neighborhood of the $\{F,E\}$ equilibrium, it lessens the costs of idiosyncratic play and reduces the number of innovators required to induce a transition. Under factor market integration, cultural and institutional innovators may enjoy an advantageous match not only with rare innovators from their own economy but also with the prevalent type of agent from the other country. Thus factor market integration provides a kind of innovation insurance, in contrast to commodity market integration which imposes an innovation penalty, because of the gains from trade that heighten the opportunity costs of the frequent mismatches that innovators may expect when paired with agents from their own country.

Suppose that some matches are made entirely with one's own nationals while others are made randomly in the global population. As pictured in Figure 5, there are now three factor markets, two of them national-specific and the third, a common pool without country identification. The common pool is populated by agents drawn at random from the two country-specific pools and hence has the same distribution of types as the meta-population (both countries combined). For both employers and employees in both countries let n be the fraction of matches made with individuals from one's own nation, the complement, $1-n$, being matches in the common pool.

One may imagine the two countries as two "villages" within which all production takes place under autarchy. But with factor market integration some (a random draw from each of the two villages) go to the cosmopolitan "city" where they make random matches with members of the other class who they encounter there. In this model n is not chosen by the individual agents; it is a characteristic of the two countries' cultures, language differences, geographical distance, immigration policies and other influences on factor movement that are exogenous from the standpoint of the individual employer or employee.

In the autarchic factor markets we have thus far assumed $n=1$. But, if $n<1$, one's expected match is n times the fraction of agents in one's own country plus $1-n$ times the distribution of types in the common pool. To see that n is a measure of the degree of national specificity of factor markets and $1-n$ is the degree of factor market integration note the following. If the countries are in the neighborhood of the equalibria $\{P,R\}$ (country 1) and $\{F,E\}$ (country 2), the country difference in an employer's probability of being paired with a Reciprocal employee is approximately $n(1-2\varepsilon)$ (see appendix A.3.1), which must be positive and increasing in n by the persistence conditions given in section 4. This same quantity $n(1-2\varepsilon)$ is the difference, conditional on being resident in country 1 or country 2, in the probability that an employee will be paired with an employer offering a Partnership contract.

To avoid considerable notational clutter for no additional insight we assume that n does not vary across countries. When factors of production are matched in the pool we assume that the product produced is determined by the nationality of the employer, reflecting the fact that the physical assets of the employer are product-specific while the skills of the worker are less so (this assumption may easily be relaxed without altering the conclusions in any relevant way). In the case of autarchy, the prices at which the output is sold are also determined by the nationality of the employer. Thus, for example, when an employee from country 2 is matched with an employer from country 1, the pair will produce the opaque good to be sold either at the prevailing international prices (in the case of trade integration) or at the autarchic prices of country 1 (in the absence of trade integration).

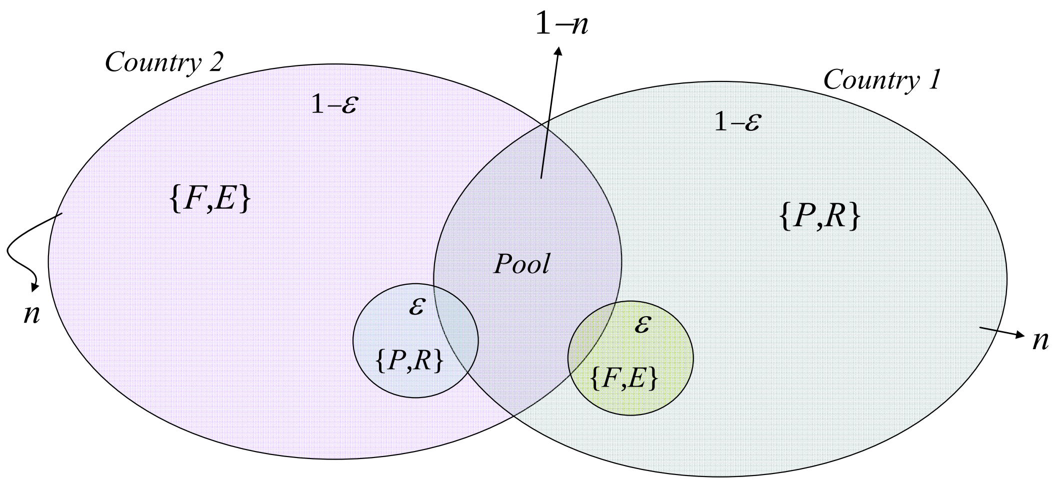

***Fig. 5*: Factor market integration.** (NOTE: ε is the expected fraction of idiosyncratic players among both employers and employees, n is the degree of national specificity of the factor markets and $1-n$ is the degree of factor market integration)

The expected payoff after factor integration is the weighted sum of the expected payoff in the national factor market plus the expected payoff in the common pool, the weights being the relative sizes of the two pools, n and $1-n$ (expected payoff equations are reported in the appendix A.3.2 for reasons of space). To determine the critical values, as before we equate expected payoffs, but we now take account of the effects of the degree of factor market integration. Thus we set $v_P(\omega_2, n) = v_F(\omega_2, n)$ and $v_R(\phi_2, n) = v_E(\phi_2, n)$, and obtain $\omega_2^*(n)$ and $\phi_2^*(n)$. (We show in appendix A.3.2 that the following results obtain using both autarchic and trade prices which means that they apply equivalently to factor market integration for autarchic or trading economies.)

First, for both employers and employees in country 2, factor market integration (reducing n) lessens the costs of idiosyncratic play, respectively for employers, $v_F(\omega_2 = 0, n) - v_P(\omega_2 = 0, n)$, and for employees, $v_E(\phi_2 = 0, n) - v_R(\phi_2 = 0, n)$. The case of employers is straightforward. The F-type best responding employers in the $\{F,E\}$ equilibrium will be disadvantaged (or unaffected) by factor market integration because they will have now a positive probability to match a reciprocal employee from country 1, who always provides quantitative labor alone under forcing contracts (as does Homo economicus), but is more costly to monitor ($\mu_R \geq \mu_E$). By contrast, when $n<1$, F-type employers who idiosyncratically offer P-contracts will enjoy a payoff-maximizing match (with a reciprocal worker) not only with the rare innovators from their own economy but also with the prevalent type of worker from the other country, who will constitute a sizeable fraction of the workers in the cosmopolitan pool. So the expected payoff to the best responder decreases (or is unchanged) and the expected payoff to the idiosyncratic player increases leading to a lessened cost of deviation (see appendix A.3.3).

The same logic applies to employees. Factor market integration increases the probability that both E-type players idiosyncratically adopting reciprocal preferences and best responding E-type workers conforming to the convention in the $\{F,E\}$ equilibrium will make a payoff-maximizing match. However, the innovators' payoff advantage from market integration is greater than the benefit received by the best responders. Both idiosyncratically playing and best responding employees in country 2 additionally benefit from the higher payoffs from being matched with a country 1 producer. In this case the worker will produce the opaque good (rather than the transparent good) to be sold either at the prevailing international prices (if trade integration is considered; in which case $Q_L^o \rho_T^o > Q_L^t \rho_T^t$) or at the autarchic prices of country 1 (in the absence of trade integration; in which case $Q_L^o \rho_1^o > Q_L^t \rho_2^t$). But taking account of both the better matching prospects and the increase in payoffs for both best responders and idiosyncratic employees, it can be shown (see appendix A.3.3) that innovators benefit from factor market integration more than best responders.

Thus, factor market integration facilitates a transition from the inferior to the superior equilibrium because it reduces the payoff disadvantage of both idiosyncratically playing employers and employees compared to those conforming to the convention, and therefore it lessens the expected costs of innovating.

Second, for the country at the inferior cultural-institutional equilibrium in country 2, it can be shown (see appendix A.3.4) that

$$\frac{\mathrm{d}\omega_2^*(n)}{\mathrm{d}n} > 0 \text{ and } \frac{\mathrm{d}\phi_2^*(n)}{\mathrm{d}n} > 0,$$

so that factor market integration (reducing n) lowers the critical fraction of innovators in both classes sufficient to induce a transition to the $\{P,R\}$ cultural-institutional convention. Thus, factor market integration facilitates transitions to the superior cultural-institutional nexus.

7. Discussion

We have shown that otherwise identical economies that differ in culture and institutions may find specialization and trade welfare-enhancing, and that trade reinforces these differences by inhibiting convergence to superior cultural-institutional arrangements, while factor market integration favors convergence.

Our paper is a contribution to the rapidly growing literature on institutions and trade (earlier contributions surveyed in Belloc, 2006). Comparative advantage based on institutional differences has been investigated for the following settings: financial systems (Beck, 2002; Kletzer and Bardhan, 1987; Ju and Wei, 2005; Matsuyama, 2005; Svaleryd and Vlachos, 2005), enforcement of contracts and property rights (Esfahani and Mookherjee, 1995; Levchenko, 2007; Nunn, 2007), intellectual property rights (Pagano, 2007), contracts and the division of labor (Acemoglu, Antràs and Helpman, 2009; Costinot, 2009), contractual incompleteness and the product cycle (Antràs, 2005), labor market flexibility and volatility (Cunat and Melitz, 2010), legal establishment and accounting systems (Vogel, 2007). In contrast to these papers, rather than studying the effects of exogenously given differences in institutions on comparative advantage and trade, we consider the impact of economic integration on the endogenous dynamics of institutions.

Other papers treating the effects of trade on institutions are Belloc (2009), Casella and Feinstein (2002), Dixit (2003), Do and Levchenko (2009) and Levchenko (2010). The main novelty of our approach with respect to this latter group of papers is our modeling of the complementary relationship between cultural preferences and institutions as a mechanism by which institutions associated with absolute disadvantage may persist indefinitely. In particular, our paper departs from and complements the work of Do and Levchenko (2009) and Levchenko (2010) in which institutional differences are a historical datum that may be modified by a cooperative lobbying game, while in our model they are implemented as an endogenously generated non-cooperative cultural-institutional equilibrium. Finally, unlike all above papers but in common with Olivier, Thoenig and Verdier (2008) and Pagano (2007), we find contrasting convergence effects of trade integration and factor market integration. But our model and these two models share little else in common, the former illustrates the dynamics of the demand for "cultural goods" that contribute to group identity, while the latter concerns intellectual property.

The co-evolution of social norms and institutions is also modeled by Francois (2008). However, in contrast to our approach, in his model institutional change is implemented by an institutional designer external to the transaction (a political actor). Furthermore, while we explore the effects of economic integration on cultural-institutional equilibria, Francois (2008) studies the effect of increasing market competition. We share with Conconi, Legros and Newman (2009) the conclusion that liberalization need not favor the evolution of efficient institutions. In contrast to ours, in their model factor market integration may induce inefficiency, and only in conjunction with good market integration are the effects of the two positive (in our model factor market integration has unambiguously positive effects). As in Krugman (1987)'s model of learning by doing, we show that a one time tariff may permanently alter a nation's comparative advantage and induce welfare gains.

The possibility that trade may induce institutional and cultural divergence rather than convergence is suggested by the experience of Europe in the late 19th century, when the institutional response to the import of cheap North American grain was radically different from country to country, resulting in a divergence with respect to tariffs and agrarian institutions (Gourevitch, 1977). Culture differences were also heightened, as the social solidarity of the subsidized Danish dairy cooperatives differed markedly from the nationalism associated with the German and French tariffs. Likewise, the centuries-long persistence of institutional differences among Western Hemisphere economies documented in Sokoloff and Engerman (2000) may be explained in part by the fact that trade allowed specialization in "plantation goods" such as sugar and cotton in some countries and "family farm" goods such as tobacco and wheat in others. Freeman (2000) and Moriguchi (2003) document a divergence in labor market institutions in open economies. The "cultural and institutional bifurcation" of China and Europe studied by Greif and Tabellini (2010) could persist even in the presence of exchange (and would favor Europe's specialization in goods in which economies of scale were more pronounced).

These cases of divergence notwithstanding, the impact of the U.S. civil war studied by Nilsson (1994) is a reminder that cultural-institutional convergence does appear to be a powerful tendency in integrated global systems. But, like the convergence of European political institutions to the national state model over the half millennium prior to the First World War (Tilly, 1990), and the contemporaneous global diffusion of institutions and cultures of European origin, it also points to the important role of military and other political forces rather than the autonomous workings of international trade *per se* in this cultural and institutional convergence process.

References

Acemoglu, D., P. Antràs, E. Helpman. 2009. Contracts and Technology Adoption, *American Economic Review* 97(3): 916-943.

Acemoglu, D., J. Robinson. 2006. *Economic Origins of Dictatorship and Democracy*. Cambridge: University of Cambridge Press.

Antràs, P. 2005. Incomplete Contracts and the Product Cycle. *American Economic Review* 95(4): 1054-1073.

Banerjee, A., L. Iyer. 2005. History, Institutions and Economic Performance: The Legacy of

Colonial Land Tenure Systems in India. *American Economic Review* 95(4): 1190-213.

Beck, T. 2002. Financial Development and International Trade. Is there a Link? *Journal of International Economics* 57: 107-131.

Belloc, M. 2009. Cross-Cultural Trade and Institutional Stability. *B.E. Journal of Theoretical Economics* (*Topics*) 9(1): article 34

Belloc, M. 2006. Institutions and International Trade: a Reconsideration of Comparative Advantage. *Journal of Economic Surveys* 20: 3-26.

Bergin, J., B.L. Lipman. 1996. Evolution With State-Dependent Mutations. *Econometrica* 64(4): 943-56.

Bewley, T.F. 1999. *Why Wages Don't Fall During a Recession*. Cambridge: Harvard University Press.

Bisin, A., T. Verdier. 2001. The Economics of Cultural Transmission and the Evolution of Preferences. *Journal of Economic Theory* 97(2): 298-319.

Bowles, S. 1998. Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions. *Journal of Economic Literature* 36(1): 75-111.

Bowles, S. 2004. *Microeconomics: Behavior, Institutions, and Evolution*. Princeton: Princeton University Press, Princeton.

Bowles, S. 2008. Policies Designed for Self Interested Citizens May Undermine 'the Moral Sentiments': Evidence from Experiments. *Science* 320: 58-83.

Bowles, S., S. Polania Reyes. 2009. Economic Incentives and Social Preferences: A Preference-based Lucas Critique of Public Policy. *University of Massachusetts Amherst WP* 2009-11.

Cameron, L. 1998. Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia. *Economic Inquiry* 37(1): 47-59.

Casella, A., J.S. Feinstein. 2002. Public Goods in Trade on the Formation of Markets and Jurisdictions. *International Economic Review* 43(2): 437-462.

Conconi, P., P. Legros, A.F. Newman. 2009. Trade Liberalization and Organizational Change. *Université Libre de Bruxelles*.

Costinot, A. 2009. On the Origins of Comparative Advantage. *Journal of International Economics* 77(2): 255-264.

Cunat, A., M. Melitz. 2010. Volatility, Labour Market Flexibility, and the Pattern of Comparative Advantage, *Journal of the European Economics Association*, forthcoming.

Dixit, A.K. 2003. Trade Expansion and Contract Enforcement, *Journal of Political Economy* 111(6): 1293-1317.

Do Q-T., A. Levchenko. 2009. Trade, Inequality and the Political Economy of Institutions. *Journal of Economic Theory* 144(4): 1489-1520.

Doepke, M., F. Zilibotti. 2008. Occupational Choice and the Spirit of Capitalism. *Quarterly Journal of Economics* 123(2): 747-793.

Esfahani, H.S., D. Mookherjee. 1995. Productivity, Contracting Modes, and Development. *Journal of Economic Development* 46: 203-231.

Falk, A., M. Kosfeld. 2006. The Hidden Costs of Control. *American Economic Review* 96(5): 1611-30.

Fehr, E., A. Falk. 2002. Psychological Foundations of Incentives. *European Economic Review* 46(4-5): 687-724.

Fehr, E., A. Klein, K.M. Schmidt. 2007. Fairness and Contract Design. *Econometrica* 75(1): 121-54.

Fenoaltea, S. 1984. Slavery and Supervision in Comparative Perspective: A Model, *Journal of Economic History* 44(3): 635-668.

Francois, P. 2008. Norms and Institution Formation. *CEPR DP* 6735.

Freeman, R. 2000. Single Peaked vs. Diversified Capitalism: the Relation Between Economic Institutions and Outcomes. *NBER WP* 7556.

Fershtman, C., O. Bar-Gill. 2005. The Limit of Public Policy: Endogenous Preferences. *Journal of Public Economic Theory* 7(5): 841-57.

Galor, O., O. Moav. 2002. Natural Selection and the Origin of Economic Growth. *Quarterly Journal of Economics* 117: 1133-91.

Gerschenkron, A. 1944. *Bread and Democracy in Germany*. Berkeley: University of California Press.

Gourevitch, P. 1977. International Trade, Domestic Coalitions, and Liberty: Comparative Responses to the Crisis of 1873-1896. *Journal of Interdisciplinary History* 8(2): 281-313.

Greif, A. 1994. Cultural Beliefs and the Organization of Society: An Historical and Theoretical Reflection on Collectivist and Individualist Societies. *Journal of Political Economy* 102(5): 912-50.

Greif, A., G. Tabellini. 2010. Cultural and Institutional Bifurcation: China and Europe Compared. *CEPR DP* 7648

Guiso, L., P. Sapienza, L. Zingales. 2009. Cultural Biases in Economic Exchange? *Quarterly Journal of Economics*, 124(3): 1095-1131.

Herrmann, B., C. Thoni, S. Gaechter. 2008. Antisocial Punishment Across Societies. *Science* 319

(7 March 2008): 1362-67.

Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, R. McElreath, M. Alvard, A. Barr, J. Ensminger, N. Smith Henrich, K. Hill, F. Gil-White, M. Gurven, F. Marlowe, J. Patton, D. Tracer. 2005. `Economic Man' in Cross-Cultural Perspective: Behavioral Experiments in 15 Small-Scale Societies. *Behavioral and Brain Sciences* 28: 795-855.

Inglehart, R. 1977. *The Silent Revolution: Changing Values and Political Styles among Western Publics*. Princeton: Princeton University Press.

Ju, J., S-J. Wei. 2005. Endowment Versus Finance: a Wooden Barrel Theory of International Trade. *IMF WP* 05/123.

Kandori, M.G., G.J. Mailath, R. Rob. 1993. Learning, Mutation, and Long Run Equilibria in Games. *Econometrica* 61(1): 29-56.

Keynes, J.M. 1933. National Self-Sufficiency. *Yale Review* 22(4): 755-769.

Kindleberger, C.P. 1962, *Foreign Trade and the National Economy*, New Haven, CT: Yale University Press.

Kletzer, K., P.K. Bardhan. 1987. Credit Markets and Patterns of International Trade. *Journal of Development Economics* 27: 57-70.

Krugman, P. 1987. The Narrow Moving Band, the Dutch Disease, and the Competitive Consequences of Mrs. Thatcher, Notes on Trade in the Presence of Dynamic Scale Economies. *Journal of Development Economics* 27: 41-55.

Levchenko, A.A. 2007. Institutional Quality and International Trade. *Review of Economic Studies* 74, pp. 791-819.

Levchenko, A.A. 2010. International Trade and Institutional Change. *University of Michigan.*

Levine, D.K. 1998. Modeling Altruism and Spitefulness in Experiments. *Review of Economic Dynamics* 1: 593-622.

Matsuyama, K. 2005. Credit Market Imperfections and Patterns of International Trade and Capital Flows. *Journal of the European Economic Association* 3: 714-723.

Moriguchi, C. 2003. Implicit Contracts, the Great Depression, and Institutional Change: A Comparative Analysis of U.S. and Japanese Employment Relations, 1920-1940. *Journal of Economic History* 63(3): 625-665.

Nilsson, E. 1994. Empirical Evidence that the Social Relations of Production Matter: the Case of the Ante-bellum U.S. South. *Cambridge Journal of Economics* 18: 259-77.

Nunn, N. 2007. Relation-Specificity, Incomplete Contracts and the Pattern of Trade. *Quarterly Journal of Economics* 122: 569-600.

Nunn, N., L. Wantchekon. 2010. The Slave Trade and the Origins of Mistrust in Africa. *Harvard University*.

Olivier, J., M. Thoenig, T. Verdier. 2008. Globalization and the Dynamics of Cultural Identity. *Journal of International Economics* 76: 356-370.

Ortiz, F. 1963. *Contrapunteo Cubano del Tabaco y el Azucar*. Barcelona: Editorial Ariel.

Pagano, U. 2007. Cultural Globalisation, Institutional Diversity and the Unequal Accumulation of Intellectual Capital, *Cambridge Journal of Economics* 31: 649–667.

Prendergast, C. 2008. Work Incentives, Motivation, and Identity. *American Economic Review: Papers & Proceedings* 98(2): 201–205.

Rabin, M. 1993. Incorporating Fairness into Game Theory and Economics. *American Economic Review* 83(5): 1281-1302.

Samuelson, P. 1948. International Trade and the Equalization of Factor Prices. *Economics Journal* 58: 163-84.

Sokoloff, K.L., S.L. Engerman. 2000. Institutions, Factor Endowments, and Paths of Development in the New World. *Journal of Economic Perspectives* 14(3): 217-32.

Spolaore, E., R. Wacziarg 2009. The Diffusion of Development. *Quarterly Journal of Economics* 124 (2): 469-530.

Svaleryd, H., J. Vlachos. 2005. Financial Markets, the Pattern of Specialization and Comparative Advantage. Evidence from OECD Countries. *European Economic Review* 49: 113-44.

Tabellini, G. 2008. The Scope of Cooperation: Values and Incentives. *Quarterly Journal of Economics* 123(3): 905-950.

Tilly, C. 1990. *Coercion, Capital, and European States, AD 990-1990*. Cambridge, Mass., USA: Blackwell.

Vogel, J. 2007. Institutions and Moral Hazard in Open Economies. *Journal of International Economics* 71: 495-514.

Young, H.P. 1993. The Evolution of Conventions. *Econometrica* 61, pp. 57-84.

Young, H.P. 1998. *Individual Strategy and Social Structure: an Evolutionary Theory of Institutions*. Princeton, NJ: Princeton University Press.

A. Mathematical appendix

A more detailed appendix (appendix B that is attached to the manuscript) will be available from the authors upon request and posted on the first author’s website.

A.1. Nash equilibria. Given our assumptions on the production process, to show that {P,R} and {F,E} are Nash equilibria we need to prove that $\rho_1^i Q_L^i > \rho_2^i Q_N^i$ for any i=o,t. Recalling that $\rho_1^o = Q_L^t/(Q_L^t + Q_L^o)$ and $\rho_2^o = Q_N^t/(Q_N^t + Q_N^o)$, $\rho_1^o Q_L^o > \rho_2^o Q_N^o$ can be rewritten as $Q_L^t Q_L^o/(Q_L^t + Q_L^o) > Q_N^o Q_N^t/(Q_N^t + Q_N^o)$, which is true because it is equivalent to $Q_L^t Q_N^t (Q_L^o - Q_N^o) + Q_L^o Q_N^o (Q_L^t - Q_N^t) > 0$. Recalling that $\rho_1^t = Q_L^o/(Q_L^t + Q_L^o)$ and $\rho_2^t = Q_N^o/(Q_N^t + Q_N^o)$, the analogous proof for $\rho_1^t Q_L^t > \rho_2^t Q_N^t$ is straightforward.

A.2. Trade integration

A.2.1 Trade integration increases the costs of deviation (We only consider country 2, extension to country 1 being straightforward).

PART A: *Employers.* The cost of deviation in the {F,E} equilibrium is given by $v_F(\omega_2 = 0) - v_P(\omega_2 = 0)$, where $v_F(\omega_2 = 0)$ and $v_P(\omega_2 = 0)$ are given by equations (4) in the text with $\omega_2 = 0$. We easily obtain $v_F(\omega_2 = 0) - v_P(\omega_2 = 0) = \rho_2^i Q_N^i / 2 - (w + \mu_E)$, which is increasing in ρ_2^i.

PART B: *Employees.* Similarly, the corresponding cost of deviation for employees is given by $v_E(\phi_2 = 0) - v_R(\phi_2 = 0)$, where $v_E(\phi_2 = 0)$ and $v_R(\phi_2 = 0)$ are given by equations (5) in the text with $\phi_2 = 0$; thereby $v_E(\varphi_2 = 0) - v_R(\varphi_2 = 0) = \alpha(\rho_2^i Q_N^i - w - \mu_R)$, which is also increasing in ρ_2^i.

A.2.2 Trade integration increases the critical values $\boldsymbol{\omega^*}$ and $\boldsymbol{\phi^*}$. PART A: The derivative of ω_j^* given in (7) in the text with respect to ρ_j^i is

$$\frac{\mathrm{d}\omega_j^*}{\mathrm{d}\rho_j^i} = \frac{\dfrac{Q_N^i}{2}(\mu_R - \mu_E) + \left(\dfrac{Q_L^i}{2} - \dfrac{Q_N^i}{2}\right)(w + \mu_E)}{\left[\dfrac{\rho_j^i Q_L^i}{2} - \dfrac{\rho_j^i Q_N^i}{2} + (\mu_R - \mu_E)\right]^2} > 0,$$

which is equation (8) in the paper and is always positive because $Q_L^i > Q_N^i$ and $\mu_R \geq \mu_E$.

PART B: The derivative of ϕ_j^* also given in (7) with respect to ρ_j^i is

$$\frac{\mathrm{d}\phi_j^*}{\mathrm{d}\rho_j^i} = \frac{Q_N^i(-\delta + \eta) + \left[(1+\alpha)\dfrac{Q_L^i}{2} - \dfrac{Q_N^i}{2}\right](w + \mu_R)}{\left\{\left[(1+\alpha)\dfrac{\rho_j^i Q_L^i}{2} - \delta\right] - \left[\left(\dfrac{\rho_j^i Q_N^i}{2} - \eta\right) - \alpha(\rho_j^i Q_N^i - w - \mu_R)\right]\right\}^2},$$

which is equation (9) in the paper and is positive iff $\left[(1+\alpha)\frac{Q_L^i}{2}-\frac{Q_N^i}{2}\right](w+\mu_R)-Q_N^i(\delta-\eta)>0.$

A.2.3 Transition-inducing tariff rate. The transition-inducing tariff is given by $\theta^* = \min[\theta_\omega^*, \theta_\phi^*]$. The after-tariff price of the imported *o*-good in country 2 is $p_T^o(1+\theta_\omega^*)$.

PART A: By equating ω_2^* to zero, using the after-tariff trade prices, and then solving for $1+\theta_\omega^*$, we obtain:

$$\frac{p_T^t}{p_T^t + p_T^o(1+\theta_\omega^*)}\frac{Q_N^t}{2}-(w+\mu_E)=0, \text{i.e. } (1+\theta_\omega^*)=\left[\frac{Q_N^t}{2(w+\mu_E)}-1\right]\frac{p_T^t}{p_T^o},$$

which is the first of equations (10) in the text.

PART B: Similarly, by equating ϕ_2^* to zero and solving for $1+\theta_\phi^*$, we have:

$$\frac{p_T^t}{p_T^t + p_T^o(1+\theta_\phi^*)}Q_N^t-(w+\mu_R)=0, \text{i.e. } (1+\theta_\phi^*)=\left(\frac{Q_N^t}{w+\mu_R}-1\right)\frac{p_T^t}{p_T^o},$$

which is the second of equations (10) in the text.

A.3. Factor market integration

A.3.1 Country probability difference of matching *R*-employees and *P*-employers. In the neighborhood of the equalibria, the probability of an employer's being paired with a Reciprocal employee conditional to being resident, respectively, in country 1 and in country 2 are $n(1-\varepsilon)+(1-n)[s_1(1-\varepsilon)+s_2\varepsilon]$ and $n\varepsilon+(1-n)[s_1(1-\varepsilon)+s_2\varepsilon]$, where s_1 and s_2 are the sizes of country 1 and country 2. It is straightforward to see that the difference between the two is $n(1-\varepsilon)-n\varepsilon=n(1-2\varepsilon)$. Similar expressions are readily found for the corresponding country difference in the probability of an employee being paired with a Partnership.

A.3.2 Critical values $\boldsymbol{\omega^*(n)}$ and $\boldsymbol{\phi^*(n)}$ under factor market integration. The expected payoff in country 2 after factor integration is the weighted sum of the expected payoff in the national factor market plus the expected payoff in the common pool, the weights being the relative sizes of the two pools (n and $1-n$). The expected payoff in the common pool, in turn, is the weighted sum of the expected payoffs from matching an individual resident in country 1 and in country 2 with weights (respectively) s_1 and s_2. Notice that in computing the expected payoffs in country 2 (equations (A1) and (A3) below) the ω and ϕ appearing in the terms referring to own country matching are the distributions of play not the distribution of types (the two differ due to idiosyncratic play). Because we assume that all employers (employees) in country 1 are Partnership types (Reciprocators), taking account of idiosyncratic play, country 2 agents who are matched in the pool with agents from country 1 will with probability $1-\varepsilon$ encounter employers (employees) offering *P*-contracts (reciprocal types), while with probability ε will match employers (employees) offering F-contracts (self-regarding). The proofs contained in this subsection and in the following two are valid using both autarchic and trade prices. Clearly, if we consider trade prices it follows that $\rho_1^t = \rho_2^t = \rho_T^t$ and $\rho_1^o = \rho_2^o = \rho_T^o$, whereas if we consider autarchic prices we have $\rho_1^t > \rho_2^t$ and $\rho_1^o < \rho_2^o$; but our conclusions do not change in substance. To avoid ambiguity we use subscript 1 and 2 denoting the country (/equilibrium).

PART A: *Employers.* The expected payoffs to employers offering P- and F-contracts are:

$$v_P(\omega_2,n)=n\left[\omega_2\frac{\rho_2^tQ_L^t}{2}+(1-\omega_2)\frac{\rho_2^tQ_N^t}{2}\right]+(1-n)\left\{s_1\left[\frac{\rho_2^tQ_L^t}{2}(1-\varepsilon)+\frac{\rho_2^tQ_N^t}{2}\varepsilon\right]+s_2\left[\omega_2\frac{\rho_2^tQ_L^t}{2}+(1-\omega_2)\frac{\rho_2^tQ_N^t}{2}\right]\right\},$$

$$v_F(\omega_2,n)=n[\rho_2^tQ_N^t-(w+\mu_E)-\omega_2(\mu_R-\mu_E)]+(1-n)\{s_1[\rho_2^tQ_N^t-(w+\mu_R)+\varepsilon(\mu_R-\mu_E)]+s_2[\rho_2^tQ_N^t-(w+\mu_E)-\omega_2(\mu_R-\mu_E)]\}. \tag{A1}$$

To obtain $\omega_2^*(n)$, we compute the value of $\omega_2(n)$ such that $v_P(\omega_2,n)=v_F(\omega_2,n)$; after some manipulation it turns out to be

$$\omega_2^*(n)=\frac{\frac{(1-n)s_1}{ns_1+s_2}[\rho_2^tQ_N^t-(w+\mu_R)+\varepsilon(\mu_R-\mu_E)]-\frac{(1-n)s_1}{ns_1+s_2}\left[\frac{\rho_2^tQ_L^t}{2}(1-\varepsilon)+\frac{\rho_2^tQ_N^t}{2}\varepsilon\right]+\left[\frac{\rho_2^tQ_N^t}{2}-(w+\mu_E)\right]}{\left(\frac{\rho_2^tQ_L^t}{2}-\frac{\rho_2^tQ_N^t}{2}\right)+(\mu_R-\mu_E)}. \tag{A2}$$

PART B: *Employees*: The expected payoffs to R- and E-employees are:

$$v_R(\phi_2,n)=n\left\{\phi_2\left[(1+\alpha)\frac{\rho_2^tQ_L^t}{2}-\delta\right]+(1-\phi_2)[w-\eta-\alpha(\rho_2^tQ_N^t-w-\mu_R)]\right\}+(1-n)\left\{s_1\left[(1+\alpha)\frac{\rho_1^oQ_L^o}{2}-\delta\right](1-\varepsilon)+\right.$$
$$\left.+s_1[w-\eta-\alpha(\rho_1^oQ_N^o-w-\mu_R)]\varepsilon+s_2\phi_2\left[(1+\alpha)\frac{\rho_2^tQ_L^t}{2}-\delta\right]+s_2(1-\phi_2)[w-\eta-\alpha(\rho_2^tQ_N^t-w-\mu_R)]\right\},$$

$$v_E(\phi_2,n)=n\left[\phi_2\left(\frac{\rho_2^tQ_N^t}{2}-\eta\right)+(1-\phi_2)(w-\eta)\right]+(1-n)\left\{s_1\left[\left(\frac{\rho_1^oQ_N^o}{2}-\eta\right)(1-\varepsilon)+(w-\eta)\varepsilon\right]+s_2\left[\phi_2\left(\frac{\rho_2^tQ_N^t}{2}-\eta\right)+(1-\phi_2)(w-\eta)\right]\right\}. \tag{A3}$$

To find $\phi_2^*(n)$, we compute the value of $\phi_2(n)$ such that $v_R(\phi_2,n)=v_E(\phi_2,n)$. We obtain

$$\phi_2^*(n)=\frac{\frac{(1-n)s_1}{ns_1+s_2}\left(\frac{\rho_1^oQ_N^o}{2}-\eta\right)(1-\varepsilon)-\frac{(1-n)s_1}{ns_1+s_2}\left\{\left[(1+\alpha)\frac{\rho_1^oQ_L^o}{2}-\delta\right](1-\varepsilon)-\alpha(\rho_1^oQ_N^o-w-\mu_R)\varepsilon\right\}+\alpha(\rho_2^tQ_N^t-w-\mu_R)}{\left\{\left[(1+\alpha)\frac{\rho_2^tQ_L^t}{2}-\delta\right]+\alpha(\rho_2^tQ_N^t-w-\mu_R)\right\}-\left(\frac{\rho_2^tQ_N^t}{2}-\eta\right)}. \tag{A4}$$

A.3.3 Factor market integration decreases the costs of deviation.

PART A: *Employers.* The cost of deviation is given by $v_F(\omega_2=0,n)-v_P(\omega_2=0,n)$, where $v_F(\omega_2=0,n)$ and $v_P(\omega_2=0,n)$ are given by equations (A1) with $\omega_2=0$. This difference is smaller than the corresponding expression under factor immobility (n=1). This is easily shown by the fact that the expected payoff to an F-contract best responding employer under factor mobility is smaller than (or equal to) that under factor immobility because $\mu_R \geq \mu_E$, whereas the expected payoff to an idiosyncratic player offering a P-contract under factor mobility is greater than under factor immobility because $\rho_2^t Q_L^t / 2 > \rho_2^t Q_N^t / 2$.

PART B: *Employees.* The cost of deviation is given by $v_E(\phi_2 = 0, n) - v_R(\phi_2 = 0, n)$, where $v_E(\phi_2 = 0, n)$ and $v_R(\phi_2 = 0, n)$ are given by equations (A3) with $\phi_2 = 0$. This difference is smaller than the corresponding expression under factor immobility (n=1). Indeed, while both the expected payoff to E-type best responding employees and the expected payoff to idiosyncratic workers adopting R-preferences increase after factor market integration, the latter increases more than the former because $(1+\alpha)\rho_1^o Q_L^o / 2 - \delta > \rho_1^o Q_N^o / 2 - \eta$ and, as it easily proven, $\rho_2^t Q_N^t > \rho_1^o Q_N^o$.

A.3.4 Factor market integration decreases the critical values $\omega^*(n)$ and $\phi^*(n)$. PART A: Notice that the denominator of (A2) (which is positive) and the last term in squared brackets in the numerator does not depend on n. Then it is easily shown that $\mathrm{d}\omega_2^*(n)/\mathrm{d}n > 0$. Indeed $\mathrm{d}[(1-n)s_1/(ns_1+s_2)]/\mathrm{d}n < 0$ and

$$[\rho_2^t Q_N^t - (w+\mu_R) + \varepsilon(\mu_R - \mu_E)] - \left[\frac{\rho_2^t Q_L^t}{2}(1-\varepsilon) + \frac{\rho_2^t Q_N^t}{2}\varepsilon\right] < 0.$$

The above inequality is true because it is equivalent to $\varepsilon < 1 - \omega_2^*$, which follows from the persistence conditions (see section 4 in the text).

PART B: Notice that the denominator of (A4) (which is positive) and the last term in the numerator does not depend on n. Then, it is easily shown that $\mathrm{d}\phi_2^*(n)/\mathrm{d}n > 0$. Indeed $\mathrm{d}[(1-n)s_1/(ns_1+s_2)]/\mathrm{d}n < 0$ and

$$\left(\frac{\rho_1^o Q_N^o}{2} - \eta\right)(1-\varepsilon) - \left\{\left[(1+\alpha)\frac{\rho_1^o Q_L^o}{2} - \delta\right](1-\varepsilon) - \alpha(\rho_1^o Q_N^o - w - \mu_R)\varepsilon\right\} < 0.$$

The above inequality is true because it is equivalent to $\varepsilon < 1 - \phi_1^*$, which follows from the persistence conditions (see section 4 in the text).

B. DETAILED MATHEMATICAL APPENDIX (not intended for publication)

This appendix will be available from the authors upon request and posted on the first author's website.

B.1 Nash equilibria. {*P*,*R*} and {*F*,*E*} are proven to be Nash equilibria as long as: (*i*) $(1+\alpha)\rho_1^i Q_L^i /2-\delta > \rho_2^i Q_N^i /2-\eta$, (*ii*) $w-\eta > w-\eta-\alpha(\rho_2^i Q_N^i - w-\mu_R)$, (*iii*) $\rho_1^i Q_L^i /2 > \rho_2^i Q_N^i - w - \mu_R$, and (*iv*) $\rho_2^i Q_N^i - w-\mu_E > \rho_2^i Q_N^i /2$. Inequality (*ii*) is self-explained. The other inequalities are verified to be true given our assumptions on the production process that $(1+\alpha)\rho_j^i Q_L^i /2-\delta > \rho_j^i Q_N^i /2-\eta$ and $\rho_2^i Q_N^i > 2(w-\mu_E)$, as long as $\rho_1^i Q_L^i > \rho_2^i Q_N^i$ for any *i*=*o*,*t*. Recalling that $\rho_1^o = Q_L^t /(Q_L^t + Q_L^o)$ and $\rho_2^o = Q_N^t /(Q_N^t + Q_N^o)$, $\rho_1^o Q_L^o > \rho_2^o Q_N^o$ can be rewritten as $Q_L^t Q_L^o /(Q_L^t + Q_L^o) > Q_N^o Q_N^t /(Q_N^t + Q_N^o)$, which is true because it is equivalent to $Q_L^t Q_N^t (Q_L^o - Q_N^o) + Q_L^o Q_N^o (Q_L^t - Q_N^t) > 0$. Recalling that $\rho_1^t = Q_L^o /(Q_L^t + Q_L^o)$ and $\rho_2^t = Q_N^o /(Q_N^t + Q_N^o)$, the analogous proof for $\rho_1^t Q_L^t > \rho_2^t Q_N^t$ is straightforward.

B.2. Trade integration

B.2.1 Critical values ω^* and ϕ^* in autarchy.

PART A: *Employers.* The expected payoffs to employers offering respectively *P*- and *F*-contracts, where *i*=*o*,*t* and *j*=1,2, are:

$$
\begin{aligned}
v_P(\omega_j) &= \omega_j \frac{\rho_j^i Q_L^i}{2} + (1-\omega_j)\frac{\rho_j^i Q_N^i}{2}, \\
v_F(\omega_j) &= \omega_j[\rho_j^i Q_N^i - (w+\mu_R)] + (1-\omega_j)[\rho_j^i Q_N^i - (w+\mu_E)] \qquad \text{(B1)} \\
&= [\rho_j^i Q_N^i - (w+\mu_E)] - \omega_j(\mu_R - \mu_E).
\end{aligned}
$$

ω_j^* is the level of ω_j such that $v_P(\omega_j) = v_F(\omega_j)$, i.e.

$$\omega_j \frac{\rho_j^i Q_L^i}{2} + (1-\omega_j)\frac{\rho_j^i Q_N^i}{2} = [\rho_j^i Q_N^i - (w+\mu_E)] - \omega_j(\mu_R - \mu_E),$$

hence

$$\omega_j^* = \frac{\dfrac{\rho_j^i Q_N^i}{2} - (w+\mu_E)}{\dfrac{\rho_j^i Q_L^i}{2} - \dfrac{\rho_j^i Q_N^i}{2} + (\mu_R - \mu_E)}, \qquad \text{(B2)}$$

which is the first of equations (7) in the paper.

PART B: *Employees.* Similarly, the expected payoffs to respectively *R*- and *E*-employees are:

$$v_R(\phi_j) = \phi_j\left[(1+\alpha)\frac{\rho_j^i Q_L^i}{2} - \delta\right] + (1-\phi_j)[w - \eta - \alpha(\rho_j^i Q_N^i - w - \mu_R)],$$
$$v_E(\phi_j) = \phi_j\left(\frac{\rho_j^i Q_N^i}{2} - \eta\right) + (1-\phi_j)(w-\eta). \tag{B3}$$

ϕ_j^* is the value of ϕ_j such that $v_R(\phi_j) = v_E(\phi_j)$, i.e.

$$\phi_j\left[(1+\alpha)\frac{\rho_j^i Q_L^i}{2} - \delta\right] + (1-\phi_j)[w - \eta - \alpha(\rho_j^i Q_N^i - w - \mu_R)] = \phi_j\left(\frac{\rho_j^i Q_N^i}{2} - \eta\right) + (1-\phi_j)(w-\eta),$$

hence

$$\phi_j^* = \frac{\alpha(\rho_j^i Q_N^i - w - \mu_R)}{\left[(1+\alpha)\frac{\rho_j^i Q_L^i}{2} - \delta\right] - \left(\frac{\rho_j^i Q_N^i}{2} - \eta\right) + \alpha(\rho_j^i Q_N^i - w - \mu_R)}, \tag{B4}$$

which is the second of equations (7) in the paper.

B.2.2 Effects of trade integration on the costs of deviation. Trade integration, i.e. an increase in ρ_j^i, increases the cost of deviating from the status quo cultural-institutional convention.

Equilibrium 1: PART A: *Employers.* Rewrite the expected payoff equations for employers offering respectively *P*- and *F*-contracts when all the employees in the previous period were Reciprocators (i.e. equations (B1) with *j*=1 and $\omega_1 = 1$):

$$v_P(\omega_1 = 1) = \frac{\rho_1^i Q_L^i}{2},$$
$$v_F(\omega_1 = 1) = \rho_1^i Q_N^i - (w + \mu_R). \tag{B5}$$

The cost of deviation in the {*P*,*R*} equilibrium is given by $v_P(\omega_1 = 1) - v_F(\omega_1 = 1)$. Using equations (B5) this is equivalent to

$$v_P(\omega_1 = 1) - v_F(\omega_1 = 1) = \frac{\rho_1^i Q_L^i}{2} - \rho_1^i Q_N^i + (w + \mu_R), \tag{B6}$$

which is increasing in ρ_1^i, because, as explained in the paper, $Q_L^i / 2 - Q_N^i$.

PART B: *Employees.* Similarly, the expected payoff equations for respectively *R*- and *E*-employees when all the employers in the previous period were offering *P*-contracts (i.e. equations (B3) with *j*=1 and $\phi_1 = 1$) may be rewritten as:

$$v_R(\phi_1 = 1) = (1+\alpha)\frac{\rho_1^i Q_L^i}{2} - \delta,$$
$$v_E(\phi_1 = 1) = \frac{\rho_1^i Q_N^i}{2} - \eta. \tag{B7}$$

The cost of deviation in the {*P*,*R*} equilibrium is thus given by $v_R(\phi_1 = 1) - v_E(\phi_1 = 1)$ which, using equations (B7), can be rewritten as

$$v_R(\phi_1 = 1) - v_E(\phi_1 = 1) = \left[(1+\alpha)\frac{\rho_1^i Q_L^i}{2} - \delta\right] - \left(\frac{\rho_1^i Q_N^i}{2} - \eta\right), \tag{B8}$$

which is also increasing in ρ_1^i, because $\rho_1^i Q_L^i > \rho_1^i Q_N^i$.

Equilibrium 2: PART A: *Employers.* Expected payoff equations for *P*- and *F*-contract employers when all the employees in the previous period were Homo economicus (i.e. equations (B1) with *j*=2 and $\omega_2 = 0$) are:

$$\begin{aligned} v_P(\omega_2 = 0) &= \frac{\rho_2^i Q_N^i}{2}, \\ v_F(\omega_2 = 0) &= \rho_2^i Q_N^i - (w + \mu_E). \end{aligned} \tag{B9}$$

The cost of deviation in the {*F*,*E*} equilibrium is given by $v_F(\omega_2 = 0) - v_P(\omega_2 = 0)$. Using equations (B9) this is equivalent to

$$v_F(\omega_2 = 0) - v_P(\omega_2 = 0) = \frac{\rho_2^i Q_N^i}{2} - (w + \mu_E), \tag{B10}$$

which is increasing in ρ_2^i.

PART B: *Employees.* Similarly, expected payoff equations for respectively *R*- and *E*-employees when all the employers in the previous period were offering *F*-contracts (i.e. equations (B3) with *j*=2 and $\phi_2 = 0$) may be rewritten as:

$$\begin{aligned} v_R(\phi_2 = 0) &= w - \eta - \alpha(\rho_2^i Q_N^i - w - \mu_R), \\ v_E(\phi_2 = 0) &= w - \eta. \end{aligned} \tag{B11}$$

The cost of deviation in the {*F*,*E*} equilibrium is given by $v_E(\phi_2 = 0) - v_R(\phi_2 = 0)$ that, using equations (B11), turns out to be

$$v_E(\phi_2 = 0) - v_R(\phi_2 = 0) = \alpha(\rho_2^i Q_N^i - w - \mu_R), \tag{B12}$$

which is also increasing in ρ_2^i.

B.2.3 Effects of trade integration on the critical values ω^* and ϕ^*. Trade integration (increase in ρ_j^i) leads to an increase in the expected number of idiosyncratic players in either class (employers and employees) sufficient to induce a transition from the {*F*,*E*} to the {*P*,*R*} equilibrium. To show this we study the sign of the derivatives of ω_j^* and ϕ_j^* with respect to ρ_j^i.

PART A: Using expression (B2), the former is

$$\frac{\mathrm{d}\omega_j^*}{\mathrm{d}\rho_j^i}=\frac{\frac{Q_N^i}{2}\left[\frac{\rho_j^iQ_L^i}{2}-\frac{\rho_j^iQ_N^i}{2}+(\mu_R-\mu_E)\right]-\left(\frac{Q_L^i}{2}-\frac{Q_N^i}{2}\right)\left[\frac{\rho_j^iQ_N^i}{2}-(w+\mu_E)\right]}{\left[\frac{\rho_j^iQ_L^i}{2}-\frac{\rho_j^iQ_N^i}{2}+(\mu_R-\mu_E)\right]^2}=\frac{\frac{Q_N^i}{2}(\mu_R-\mu_E)+\left(\frac{Q_L^i}{2}-\frac{Q_N^i}{2}\right)(w+\mu_E)}{\left[\frac{\rho_j^iQ_L^i}{2}-\frac{\rho_j^iQ_N^i}{2}+(\mu_R-\mu_E)\right]^2}>0,$$

which is equation (8) in the paper and is always positive because $Q_L^i > Q_N^i$ and $\mu_R \geq \mu_E$.

PART B: Analogously, using (B4), the latter can be written as

$$\frac{\mathrm{d}\phi_j^*}{\mathrm{d}\rho_j^i}=\frac{\alpha Q_N^i\left\{\left[(1+\alpha)\frac{\rho_j^iQ_L^i}{2}-\delta\right]-\left[\left(\frac{\rho_j^iQ_N^i}{2}-\eta\right)-\alpha(\rho_j^iQ_N^i-w-\mu_R)\right]\right\}-\alpha(\rho_j^iQ_N^i-w-\mu_R)\left[(1+\alpha)\frac{Q_L^i}{2}-\frac{Q_N^i}{2}+\alpha Q_N^i\right]}{\left\{\left[(1+\alpha)\frac{\rho_j^iQ_L^i}{2}-\delta\right]-\left[\left(\frac{\rho_j^iQ_N^i}{2}-\eta\right)-\alpha(\rho_j^iQ_N^i-w-\mu_R)\right]\right\}^2}$$

$$=\frac{Q_N^i(-\delta+\eta)+\left[(1+\alpha)\frac{Q_L^i}{2}-\frac{Q_N^i}{2}\right](w+\mu_R)}{\left\{\left[(1+\alpha)\frac{\rho_j^iQ_L^i}{2}-\delta\right]-\left[\left(\frac{\rho_j^iQ_N^i}{2}-\eta\right)-\alpha(\rho_j^iQ_N^i-w-\mu_R)\right]\right\}^2},$$

which is equation (9) in the paper and is positive iff $\left[(1+\alpha)\frac{Q_L^i}{2}-\frac{Q_N^i}{2}\right](w+\mu_R)-Q_N^i(\delta-\eta)>0.$

B.2.4 Transition-inducing tariff rate. $\theta^* > 0$ is the tariff protecting the opaque good in country 2 such that a cultural-institutional transition from the $\{F,E\}$ to the $\{P,R\}$ convention will occur. Given the international price ratio, θ_ω^* and θ_ϕ^* are the ad-valorem tariff rates such that, respectively, $\omega_2^*=0$ and $\phi_2^*=0$. The transition-inducing tariff is given by $\theta^*=\min[\theta_\omega^*,\theta_\phi^*]$. The after-tariff price of the imported *o*-good in country 2 is $p_T^o(1+\theta_\omega^*)$.

PART A: By equating (B2) to zero, setting *i*=*t*, using the after-tariff trade prices, and then solving for $1+\theta_\omega^*$, we obtain:

$$\frac{p_T^t}{p_T^t+p_T^o(1+\theta_\omega^*)}\frac{Q_N^t}{2}-(w+\mu_E)=0,\text{ i.e. }(1+\theta_\omega^*)=\left[\frac{Q_N^t}{2(w+\mu_E)}-1\right]\frac{p_T^t}{p_T^o},$$

which is the first of equations (10) in the paper.

PART B: Similarly, using (B4) and solving for $1+\theta_\phi^*$, we have:

$$\frac{p_T^t}{p_T^t+p_T^o(1+\theta_\phi^*)}Q_N^t-(w+\mu_R)=0,\text{ i.e. }(1+\theta_\phi^*)=\left(\frac{Q_N^t}{w+\mu_R}-1\right)\frac{p_T^t}{p_T^o},$$

which is the second of equations (10) in the paper.

B.3. Factor market integration

B.3.1 Critical values $\omega^*(n)$ and $\phi^*(n)$ under factor market integration.

PART A: *Employers.* The expected payoffs to employers offering *P*- and *F*-contracts after factor market integration are (notice the superscript referring to the good and the subscript referring to the country do not change in the pool because, as explained in the paper, the nationality of the employer determines the good produced and the prices at which the output is sold):

$$v_P(\omega_2,n)=n\left[\omega_2\frac{\rho_2^tQ_L^t}{2}+(1-\omega_2)\frac{\rho_2^tQ_N^t}{2}\right]+(1-n)\left\{s_1\left[\frac{\rho_2^tQ_L^t}{2}(1-\varepsilon)+\frac{\rho_2^tQ_N^t}{2}\varepsilon\right]+s_2\left[\omega_2\frac{\rho_2^tQ_L^t}{2}+(1-\omega_2)\frac{\rho_2^tQ_N^t}{2}\right]\right\},$$

$$v_F(\omega_2,n)=n[\rho_2^tQ_N^t-(w+\mu_E)-\omega_2(\mu_R-\mu_E)]+(1-n)\{s_1[\rho_2^tQ_N^t-(w+\mu_R)+\varepsilon(\mu_R-\mu_E)]+s_2[\rho_2^tQ_N^t-(w+\mu_E)-\omega_2(\mu_R-\mu_E)]\}$$

(B13)

To obtain $\omega_2^*(n)$, we compute the value of $\omega_2(n)$ such that $v_P(\omega_2,n)=v_F(\omega_2,n)$. It follows:

$$(ns_1+s_2)\left[\omega_2\frac{\rho_2^tQ_L^t}{2}+(1-\omega_2)\frac{\rho_2^tQ_N^t}{2}\right]+(1-n)s_1\left[\frac{\rho_2^tQ_L^t}{2}(1-\varepsilon)+\frac{\rho_2^tQ_N^t}{2}\varepsilon\right]=$$

$$=(ns_1+s_2)[\rho_2^tQ_N^t-(w+\mu_E)-\omega_2(\mu_R-\mu_E)]+(1-n)s_1[\rho_2^tQ_N^t-(w+\mu_R)+\varepsilon(\mu_R-\mu_E)],$$

whereby,

$$\omega_2(ns_1+s_2)\left(\frac{\rho_2^tQ_L^t}{2}-\frac{\rho_2^tQ_N^t}{2}\right)+\omega_2(ns_1+s_2)(\mu_R-\mu_E)=-(ns_1+s_2)\frac{\rho_2^tQ_N^t}{2}+$$

$$-(1-n)s_1\left[\frac{\rho_2^tQ_L^t}{2}(1-\varepsilon)+\frac{\rho_2^tQ_N^t}{2}\varepsilon\right]+(ns_1+s_2)[\rho_2^tQ_N^t-(w+\mu_E)+(1-n)s_1[\rho_2^tQ_N^t-(w+\mu_R)+\varepsilon(\mu_R-\mu_E)].$$

Finally, after manipulation, we obtain

$$\omega_2^*(n)=\frac{\frac{(1-n)s_1}{ns_1+s_2}\left[\rho_2^tQ_N^t-(w+\mu_R)+\varepsilon(\mu_R-\mu_E)\right]-\frac{(1-n)s_1}{ns_1+s_2}\left[\frac{\rho_2^tQ_L^t}{2}(1-\varepsilon)+\frac{\rho_2^tQ_N^t}{2}\varepsilon\right]+\left[\frac{\rho_2^tQ_N^t}{2}-(w+\mu_E)\right]}{\left(\frac{\rho_2^tQ_L^t}{2}-\frac{\rho_2^tQ_N^t}{2}\right)+(\mu_R-\mu_E)}. \quad \text{(B14)}$$

PART B: *Employees*: The expected payoffs to *R*- and *E*-employees after factor market integration are (notice that the superscript referring to the good and the subscript referring to the country change in the pool because, as explained in the paper, the nationality of the employer determines the good produced and the prices at which the output is sold):

$$v_R(\phi_2,n)=n\left\{\phi_2\left[(1+\alpha)\frac{\rho_2^tQ_L^t}{2}-\delta\right]+(1-\phi_2)[w-\eta-\alpha(\rho_2^tQ_N^t-w-\mu_R)]\right\}+(1-n)\left\{s_1\left[(1+\alpha)\frac{\rho_1^oQ_L^o}{2}-\delta\right](1-\varepsilon)+\right.$$

$$\left.+s_1[w-\eta-\alpha(\rho_1^oQ_N^o-w-\mu_R)]\varepsilon+s_2\phi_2\left[(1+\alpha)\frac{\rho_2^tQ_L^t}{2}-\delta\right]+s_2(1-\phi_2)[w-\eta-\alpha(\rho_2^tQ_N^t-w-\mu_R)]\right\},$$

$$v_E(\phi_2,n)=n\left[\phi_2\left(\frac{\rho_2^tQ_N^t}{2}-\eta\right)+(1-\phi_2)(w-\eta)\right]+(1-n)\left\{s_1\left[\left(\frac{\rho_1^oQ_N^o}{2}-\eta\right)(1-\varepsilon)+(w-\eta)\varepsilon\right]+s_2\left[\phi_2\left(\frac{\rho_2^tQ_N^t}{2}-\eta\right)+(1-\phi_2)(w-\eta)\right]\right\}.$$

(B15)

To obtain $\phi_2^*(n)$, we compute the value of $\phi_2(n)$ such that $v_R(\phi_2,n)=v_E(\phi_2,n)$. We can write

$$(ns_1+s_2)\left\{\phi_2\left[(1+\alpha)\frac{\rho_2^tQ_L^t}{2}-\delta\right]+(1-\phi_2)[w-\eta-\alpha(\rho_2^tQ_N^t-w-\mu_R)]\right\}+(1-n)\left\{s_1\left[(1+\alpha)\frac{\rho_1^oQ_L^o}{2}-\delta\right](1-\varepsilon)+\right.$$

$$\left.+s_1[w-\eta-\alpha(\rho_1^oQ_N^o-w-\mu_R)]\varepsilon\right\}=(ns_1+s_2)\left[\phi_2\left(\frac{\rho_2^tQ_N^t}{2}-\eta\right)+(1-\phi_2)(w-\eta)\right]+(1-n)s_1\left[\left(\frac{\rho_1^oQ_N^o}{2}-\eta\right)(1-\varepsilon)+(w-\eta)\varepsilon\right],$$

whereby

$$\phi_2(ns_1+s_2)\left\{\left[(1+\alpha)\frac{\rho_2^tQ_L^t}{2}-\delta\right]-[w-\eta-\alpha(\rho_2^tQ_N^t-w-\mu_R)]\right\}-\phi_2(ns_1+s_2)\left[\left(\frac{\rho_2^tQ_N^t}{2}-\eta\right)-(w-\eta)\right]=-(ns_1+s_2)[w-\eta-\alpha(\rho_2^tQ_N^t-w-\mu_R)]+$$

$$-(1-n)\left\{s_1\left[(1+\alpha)\frac{\rho_1^oQ_L^o}{2}-\delta\right](1-\varepsilon)+s_1[w-\eta-\alpha(\rho_1^oQ_N^o-w-\mu_R)]\varepsilon\right\}+(ns_1+s_2)(w-\eta)+(1-n)s_1\left[\left(\frac{\rho_1^oQ_N^o}{2}-\eta\right)(1-\varepsilon)+(w-\eta)\varepsilon\right].$$

Finally, we obtain

$$\phi_2^*(n)=\frac{\frac{(1-n)s_1}{ns_1+s_2}\left(\frac{\rho_1^oQ_N^o}{2}-\eta\right)(1-\varepsilon)-\frac{(1-n)s_1}{ns_1+s_2}\left\{\left[(1+\alpha)\frac{\rho_1^oQ_L^o}{2}-\delta\right](1-\varepsilon)-\alpha(\rho_1^oQ_N^o-w-\mu_R)\varepsilon\right\}+\alpha(\rho_2^tQ_N^t-w-\mu_R)}{\left\{\left[(1+\alpha)\frac{\rho_2^tQ_L^t}{2}-\delta\right]+\alpha(\rho_2^tQ_N^t-w-\mu_R)\right\}-\left(\frac{\rho_2^tQ_N^t}{2}-\eta\right)}.\text{(B}$$

16)

B.3.2 Effects of factor market integration on the costs of deviation. The cost of deviation from the best response convention in the $\{F,E\}$ cultural-institutional equilibrium for both employers and employees decreases after factor market integration (extension to the $\{P,R\}$ convention is straightforward).

PART A: *Employers.* First, we write the expected payoff equations for employers under factor market integration when all the employees in the previous period were self-regarding. These are given by equations (B13) with $\omega_2=0$,

$$v_P(\omega_2=0,n)=n\frac{\rho_2^tQ_N^t}{2}+(1-n)\left\{s_1\left[\frac{\rho_2^tQ_L^t}{2}(1-\varepsilon)+\frac{\rho_2^tQ_N^t}{2}\varepsilon\right]+s_2\frac{\rho_2^tQ_N^t}{2}\right\},$$

$$v_F(\omega_2=0,n)=n[\rho_2^tQ_N^t-(w+\mu_E)]+(1-n)\left\{s_1\left[\rho_2^tQ_N^t-(w+\mu_R)+\varepsilon(\mu_R-\mu_E)\right]+s_2[\rho_2^tQ_N^t-(w+\mu_E)]\right\}.$$

(B17)

The cost of deviation is given by $v_F(\omega_2 = 0, n) - v_P(\omega_2 = 0, n)$. This difference is smaller than the corresponding expression under factor immobility (n=1) given in (B10) (notice that if trade is considered i=t by specialization, whereas if autarchy is considered the value of the output is invariant across sectors i=o,t). This is easily shown by the fact that the expected payoff to an F-contract best responding employer under factor mobility (second of equations (B17)) is smaller than (or equal as) that under factor immobility (second of equations (B9) with i=t) because $\mu_R \geq \mu_E$, whereas the expected payoff to an idiosyncratic player offering a P-contract under factor mobility (first of equations (B17)) is greater than under factor immobility (first of equations (B9) with i=t) because $\rho_2^t Q_L^t / 2 > \rho_2^t Q_N^t / 2$.

PART B: *Employees.* The expected payoff equations for employees under factor mobility when all the employers in the previous period were offering F-contracts, i.e. equations (B15) with $\phi_2 = 0$, may be written as:

$$
\begin{aligned}
v_R(\phi_2=0,n) &= n[w-\eta-\alpha(\rho_2^t Q_N^t - w - \mu_R)] + (1-n)\left\{ s_1\left[(1+\alpha)\frac{\rho_1^o Q_L^o}{2} - \delta\right](1-\varepsilon) + s_1[w-\eta-\alpha(\rho_1^o Q_N^o - w - \mu_R)]\varepsilon + \right. \\
&\quad \left. + s_2[w-\eta-\alpha(\rho_2^t Q_N^t - w - \mu_R)] \right\} \\
v_E(\phi_2=0,n) &= n(w-\eta) + (1-n)\left\{ s_1\left[\left(\frac{\rho_1^o Q_N^o}{2} - \eta\right)(1-\varepsilon) + (w-\eta)\varepsilon\right] + (w-\eta) \right\}.
\end{aligned}
$$

(B18)

The cost of deviation is given by $v_E(\phi_2 = 0, n) - v_R(\phi_2 = 0, n)$, which is smaller than the corresponding expression under factor immobility (n=1) given in (B12). Indeed, while both the expected payoff to E-type best responding employees (second of equations (B18)) and the expected payoff to idiosyncratic workers adopting R-preferences (first of equations (B18)) increase after factor market integration, the latter increases more than the former because $(1+\alpha)\rho_1^o Q_L^o / 2 - \delta > \rho_1^o Q_N^o / 2 - \eta$ and, as it easily proven, $\rho_2^t Q_N^t > \rho_1^o Q_N^o$.

B.3.3 Effects of factor market integration on the critical values $\boldsymbol{\omega^*(n)}$ and $\boldsymbol{\phi^*(n)}$. Factor market integration leads to a decrease in the expected number of idiosyncratic players in either class (employers and employees) sufficient to induce a transition from the {F,E} to the {P,R} cultural-institutional convention. To show this, we study the sign of the derivative of $\omega_2^*(n)$ and $\phi_2^*(n)$, given respectively by (B14) and (B16), with respect to n.

PART A: To study the sign of $\mathrm{d}\omega_2^*(n)/\mathrm{d}n$, notice that the denominator of (B14) (which is positive) and the last term in squared brackets in the numerator does not depend on n. Then it is easily shown that $\mathrm{d}\omega_2^*(n)/\mathrm{d}n > 0$. Indeed $\mathrm{d}[(1-n)s_1/(ns_1 + s_2)]/\mathrm{d}n < 0$ and

$$\left[\rho_2^t Q_N^t - (w + \mu_R) + \varepsilon(\mu_R - \mu_E)\right] - \left[\frac{\rho_2^t Q_L^t}{2}(1-\varepsilon) + \frac{\rho_2^t Q_N^t}{2}\varepsilon\right] < 0.$$

The above inequality is true because it can be rewritten as

$$\varepsilon < 1 - \frac{\frac{\rho_2^t Q_N^t}{2} - w - \mu_E}{\frac{\rho_2^t Q_L^t}{2} - \frac{\rho_2^t Q_N^t}{2} + (\mu_R - \mu_E)} = 1 - \omega_2^*, \qquad \text{(B19)}$$

which follows from the persistence conditions (see section 4 in the paper).

PART B: To study the sign of $\mathrm{d}\phi_2^*(n)/\mathrm{d}n$, notice that the denominator of (A4) (which is positive) and the last term in the numerator does not depend on *n*. Then, it is easily shown that $\mathrm{d}\phi_2^*(n)/\mathrm{d}n > 0$. Indeed $\mathrm{d}[(1-n)s_1/(ns_1 + s_2)]/\mathrm{d}n < 0$ and

$$\left(\frac{\rho_1^o Q_N^o}{2} - \eta\right)(1-\varepsilon) - \left\{\left[(1+\alpha)\frac{\rho_1^o Q_L^o}{2} - \delta\right](1-\varepsilon) - \alpha(\rho_1^o Q_N^o - w - \mu_R)\varepsilon\right\} < 0.$$

The above inequality is true because it can be rewritten as

$$\varepsilon < 1 - \frac{\alpha(\rho_1^o Q_N^o - w - \mu_R)}{\left[(1+\alpha)\frac{\rho_1^o Q_L^o}{2} - \delta\right] - \left(\frac{\rho_1^o Q_N^o}{2} - \eta\right) + \alpha(\rho_1^o Q_N^o - w - \mu_R)} = 1 - \phi_1^*, \qquad \text{(B20)}$$

which follows from the persistence conditions (see section 4 in the paper).