Bilateral trade agreements and the feasibility of multilateral free trade*

Kamal Saggi† and Halis Murat Yildiz‡

First draft: June 2004. This draft: July 2005

Abstract

What is the relationship between preferential and multilateral trade liberalization? Does the option to form free trade agreements (FTAs) reduce the likelihood of obtaining global free trade? In a three-country model of intraindustry trade, we show that when the degree of cost asymmetry between countries is small, global free trade is *less* likely to occur when FTAs are permissible (*stumbling bloc effect*). However, when the degree of cost asymmetry is sufficiently high, global free trade is *infeasible* whereas welfare improving FTAs are *feasible*. In fact, there exist circumstances where (*i*) FTAs can even yield higher world welfare than global free trade and (*ii*) multilateral free trade is an equilibrium *only if* countries have the option to form FTAs (*building bloc effect*).

Keywords: Multilateral Trade Liberalization, Free Trade Agreements, GATT, Intraindustry Trade, Oilgopoly. **JEL Classifications**: F13, F12.

*We thank seminar audiences at Bilkent University, McGill University, Sabanci University, Syrcause University, Southern Methodist University, University of Bologna, and York University for helpful comments. All errors are our own.

†Department of Economics, Southern Methodist University, Dallas, TX 75275-0496. Phone: 214-768-3274; fax: 214-768-1821; e-mail: ksaggi@smu.edu.

‡Department of Economics, Ryerson University, 350 Victoria Street, Toronto, ON, Canada M5B 2K3. Phone: 416-979-5000 (ext 6689); fax: 416-979-5289; e-mail: hyildiz@ryerson.ca.

1. Introduction

By their very nature, free trade agreements (FTAs) require member countries to grant tariff concessions to each other that they typically do not extend to non-members. Ever since Jacob Viner (1950)'s classic analysis, the static distortions created by such preferential trade liberalization have received significant attention from economists and policy-makers alike. Furthermore, in recent years there has been widespread concern regarding the potential adverse effects of FTAs on the process of multilateral trade liberalization (the *raison d'etre* of the World Trade Organization (WTO)). This concern appears to be rather well-founded – so widespread are preferential trade agreements (PTAs) today that most favored nation (MFN) treatment has begun to appear more of an exception rather than a core rule of the WTO. As per the WTO, over 170 PTAs are in force today and most countries participate in some type of a PTA or another.[1]

How does the pursuit of FTAs interact with the process of multilateral trade liberalization? Would global free trade be easier to achieve if countries were to pursue trade liberalization only multilaterally?[2] To address these questions, we evaluate incentives for multilateral trade liberalization under two scenarios: one where countries can pursue either bilateral or multilateral trade liberalization and another where only the multilateral option is available to them. Formally, we analyze the coalition proof Nash equilibria of two tariff games (called an FTA

[1] Summary facts on the proliferation of PTAs (also called RTAs – regional trade agreements) are available at the WTO's web-site: http://www.wto.org/.

[2] Our first question is related (but not completely equivalent) to the question posed by Jagdish Bhagwati (1991): "Are preferential trade agreements (PTAs) building or stumbling blocks for multilateral trade liberalization?" The difference is that according to our view, both preferential and multilateral trade liberalization are endogenous and GATT Article XXIV – which sanctions PTAs so long as they abide by certain conditions – is the underlying exogenous factor.

game and a No FTA game) between three countries that differ with respect to their production costs. The underlying framework is one of intraindustry trade under oligopoly.

The FTA game proceeds as follows. In the first stage, each country announces the set of countries with whom it wants to form a trade agreement. An FTA between two countries requires them to abolish tariffs on each other and it arises iff they both announce each other's name. Similarly, free trade emerges iff all countries call each other's names. In second stage of the game, those countries that do not announce in favor of free trade, choose their tariffs to maximize national welfare (defined as the sum of the local firm's profits, consumer surplus, and tariff revenue). Finally, firms compete in the product market in a Cournot fashion where individual country markets are assumed to be segmented. In contrast to the FTA game, in the first stage of the No FTA game, countries can choose between only two alternatives: they can either announce in favor of free trade or no trade agreement at all (wherein they impose their individually optimal tariffs on each other). The rest of the No FTA game proceeds just like the FTA game.

From each country's perspective, an FTA embodies the following trade-off. On the one hand, joining an FTA lowers a country's domestic surplus relative to the case where it can use its optimally chosen tariff(s). On the other hand, being part of an FTA increases export profits in the markets of other member countries. Analysis of the coalition proof Nash equilibria of the two games delivers several interesting results. First, despite the absence of any political economy considerations in our model, free trade can fail to be an equilibrium even when FTAs are *not* permissible. Second, while the option to form FTAs necessarily reduces the likelihood of obtaining a free trade equilibrium when countries are symmetric (*stumbling bloc effect*), such need not be the case under asymmetry – in particular, there exist circumstances where multilateral free trade is an equilibrium *only if* countries have the option to form FTAs (*building bloc effect*). Third, as Summers (1991) argued, we show that welfare improving FTAs can be feasible when global free trade is not.[3] In fact, a pattern of FTAs where a low cost country forms independent FTAs with two high cost ones can yield higher aggregate welfare than global free trade.

In a similar three-country model, Krishna (1998) has shown that an FTA between two countries reduces their incentives to liberalize trade with respect to the third country.[4] Unlike us, Krishna (1998) does not derive equilibrium FTAs and assumes that tariffs are exogenously given. In a recent paper, Aghion et. al. (2004) examine a leading country's choice between sequential and multilateral bargaining of free trade agreements. Like us, Aghion et. al. (2004) also identify building and stumbling bloc effects of FTAs. However, there are important differences between their approach and ours. First, in our model, *all* countries are free to negotiate FTAs and not just a single leading country.[5] Also, countries are free to form a pair of bilateral FTAs in our model and are not required to choose between joining a single grand coalition or staying out. Second, tariffs plays a

[3] The literature (see, for example, Lawrence, 1996) has noted that FTAs may allow members to implement "deeper integration" relative to what is possible multilaterally. For example, the North American Free Trade Agreement was able to achieve substantial liberalization in the area of foreign investment whereas multilateral investment liberalization (as evidenced by efforts undertaken at the WTO as well as by several European countries) has been rather difficult to achieve.

[4] The literature on PTAs is rather extensive and we only discuss closely related papers. The reader is referred to Bhagwati et. al. (1999) for a collection of many of the important papers in the area.

[5] Aghion et. al. (2004) do consider extensions where the leadership role is assigned to other countries if the first leader's offer is not accepted by the followers but they focus on deriving necessary conditions for a free trade equilibrium.

crucial role in determining the nature and magnitude of externalities created by FTAs. For example, we show that an FTA member's tariff on the non-member is lower than its Nash tariff under no agreement and whether a non-member enjoys positive or negative externalities from the formation of an FTA depends upon the FTAs external tariffs. Third, our analysis of building and stumbling bloc effects of FTAs complements theirs in two important respects (*i*) we assume governments maximize aggregate social welfare whereas their examples illustrating the effects of FTAs assume governments care only about producer surplus and (*ii*) unlike them we do not allow transfers between different coalitions. Point (*ii*) is important because when transfers are possible and there is grand coalition superadditivity along with the absence of externalities, global free trade necessarily emerges in equilibrium regardless of whether the leading country chooses a sequential or multilateral approach.[6] In our model, even when free trade is Pareto optimal (as it is under symmetry), the externalities created by FTAs can keep free trade from obtaining as an equilibrium.

Our approach is also related to that of Riezman (1999) who also asks whether the FTA option facilitates or hinders the achievement of free trade. However, while we analytically derive the coalition proof Nash equilibria of two non-cooperative games, Riezman (1999) uses the cooperative solution concept of the core and illustrates his results via numerical examples. Second, our model allows us to focus on asymmetries between countries in a way that cannot be done in the inter-industry

[6]Grand coalition superadditivity holds if the *joint* payoff of the three countries is larger under free trade than under no FTAs whatsoever or a bilateral FTA between any two countries. When this condition fails, Aghion et. al. (2004) show that the nature of externalities created by FTAs assumes a crucial role: when such externalities are negative, FTAs necessarily facilitate the achievement of global free trade whereas when they are positive, they hamper it.

trade framework utilized by Riezman (1999). As has already been noted, in our model cost asymmetry between countries plays a crucial role in highlighting conditions under which multilateral free trade is infeasible whereas FTAs are feasible as well as desirable. It is noteworthy that both Krugman (1991) and Grossman and Helpman (1995) noted that asymmetries across countries can potentially play a crucial role in determining incentives for bilateral and multilateral trade liberalization.

The stumbling versus building bloc question posed by Bhagwati (1991) has also been analyzed in the literature through models of repeated interaction between countries – see Bagwell and Staiger (1997a and 1997b), Bond et. al. (2001), Bond and Syropoulos (1996), and Saggi (2005). In such models, cooperation is self-enforcing in the sense that each country balances the current benefit of deviating from the cooperative tariff against the future losses caused by the breakdown of multilateral cooperation that follows any defection. In our model, a trade agreement needs to be self-enforcing in the sense that must be immune to credible coalitional deviations by both members and non-members. Also, we add value to this literature by treating both bilateral and multilateral liberalization as endogenous.

Levy (1997) focuses on political economy considerations and finds that in the monopolistic competition model of intraindustry trade in differentiated goods, FTAs can supplant multilateral trade liberalization. Unlike the present paper, tariffs play no real role in Levy's analysis since he only examines the choice between free trade (bilateral as well as multilateral) and autarky. Freund (2000) shows that (exogenous) multilateral trade liberalization encourages the formation of FTAs.

In our model, as well as in models where tariffs are driven by terms of trade considerations (such as Bagwell and Staiger, 1997), FTAs help remove negative externalities that countries impose on each other via their individually optimal tariffs. A completely different perspective on international trade agreements is provided by Maggi and Rodriguez-Clare (1997) who argue that in the presence of domestic protectionist pressures, such agreements can help a government credibly commit to free trade and that such commitment can improve the allocation of domestic resources. On the other hand, McLaren (1997) has shown that in the presence of sunk investment costs, anticipation of a free trade agreement with a big country can result in forward looking investments by individual investors in a small country that can lower its bargaining ability as well as welfare (relative to autarky).

2. Model

In this section, we present our oligopoly model of international trade in which each country has a unilateral incentive to impose rent extracting tariffs on its trading partners (unless it commits not to do so via an FTA). There are three countries (denoted by $i = a, b, c$) and two goods: x and y. Preferences over the two goods are quasilinear: $U(x, y) = u(x) + y$.

Good x is produced by a single profit-maximizing firm in each country at a constant marginal cost in terms of the numeraire good y.[7] Taking trade policies as given, firms compete in quantities (Cournot competition) and make independent decisions regarding how much to sell in each market (i.e. markets are segmented as in Brander and Krugman (1983) and Brander and Spencer (1984)).

[7]The gains from trade stem from reduced market power and monopoly is a simple way of representing market power.

2.1. Production and trade

Due to market segmentation, it is sufficient to focus on only one country's market. If country i does not belong to any FTA, exporters to its market face a specific tariff t_i per unit of exports. Country j's effective marginal cost of exporting to country i, equals $\zeta_j + t_i$ where $\zeta_j \geq 0$ equals its marginal cost of production for good x and $j \neq i$. By assumption, countries impose no taxes on local firms and the numeraire good (that may be traded internationally in order to balance trade).

Let x_{ji} denote country j's exports to country i; x_{ii} the sales of firm i in country i; and $x_i = x_{ii} + \sum_j x_{ji}$ denote total sales of good x in country i. Country j's profit function for exports to country i, denoted by Π_{ji}, can be written as:

$$\Pi_{ji} = [p_i(x_i) - \zeta_j - t_i]x_{ji} \tag{2.1}$$

First order conditions (FOCs) for profit maximization for exporters are

$$p_i + p_i' x_{ji} = \zeta_j + t_i \tag{2.2}$$

The above FOCs together with an analogous condition for the local firm can be easily solved for equilibrium output levels. The following comparative statics are standard and we assume that they hold:

$$\frac{dx_{ji}}{dt_i} < 0 < \frac{dx_{ii}}{dt_i}; \text{ and } \frac{dx_i}{dt_i} < 0 \tag{2.3}$$

In other words, country i's tariff lowers exports of countries j and k (x_{ji}) to its market; increases the sales of its local firm (x_{ii}); and lowers the total output (x_i) sold in its market.

Welfare of country i as a function of the global tariff vector $\mathbf{t} = (t_i, t_j, t_k)$ is defined as the sum of its domestic surplus $S_i(t_i)$ and total export profits:

$$W_i(\mathbf{t}) \equiv S_i(t_i) + \sum_j \Pi_{ij}(t_j) \tag{2.4}$$

where $S_i(t_i)$ is defined as

$$S_i(t_i) \equiv u(x_i) - p_i x_i + t_i \sum_j x_{ji} + \Pi_{ii} \tag{2.5}$$

where $u(x_i) - p_i x_i$ is consumer surplus in country i; $\Pi_{ii} = (p_i - \zeta_i)x_{ii}$ equals firm i's profits in its own market; and $\Pi_{ij} = (p_i - \zeta_i - t_j)x_{ij}$ its profits in foreign market j, where $j \neq i$. World welfare is defined the sum of the welfare of individual countries:

$$WW(\mathbf{t}) = W_i(\mathbf{t}) + W_j(\mathbf{t}) + W_k(\mathbf{t}) \tag{2.6}$$

3. Optimal tariffs

As per GATT, we assume that in the absence of any trade agreements each country imposes a non-discriminatory (MFN) tariff on its trade partners. Since markets are segmented and marginal costs are constant, strategic independence of trade policies obtains and own tariffs do not affect export profits. Given this, country i's tariff choice problem reduces to the maximization of its domestic surplus:

$$\max_{t_i} S_i(t_i) \equiv u(x_i) - p_i x_i + t_i \sum_j x_{ji} + (p_i - \zeta_i)x_{ii} \tag{3.1}$$

where $j \neq i$. Using $u' = p_i$, the first order condition for the above problem is given by

$$\frac{dS_i(t_i)}{dt_i} = -\frac{dp_i}{dt_i}x_i + \sum_j x_{ji} + t_i \sum_j \frac{dx_{ji}}{dt_i} + \frac{dp_i}{dt_i}x_{ii} + (p_i - \zeta_i)\frac{dx_{ii}}{dt_i} = 0 \tag{3.2}$$

where

$$\frac{dp_i}{dt_i} = p_i' \frac{dx_i}{dt_i} \tag{3.3}$$

Equation (3.2) can be rewritten as

$$\frac{dS_i(t_i)}{dt_i} = \left(1 - \frac{dp_i}{dt_i}\right) \sum_j x_{ji} + t_i \sum_j \frac{dx_{ji}}{dt_i} + (p_i - \zeta_i) \frac{dx_{ii}}{dt_i} = 0 \tag{3.4}$$

As Brander and Spencer (1984) have noted, the above equation implicitly defines country i's optimal tariff t_i^* under no agreement:

$$t_i^* = -\left[\frac{\left(1 - \frac{dp_i}{dt_i}\right) \sum_j x_{ji} + (p_i - \zeta_i) \frac{dx_{ii}}{dt_i}}{\sum_j \frac{dx_{ji}}{dt_i}} \right] \tag{3.5}$$

Since $\frac{dx_{ji}}{dt_i} < 0$ and $\frac{dx_{ii}}{dt_i} > 0$ (see equation 2.3), it follows that $t_i^* > 0$ if

$$\frac{dp_i}{dt_i} < 1 + \frac{(p_i - \zeta_i) \frac{dx_{ii}}{dt_i}}{\sum_j x_{ji}} \tag{3.6}$$

As a result, a sufficient condition for $t_i^* > 0$ is that $\frac{dp_i}{dt_i} < 1$ – an increase in t_i should cause the domestic price to increase less than the tariff.

If country i signs an FTA with country j, it solves the following problem:

$$\max_{t_{ki}} S_i(t_{ji}, t_{ki}) \text{ where } t_{ji} = 0 \tag{3.7}$$

Following the derivations for t_i^*, the first order condition for the above problem is given by

$$\frac{dS_i(0, t_{ki})}{dt_{ki}} = -\frac{dp_i}{dt_{ki}} x_i + x_{ki} + t_{ki} \frac{dx_{ki}}{dt_{ki}} + \frac{dp_i}{dt_{ki}} x_{ii} + (p_i - \zeta_i) \frac{dx_{ii}}{dt_{ki}} = 0 \tag{3.8}$$

which implies that country i's optimal tariff on the non-member country k is implicitly defined by the following equation:

$$t_{ki}^{f} = -\left[\frac{-\frac{dp_i}{dt_{ki}}\sum_j x_{ji} + x_{ki} + (p_i - \zeta_i)\frac{dx_{ii}}{dt_{ki}}}{\frac{dx_{ki}}{dt_{ki}}}\right] \tag{3.9}$$

Since $\frac{dx_{ki}}{dt_{ki}} < 0$, it follows that $t_{ki}^{f} > 0$ if

$$\frac{dp_i}{dt_i} < \frac{x_{ki}}{\sum_j x_{ji}} + \frac{(p_i - \zeta_i)\frac{dx_{ii}}{dt_i}}{\sum_j x_{ji}} \tag{3.10}$$

To gain further insight into how an FTA between countries i and j affects country i's tariff on country k, suppose $\zeta_i = \zeta$ for all i and consider conditions (3.6) and (3.10) at $t_i = 0$. Under these assumptions, all firms produce equal amounts and $\sum_j x_{ji} = 2x_{ki}$. A comparison of (3.6) and (3.10) immediately implies that a *stronger sufficient* condition is needed for the tariff on the non-member to be positive under an FTA relative to the case where there is no trade agreement between countries i and j. In fact, one can directly show that the first order condition for an FTA's tariff at the optimal tariff under no agreement (i.e. at t_i^*) is negative. Denoting $t_{ki} = t_i$ we have:

$$\left.\frac{dS_i(0,t_i)}{dt_i}\right|_{t_i=t_i^*} = -\frac{dp_i}{dt_i}\sum_j x_{ji} + x_{ki} - \left[\frac{\left(1 - \frac{dp_i}{dt_i}\right)\sum_j x_{ji} + (p_i - \zeta_i)\frac{dx_{ii}}{dt_i}}{\sum_j \frac{dx_{ji}}{dt_i}}\right]\frac{dx_{ki}}{dt_i} + (p_i - \zeta_i)\frac{dx_{ii}}{dt_i} \tag{3.11}$$

The above simplifies to

$$\left.\frac{dS_i(0,t_i)}{dt_i}\right|_{t_{ki}=t_i=t_i^*} = -\frac{x_{ji}\left[\frac{dp_i}{dt_i}\frac{dx_{ji}}{dt_i}+\frac{dx_{ki}}{dt_i}\right]}{\sum\limits_j \frac{dx_{ji}}{dt_i}} + (p_i-\zeta_i)\frac{dx_{ii}}{dt_i}\frac{dx_{ji}}{dt_i}\frac{1}{\sum\limits_j \frac{dx_{ji}}{dt_i}} + \frac{x_{ki}\left[(1-\frac{dp_i}{dt_i})\frac{dx_{ji}}{dt_i}\right]}{\sum\limits_j \frac{dx_{ji}}{dt_i}} \tag{3.12}$$

Under cost symmetry, $x_{ji} = x_{ii} > x_{ki}$ since the tariff t_i applies to only country k. Given this if (i) $\frac{dp_i}{dt_i} < 1$ and (ii) $\left|\frac{dx_{ki}}{dt_i}\right| > \frac{dx_{ji}}{dt_i}$, then it must be that $\frac{dp_i}{dt_i}\frac{dx_{ji}}{dt_i} + \frac{dx_{ki}}{dt_i} < 0$. As a result, conditions (i) and (ii) are sufficient to ensure that $\left.\frac{dS_i(0,t_i)}{dt_i}\right|_{t_{ki}=t_i=t_i^*} < 0$ implying that the tariff t_i^* is *too high* from the viewpoint of maximizing country i's welfare under a bilateral FTA between countries i and j.

The result that $t_{ki}^f < t_i^*$ also holds in the competitive trade model of Bagwell and Staiger (1997a) and they call it the *tariff complementarity* effect.[8] In this context, it is interesting to note that Article XXIV of GATT does not permit member countries of preferential trade agreements to raise tariffs on non-members. The above analysis implies that least for the case of FTA, such a requirement may be redundant.

Now that we have derived optimal MFN and FTA tariffs, we are now in a position to analyze the process of FTA formation and its interaction with multilateral trade liberalization.

[8] In a detailed micro level study, Chang and Winters (2002) found that tariffs on non-members actually declined after the formation of MERCOSUR (the major Latin American customs union). Similar evidence is cited in Bohara et. al. (2004) who argue that their results support Richardson's (1993) political economy model of endogenous protection.

4. Endogenous FTAs

Consider the following three stage game (called the FTA game). In the first stage, each country announces whether or not it wants to form an FTA with each of its trading partners (country i's announcement is denoted by α_i). Next, countries simultaneously choose their tariffs given the agreement(s) formed in the first stage. Finally, firms compete in product markets.

Under the FTA game, a country's strategy set consists of four possible announcements. For example, country a's strategy set Ω_a^F is:

$$\Omega_a^F = \{\{\phi, \phi\}, \{b, \phi\}, \{\phi, c\}, \{b, c\}\} \tag{4.1}$$

where $\{\phi, \phi\}$ is an announcement in favor of no agreement with either country of its trade partners (i.e. b and c). In order to conserve notation, each trade policy regime is denoted as follows: (i) No agreement $\langle\{\Phi\}\rangle$ obtains when no two announcements match or when everyone announces ϕ; (ii) an FTA between countries i and j denoted by $\langle\{ij\}\rangle$ is formed iff countries i and j announce each other's name $j\epsilon\alpha_i$ and $i\epsilon\alpha_j$; (iii) two independent FTAs in which i is the common member denoted by $\langle\{ij, ik\}\rangle$ are formed iff (1) $j\epsilon\alpha_i$ and $i\epsilon\alpha_j$ and (2) $k\epsilon\alpha_i$ and $i\epsilon\alpha_k$; and (iv) free trade, denoted by $\langle\{F\}\rangle$, obtains iff all countries announce each others' names.[9]

As is clear from the structure of the FTA game, an FTA member can sign an independent FTA with the non-member without needing consent of the other

[9] Formally, free trade $\langle\{abc\}\rangle$ obtains iff $\alpha_a = \{b, c\}$, $\alpha_b = \{a, c\}$, and $a_c = \{a, b\}$.

member.[10] Consider the following announcements:

$$\alpha_a = \{b, \phi\},\ \alpha_b = \{a, c\},\ \alpha_c = \{\phi, b\} \tag{4.2}$$

The above strategy vector gives rise to two independent FTAs $\langle\{ab, bc\}\rangle$ of which country b is the common member. Note also that different strategy vectors may yield the same agreement(s) when countries' announcements do not match. To see this, suppose the strategy vector is given by:

$$\alpha_a = \{b, c\},\ \alpha_b = \{a, c\},\ \alpha_c = \{\phi, b\} \tag{4.3}$$

Even though country a announces country c, country c wants to form an FTA only with country b.[11] As a result, the strategy vector in (4.3) yields the same agreements $\langle\{ab, bc\}\rangle$ as the one in (4.2). In order to eliminate redundant announcements, assume that each FTA announcement costs ε (where $\varepsilon > 0$ is arbitrarily small).

Our method of analysis is to compare the FTA game with the following No FTA game. In the first stage of the No FTA game, each country announces either in favor of or against free trade.[12] If all countries announce in favor, free trade emerges. If not, there is no agreement and the status quo of optimal tariffs prevails. Next, given the agreement(s) formed in the first stage, countries choose their tariffs. Finally, firms compete in the product market.

It is straightforward that no agreement $\langle\{\Phi\}\rangle$ is a Nash equilibrium of the FTA game since no country i has an incentive to announce country j's name if

[10] As indicated in Furusawa and Konishi (2003), this distinction creates an important difference between an FTA and a customs union and leads to a sharp contrast to Yi (1996).

[11] Under the open membership rule analyzed by Yi (1996), membership is open to all countries. However, this rule does not seem to be appealing for discussions of FTAs since the formation of an FTA requires consent from both sides.

[12] The strategy set of country i in the No FTA game is $\Omega_i = \{\{\phi, \phi\}, \{j, k\}\}$, $j \neq k \neq i$.

the latter does not announce its name in return. In fact, as is well known that such games admit multiple Nash equilibria. To deal with this multiplicity and to capture the process of FTA formation in a more realistic fashion, we focus attention on Nash equilibria that are coalition proof (i.e. are immune to credible coalitional deviations). Following Dutta and Mutuswami's (1997) terminology, we refer to coalition proof Nash equilibria as *stable* equilibria.

It is useful to note that in the No FTA game, a unilateral deviation from free trade by any country results in no agreement whereas in the FTA game the same deviation results in the deviating country becoming a non-member country (with an FTA between the other two countries). Since the welfare of a non-member under an FTA can be lower than its welfare under no agreement, it is *not* immediately obvious under which game the unilateral incentive to deviate from free trade is stronger. Thus, even though the set of possible announcement deviations from free trade under the No FTA game is a strict subset of those under the FTA game, it does not follow that free trade is more likely to be a stable equilibrium of the No FTA game.

Let $\mathcal{T}$ be the set of all feasible trade policy regimes:

$$\mathcal{T} = \{\langle\{\Phi\}\rangle, \langle\{ab\}\rangle, \langle\{ac\}\rangle, \langle\{bc\}\rangle, \langle\{ab, ac\}\rangle, \langle\{ab, bc\}\rangle, \langle\{ac, bc\}\rangle, \text{ or } \langle\{F\}\rangle\}$$

and let r and v be any two elements of $\mathcal{T}$. Let $w_i(r)$ denote country i's welfare under regime r and let $\Delta w_i(r-v)$ define the difference between country i's welfare under regimes r and v:

$$\Delta w_i(r - v) \equiv w_i(r) - w_i(v)$$

For notational simplicity, let

$$\Delta w_i(r) \equiv w_i(r) - w_i(F)$$

denote the difference between country i's welfare under trade regime r and free trade.

Before proceeding further we clarify a notational point. Through-out the paper, we use small letters to denote functions of trade regimes whereas we use capital letters to denote functions of tariffs that prevail during those regimes. For example, country i's welfare under the FTA $\langle\{ij\}\rangle$ can be written as $w_i(ij)$ or equivalently as $W(0, t_{ki}^f)$. Similar notation applies to domestic surplus and export profit functions.

The analysis proceeds as follows. First, we argue that when countries are symmetric, free trade is the unique stable equilibrium of the No FTA game (Proposition 1). Then, we provide conditions under which the same is true under the FTA game (Proposition 2). Turning to the case where countries are asymmetric, we then derive conditions under which the FTA option undermines multilateral free trade (Proposition 3) and when it facilitates the obtainment of free trade in the sense that if FTAs are not permitted then free trade fails to be an equilibrium (Proposition 4). Next, we show that when multilateral free trade is infeasible, two symmetric low cost countries gain from an FTA and that such an FTA can also improve world welfare relative to no agreement. (Proposition 5). Thus, the option to form FTAs has the potential to deliver welfare gains that may be foregone when the choice is only between global free trade or no agreement. Later in the paper we show that this conjecture holds when demand is linear. Finally, for the case of linear demand, we graphically illustrate stable equilibria of the two games.

4.1. The FTA option under symmetry

In this section, we assume that the three countries are symmetric: $\zeta_i = \zeta$ for all i. Since optimal tariffs are equal under symmetry, let t^* denote a country's optimal tariff under no agreement $\langle\{\Phi\}\rangle$ and let t^f denote an FTA member's tariff on the non-member. We begin with the No FTA game under symmetry:

Proposition 1: *Under symmetry, the following hold: (i) Free trade yields higher welfare than any other trade regime and (ii) it is the unique stable equilibrium of the No FTA game.*

Part (i) of proposition 1 is proved in the appendix. The logic for part (ii) is as follows. Due to symmetry, under both no agreement and free trade, all counties have the same welfare. But if world welfare is higher under free trade than under no agreement (i.e. $ww(F) > ww(\Phi)$), it follows immediately that each country is better off under free trade than under no agreement. As a result, under symmetry no country has an incentive to deviate from free trade since *any* other announcement leads to no agreement where it (and everyone else) is worse off.

We now turn to the FTA game. We will show that the magnitude of the tariff imposed by an FTA member on the non-member (i.e. t^f) plays a crucial role in determining whether or not free trade emerges as an equilibrium. Since a country imposes no taxes on its own firm, we only need to keep track of the tariffs it imposes on foreign firms. In what follows, in functions $S(.)$ and $W(.)$, we list the tariffs faced by countries in ascending alphabetical order. In the export profit function $\Pi(.)$, the first argument is the tariff faced by a country while the second argument is the tariff faced by its rival exporter.

Consider (non-member) country c's welfare under the FTA $\langle\{ab\}\rangle$ relative to free trade $\langle\{F\}\rangle$:

$$\Delta w_c(ab) = \Delta s_c(ab) + \Delta\pi_c(ab) \tag{4.4}$$

where $\Delta s_c(ab)$ in (4.4) equals the amount by which domestic surplus of country c under $\langle\{ab\}\rangle$ exceeds that under free trade $\langle\{F\}\rangle$:

$$\Delta s_c(ab) \equiv S_c(t^*, t^*) - S_c(0,0) > 0$$

while $\Delta\pi_c(ab)$ measures the loss in its export profits:

$$\Delta\pi_c(ab) \equiv \sum_{i\neq c}\Pi_{ci}(t^f, 0) - \sum_{i\neq c}\Pi_{ci}(0,0) < 0.$$

When $t^f \longrightarrow 0$, $\Delta\pi_c(ab)$ converges to zero. Since $\Delta s_c(ab) > 0$, we have

$$\lim_{t^f \longrightarrow 0} \Delta w_c(ab) > 0 \tag{4.5}$$

Next we show that when $t^f \longrightarrow t^*$, the loss in export profits $\Delta\pi_c(ab)$ must outweigh the gain in domestic surplus $\Delta s_c(ab)$. Since free trade maximizes world welfare (see proposition 1), we must have:

$$S_c(t^*, t^*) - S_c(0,0) + \left[\sum_{i\neq c}\Pi_{ci}(t^*, t^*) - \sum_{i\neq c}\Pi_{ci}(0,0)\right] < 0. \tag{4.6}$$

Also since $\Pi_{ci}(t^*, t^*) > \Pi_{ci}(t^*, 0)$, from (4.4) and (4.6) it follows that

$$\lim_{t^f \longrightarrow t^*} \Delta w_c(ab) < 0 \tag{4.7}$$

Given inequalities (4.5) and (4.7) and $\frac{\partial \Delta w_c(ab)}{\partial t^f} < 0$, there must exist a critical threshold $\underline{t}$ such that:[13]

$$\Delta w_c(ab) \geq 0 \text{ iff } t^f \leq \underline{t} \tag{4.8}$$

[13]It is easy to show that under linear demand, when $\varsigma_i = 0$ for all i, $t^f > \underline{t}$.

As a result, if $t^f \leq \underline{t}$ free trade is not immune to unilateral announcement deviations (and is therefore not a Nash equilibrium). The intuition for this result is as follows. When a country faces low tariffs as a non-member (i.e. t^f is small), it has a unilateral incentive to deviate from free trade because it gains substantially in its domestic market from being able to charge its optimal tariff (t^*) whereas it does not lose much in export markets due to the low tariffs of FTA members ($t^f < \underline{t}$). In other words, each country has a strong incentive to *free ride* on trade liberalization undertaken by others by gaining access to their markets without having to liberalize in return and this free rider problem prevents free trade from emerging as an equilibrium. The above proposition clarifies that if the formation of FTAs alters tariffs of member countries, such tariff changes have a crucial impact on the incentives of non-members to participate in global free trade.[14]

What happens when $t^f > \underline{t}$? The analysis above implies that when $t^f > \underline{t}$ no country has a unilateral incentive to break off links with both of its trading partners. *But is there an incentive to break-off only one link*? In other words, does a country (say c) have an incentive to deviate from $\langle\{a, b\}\rangle$ to $\langle\{a, \phi\}\rangle$? It turns out that this is is not the case. If country c breaks its link with country b, export profits of country a increase in both markets because its rival exporter (i.e. country b) face tariffs whereas it itself does not. Furthermore, the domestic surplus of country a does not change relative to free trade since its own tariff equals zero. As a result, if country c breaks its link with country b, then country

[14] An analogous result obtains in the repeated game models of Bagwell and Staiger (1997a) and Saggi (2005) but their underlying logic is quite different – in such models, cooperation is easier to sustain via harsher punishments and tariff complementarity can undermine cooperation by lowering the static Nash tariffs of FTA members (by assumption, Nash tariffs are used to punish deviations).

a's welfare necessarily increases relative to free trade:

$$\Delta w_a(ab, ac) > 0 \tag{4.9}$$

Since world welfare is lower under $\langle\{ab, ac\}\rangle$ relative to free trade: $ww(ab, ac) < ww(\digamma)$ (proposition 1), the sum of countries c and b's welfare must decline:

$$\Delta w_b(ab, ac) + \Delta w_c(ab, ac) < 0 \tag{4.10}$$

But when both countries b and c have an independent FTA with country a, their welfare is equal: $w_b(ab, ac) = w_c(ab, ac)$. It follows immediately that both must be worse off under $\langle\{ab, ac\}\rangle$ relative to free trade $\langle\{\digamma\}\rangle$:

$$\Delta w_b(ab, ac) = \Delta w_c(ab, ac) < 0 \tag{4.11}$$

Hence, no country has an incentive to break off one of its free trade links. As a result, if $t^f > \underline{t}$ then free trade is a Nash equilibrium. But is it stable? We now provide conditions under which free trade is immune to all coalitional deviations (whether they are credible or not) – i.e. it is the strong Nash equilibrium of the FTA game.

Since no country has no unilateral incentive to deviate from free trade, there can be no joint deviation of a pair of countries, say a and b, from free trade $\langle\{\digamma\}\rangle$ to $\langle\{ac, bc\}\rangle$. Therefore, we need to examine only two possible coalitional deviations: (D1) the joint deviation of two countries, say a and b, from their respective announcements $\{b, c\}$ and $\{a, c\}$ to $\{\phi, \phi\}$ and $\{\phi, \phi\}$; and (D2) the joint deviation of two countries, say a and b, from their respective announcements $\{b, c\}$ and $\{a, c\}$ to $\{b, \phi\}$ and $\{a, \phi\}$.[15] Since D1 is ruled out by proposition 1, we

[15]Note that, due to symmetry, the multilateral deviation of all three countries from $\langle\{\digamma\}\rangle$ to $\langle\{\Phi\}\rangle$ yields exactly the same condition.

do not need to consider it further. We now provide a sufficient condition under which D2 does not occur. Consider the comparison of country a's welfare under the FTA $\langle\{ab\}\rangle$ relative to free trade $\langle\{F\}\rangle$:

$$\Delta w_a(ab) = \Delta s_a(ab) + \Delta\pi_a(ab) \tag{4.12}$$

where

$$\Delta s_a(ab) \equiv S_a(0, t^f) - S_a(0, 0) > 0 \tag{4.13}$$

while

$$\Delta\pi_a(ab) \equiv \Pi_{ab}(0, t^f) + \Pi_{ac}(t^*, t^*) - \sum_{i \neq a} \Pi_{ai}(0, 0) \tag{4.14}$$

Note that under symmetry the sum of domestic surplus of country a and the total export profits of b and c in country a are lower under $\langle\{ab\}\rangle$ relative to free trade $\langle\{F\}\rangle$:[16]

$$s_a(ab) + \sum_{i \neq a} \pi_{ia}(ab) < s_a(F) + \sum_{i \neq a} \pi_{ia}(F) \tag{4.15}$$

which can be rewritten as

$$\Delta s_a(ab) + \Delta\pi_{\sim a}(ab) < 0 \tag{4.16}$$

where

$$\Delta\pi_{\sim a}(ab) \equiv \sum_{i \neq a} \pi_{ia}(ab) - \sum_{i \neq a} \pi_{ia}(F) = \Pi_{ba}(0, t^f) + \Pi_{ca}(t^f, 0) - \sum_{i \neq a} \Pi_{ia}(0, 0) \tag{4.17}$$

As a result, whenever $\Delta\pi_{\sim a}(ab)$ in (4.16) is smaller than $\Delta\pi_a(ab)$ in (4.12) it must be that $\Delta w_a(ab) < 0$. Subtracting $\Delta\pi_a(ab)$ from $\Delta\pi_{\sim a}(ab)$ and using the fact that

[16] This assertion can be proven along the lines of proposition 1 (see the appendix). The only difference here is that an FTA member imposes a tariff on only the non-member. Hence, the sum of total welfare of a member and the export profits earned by the other countries in its market must be reduced due to its tariff.

under symmetry $\sum_{i\neq a}\Pi_{ia}(0,0)=\sum_{i\neq a}\Pi_{ai}(0,0)$ and $\Pi_{ab}(0,t^f)=\Pi_{ba}(0,t^f)$ gives

$$\Delta\pi(ab)\equiv\pi_{\tilde{}a}(ab)-\pi_a(ab)=\Pi_{ac}(t^*,t^*)-\Pi_{ca}(t^f,0) \tag{4.18}$$

Note that

$$\lim_{t^f\longrightarrow 0}\Delta\pi(ab)=\Pi_{ac}(t^*,t^*)-\Pi_{ca}(0,0)<0 \tag{4.19}$$

whereas

$$\lim_{t^f\longrightarrow t^*}\Delta\pi(ab)=\Pi_{ac}(t^*,t^*)-\Pi_{ca}(t^*,0)>0 \tag{4.20}$$

Since $\Delta\pi(ab)$ is continuously increasing in t^f, there exists a critical threshold tariff $\bar{t}$ such that:

$$\Delta\pi(ab)\leq 0 \text{ if } t^f\leq\bar{t} \tag{4.21}$$

Thus, $t^f\leq\bar{t}$ is sufficient for D2 *not* to occur. As a result, when $\underline{t}\leq t^f\leq\bar{t}$, free trade is immune to all possible deviations and is therefore a strong Nash equilibrium (which implies that it is also a coalition proof Nash equilibrium).[17] The intuition here is as follows. We already know that when $\underline{t}\leq t^f$, free trade is immune to unilateral deviations. The condition $t^f\leq\bar{t}$ ensures that no two countries have an incentive to jointly exclude a third country by breaking off their respective links because their optimal external tariffs on the excluded country are not high enough to justify such an exclusion.

Finally, suppose $t^f>\bar{t}$ and consider the joint deviation of countries a and b from the announcements $\{b,c\}$ and $\{a,c\}$ to $\{b,\phi\}$ and $\{a,\phi\}$ respectively (i.e. deviation D2). Free trade $\langle\{F\}\rangle$ is a stable equilibrium if, taking the announcement $\{a,b\}$ of county c as fixed, either of the two deviating countries (a or b), has an incentive to further deviate to another announcement thereby making their

[17] We show in the appendix that $\bar{t}>\underline{t}$.

original coalitional deviation non-credible. To this end, consider the further deviation of country a from the announcement $\{b, \phi\}$ to $\{b, c\}$. This is equivalent to a deviation from the FTA $\langle\{ab\}\rangle$ to a pair of FTAs $\langle\{ab, ac\}\rangle$ in which a is the common member. We show that this deviation will indeed occur if[18]

$$\Pi_I(0, t^f) > \frac{2}{3}\Pi_I(0, 0) + \frac{1}{3}\Pi_I(t^*, t^*) \tag{4.22}$$

where $\Pi_I(0, t^f) = \sum_i \Pi_{ia}(0, t^f)$ denotes total profits earned by the global industry in country a's market under the FTA $\langle\{ab\}\rangle$; $\Pi_I(0, 0)$ denotes total profits of the global industry in any one market under global free trade; and $\Pi_I(t^*, t^*)$ equals total profits of the global industry in any one market under no agreement. Intuitively, if total profits of the global industry are higher under a bilateral FTA $\langle\{ab\}\rangle$ than a weighted average of the global profits under free trade (with 2/3rd weight) and that under no agreement (with 1/3rd weight), a country has an incentive to deviate from a bilateral FTA to a pair of bilateral FTAs.

Suppose condition (4.22) holds. By contradiction, suppose country a's welfare is lower under $\langle\{ab, ac\}\rangle$ relative to $\langle\{ab\}\rangle$:

$$w_a(ab, ac) - w_a(ab) = S_a(0, 0) + \sum_{i \neq a} \Pi_{ai}(0, t^f) - [S_a(0, t^f) + \Pi_{ab}(0, t^f) + \Pi_{ac}(t^*, t^*)] < 0 \tag{4.23}$$

We know that under symmetry, the sum of domestic surplus of country a and the total export profits of b and c in country a are lower under $\langle\{ab\}\rangle$ relative to $\langle\{ab, ac\}\rangle$:

$$S_a(0, t^f) + \Pi_{ba}(0, t^f) + \Pi_{ca}(t^f, 0) - [S_a(0, 0) + \sum_{i \neq a} \Pi_{ia}(0, 0)] < 0 \tag{4.24}$$

[18] If this condition fails, a pair of bilateral FTAs will emerge as the equilibrium.

Since $\Pi_{ia}(.) = \Pi_{ai}(.) = \Pi_a(.)$ under symmetry, adding inequalities (4.23) and (4.24) yields:

$$2\Pi_a(0,t^f) - 2\Pi_a(0,0) < \Pi_a(t^*,t^*) - \Pi_a(t^f,0) \tag{4.25}$$

which is the same as

$$\Pi_I(0,t^f) < \frac{2}{3}\Pi_I(0,0) + \frac{1}{3}\Pi_I(t^*,t^*) \tag{4.26}$$

but this contradicts condition (4.22). As a result, country a further deviates from $\{b,\phi\}$ to $\{b,c\}$. Since the original deviation of country a is *not* self-enforcing, free trade is a stable equilibrium of the FTA game. The following result can now be stated:

Proposition 2: *Given symmetry, the following hold:*

(i) if $t^f < \underline{t}$*, free trade is not a Nash equilibrium of the FTA game;*

(ii) if either (1) $\underline{t} \leq t^f \leq \bar{t}$ *or (2)* $t^f > \bar{t}$ *and* $\Pi_I(0,t^f) > \frac{2}{3}\Pi_I(0,0) + \frac{1}{3}\Pi_I(t^*,t^*)$ *holds, free trade is a stable equilibrium of the FTA game.*

4.2. FTAs among asymmetric countries

How does asymmetry alter the prospects of global free trade? Is it possible that under asymmetry two countries are unwilling to engage in free trade but willing to enter into a bilateral FTA? Even more interestingly, can the FTA option make free trade more likely to obtain? We now turn to these questions.

First note that the higher a country's production cost of good x, the smaller its volume of exports and the larger its volume of imports: these arguments follow immediately from the nature of Cournot competition (see equation 2.2). Because of their larger volume of imports, higher cost countries have relatively more to gain from using tariffs. Similarly, due to the smaller volume of their exports, higher cost countries have less to lose from other countries' tariffs. Based on this intuition, we make the following assumption:

Assumption 1 (A1):

$$\frac{\partial \Delta w_i(r-v)}{\partial \zeta_i} > 0,\ \frac{\partial \Delta w_i(r-v)}{\partial \zeta_m} < 0 \ \textit{and}\ \frac{\partial \Delta w_i(r-v)}{\partial \zeta_n} > 0$$

where m is an FTA partner of country i under regime v (but not regime r) whereas n is either a partner or not under both regimes r and v (i.e. its status with respect to country i is the same under regimes r and v).

To get further insight behind A1, consider regimes $\langle\{\Phi\}\rangle$ and $\langle\{ij\}\rangle$ from country i's perspective. The intuition behind the first two parts of A1 (i.e. $\frac{\partial \Delta w_i(\Phi-ij)}{\partial \zeta_j} < 0 < \frac{\partial \Delta w_i(\Phi-ij)}{\partial \zeta_i}$) has already been stated above. The third part of A1 (i.e. $\frac{\partial \Delta w_i(\Phi-ij)}{\partial \zeta_k} < 0$) reflects the idea that granting preferential access to country j hurts country i relatively more when country k is a higher cost competitor. The idea is that country k's exports are low if its cost is high and the strategic advantage of protecting its local firm from country j's firm is high from country i's perspective. Hence country i's welfare gain of bilateral liberalization with country j declines with the cost of country k.

To highlight the role of asymmetry, it is instructive to focus on the case where two countries have symmetric and low costs relative to the third. Accordingly, throughout the analysis under asymmetry, let $\zeta_a = \zeta_b = 0$ and $\zeta_c = \zeta > 0$. It is useful to begin with the No FTA game.

4.2.1. Feasibility of free trade in the absence of FTAs

We begin with the high cost country's (i.e. c's) perspective. From proposition 1 we know that as ζ approaches zero, the welfare of country c under no agreement

$\langle\{\Phi\}\rangle$ is lower relative to free trade $\langle\{F\}\rangle$:

$$\lim_{\zeta\longrightarrow 0} \Delta w_c(\Phi) < 0 \tag{4.27}$$

Define ζ^P to be a prohibitive cost level such that when $\zeta = \zeta^P$ the export profits of country c in each foreign market equal zero under free trade:

$$\lim_{\zeta\longrightarrow\zeta^P} \pi_{ca}(F) = \lim_{\zeta\longrightarrow\zeta^P} \pi_{cb}(F) = 0$$

Since domestic surplus of each country is higher under no agreement $\langle\{\Phi\}\rangle$ than under free trade $\langle\{F\}\rangle$ the following is immediate:

$$\lim_{\zeta\longrightarrow\zeta^P} \Delta w_c(\Phi) > 0 \text{ since } s_c(\Phi) > s_c(F) \tag{4.28}$$

Inequalities (4.27), (4.28), and A1 imply that there exists a critical threshold cost level (ζ^ϕ) such that:

$$\Delta w_c(\Phi) \geq 0 \text{ iff } \zeta \geq \zeta^\phi \tag{4.29}$$

Consequently, a result analogous to proposition 1 obtains under asymmetry:

Proposition 1B: *Free trade is the unique stable equilibrium of the No FTA game iff* $\zeta \leq \zeta^\phi$.

The crucial point is that global free trade is infeasible even if countries do not have the option to form an FTAs as long as there is sufficient asymmetry (i.e. $\zeta \geq \zeta^\phi$) between them. Following the definition of ζ^ϕ, let ζ^s define the critical threshold cost level at which country c (the high cost country) is indifferent between regime s and free trade $\langle\{F\}\rangle$:

$$\Delta w_c(s) = w_c(s) - w_c(F) \geq 0 \text{ iff } \zeta \geq \zeta^s \tag{4.30}$$

where $s = \{\langle\{\Phi\}\rangle, \langle\{ab\}\rangle, \langle\{ab, ac\}\rangle, \text{or } \langle\{ab, bc\}\rangle\}$. Arguments analogous to those that underlie the existence of ζ^{ϕ} ensure that these critical cost thresholds also exist for the other trade policy regimes.[19]

4.2.2. When do FTAs act as stumbling blocs?

To see when and how FTAs can hamper the emergence of global free trade, we begin with the following assumption:

Assumption 2 (A2): $\zeta^{ab,ac} > \zeta^{ab}$.

A2 is equivalent to the assumption $w_c(ab) > w_c(ab, ac)$ for all ζ. In other words, the high cost country (c) is better off breaking its (only existing) link with a low cost country (a) when the latter also has a link with the other low cost country (b). Intuitively, A2 captures the idea that the high cost country (c) has a strong incentive to free ride on the FTA between countries a and b – if it breaks its link with country a, it faces the FTA tariff τ^f in export markets whereas it gets to impose its (unconstrained) optimal tariff t_c^*. Note that given the higher cost of country c, the tariffs of FTA members are likely to be low due to the low volume of country c's exports. In this sense, A2 is not a particularly strong assumption. Before stating our main result, we note the following:

Lemma 1: $\tau^f \leq \bar{t}$ iff $\zeta^{ab} \leq \zeta^{\phi}$.

Since $\sum_{i \neq c} \Pi_{ci}(\tau^f, 0) \geq \sum_{i \neq c} \Pi_{ci}(t^*, t^*)$ iff $\tau^f \leq \bar{t}$ it follows that $w_c(\Phi) \leq w_c(ab)$ iff $\tau^f \leq \bar{t}$. Recall that, by definition, $\zeta^{ab} \leq \zeta^{\phi}$ iff $w_c(\Phi) \leq w_c(ab)$.

Proposition 3: *Suppose A2 holds. Then, multilateral free trade is not a Nash equilibrium of the FTA game if either of the following two statements holds:*

[19] Note that the threshold ζ^{ab} exists only when τ^f (the external tariff of member countries a and b under asymmetry) exceeds $\underline{t}$ whereas the other thresholds exist for all τ^f.

(i) $\tau^f < \underline{t}$ *or*

(ii) $\underline{t} \leq \tau^f < \bar{t}$ *and* $\zeta^{ab} < \zeta$.

Part part (i) of proposition 3 simply states that part (i) of proposition 2 holds under asymmetry and the underlying intuition is the same as before: low FTA tariffs encourage free-riding and therefore undermine global free trade.

To prove part (ii), first consider the unilateral deviation of the high cost country (c) from $\{a, b\}$ to $\{a, \phi\}$.[20] Under symmetry, from proposition 2 we know that in the FTA game the following holds:

$$\Delta w_c(ab, ac) = w_c(ab, ac) - w_c(F) \leq 0 \tag{4.31}$$

Since $s_c(ab, ac) > s_c(F)$, it must be that

$$\lim_{\zeta \longrightarrow \zeta^P} \Delta w_c(ab, ac) > 0 \tag{4.32}$$

Inequalities (4.31), (4.32), and A1 imply that there exists a critical threshold cost level $(\zeta^{ab,ac})$ such that:

$$\Delta w_c(ab, ac) > 0 \text{ iff } \zeta > \zeta^{ab,ac}$$

Now consider country c's deviation from $\{a, b\}$ to $\{\phi, \phi\}$. We know that:

$$\Delta w_c(ab) > 0 \text{ iff } \zeta > \zeta^{ab}$$

Given Lemma 1 and A2, it is immediate that if $\underline{t} \leq \tau^f < \bar{t}$ and $\zeta^{ab} < \zeta$ then multilateral free trade is not a Nash equilibrium of the FTA game. We know from Proposition 1B that free trade is a stable equilibrium of the No FTA game iff $\zeta < \zeta^{\phi}$. Thus, if $\zeta < \zeta^{\phi}$, *statements (i) and (ii) of Proposition 3 provide sufficient conditions for an FTA to act as a stumbling bloc.*

[20] Since countries a and b are symmetric, the unilateral deviation of country c from $\{a, b\}$ to $\{\phi, b\}$ is equivalent to that from $\{a, b\}$ to $\{a, \phi\}$.

4.2.3. How FTAs can make multilateral free trade feasible

Here we argue that the FTA option can serve as a building bloc in a rather *strong* sense – there exist circumstances where free trade is an equilibrium of the FTA game whereas it is not so of the No FTA game. Intuitively, such a possibility arises when the tariff of an FTA member on the non-member is relatively high. Under such a situation, the high cost country prefers no agreement to multilateral free trade which is in turn preferred to an FTA between the other two countries. Since the high cost country can do nothing to prevent an FTA between the other two, it ends up favoring free trade over no agreement:

Proposition 4: *Suppose A2 holds and* $\tau^f > \bar{t}$. *When* $\zeta < \zeta^{ab}$, *multilateral free trade is a stable equilibrium of the FTA game.*

Combining Lemma 1 and Propositions 1B and 4, it follows that *FTAs act as a building bloc when* $\zeta^{\phi} < \zeta < \zeta^{ab}$.

4.2.4. Bilateral trade liberalization

While the FTA option can sometimes facilitate the achievement of global free trade, when the FTA members' tariffs on the non-member are rather low, a sufficiently high cost country prefers to face those tariffs (while using its optimal tariff) rather than grant free access to its market in return for access abroad. Under such a situation, the interesting question is if and when an FTA between two low cost countries is a stable outcome.

Proposition 5: *If* $\tau^f < \bar{t}$ *and* $\zeta > \zeta^{\phi}$, *an FTA between the two low cost countries* $\langle\{ab\}\rangle$ *is a stable equilibrium of the FTA game.*[21]

[21] If $\tau^f > \bar{t}$ we need $\zeta > \zeta^{ab}$ for proposition 5 to hold. Note that this scenario also completes proposition 4.

The above result lends support to the argument that bilateral trade liberalization may be feasible when multilateral trade liberation is not. In fact, a stronger argument in favor of FTAs can be made: it is possible for a pair of FTAs to yield higher world welfare than multilateral free trade. This surprising outcome obtains when a pair of FTAs diverts production from higher cost countries to the lowest cost one.

Proposition 6: *Suppose* $\zeta_i = \zeta > \zeta_a = 0$ *where* $i \neq a$. *The pair of FTAs* $\langle\{ab, ac\}\rangle$ *can yield higher world welfare than global free trade.*

Suppose countries b and c form individual FTAs with country a and let $\zeta_i = \zeta > \zeta_a = 0$. Under such a situation, countries b and c impose zero tariffs on country a and the tariff τ_P^f on each other where τ_P^f solves

$$\tau_P^f = \arg\max \ S_b(0,t) = \arg\max \ S_c(0,t)$$

so that

$$\left.\frac{dS_b(0,t)}{dt}\right|_{t=\tau_P^f} = \left.\frac{dS_c(0,t)}{dt}\right|_{t=\tau_P^f} = 0 \tag{4.33}$$

Now consider the impact of FTA tariffs under the trade regime $\langle\{ab, ac\}\rangle$ on world welfare . Using $\frac{dS_a(0,0)}{dt} = 0$ and equation (4.33) we can write

$$\left.\frac{dWW(0,t)}{dt}\right|_{t=\tau_P^f} = \sum_{i\neq a}\frac{d\Pi_{ai}(0,t)}{dt} + \frac{d\Pi_{bc}(t,0)}{dt} + \frac{d\Pi_{cb}(t,0)}{dt}$$

In other words, *when FTA tariffs are optimally chosen by member countries, they increase world welfare iff they increase the total export profits in the world economy.*

Note that

$$\frac{d\Pi_{ai}(0,t)}{dt} = p'\frac{dx_i}{dt}x_{ai} + p\frac{dx_{ai}}{dt} = p'\left(\frac{d(x_{bi}+x_{ci})}{dt}\right)x_{ab} > 0 \text{ where } i \neq a$$

Similarly,

$$\frac{d\Pi_{bc}(t,0)}{dt} = \left[p'\left(\frac{d(x_{ac}+x_{cc})}{dt}\right)-1\right]x_{bc} < 0$$

Also,

$$\frac{d\Pi_{ab}(0,t)}{dt} = \frac{d\Pi_{ac}(0,t)}{dt}$$

At $t = \tau_P^f$, the first order condition for world welfare maximization can be written as

$$\frac{1}{2}\left.\frac{dWW(0,t)}{dt}\right|_{t=\tau_P^f} = p'x_{ab}\left[\frac{d(x_{bb}+x_{cb})}{dt}\right]_{t=\tau_P^f} + x_{bc}\left[p'\left(\frac{d(x_{ac}+x_{cc})}{dt}\right)-1\right]_{t=\tau_P^f}$$

Under linear demand, the above simplifies to

$$\frac{1}{2}\left.\frac{dWW(0,t)}{dt}\right|_{t=\tau_P^f} = \frac{x_{ab}-3x_{bc}}{2} \geq 0 \text{ iff } x_{ab} \geq 3x_{bc}.$$

so that under linear demand FTA tariffs under the regime $\langle\{ab, ac\}\rangle$ are optimally set from the viewpoint of global welfare maximization iff $x_{ab} = 3x_{bc}$.

5. Characterization of equilibrium FTAs

To fully characterize equilibrium FTAs, assume that $u(x_i)$ is quadratic:

$$u(x_i) = \beta x_i - \frac{x_i^2}{2}$$

so that country i's inverse demand function is given by:

$$p_i(x_i) = \beta - x_i \tag{5.1}$$

We first illustrate the stumbling bloc effect of FTAs by comparing the regions over which the two games deliver multilateral free trade as a stable equilibrium (see Figure 1).

——Insert Figure 1 here——

As is clear from Figure 1, free trade is a stable outcome under the No FTA game over a much larger region. Thus, under linear demand, the FTA option has only a stumbling effect for the prospects of achieving multilateral free trade.[22] Note that under both games the region over which free trade $\langle\{F\}\rangle$ is stable is determined by the unilateral deviation of the high cost country (c) from free trade announcement $\{a, b\}$ to $\{\phi, \phi\}$. This deviation implies the underlying trade regime changes from $\langle\{F\}\rangle$ to $\langle\{ab\}\rangle$ under the FTA game whereas it changes from $\langle\{F\}\rangle$ to $\langle\{\Phi\}\rangle$ under the No FTA game. Since the high cost country (c) benefits from the tariff complementarity effect under $\langle\{ab\}\rangle$ but not under $\langle\{\Phi\}\rangle$, its incentive to unilaterally deviate from free trade is greater under the FTA game relative to the No FTA game.

Next, we show that the option to form FTAs has the potential to deliver welfare gains that may be foregone when the choice is only between free trade or no agreement (see proposition 5).

——Insert Figure 2 here——

Figure 2 illustrates that when multilateral free trade is infeasible, no agreement $\langle\{\Phi\}\rangle$ is the stable agreement of the No FTA game whereas under the FTA game: (i) $\langle\{ab\}\rangle$ is a stable equilibrium when c is sufficiently high cost and low cost countries (a and b) are relatively symmetric and (ii) $\langle\{bc\}\rangle$ is a stable equilibrium when a is sufficiently low cost and b and c are relatively symmetric.

——Insert Figure 3 here——

[22] Note that under linear demand, the FTA option fails to serve as a building bloc (as in proposition 4) since for linear demand τ^f does not exceed $\bar{t}$.

Figure 3 depicts stable FTAs under the FTA game. Note that global free trade arises as the stable equilibrium when all countries are relatively symmetric. Moreover, as indicated above, when free trade is not a stable equilibrium, the FTA $\langle\{ab\}\rangle$ is a stable equilibrium of the FTA game when c is sufficiently high cost and low cost countries (a and b) are relatively symmetric. Note also that FTA between two high cost countries $\langle\{bc\}\rangle$ is stable when i is sufficiently low cost and b and c are relatively symmetric.

Over regions A and B in Figure 3 multiple equilibria obtain: all three agreements $\langle\{ab\}\rangle$, $\langle\{ac\}\rangle$ and $\langle\{F\}\rangle$ are stable over region A while only $\langle\{ab\}\rangle$ and $\langle\{ac\}\rangle$ are stable over region B. To understand the source of this multiplicity, consider region A (a similar logic applies for region B). Over region A, the only possible deviation from free trade $\langle\{F\}\rangle$ is the joint deviation of countries b and c from the announcements $\{a, c\}$ and $\{a, b\}$ to $\{\phi, c\}$ and $\{\phi, b\}$. This deviation is not self-enforcing since taking a's announcement as given, country c has an incentive to further deviate from $\{\phi, b\}$ to $\{a, b\}$. On the other hand, there are two non-self-enforcing deviations from $\langle\{ab\}\rangle$. First, the joint deviation of countries a and c from the announcements $\{b, \phi\}$ and $\{\phi, \phi\}$ to $\{\phi, c\}$ and $\{a, \phi\}$ is *not* self-enforcing since country a has an incentive to further deviate from $\{\phi, c\}$ to $\{b, c\}$. Second, the joint deviation of all countries from $\{b, \phi\}$, $\{a, \phi\}$ and $\{\phi, \phi\}$ to $\{b, c\}$, $\{a, c\}$ and $\{a, b\}$ is also *not* self-enforcing since taking i's announcement as given countries b and c have incentives to deviate from $\{a, c\}$ and $\{a, b\}$ to $\{\phi, c\}$ and $\{\phi, b\}$. As a result, $\langle\{ab\}\rangle$ is a stable equilibrium. Finally, over the same region, the joint deviation of all countries from $\{\phi, c\}$, $\{\phi, \phi\}$ and $\{a, \phi\}$ to $\{b, c\}$, $\{a, c\}$ and $\{a, b\}$ is *not* self-enforcing since countries b and c have incentives to deviate from $\{a, c\}$ and $\{a, b\}$ to $\{\phi, c\}$ and $\{\phi, b\}$. As a result, $\langle\{ac\}\rangle$ is also a stable equilibrium.

Finally, figure 4 provides additional results regarding the welfare implications of FTA option.[23]

——Insert Figure 4 here——

Three distinct regions are shown in region I: $\langle\{ab\}\rangle$ is the stable equilibrium under the FTA game while free trade $\langle\{F\}\rangle$ obtains under the No FTA game; region II: $\langle\{ab\}\rangle$ is the stable agreement under the FTA game whereas no agreement $\langle\{\Phi\}\rangle$ is the stable outcome under the No FTA game; region III: $\langle\{bc\}\rangle$ is the stable equilibrium under the FTA game whereas no agreement $\langle\{\Phi\}\rangle$ obtains under the No FTA game.

Proposition 7: *Suppose demand is linear. A comparison of the stable agreements under FTA game and the No FTA game yields the following results: (i) over region I, FTA option has beneficial welfare effects for the highest cost country (c) but harmful welfare effects for the other two countries as well as the world as a whole; and (ii) over region II and region III, the FTA option has beneficial welfare effects for all countries (and therefore the world as a whole).*

Why does the highest cost country benefits from the FTA option? This is because granting market access to the highest cost country does not impose too high a cost on the lower cost countries since they are at a competitive advantage. However, the willingness of low cost countries to enter into trade agreements with the high cost country confers an advantage upon the high cost country who ends up exploiting it in equilibrium. In fact, over region I the highest cost country

[23] For simplicity, we assume that the 'natural' trading block $\langle\{ab\}\rangle$ rather than $\langle\{ac\}\rangle$ obtains as the stable agreement over region A in figure 3.

always prefers $\langle\{ab\}\rangle$ to $\langle\{F\}\rangle$ and $\langle\{\Phi\}\rangle$ under the FTA game and No FTA game respectively. However, over region I, countries a and b prefer free trade $\langle\{F\}\rangle$ that is not stable under the FTA game. As a result, these countries are worse off with the FTA option over region I and the lower the cost of a country the larger the loss it suffers from the FTA option. Over region II, when $\langle\{ab\}\rangle$ is a stable equilibrium, countries countries a and b have no unilateral incentive to deviate to $\langle\{\Phi\}\rangle$. Over the same region, country c also prefers $\langle\{ab\}\rangle$ to $\langle\{\Phi\}\rangle$ because of the tariff complementarity effect. Consequently, world welfare is higher under $\langle\{ab\}\rangle$ than under $\langle\{\Phi\}\rangle$. An analogous explanation applies for the FTA $\langle\{bc\}\rangle$ over region III.

6. Concluding remarks

This paper contributes to the long-standing debate regarding the relationship between preferential and multilateral trade liberalization by analyzing two trade policy games: one where countries can choose between both types of trade liberalization and another where they can only pursue the multilateral route. We show that the effect of the option to form FTAs on the likelihood of achieving global free trade depends upon the external tariffs of FTA members as well as on the degree of asymmetry between countries. Somewhat ironically, relatively low FTA tariffs create incentives for free-riding on the part of non-member countries and can work *against* the goal of achieving global free trade. On the other hand, when FTAs do not significantly liberalize trade with respect to others, there exist circumstances where multilateral free trade is an equilibrium only when countries have the option to form FTAs.

We also find that FTAs can deliver welfare improving trade liberalization when multilateral free trade is infeasible. In this sense, GATT Article XXIV is desirable – it may indeed be better to have some preferential trade liberalization when multilateral liberalization is infeasible. In fact, the underlying asymmetry in our model leads to an interesting new insight: a pair of bilateral FTAs between a low cost country and two high cost ones can be welfare preferred to free trade.

Our analysis helps sharpen the stumbling versus building bloc debate regarding FTAs by highlighting conditions under which each of the two effects is more likely to obtain. When countries are relatively symmetric, the option to pursue FTAs does more harm than good. On the other hand, under asymmetry free trade is harder to sustain and FTAs can actually be desirable from a world welfare perspective.

7. Appendix

All supporting calculations and proofs not provided in the text are given below.

Proposition 1

Differentiating world welfare with respect to t_i gives:

$$\frac{dWW}{dt_i} = \frac{dS_i}{dt_i} + \sum_i \frac{d\Pi_{ij}(t_i)}{dt_i} \text{ where } j \neq i. \tag{7.1}$$

Using $u' = p_i$ and $x_i = x_{ii} + \sum_j x_{ji}$, we have

$$\frac{dS_i}{dt_i} = -\sum_j x_{ji}\left[\frac{dp_i}{dt_i} - 1\right] + t_i \sum_j \frac{dx_{ji}(t_i)}{dt_i} + [p_i - \zeta]\frac{dx_{ii}}{dt_i} \tag{7.2}$$

Also note that

$$\sum_j \frac{d\Pi_{ji}(t_i)}{dt_i} = \sum_j x_{ji}\left[\frac{dp_i}{dt_i} - 1\right] + \sum_j \frac{dx_{ji}(t_i)}{dt_i}[p_i - \zeta - t_i] \tag{7.3}$$

which implies

$$\frac{dWW}{dt_i} = [p_i - \zeta]\frac{dx_{ii}}{dt_i} + \sum_j \frac{dx_{ji}(t_i)}{dt_i}[p_i - \zeta]$$

Using $x_i = x_{ii} + \sum_j x_{ji}$, the following is immediate:

$$\frac{dWW}{dt_i} = (p_i - \zeta)\frac{dx_i}{dt_i} < 0 \text{ since } \frac{dx_i}{dt_i} < 0 \tag{7.4}$$

As a result, under symmetry, world welfare is maximized under free trade. Following an analogous procedure, we can show that even under a bilateral FTA, it is socially optimal to set the FTA's external tariff to zero.

Proposition 2

Here we show by contradiction that $\bar{t} > \underline{t}$. Suppose $\bar{t} \leq \underline{t}$. Consider $t^f = \bar{t}$. Then from condition (4.21), the following holds:

$$\sum_{j \neq i} \Pi_{ij}(t^f, 0) = \sum_{j \neq i} \Pi_{ij}(t^*, t^*) \tag{7.5}$$

and since $t^f = \bar{t} \leq \underline{t}$, from deviation (II) in (4.8) we have:

$$\Delta w_i(jk) = \Delta s_i(jk) + \Delta \pi_{ij}(jk) > 0 \tag{7.6}$$

where

$$\Delta s_i(jk) \equiv S_i(t^*, t^*) - S_i(0, 0)$$

and

$$\Delta \pi_{ij}(jk) \equiv \sum_{j \neq i} \Pi_{ij}(t^f, 0) - \sum_{j \neq i} \Pi_{ij}(0, 0)$$

Using (7.5) and the fact that $\Delta s_i(jk) = \Delta s_i(\Phi)$, we can rewrite $\Delta w_i(jk)$ in (7.6) as follows:

$$\Delta w_i(jk) = \Delta s_i(\Phi) + \Delta \pi_{ij}(\Phi) > 0 \tag{7.7}$$

However we know from Proposition 1 that the above inequality is impossible. As a result, it must be that $\bar{t} > \underline{t}$. This completes the proof.

Proposition 4

Unilateral deviations from free trade: Since $\zeta < \zeta^{ab,ac} = \zeta^{ab,bc}$, country c has no incentive to deviate unilaterally from $\{a, b\}$ to $\{a, \phi\}$ or $\{\phi, b\}$. Similarly, since $\zeta < \zeta^{ab}$ country c does not deviate unilaterally from $\{a, b\}$ to $\{\phi, \phi\}$.

Now, consider unilateral deviation incentives of low cost countries (a and b). Since countries a and b are symmetric, hereafter we only consider the unilateral deviations of country a. We first argue that country a has no incentive to unilaterally deviate from $\{b, c\}$ to $\{b, \phi\}$. Note that under symmetry

$$\Delta w_a(ab, bc) = w_a(ab, bc) - w_a(F) < 0 \text{ if } \zeta = 0 \tag{7.8}$$

From A1, $\frac{\partial \Delta w_a(ab,bc)}{\partial \zeta} < 0$. As a result, country a does not deviate from $\{b, c\}$ to $\{b, \phi\}$:

$$\Delta w_a(ab, bc) < 0 \text{ for all } \zeta \tag{7.9}$$

Similarly, country a has no incentive to unilaterally deviate from $\{b, c\}$ to $\{\phi, c\}$. Note that under symmetry

$$\Delta w_a(ac, bc) = w_a(ac, bc) - w_a(F) < 0 \text{ if } \zeta = 0 \tag{7.10}$$

Also,

$$\lim_{\zeta \longrightarrow \zeta^P} \Delta w_a(ac, bc) < 0 \tag{7.11}$$

Using the above inequality and A1, we have

$$\Delta w_a(ac, bc) < 0 \text{ for all } \zeta \tag{7.12}$$

Using an argument very similar to above, we can show that

$$\Delta w_a(bc) < 0 \text{ for all } \zeta \tag{7.13}$$

Thus, free trade is a Nash equilibrium of the FTA game. Next, we consider coalitional deviations from free trade.

Coalitional deviations: Consider the joint deviation of a low cost country (a) and the high cost country (c) from their respective announcements $\{b, c\}$ and $\{a, b\}$ to $\{\phi, c\}$ and $\{a, \phi\}$. We argue that this deviation is *not* self-enforcing, since taking country b's announcement as given, country a has an incentive to further deviate from $\{\phi, c\}$ to $\{b, c\}$.

We know that

$$\lim_{\zeta \longrightarrow \zeta^P} \Delta w_a(ac - (ab, ac)) < 0 \tag{7.14}$$

The above inequality and A1 imply that

$$\Delta w_a(ac - (ab, ac)) < 0 \text{ for all } \zeta \tag{7.15}$$

As a result, country a has an incentive to deviate further from $\{\phi, c\}$ to $\{b, c\}$ and the initial deviation is not self-enforcing.

Next, consider the joint deviation of low cost countries a and b from their respective announcements $\{b, c\}$ and $\{a, c\}$ to $\{b, \phi\}$ and $\{a, \phi\}$. This deviation is not self-enforcing because country a has an incentive to further deviate from $\{b, \phi\}$ to $\{b, c\}$. From A1, A2 and Proposition 1, it follows that

$$\Delta w_a(ab - (ab, ac)) < 0 \text{ for all } \zeta \tag{7.16}$$

As a result, the initial deviation is not self-enforcing.

Finally, it is trivial to establish that the two low cost countries have no incentive to deviate from their free trade announcements to $\{\phi, \phi\}$. As a result, multilateral free trade $\langle\{F\}\rangle$ is a stable equilibrium of the FTA game.

Proposition 5

We prove the above proposition in two parts: (i) $\tau^f < \underline{t}$ and (ii) $\underline{t} \leq \tau^f < \overline{t}$. We begin with part (a) and first show that neither country a nor b has an incentive to unilaterally deviate from $\{b, \phi\}$ and $\{a, \phi\}$ to $\{\phi, \phi\}$. Due to symmetry, we consider only country a. We know

$$\lim_{\zeta \longrightarrow \zeta^P} \Delta w_a(ab - \Phi) = w_a(ab) - w_a(\Phi) > 0 \tag{7.17}$$

and from A1 we know that $\frac{\partial \Delta w_a(ab - \Phi)}{\partial \zeta} < 0$. As a result, $\Delta w_a(ab - \Phi) > 0$ for all ζ.

Now consider coalitional deviations from $\langle\{ab\}\rangle$. Since the high cost country (c) prefers $\langle\{ab\}\rangle$ to $\langle\{F\}\rangle$ when $\zeta > \zeta^{ab}$, there can be no coalitional deviation from $\langle\{ab\}\rangle$ to $\langle\{F\}\rangle$.

Next, consider a joint deviation of countries a and c from their respective announcements $\{b, \phi\}$ and $\{\phi, \phi\}$ to $\{\phi, c\}$ and $\{a, \phi\}$.[24] Note that

$$\Delta w_c(ab - ac) = w_c(ab) - w_c(ac) > 0 \text{ if } \zeta = 0 \tag{7.18}$$

From A1, $\frac{\partial \Delta w_c(ab - ac)}{\partial \zeta} > 0$. Therefore, the following is immediate:

$$\Delta w_c(ab - ac) > 0 \text{ for all } \zeta \tag{7.19}$$

As a result, the high cost country (c) has no incentive to deviate jointly with country a (symmetrically with country b) from the initial announcements $\{b, \phi\}$

[24] Since countries a and b are symmetric, the deviation from $\langle\{ab\}\rangle$ to $\langle\{bc\}\rangle$ is equivalent to that from $\langle\{ab\}\rangle$ to $\langle\{ac\}\rangle$.

and $\{\phi,\phi\}$ to $\{\phi,c\}$ and $\{a,\phi\}$. Using very similar arguments we can rule out the joint deviation of countries a and c from their announcements $\{b,\phi\}$ and $\{\phi,\phi\}$ to $\{b,c\}$ and $\{a,\phi\}$.

Finally, consider the joint deviation of countries a, b, and c from their respective announcements $\{b,\phi\}$, $\{a,\phi\}$ and $\{\phi,\phi\}$ to $\{\phi,c\}$, $\{\phi,c\}$ and $\{a,b\}$. In order to show that this joint deviation is not self-enforcing, it is enough to show that countries a and b (subset of initially deviating countries) have an incentive to further deviate from $\{\phi,c\}$ and $\{\phi,c\}$ to $\{b,c\}$ and $\{a,c\}$. Note that under symmetry:[25]

$$\Delta w_a(ac,bc) = w_a(ac,bc) - w_a(F) < 0 \text{ if } \zeta = 0 \tag{7.20}$$

Also,

$$\lim_{\zeta \longrightarrow \zeta^P} \Delta w_a(ac,bc) < 0$$

and from A1 $\frac{\partial \Delta w_a(ac,bc)}{\partial \zeta} > 0$ we have:

$$\Delta w_a(ac,bc) < 0 \text{ for all } \zeta \tag{7.21}$$

As a result, the initial deviation is not self enforcing. Therefore, $\langle\{ab\}\rangle$ is a stable equilibrium of the FTA game if $\tau^f < \underline{t}$ and $\zeta > \zeta^\phi$. The proof of part (a) is done.

Now consider part (ii) where $\underline{t} \leq \tau^f < \overline{t}$. As shown in part ($i$), $\langle\{ab\}\rangle$ is a Nash equilibrium of the FTA game since neither country a nor b has an incentive to unilaterally deviate from $\{b,\phi\}$ and $\{a,\phi\}$ to $\{\phi,\phi\}$. Moreover, since country c prefers $\langle\{ab\}\rangle$ to $\langle\{F\}\rangle$ when $\zeta > \zeta^\phi > \zeta^{ab}$, there is no deviation from $\langle\{ab\}\rangle$ to $\langle\{F\}\rangle$.

Next, consider a joint deviation of countries a and c from $\{b,\phi\}$ and $\{\phi,\phi\}$ to $\{\phi,c\}$ and $\{a,\phi\}$. Suppose this deviation occurs. If so, country a has an incentive to deviate further from $\{\phi,c\}$ to $\{b,c\}$. From Proposition 2, the following is immediate under symmetry:

$$\Delta w_a(ac-(ab,ac)) = w_a(ac) - w_a(ab,ac) < 0 \text{ if } \zeta = 0 \tag{7.22}$$

As before, it is straightforward to argue that

$$\Delta w_a(ac-(ab,ac)) < 0 \text{ for all } \zeta \tag{7.23}$$

so that the initial deviation is not self-enforcing.

[25]Since countries a and b are symmetric, this holds for country b as well.

Consider now the joint deviation of countries a and c from $\{b, \phi\}$ and $\{\phi, \phi\}$ to $\{b, c\}$ and $\{a, \phi\}$. From A2, we know that $\zeta^{ab,ac} = \zeta^{ba,bc} > \zeta^{ab}$. Therefore, the following is immediate:

$$\Delta w_c(ab - (ab, ac)) = w_c(ab) - w_c(ab, ac) > 0 \text{ if } \zeta = \zeta^{ab} \tag{7.24}$$

From A1, $\frac{\partial \Delta w_c(ab-(ab,ac))}{\partial \zeta} > 0$. As a result, the high cost country (c) has no incentive to deviate from $\{\phi, \phi\}$ to $\{a, \phi\}$:

$$\Delta w_c(ab - (ab, ac)) > 0 \text{ for all } \zeta > \zeta^{\phi} \tag{7.25}$$

Finally, consider the deviation of all countries from $\{b, \phi\}$, $\{a, \phi\}$ and $\{\phi, \phi\}$ to $\{\phi, c\}$, $\{\phi, c\}$, and $\{a, b\}$. Suppose that this deviation occurs. If so, taking country c's announcement as given, countries a and b have a joint incentive to deviate further from $\{\phi, c\}$ and $\{\phi, c\}$ to $\{b, c\}$ and $\{a, c\}$. Since countries a and b are symmetric, consider country a only. Using Proposition 2, we have:

$$\Delta w_a(ac, bc) = w_a(ac, bc) - w_a(F) < 0 \text{ if } \zeta = 0 \tag{7.26}$$

Using A2 and evaluating $\Delta w_a(ac, bc)$ as $\zeta \longrightarrow \zeta^P$, it is straightforward to argue that

$$\Delta w_a(ac, bc) < 0 \text{ for all } \zeta \tag{7.27}$$

As a result, the initial deviation is not self-enforcing. Therefore, $\langle\{ab\}\rangle$ is a stable equilibrium of the FTA game if $\underline{t} \leq \tau^f < \bar{t}$ and $\zeta > \zeta^{\phi}$.

References

[1] Aghion, Philippe, Pol Antràs, and Elhanan Helpman, 2004. "Negotiating Free Trade." NBER Working Paper 10721.

[2] Bagwell, Kyle, and Robert. W. Staiger, 1997a. "Multilateral Tariff Cooperation During the Formation of Free Trade Areas." *International Economic Review* 38, 291-319.

[3] Bagwell, Kyle and Robert W. Staiger, 1997b, "Multilateral Tariff Cooperation during the Formation of Customs Unions," *Journal of International Economics* 42, 91-123.

[4] Bernheim, Douglas B., Bezalel Peleg and Michael Whinston, 1987. "Coalition-proof Nash Equilibria I. Concepts." *Journal of Economic Theory* 42, 1-12.

[5] Bhagwati, Jagdish. *The World Trading System at Risk*, 1991, Princeton University Press, Princeton, NJ.

[6] Bhagwati, Jagdish, Arvind Panagariya, and Pravin Krishna, eds., *Trading Blocs*, 1990, The MIT Press, Cambridge, MA.

[7] Bhagwati, Jagdish, Arvind Panagariya. "Preferential Trading Areas and Multilateralism – Strangers, Friends, or Foes?" in Bhagwati, Jagdish, Arvind Panagariya, and Pravin Krishna, eds., *Trading Blocs*, 1999, The MIT Press, Cambridge, MA.

[8] Bohara, Alok K., Gawande, Kishore and Pablo Sanguinetti, 2004. "Trade Diversion and Declining Tariffs: Evidence from MERCOSUR." *Journal of International Economics* 64, 65-88.

[9] Bond, Eric W., and Constantinos Syropoulos, 1996. "The Size of Trading Blocs: Market Power and World Welfare Effects." *Journal of International Economics* 40, 411-437.

[10] Bond, Eric W., Syropoulos, Constantinos, and Winters, L. Alan, 2001. "Deepening of Regional Integration and Multilateral Trade Agreements." *Journal of International Economics* 53, 335-362.

[11] Brander, James A., and Barbara J. Spencer. "Tariff Protection and Imperfect competition." In ed. H. Kierzkowski *Monopolistic Competition and International Trade*, 1984, Oxford University Press, Oxford.

[12] Cadot, Olivier, Jaime de Melo, and Marcelo Olarreaga, 1999. "Regional Integration and Lobbying for Tariffs Against Nonmembers." *International Economic Review* 40, 635-658.

[13] Chang, Won and Winters, L. Alan, 2002. "How Regional Blocs Affect Excluded Countries: The Price Effects of MERCOSUR." *American Economic Review* 92, 889-904.

[14] Dutta, Bhaskar and Suresh Mutuswami, 1997. "Stable Networks." *Journal of Economic Theory* 76, 322-344.

[15] Ethier, Wilfred J., 1998. "Regionalism in a Multilateral World." *Journal of Political Economy* 106, 1214-1245.

[16] Freund, Caroline, 2000. "Multilateralism and the Endogenous Formation of Preferential Trade Agreements." *Journal of International Economics* 52, 359-376.

[17] Grossman, Gene M. and Elhanan Helpman, 1995. "The Politics of Free-Trade Agreements." *American Economic Review* 85, 667-690.

[18] Furusawa Taiji and Hideo Konishi, 2003. "Free Trade Networks." Unpublished manuscript.

[19] Knetter, Michael M., 1993 "International Comparisons of Pricing to Market Behavior." *American Economic Review* 83, 473-486.

[20] Krishna, Pravin, 1998. "Regionalism and Multilateralism: A Political Economy Approach." *The Quarterly Journal of Economics* 113, 227-251.

[21] Krugman, Paul R. "The Move Toward Free Trade Zones." in *Policy Implications of Trade and Currency Zones: A Symposium Sponsored by the Federal Reserve Bank of Kansas City*, Federal Reserve Bank of Kansas City, Kansas City, 7-41, 1991.

[22] Lawrence, Robert. *Regionalism, Multilateralism and Deeper Integration*, 1996, Washington, D.C.

[23] Levy, Philip I., 1997. "A Political-Economic Analysis of Free Trade Agreements." *American Economic Review* 87, 506-519.

[24] Maggi, Giovanni and Andres Rodiguez-Clare, 1998. "The Value of Trade Agreements in the Presence of Political Pressures." *Journal of Political Economy* 106, 574-601.

[25] McLaren, John, 1997. "Size, Sunk Costs, and Judge Bowker's Objection to Free Trade." *American Economic Review* 87, 400-420.

[26] Richardson, Martin, 1995. "Tariff Revenue Competition in a Free Trade Area." *European Economic Review* 39, 1427-1437.

[27] Riezman, Raymond, 1999. "Can Bilateral Trade Agreements Help Induce Free Trade?" *Canadian Journal of Economics* 32, 751-766.

[28] Saggi, Kamal, 2005. "Preferential Trade Agreements and Multilateral Tariff Cooperation." *International Economic Review*, forthcoming.

[29] Summers, Lawrence H., "Regionalism and the World Trading System," in *Policy Implications of Trade and Currency Zones*, 1991, Kansas City: Federal Reserve Bank.

[30] Jacob, Viner, 1950. *The Custom Union Issue*. New York: Carnegie Endowment for International Peace.

[31] Yi, Sang-Seung, 1996. "Endogenous Formation of Customs Unions under Imperfect Competition: Open Regionalism is Good." *Journal of International Economics* 41, 153-177.

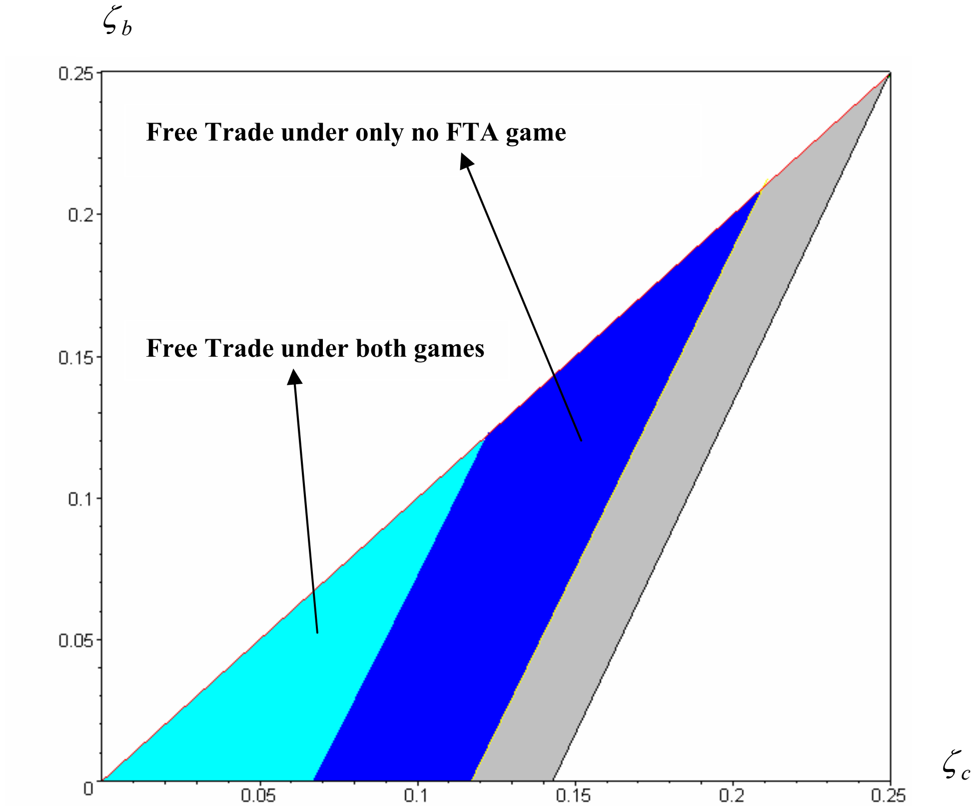

Figure 1: Multilateral Free Trade under both games

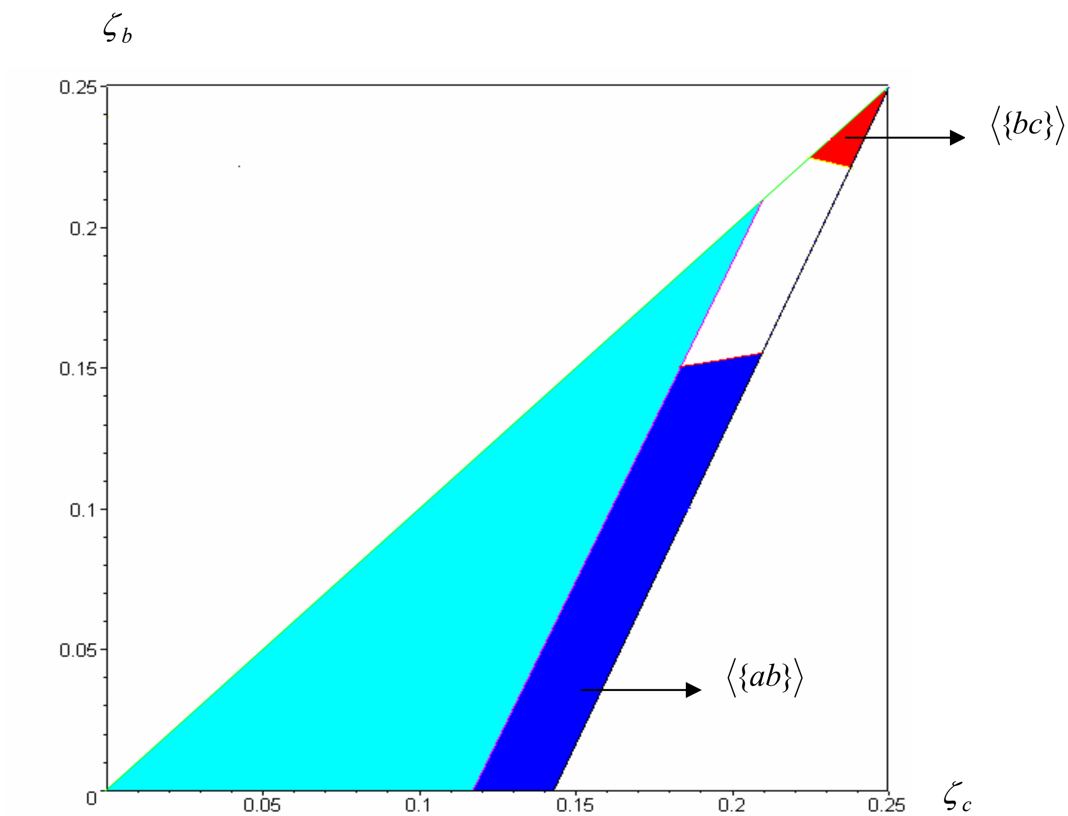

Figure 2: Bilateral FTAs when Free Trade is infeasible

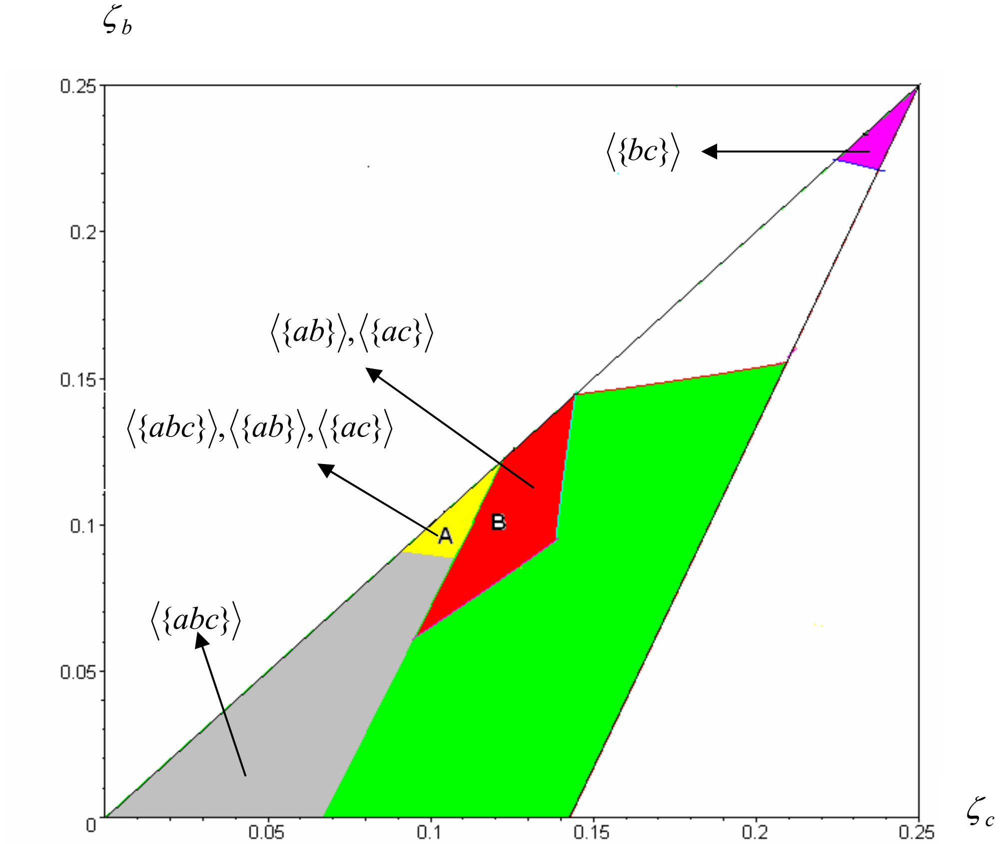

Figure 3: Stable Agreements under the FTA Game

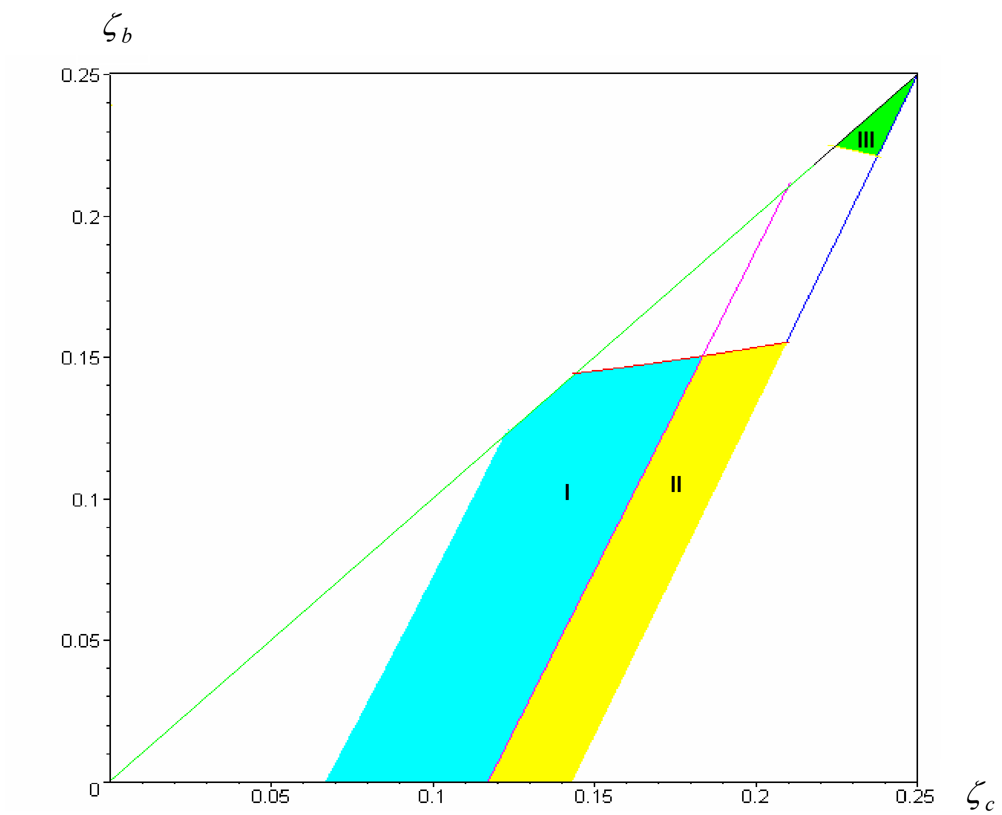

Figure 4: Stable Agreements under the FTA Game vs. No FTA game