External Trade Diversion, Exclusion Incentives and the Nature of Preferential Trade Agreements

Missios, Paul and Saggi, Kamal and Yildiz, Halis Murat

19 September 2014

External Trade Diversion, Exclusion Incentives and the Nature of Preferential Trade Agreements

Paul Missios*, Kamal Saggi† and Halis Murat Yildiz‡

Abstract

We examine whether the pursuit of free trade agreements (FTAs) affects the prospects of global free trade differently than the pursuit of Customs Unions (CUs). Our analysis is driven by a fundamental difference between these two types of preferential trade agreements (PTAs): while CU members impose jointly optimal common tariffs on non-members, members of an FTA adopt individually optimal external tariffs. This implies that (a) FTAs are relatively more flexible than CUs since an FTA member is free to undertake further trade liberalization with respect to non-members whereas a CU member can do so only if all CU members wish to do the same and (b) coordination during tariff setting allows CU members to pool their market power. In our comparative advantage based three country framework, the formation of either type of PTA induces the non-member to *lower* its external tariffs due to the reduction in the volume of exports flowing from members to the non-member (we call this *external trade diversion*). While the pursuit of CUs prevents free trade from emerging in equilibrium, the pursuit of FTAs does not. This key result is driven by the relative flexibility of FTAs; the higher market power of CUs by itself does not undermine the objective of reaching global free trade. Even if CUs are prohibited from raising their tariffs above pre-existing levels, free trade still fails to obtain if the pursuit of PTAs takes the form of CUs.

Keywords: Free Trade Agreement, Customs Union, Hub and Spoke Agreements, Free Trade, Optimal Tariffs. *JEL Classifications*: F13, F12.

*Department of Economics, Ryerson University, 350 Victoria Street, Toronto, ON, Canada M5B 2K3. Phone: 416-979-5000 (ext 6186); e-mail: pmissios@ryerson.ca.

†Department of Economics, Vanderbilt University, Nashville, TN 37235-1828. Phone: 615-322-3237; e-mail: kamal.saggi@vanderbilt.edu.

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

1 Introduction

Preferential trade agreements (PTAs) are more popular than ever before while multilateral trade liberalization seems to have come to a stand still. The ever increasing popularity of PTAs can be gauged from the fact that as of 15 January 2012, 511 notifications of PTAs (counting goods and services separately) had been received by the GATT/WTO. Of these notifications, 319 PTAs are already in force with others scheduled for implementation in the near future. By contrast, the last round of multilateral trade negotiations – i.e. the Doha Round – failed to come to a fruitful conclusion despite eleven years of intense negotiations. Economists and policy-makers have long suspected that the contrasting fortunes of these two types of trade liberalization may be inter-related. More specifically, there is widespread concern that the formation of PTAs has served to undermine multilateral liberalization. Our objective in this paper is to investigate this issue with a finer lens by comparing the implications of the two most popular PTAs – i.e. free trade agreements (FTAs) and customs unions (CUs) – for global free trade.[1]

As is well known, the central difference between an FTA and a CU is that members of a CU impose jointly optimal common tariffs on non-members whereas FTA members adopt individually optimal external tariffs. This difference in tariff setting behavior between the two types of PTAs has two important consequences. First, while an FTA member is free to enter into additional FTAs with non-member countries that do not include its existing FTA partners, a CU member can do so only if all members agree to participate in the new agreement. For example, the United States has signed several FTAs since the ratification of NAFTA in 1995 that do not include Mexico and Canada as partners whereas the enlargement of the European Union over the years has been subject to the approval of all existing members. In other words, FTAs are more *flexible* than CUs. Second, as was noted in Bagwell and Staiger (1997a) and Bond and Syropolous (1996), the coordination of tariffs within a CU allows members to *pool their market power*, thereby leading them to impose relatively higher external tariffs than FTAs.

The objective of this paper is to isolate the implications of these fundamental differences between FTAs and CUs for the prospects of free trade in the global economy. Our approach to the formation of PTAs follows that of Saggi and Yildiz (2010) and Saggi et. al. (2013) under which PTAs emerge endogenously as the outcome of a game of trade liberalization

[1] Roughly 90% of the existing PTAs take the form of FTAs, with CUs comprising the rest (Freund and Ornelas, 2010). However, the existing CUs do involve major trading areas of the world: the EU and much of Latin America (where MERCOSUR resides).

between three countries that are free to pick their PTA partners as well as their tariff levels. We consider two games: one in which PTAs takes the form of an FTA and another where they are CUs. In addition to deriving Nash equilibria of these games of PTA formation, we isolate Nash equilibria that are stable or coalition proof. By this we mean that they are immune to self-enforcing coalitional deviations, where a coalitional deviation is said to be self-enforcing if a proper subset of players in the deviating coalition have no incentive to undertake a further deviation (see Bernheim et. al., 1987).

Our model addresses the important issue of when and why countries prefer to liberalize preferentially as opposed to doing so multilaterally. By contrast, existing literature takes one of two approaches: it either (i) takes PTAs to be exogenously given and compares incentives for multilateral liberalization in their presence to those in their absence or (ii) simply asks whether a pair of countries benefit from entering into a PTA given that there exists no trade agreement between them, a comparison that does not tell us much about when and why they deliberately choose to *exclude* others from their mutual trade liberalization. To address the exclusion incentive in a convincing manner, one needs a model that gives all countries an active voice during negotiations so that one can determine whether some countries prefers to exclude others from the trade agreement even though they themselves wish to be included. We provide such a model and use it to assess the strength of the exclusion incentive under FTAs and CUs as well as the ability of member countries to exercise it in equilibrium.

Building on a trading structure similar to that of Horn et al. (2010), we endogenize the formation of PTAs in a three country comparative advantage based model in which each country exports a unique good to the other two. The model identifies a novel type of tariff complementarity: when two countries form a PTA (either an FTA or a CU), the excluded country finds it optimal to voluntarily *reduce* its tariffs on member countries.[2] As a result, PTA members benefit not only from their mutual trade liberalization but also from the unilateral liberalization that is induced in the non-member country. This finding is reminiscent of the reciprocated unilateralism result of Krishna and Mitra (2005) who showed that unilateral tariff liberalization by a (large) country can result in reciprocal tariff reduction by its smaller trading partner.[3] In their model, unilateral liberalization by

[2]By contrast, existing literature has tended to focus on how the formation of a PTA can induce member countries to voluntarily lower their external tariffs on non-members. See Bagwell and Staiger (1997a, 1997b and 1998), Bond et al. (2004), and Estevadeordal et al. (2008).

[3]See also Coates and Ludema (2001) for a theory of trade policy leadership based on repeated interaction between a large and a small country. In their model, a large country can undertake unilateral liberalization in order to induce reciprocal trade liberalization by the small country.

the large country encourages the formation of an export lobby in the small country which then competes effectively with the import-competing lobby to lower tariffs and export taxes.

It is worth emphasizing that our model highlights the consequences of a type of trade diversion that has generally been overlooked in the literature. Traditionally, trade diversion is defined as the increase in trade between PTA members that occurs at the expense of exports of non-members to the PTA, i.e., the traditional notion of trade diversion refers to the reduction in the volume of *imports* that are sourced by PTA members from non-members. By contrast, in our model, the formation of a PTA reduces the volume of *exports* of PTA members to the non-member - a phenomenon we refer to as *external trade diversion*. Indeed, such external trade diversion is precisely what makes it optimal for the non-member to voluntarily lower its tariffs on PTA members.[4] To the best of our knowledge, our model is the first to capture the effects of such external trade diversion on (1) the tariff structure of non-member country (2) the welfare of members and non-members (3) the nature of PTAs that arise in equilibrium and (4) the implications of FTAs and CUs for global free trade, issues that cannot be adequately addressed by models that hold the tariffs of non-members constant (by taking them as exogenously given) or assume a trading structure under which PTA formation does not affect them.[5]

The fact that external trade liberalization by a CU member is conditional on the approval of other members implies that, relative to an FTA, a CU is less susceptible to *opportunistic unilateral deviations* by member countries. To see this clearly, suppose there are three countries (i, j, and k) and countries i and j have an FTA with each other. Then, country i is free to negotiate a *separate* FTA with country k that does not include country j. Such a pattern of bilateral FTAs is usually referred to as a *hub and spoke arrangement*, with country i as the hub and countries j and k as the spokes. By contrast, if countries i and j are in a CU, the only way one of them can undertake further trade liberalization with country k is if they both agree to reduce their common external tariff, perhaps in return for some tariff reductions on country k's part. As a result, a hub and spoke type of trading arrangement is simply *infeasible* under a CU. We show that this crucial difference between a CU and an FTA has important consequences for multilateral trade liberalization. In particular, in our three country model of symmetric countries, we find that while free trade

[4]Indeed, in our model, this logic would also apply to unilateral liberalization by any country so long as it is preferential in nature (i.e. applies only to one country).

[5]See, for example, Bagwell and Staiger (1997a, 1997b, 1998), Krishna (1998), Ornelas (2005a and 2005b), and Saggi and Yildiz (2010). In general, the literature has tended to focus on the reduction of market access experienced by non-members due to the formation of a PTA.

obtains as the unique stable outcome when the alternative to multilateral liberalization are bilateral FTAs, such is not the case when the alternative takes the form of CUs: under the latter scenario, a CU between two countries emerges as the stable outcome even though all three countries are symmetric. Thus, in our model, CUs undermine global free trade whereas FTAs do not even though the exclusion incentive exists under both types of PTAs in the sense that the welfare of members of both types of PTAs is strictly higher than that under free trade.

We employ a three stage game. In the first stage, each country simultaneously announces the names of countries with whom it wants to sign a PTA. Here it is important to note that we employ a game of announcements or proposals for a PTA link. In our game, a country does not announce in favor of a specific trade agreement but rather names partners with whom it wants to form a PTA (FTA or CU). If no announcements match or the only matching announcement is in favor of no agreement, then no agreement would prevail. A bilateral PTA would form if and only if two countries announce each others' names and there is no other matching announcement. In the FTA game, due to independent external tariff determination, there is a possibility of two independent FTAs (where the hub country has an independent FTA with each of the spoke countries while spoke countries do not have an FTA with each other) and this would arise (i) if a country (this country would be hub) announces in favor of signing an FTA with the other two countries while (ii) the other two countries (they would be spokes) announce only in favor of signing an FTA with the former country (not with each other). Finally, free trade would emerge if all countries announce in favor forming a PTA with all the other countries. Next, given the trade policy regime, countries choose their optimal tariffs. In the final stage of the game, production, consumption and international trade take place.

To understand the intuition behind this key result, first consider the game where PTAs take the form of FTAs. Suppose all countries (say i, j and k) announce in favor forming a PTA with all the other countries. Under such a case, due to exclusion incentives, two of these countries (say i and j), taking the announcement of their complement as fixed (i.e. country k still announces in favor of an FTA with i and j), would jointly deviate to announcements where they only call in favor of an FTA with each other. For this deviation to be self enforcing, a proper subset of the initially deviating countries (means i or j) should not have an incentive to further deviate to another announcement, taking the announcement of complement (country k) as fixed. However, each of the initially deviating country (i or j) has an incentive to further deviate to an announcement in favor of FTAs with both

countries since being a hub is better than being an FTA member. Therefore, the initial deviation is not self-enforcing and thus free trade is coalition-proof Nash equilibrium under the FTA game. By contrast, due to common external tariff determination in the CU game, announcing in favor of a CU with the other two countries is tantamount to announcing in favor of global free trade. Since two independent CUs (a hub and spoke type arrangement) is not even feasible, the above explained further announcement deviation would not happen and thus the initial deviation would be self-enforcing. Thus, free trade fails to be stable and a bilateral CU arises as the unique stable equilibrium of the CU game. As a result, whereas the exclusion incentive is reflected in the equilibrium of the CU game, it goes unexpressed in the FTA game due to the lure of a hub and spoke arrangement and the flexibility that FTA members have.

To isolate the market power effect of a CU from their relative lack of flexibility, we also consider a scenario where CU members are not allowed to raise their tariffs above pre-existing levels. This experiment is also well motivated on policy grounds: Article XXIV of GATT – the key clause that sanctions PTAs in the WTO – forbids member countries of a PTA from raising tariffs on non-members. When CUs are constrained in this manner, we find that they continue to arise as a stable equilibrium thereby preventing the realization of global free trade. Thus, the restriction on external tariffs mandated by Article XXIV simply serves the soften the impact of a CU on outsiders; it does not eliminate the exclusion incentive that gives rise to the CU in the first place. Thus, it is the relative flexibility of FTAs over CUs that helps the prospects of global free trade and not their weaker market power.

2 Trade model

Our model of trade agreements is an adapted version of the two-country model of Horn, Maggi, and Staiger (2010). We consider a perfectly competitive world with three large countries: $z = i, j$, and k and three (non-numeraire) goods: $g = I$, J, and K and a numeraire good v_0. On the demand side, the representative citizen's utility function is linear in the numeraire good and separable in the non-numeraire goods:

$$U(\mathbf{v}, v_0) = u(\mathbf{v}) + v_0, \tag{1}$$

where $\mathbf{v} = [v_I, v_J, v_K]$ is the consumption vector for the three non-numeraire goods, v_0 denotes the consumption of the numeraire good, and $u(\mathbf{v})$ is quadratic and additively

separable in the three non-numeraire goods. The demand for good g in country z is then given by

$$d_z^g(p_z^g) = \alpha - p_z^g. \tag{2}$$

where p_z^g denotes the consumer price of good g in country z. Assuming that the population in each country is a continuum of measure one, we can write the consumer surplus associated with good g in country z as:

$$CS_z^g(p_z^g) = u_z^g[d_z^g(p_z^g)] - p_z^g d_z^g(p_z^g) \tag{3}$$

On the supply side, as in Horn et. al. (2010), labour (l) is the only factor of production which is employed in the production of the numeraire good that is produced one-for-one from labor. The supply of labor is assumed to be large enough that the numeraire good is always produced in a positive amount; therefore the equilibrium wage is equal to one.

Each non-numeraire good is produced from labor with diminishing returns. In particular, we assume the following production function for non-numeraire good g in country z: $Q_z^g = \sqrt{2\lambda_z^g l_g}$, where Q_z^g is the production of good g in country z and l_g is the labor employed in the production of good g. The supply function of good g in country z is as follows:

$$s_z^g(q_z^g) = \lambda_z^g q_z^g \tag{4}$$

where q_z^q denotes the producer price for good g in country z.

We assume that there exists a symmetric comparative advantage structure across countries: $\lambda_i^I = \lambda_j^J = \lambda_k^K = 1 + \lambda$ while $\lambda_i^J = \lambda_i^K = \lambda_j^I = \lambda_j^K = \lambda_k^I = \lambda_k^J = 1$. In other words, each country has a comparative advantage in one good while having a comparative disadvantage in the other two goods: each country exports the good that is indexed by the same uppercase letter as the identity of the country. For example, country i exports good I while importing good J from country j and good K from country k. Thus, there are two *competing importers* for each non-numeraire good and the model is Ricardian in nature with diminishing returns in the production of each good. Country z's producer surplus in good g as follows:

$$PS_z^g(q_z^g) = \int s_z^g(q_z^g) dq_z^g = \frac{1}{2}\lambda_z^g (q_z^g)^2 \tag{5}$$

As a representative scenario for all goods and countries, consider good I (i.e. the good in which country i is has a comparative advantage). Let t_{ji} be the tariff imposed by country

j on its imports of good I from country i.[6] Given that all countries are large, world price of good I depends on the tariffs chosen by countries j and k but to simplify notation we suppress the dependence of prices on tariffs and simply denote the price of good I by p_i^I.

Due to the absence of any tariff in country i on good I, the consumer and producer prices of good I in country i are equal: $q_i^I = p_i^I$. As there is no domestic taxation for the import competing sectors, producer and consumer prices are also equal: $q_i^g = p_i^g$, where $g \neq I$. Finally, ruling out prohibitive tariffs yields the following no-arbitrage conditions for good I:

$$p_j^I = p_i^I + t_{ji} \text{ and } p_k^I = p_i^I + t_{ki} \tag{6}$$

Let m_j^I and m_k^I be the imports of good I by countries j and k:

$$m_j^I = d(p_j^I) - s_j^I(q_j^I) \text{ and } m_k^I = d(p_k^I) - s_k^I(q_k^I) \tag{7}$$

Similarly, let x_j^I and x_k^I denote country i's exports (of good I) to countries j and k where

$$x_j^I = s_i^I(q_i^I) - d(p_i^I) \tag{8}$$

and

$$x_k^I = s_i^I(q_i^I) - d(p_i^I) \tag{9}$$

The equilibrium world price of each traded good g is determined by the market clearing conditions. Market clearing for good I requires that country i's export to a country equals the imports of that country:

$$x_z^I = m_z^I \text{ where } z = j, k \tag{10}$$

Before proceeding with the derivation of optimal tariffs, it is useful to highlight some important features of the model. Since each country exports a unique good in the model, a country's tariff on one of its trading partners has no impact on the volume of its imports from its second trading partner. This implies that a country's external tariffs are independent of one another (since they apply to different goods, each with its own demand function).[7] Second, if two countries liberalize trade only towards one another, they import more from each other and start exporting less to the third country – a phenomenon which we call *external trade diversion*. As we will see below, this reduction in the volume of

[6] We assume that tariff revenues for each good are redistributed unifomly to all individuals.

[7] One consequence of this feature is that the MFN principle of non-discrimination plays no role in our model. Discriminatory tariffs obtain in our model if two countries choose to liberalize on a preferntial basis.

exports to the third country in turn has implications for its optimal tariffs.

Country z's welfare is defined as the sum of consumer surplus, producer surplus, and tariff revenue over all goods:

$$w_z = \sum_g CS_z^g(p_z^g) + \sum_g PS_z^g(q_z^g) + \sum_{h=j,k} t_{zh} m_z^H \tag{11}$$

In the absence of any trade agreement, each country chooses its tariffs to maximize its welfare. To derive optimal tariffs, we follow the approach of Feenstra (2004) and Broda et. al (2008). Consider country k's tariff problem. Differentiating w_k with respect to t_{ki}, we obtain:

$$\frac{\partial w_k}{\partial t_{ki}} = t_{ki} \frac{\partial m_k^I}{\partial p_k^I} \frac{\partial p_k^I}{\partial t_{ki}} - m_k^I \frac{\partial p_i^I}{\partial t_{ki}} \tag{12}$$

The first term of the above first order condition is the efficiency cost of the tariff (i.e. the marginal deadweight loss from the tariff) while the second term is the terms of trade effect, that is, the reduction in the price of good i that accrues to country i (p_i^I) multiplied by the quantity of country k's imports from country i. The optimal ad-valorem tariff is computed where (12) equals zero:

$$\frac{\partial w_k}{\partial t_{ki}} = 0 \Rightarrow \frac{t_{ki}}{p_i^I} = \frac{\frac{\partial p_i^I}{\partial t_{ki}} \frac{m_k^I}{p_i^I}}{\frac{\partial m_k^I}{\partial p_k^I} \frac{\partial p_k^I}{\partial t_{ki}}} \tag{13}$$

The above expression can be interpreted in two different ways. Note first that since $m_k^I = x_k^I$ we must have

$$\frac{\partial m_k^I}{\partial p_k^I} \frac{\partial p_k^I}{\partial t_{ki}} = \frac{\partial x_k^I}{\partial t_{ki}}$$

Substituting this into (13) shows that country k's optimal ad-valorem tariff equals the inverse of the elasticity of the export supply curve of country i to country k, denoted by ε_{ik}:

$$\frac{t_{ki}}{p_i^I} = \frac{1}{\varepsilon_{ik}} = \left[\frac{\partial x_k^I}{\partial p_i^I} \frac{p_i^I}{x_i^I} \right]^{-1} \tag{14}$$

For an alternative interpretation of the optimal tariff, we can rearrange (13) and write the optimal tariff formula for country k as follows:

$$\frac{t_{ki}}{p_i^I} = \frac{1}{\mu_k^I} \frac{\partial p_i^I}{\partial t_{ki}} \left[\frac{\partial p_k^I}{\partial t_{ki}} \right]^{-1} \text{ where } \mu_k^I = \frac{\partial m_k^I}{\partial p_k^I} \frac{p_k^I}{m_k^I} \tag{15}$$

As can be seen from above, the optimal tariff is also equal to the inverse of the elasticity

of import demand of good I in k (μ_k^I), times the ratio of the change in the relative world price and domestic price of imports. Given that import demand elasticity $\mu_k^I < 0$, the fact that country k's tariff on good I drives down the local price of the good in country i (i.e. $\frac{\partial p_i^I}{\partial t_{ki}} < 0$) while raising it locally ($\frac{\partial p_k^I}{\partial t_{ki}} > 0$), the optimal tariff imposed by country k is positive. Note that the term $\frac{\partial p_i^I}{\partial t_{ki}}$ captures the *terms of trade gain* of the tariff since it informs us how country k's tariff on good I affects the price collected by country i while $\frac{\partial p_k^I}{\partial t_{ki}}$ refers to the *pass through* of the tariff since it tells us how the domestic price of good I in country k varies with its tariff on country i.[8]

Using the demand and supply functions in equations (2) and (4) as well as equations (6) through (10), the equilibrium prices of good I in country i and in importing country j equal:

$$p_i^I = \frac{3\alpha - 2\sum_{z \neq i} t_{zi}}{\lambda + 6} \text{ and } p_j^I = \frac{3\alpha - 2t_{ki} + (4+\lambda)t_{ji}}{\lambda + 6} \tag{16}$$

As is clear from equation (16), the price of good I in country i decreases in the degree of comparative advantage λ (supply effect) and the tariffs it faces in export markets (terms of trade effect). Similarly, the prices of good I in country j increases with its own tariff whereas it decreases with the tariff of the rival importer (i.e. country k). Using the above price equations, we can explicitly calculate the terms of trade gain and the pass through of import tariffs the tariffs.[9] We have:

$$\frac{\partial p_i^I}{\partial t_{ji}} = \frac{\partial p_i^I}{\partial t_{ki}} < 0 \tag{17}$$

i.e. the tariffs imposed by countries j and k lower the price collected by country i's exporters and

$$\frac{\partial p_j^I}{\partial t_{ji}} = \frac{\partial p_k^I}{\partial t_{ki}} < 1 \tag{18}$$

i.e. the pass through from tariffs to local prices in importing countries is incomplete – i.e. the local price in a country does not increase one-to-one with its import tariff.

The first order condition in (12) can be written as:

$$\frac{\partial w_k}{\partial t_{ki}} = \frac{2[\alpha\lambda - t_{ki}(\lambda+4)(\lambda+8) + 4t_{ji}]}{(\lambda+6)^2} \tag{19}$$

[8] Broda et al. (2008) provide evidence that the importers with market power indeed use it in setting their non-cooperative trade policy.

[9] Using the prices in (16), the volume of trade as a function of tariffs can be easily calculated.

It is immediate from the above first order condition that we have positively sloped reaction functions, i.e., tariffs imposed by different countries on the same good (i.e. their common import) are *strategic complements* in our model:

$$\frac{\partial t_{ki}}{\partial t_{ji}} > 0 \tag{20}$$

The intuition for why tariffs of different countries end up being strategic complements in our model is easy to see: an increase in the tariff country j imposes on country i increases the volume of country i's exports to country k thereby increasing the latter's ability to manipulate its terms of trade. Simultaneous solution of first order conditions for countries j and k leads to the following optimal Nash tariffs (which are equal due to symmetry):

$$t^*_{ji} = t^*_{ki} = t^* = \frac{\alpha\lambda}{\lambda^2 + 12\lambda + 28} \tag{21}$$

3 Endogenous preferential trade agreements

We now describe our three stage game of trade liberalization. In the first stage, each country simultaneously announces the names of countries with whom it wants to sign a PTA. This stage determines the underlying trade policy regime. Next, given the trade policy regime, countries choose their optimal tariffs. Finally, consumption, production and international trade take place. As noted before, a key difference between the two types of PTAs is the relative flexibility of FTAs: while an FTA member is free to sign another FTA with an existing non-member without needing consent of an existing FTA partner, a CU member cannot do so due to the requirement of common external tariffs. To capture the implications of this important difference between an FTA and a CU, we compare the equilibrium outcome of the two games, beginning with the FTA game.

3.1 Free trade agreements

In the first stage of the FTA game, each country simultaneously announces the names of countries with whom it wants to sign a free trade agreement (FTA). Here it is important to note that we employ a game of announcements or proposals for a PTA link. In our game, a country does not announce in favor of a specific trade agreement but rather names partners with whom it wants to form a PTA (FTA or CU).**(Footnote here maybe stating that announcing in favor of a specific agreement does not make too much**

sense). Country i's announcement is denoted by σ_i and its strategy set Ω_i consists of four possible announcements: $\Omega_i = \{\{\phi,\phi\},\{j,\phi\},\{\phi,k\},\{j,k\}\}$, where $\{\phi,\phi\}$ denotes an announcement in favor of no FTA with either trading partners, $\{j,\phi\}$ in favor of an FTA with only country j; $\{\phi,k\}$ in favor of an FTA with only country k; and $\{j,k\}$ in favor of FTAs with both of them.

The FTA game can yield the following outcomes: (i) if no two announcements match or the only matching announcements are $\{\phi,\phi\}$, then no agreement $\langle\Phi\rangle$ would prevail; (ii) an FTA between countries i and j denoted by $\langle ij\rangle$ would form if and only if two countries announce each others' name and there is no other matching announcement: i.e., $i \in \sigma_j$ and $i \in \sigma_j$ while $i \notin \sigma_k$ and/or $k \notin \sigma_i$ and $j \notin \sigma_k$ and/or $k \notin \sigma_j$; (iii) two independent FTAs (hub and spoke trading regime) in which i is the common member denoted by $\langle ij, ik\rangle$ (or simply $\langle ih\rangle$) are formed if and only if country i announces in favor of signing an FTA with countries j and k while countries j or/and k announce only in favor of signing an FTA with country i: i.e. $j \in \sigma_i$ and $i \in \sigma_j$ and $k \in \sigma_i$ and $i \in \sigma_k$ while $k \notin \sigma_j$ and/or $j \notin \sigma_k$ and (iv) free trade, denoted by $\langle F\rangle$, would obtain if and only if all countries announce each others' names.[10]

In what follows, we allow countries to undertake coalitional deviations and isolate Nash equilibria that are coalition proof. Following Dutta and Mutuswami's (1997) terminology, we refer to coalition proof Nash equilibrium as *stable* equilibrium. Following Bernheim et. al. (1987): "... *an agreement is coalition-proof if and only if it is Pareto efficient within the class of self-enforcing agreements. In turn, an agreement is self-enforcing if and only if no proper subset (coalition) of players, taking the actions of its complement as fixed, can agree to deviate in a way that makes all of its members better off.*" Therefore, a Nash equilibrium is stable if and only if it is immune to all self-enforcing coalitional deviations. The application of this solution concept is eminently desirable in the present context since countries considering bilateral trade agreements certainly have the capacity to communicate with one another without necessarily having the ability to make binding commitments regarding their plans.

If two countries form an FTA $\langle ij\rangle$, they remove their tariffs on each other ($t_{ij} = t_{ji} =$

[10] Note that the FTA $\langle ij\rangle$ obtains so long as country i and j call only each other, regardless of the nature of country k's announcement. Thus, if $\sigma_i = \{j,\phi\}$ and $\sigma_j = \{i,\phi\}$, then country k would be indifferent between $\sigma_k = \{\phi,\phi\}$, $\{i,\phi\}$, $\{\phi,j\}$ and $\{i,j\}$ because its announcement has no bearing upon the outcome. Under such a situation, we assume that country k makes the most parsimonious announcement among the three, $\sigma_k = \{\phi,\phi\}$. The intuitive justification for this assumption is that an FTA proposal is likely to be costly in the real world and a country that receives no proposals from others would be better off not making any proposals of its own.

0) and impose their individually optimal external tariffs on the non-member country by solving:

$$\left.\frac{\partial w_i}{\partial t_{ik}}\right|_{t_{ij}=0} = 0 \text{ and } \left.\frac{\partial w_j}{\partial t_{jk}}\right|_{t_{ji}=0} = 0 \tag{22}$$

Since member countries i and j are competing importers of good K, the elimination of their internal tariffs due to the FTA has *no impact* on their tariffs on imports of good K and thus the optimal external tariff of FTA members, denoted by $t(ij)$, is the same as that under no agreement, denoted by t^*. We have $t(ij) = t^*$.[11]

On the other hand, since countries compete over imports, the formation of an FTA between two countries changes the first order conditions of the tariff choice problem of the non-member country. We have:

$$\left.\frac{\partial w_k}{\partial t_{ki}}\right|_{t_{ji}=0} = \frac{2[\alpha\lambda - t_{ki}(\lambda+4)(\lambda+8)]}{(\lambda+6)^2} \tag{23}$$

Since we know from (20) that tariffs on the same good imposed by different importers are strategic complements, the reduction of t_{ij} and t_{ji} to zero induces country k to lower its external tariff on the imports from country i and country j. We have:

$$t_k(ij) = \frac{\alpha\lambda}{(\lambda+4)(\lambda+8)} < t^*$$

We can now state:

Proposition 1: *The external trade diversion caused by a bilateral FTA induces the non-member to lower its tariffs on members (i.e. $t_k(ij) < t^*$) while it has no effect on the external tariffs of members (i.e. $t(ij) = t^*$).*

The general intuition behind the above proposition is as follows. Since the removal of internal tariffs under the FTA $\langle ij \rangle$ leads to an increase in the imports of country j from country i, the export supply curve of country i to country k shifts to the left. As a result, the equilibrium world price of good I rises while the equilibrium exports of country i to country k decline. Since export supply curves are linear in our model, the elasticity of export supply curves of countries i and j facing country k rises due to the formation of the bilateral FTA $\langle ij \rangle$:

$$\varepsilon_{ik}(ij) - \varepsilon_{ik}(\Phi) = \frac{2}{\lambda} > 0 \tag{24}$$

[11] Thus, the model does not exhibit the type of tariff complementarity described in Bagwell and Staiger (1997a, 1997b) and some other models of PTAs.

which in turn implies that country k's optimal tariff on imports from countries i and j under the FTA $\langle ij \rangle$ is lower than its optimal Nash tariff t^*.

For an alternative interpretation of Proposition 1, consider the optimal tariff formula in (15). Note that the FTA $\langle ij \rangle$ leads to an upward movement along the import demand curves of country k for goods i and j leading to a higher elasticity of import demand (common due to symmetry). Thus, the removal of the internal tariffs under the FTA $\langle ij \rangle$ leads to a lower terms of trade gain and a higher tariff pass through in country k:

$$\frac{\partial p_i^I}{\partial t_{ki}} \mid_{t_{ji}=0} > \frac{\partial p_i^I}{\partial t_{ki}} \mid_{t_{ji}=t^*} \text{ and } \frac{\partial p_k^I}{\partial t_{ki}} \mid_{t_{ji}=0} > \frac{\partial p_k^I}{\partial t_{ki}} \mid_{t_{ji}=t^*}$$

both of which tend to lower its optimal tariff. Analogous reasoning explains why the two spoke countries end up imposing lower tariffs on each other relative to the status quo: under the hub and spoke arrangement, each spoke exports more to the hub and less to the other spoke relative to the status quo, which in turn lowers the ability of both spokes to manipulate their terms of trade vis-à-vis one another.

Let country i's welfare as a function of trade agreement a be denoted by $w_i(a)$ and let $\Delta w_i(a-b)$ denote the difference between country i's welfare under trade agreements a and b: $\Delta w_i(a-b) \equiv w_i(a) - w_i(b)$. Also, let m denote a member country of the bilateral FTA $\langle ij \rangle$ so that $m = i$ or j.

We establish the following result below:

Lemma 1: *A pair of countries have an incentive to form a bilateral FTA and the formation of such an FTA makes the non-member country worse-off.*

First note that the formation of a bilateral FTA does not affect the non-member's producer surplus in export markets since external tariffs of FTA members are the same as those under the status quo. Furthermore, since the external trade diversion caused by an FTA lowers the non-member's ability to manipulate its terms of trade vis-à-vis member countries (which is reflected in turn in its reduced external tariff), the non-member is worse off relative to the status quo. Since aggregate world welfare increases due to the trade liberalization undertaken by FTA member countries, we can conclude that the formation of an FTA makes member countries better off at the expense of the non-member:

$$\Delta w_m(ij - \Phi) > 0 > \Delta w_k(ij - \Phi) \text{ for } m = i, j \tag{25}$$

From the first inequality above it immediately follows that, starting at no agreement $\langle \Phi \rangle$, two countries (say i and j) always have an incentive to announce each other's name that

would lead to a deviation from no agreement $\langle \Phi \rangle$ to a bilateral FTA $\langle ij \rangle$.

Next, we examine the welfare of a member country m and the non-member country under the FTA $\langle ij \rangle$ relative to the hub and spoke trading regime $\langle mh \rangle$ where the hub country has an independent FTA with each of the two spoke countries who do not have an FTA with one other. As one might expect, taking the announcement of their complement as fixed, each member country (i or j) of an FTA $\langle ij \rangle$ and the non-member country (k) have incentives to jointly deviate to announcements where they call in favor of an FTA with each other. This deviation would lead to a hub and spoke regime where deviating FTA member becomes a hub that forms two independent FTAs:

$$\Delta w_m(mh - ij) > 0 \text{ for } m = i, j \tag{26}$$

Such a move on the part of a member country of an existing FTA makes the other member worse off:

$$\Delta w_{\widetilde{m}}(mh - ij) < 0 \text{ for } m, \widetilde{m} = i, j \tag{27}$$

It is important to note that the source of the adverse impact on the pre-existing FTA member that does not become the hub is the external trade diversion caused by the formation of the second FTA which reduces the volume of exports flowing from the non-member to its market which in turn reduces its optimal external tariff. The flip side of this result is that it is better to be a spoke than to be a non-member under the FTA $\langle ij \rangle$:

$$\Delta w_k(mh - ij) > 0 \tag{28}$$

Finally, consider the welfare of individual countries under the hub and spoke regime relative to free trade. We find that while each spoke country is better off under free trade, the hub country is worse off:

$$\Delta w_{\widetilde{m}}(F - mh) > 0 > \Delta w_m(F - mh) \tag{29}$$

A comparison of hub country m's welfare under $\langle mh \rangle$ relative to $\langle F \rangle$ yields the following: (i) the hub country's producer surplus is equal under the two regimes since its producers face zero tariffs under both regimes and are afforded no protection in the home market (ii) its domestic welfare is higher under $\langle mh \rangle$ relative to $\langle F \rangle$ since it benefits from the positive terms of trade effects of tariffs that the spokes impose on each other – it is able to import goods from both spokes at prices that are below those under free trade. As a result, the

hub country is strictly better off under relative to free trade.

To see why the spokes are worse off under relative to free trade, it is sufficient to note that aggregate global welfare is strictly higher under free trade. Given that the hub country is strictly better off relative to free trade and the fact that the welfare of the two spokes is equal (due to symmetry), both spokes must be worse off. We are now ready to state one of our key results (see the appendix for proof):

Proposition 2: *Even though a pair of countries prefer a bilateral FTA to free trade (i.e.* $\Delta w_m(F-ij)<0$*), the only stable equilibrium of the FTA game is free trade.*

The key message of this result is that even though a pair of countries benefit from excluding the third country from their trade agreement, *they are unable to exercise this exclusion incentive in equilibrium.* The forces that give rise to the exclusion incentive can be understood as follows. Relative to free trade, each member country of an FTA has the ability to manipulate its terms of trade vis-à-vis the non-member while also being able to free ride on the terms of trade effect of the non-member's tariff on their FTA partner. These benefits of exclusion are somewhat tempered by the fact that the deviation from free trade to an FTA results in member countries facing positive tariffs in the non-member's market, although these tariffs are lower than those under the status quo owing to the external trade diversion caused by the FTA. However, the two positive effects of exclusion on FTA members dominate the negative effect so that a pair of countries benefit if they can successfully exclude the third country from their trade agreement.[12] It is worth emphasizing the role that the flexible nature of FTAs plays in ensuring that the exclusion incentive goes unexercised in the FTA game. Since the formation of an FTA induces the non-member to lower its tariffs on members while leaving external tariffs of countries that become FTA members unchanged (Proposition 1), two countries have an incentive to exclude the third country. However, since the most preferred arrangement of each country – i.e. being a hub under a hub and spoke arrangement – is permissible under FTAs and the non-member prefers to be a spoke under a hub and spoke regime, no two countries are able to exclude the third by forming a bilateral FTA. The lure of creating a hub and spoke arrangement ends up undermining bilateral FTAs thereby delivering free trade as the only stable outcome.

What if countries pursue CUs instead of FTAs? Next, we examine this possibility.

[12] Of course, since world welfare is higher under free trade relative to an FTA, the gains that members enjoy come at the expense of the non-member:

$$\Delta w_k(ij-F)<0 \tag{30}$$

3.2 Customs Unions

Suppose the PTA under consideration is a CU as opposed to an FTA. As under the FTA game, at the first stage of the CU formation game each country announces the names of countries with whom it wants to form a CU. Country i's announcement is denoted by σ_i and its strategy set Ω_i consists of four possible announcements:

$$\Omega_i = \{\{\phi, \phi\}, \{j, \phi\}, \{\phi, k\}, \{j, k\}\} \tag{31}$$

where $\{\phi, \phi\}$ denotes an announcement in favor of no CU with either trading partners, $\{j, \phi\}$ in favor of a CU with only country j; $\{\phi, k\}$ in favor of a CU with only country k; and $\{j, k\}$ in favor of a CU that includes all three countries which is tantamount to announcing in favor of global free trade.

The following policy regimes can arise in the customs union game: (i) no agreement $\langle\Phi\rangle$ would prevail when no two announcements match or the only matching announcements are $\{\phi, \phi\}$; (ii) a CU between countries i and j denoted by $\langle ij^u\rangle$ would be formed if and only if two countries announce each others' name and there is no other matching announcement: i.e., $i \in \sigma_j$ and $i \in \sigma_j$ while $i \notin \sigma_k$ and/or $k \notin \sigma_i$ and $j \notin \sigma_k$ and/or $k \notin \sigma_j$; (iv) free trade, would obtain if and only if all countries announce the name of the other two countries. Recall that the equivalent of a hub and spoke trading regime cannot arise under the CU game due to the fact that CU members coordinate their external tariffs.

Next, we consider the formation of a CU between countries i and j, denoted by $\langle ij^u\rangle$. Like FTA members, CU members remove tariffs on each other. However, unlike FTA members, CU members impose a jointly optimal external tariff on the non-member. Under the CU $\langle ij^u\rangle$ members solve:

$$\max_t \; w_i(ij^u) + w_j(ij^u) \text{ subject to } t_{ij} = t_{ji} = 0 \tag{32}$$

As the CU forms, the common market becomes larger relative to their individual markets and members' international market power increases. Since the terms-of-trade externalities across members are internalized by a CU and tariffs on the same good across countries are complementary in our model, the optimum external tariffs of members rise following the formation of CU.[13]

To gain further insight, let m^K be the total import demand of good K in countries i

[13] Olarreaga et al. (1999) provide evidence that the terms-of-trade externalities among Mercosur's members were internalized in its external tariffs.

and j:

$$m^K = \sum_{z=i,j} \left[d(p_z^K) - s_z^K(q_z^K) \right] \tag{33}$$

while x^K denote the exports of good K (from country k) to the common market of countries i and j:[14]

$$x^K = s_k^K(q_k^K) - d(p_k^K) \tag{34}$$

The equilibrium world price of good K is determined by the market clearing condition:

$$x^K = m^K \tag{35}$$

Using the market clearing prices and quantities, the first order condition for the welfare maximization problem in (32) can be written as:

$$\frac{\partial[w_i(ij^u) + w_j(ij^u)]}{\partial t} = \frac{4[2\alpha\lambda - t(\lambda+2)(\lambda+10)]}{(\lambda+6)^2} = 0 \tag{36}$$

Solving this yields the optimal tariff of the CU $\langle ij^u \rangle$:[15]

$$t(ij^u) = \frac{2\alpha\lambda}{\lambda^2 + 12\lambda + 20} \tag{38}$$

Since both countries import the same good from the non-member country, the market power effect of a CU emphasized by Bagwell and Staiger (1997a) arises here. A CU allows members to pool their market power and extract a larger terms of trade gain from the non-member leading to an increase in their tariffs:

$$t(ij^u) > t(ij) = t^* \tag{39}$$

It is also immediate that since member countries reduce their internal tariffs to zero, like an FTA, a bilateral CU also induces the non-member to lower its tariffs on members (and

[14] Note that with a CU we are back to two country set-up of Feenstra (1994) and Broda et al. (2008).

[15] It is straightforward to confirm that the optimal ad-valorem tariff of the CU equals the inverse of the elasticity of country k's export supply curve:

$$\frac{t}{p_k^K} = \frac{1}{\varepsilon_k(ij^u)} \tag{37}$$

exactly to the same level):[16]

$$t_k(ij^u) = t_k(ij) < t^* \tag{40}$$

It is straightforward to establish that the external trade diversion caused by a CU has qualitative similar effects as an FTA:

Proposition 3: *The formation of a CU induces members to raise their external tariffs (i.e.* $t(ij^u) > t^*$*) and the non-member to lower them:* $t_k(ij^u) < t^*$.

To see the general picture behind the above result more clearly, suppose that we start with the status quo where all countries impose the optimal tariff t^* on each other. It is obvious that the export supply of country k to the CU's common market lies to the right of its individual export supply curves to the markets of countries i and j since it applies to a common larger market. Similarly, following the formation of the CU $\langle ij^u \rangle$, the common import demand of member countries is larger relative to individual import demands. Therefore, market clearing under the CU $\langle ij^u \rangle$ occurs at a lower $\frac{p_k^K}{x_k^K}$ ratio relative to the status quo. Since export supply curves are linear, the elasticity of country k's export supply curve falls due to the formation of the CU $\langle ij^u \rangle$. As a result, the optimal common external tariff under CU $\langle ij^u \rangle$ is higher than the individually optimal tariffs of countries i and j.

It is worth noting that Article XXIV of GATT requires PTA members to not raise their external tariffs on non-member countries. For now, we ignore this tariff restriction imposed by Article XXIV and assume that a CU can impose its optimal tariffs. At the end of this section, we examine the implications of this tariff restriction on welfare and on the prospects of global free trade.

Since CU members *pool their market power*, they impose higher external tariffs than FTA members and extract a larger terms of trade gain from the non-member. As a result, the welfare of a CU member is higher than that of an FTA member:

$$\Delta w_m(ij^u - ij) > 0 \text{ where } m = i, j. \tag{41}$$

Then, combining (25) and (41), we can argue that, relative to no agreement, member countries are better off under a CU while the non-member country is worse-off. This implies that a CU is a Nash equilibrium since neither member country (i or j) has an incentive to unilaterally change their announcements that would only lead to no agreement

[16] Thus, Proposition 1 holds regardless of whether the PTA is an FTA or a CU.

$\langle\Phi\rangle$:

$$\Delta w_m(ij^u - \Phi) > 0 \tag{42}$$

Using a similar logic, since the non-member faces higher tariffs in the export markets under a CU relative to FTA while its external tariffs are the same, the non-member country is worse off under a CU relative to an FTA:

$$\Delta w_k(ij^u - ij) < 0 \tag{43}$$

Combining (48) and (43), it follows that the non-member country is worse-off relative to free trade:

$$\Delta w_k(F - ij^u) > 0 \tag{44}$$

Therefore, free trade is also a Nash equilibrium since a country (say k) has no incentive to unilaterally deviate from its announcement $\{i, j\}$ to $\{\phi, \phi\}$ that would result a deviation from $\langle F\rangle$ to $\langle ij^u\rangle$. Finally, as in the FTA game, no agreement $\langle\Phi\rangle$ is a Nash equilibrium since no country has an incentive to announce another's name if the latter does not announce its name in return.

Following our discussion of the coalition-proof Nash equilibrium in the FTA game, it is easy to see that the only two candidates for stable equilibrium of the CU game are a bilateral CU $\langle ij^u\rangle$ and free trade $\langle F\rangle$. Which, if any, of these Nash equilibria are coalition-proof? Consider free trade first. As in the FTA game, since world welfare is the highest under $\langle F\rangle$, each country prefers $\langle F\rangle$ to $\langle\Phi\rangle$ and thus we can immediately rule out any coalitional announcement deviations that would lead to a deviation from $\langle F\rangle$ to $\langle\Phi\rangle$. The only issue is, taking the announcement of their complement (country k) fixed, whether two countries (say i and j) have an incentive to coalitionally deviate from their announcements $\{j, k\}$ and $\{i, k\}$ to $\{j, \phi\}$ and $\{i, \phi\}$ respectively, leading to a deviation from free trade $\langle F\rangle$ to a bilateral CU $\langle ij^u\rangle$. In fact, we can show the following:

Lemma 2: *The exclusion incentive is stronger for a CU relative to an FTA:*

$$\Delta w_m(ij^u - F) > \Delta w_m(ij - F) > 0 \tag{45}$$

The key difference relative to the FTA game is that the joint deviation of countries i and j from from their announcements $\{j, k\}$ and $\{i, k\}$ to the announcements $\{j, \phi\}$ and $\{i, \phi\}$ (converting free trade $\langle F\rangle$ to a bilateral CU $\langle ij^u\rangle$) is now self-enforcing since neither of the initially deviating countries (i or j) has an incentive to further deviate unilaterally, taking

country k's announcement fixed: $\sigma_k = \{i, j\}$. Thus, since the initial joint announcement deviation of two countries is self-enforcing, free trade $\langle F \rangle$ fails to be a coalition-proof Nash equilibrium under the CU game.

Finally, we know from (42) and (45) that countries i and j countries have no incentives (unilaterally or coalitionally) to deviate from their respective announcements $\{j, \phi\}$ and $\{i, \phi\}$ since such a move is not beneficial. Thus, we have the following key result:

Proposition 4: *The only stable agreement of the CU game is the bilateral CU* $\langle ij^u \rangle$.[17]

The difference in results reported in Propositions 2 and 4 is driven by the relatively flexible nature of FTAs compared to CUs. When two countries (i and j) jointly exclude the third country from free trade under the FTA game, each member has an incentive to further deviate from the bilateral FTA and sign an independent FTA with the third country thereby making itself a hub. The existence of this further deviation acts as a deterrent for the other initially deviating country (say country j) since it is worse off as a spoke under $\langle ih \rangle$ relative to free trade and thus the initial joint deviation from free trade to a bilateral FTA does not occur. However, unlike the FTA game, no such deterrent exists under the CU game since a CU member cannot form an independent agreement with the external country without the consent of its CU partner due to common determination of external tariffs. Therefore, our model suggests that the pursuit of PTAs is compatible with the goal of achieving global free trade only when PTAs take the form of FTAs.

We now turn to isolating the implications of the pooling of market power under a CU from their relative flexibility compared to FTAs. We do this by holding the external tariff of the CU constant at the pre-existing level. This is well motivated experiment since Article XXIV of GATT – the key WTO clause that sanctions PTAs – forbids member countries of a PTA from raising tariffs on non-members. In our model, in the absence of such a restriction, CU members indeed raise their tariffs on outsiders due to the pooling of their market power whereas FTA members do not. What is the equilibrium outcome of the CU game when CU members are unable to raise their tariffs above the levels that prevail under the status quo?

Let $\langle \overline{ij^u} \rangle$ denote the CU between countries i and j when the tariff restriction of Article XXIV binds. If CU members are prohibited from raising tariffs on non-members, thus the

[17] This is in sharp contrast to the results obtained in the competing exporters model of Saggi et al. (2013). When countries are competing exporters, while the pooling of market power exists under the CU, external trade diversion and the tariff complementarity effect in the non-member country's market that results from it do not arise and this weakens the exclusion incentive in that framework. As a result, in Saggi et al. (2013) a CU simply does not arise in equilibrium when countries are symmetric in size.

external tariffs under a CU and FTA are equal:

$$t_i(ij) = t_i(\overline{ij^u}) = t^* \tag{46}$$

Given that they impose the same external tariffs and face the same tariffs in export markets, member countries obtain the same welfare levels under a CU and an FTA i.e. $w_m(\overline{ij^u}) = w_m(ij)$ for $m = i, j$. It turns out that countries i and j still have incentives to jointly deviate from from from their respective announcements $\{j, k\}$ and $\{i, k\}$ to the announcements $\{j, \phi\}$ and $\{i, \phi\}$ (converting free trade $\langle F \rangle$ to a bilateral CU $\left\langle \overline{ij^u} \right\rangle$) even when the external tariffs are restricted by Article XXIV:

$$\Delta w_m(F - \overline{ij^u}) < 0 \text{ for } m = i, j \tag{47}$$

As before, this deviation is self-enforcing and thus free trade fails to be coalition-proof under the restricted CU game. Thus, even when CU members are prevented from raising their tariffs on outsiders, a bilateral CU is still the unique coalition-proof Nash equilibrium of the CU game. Thus, even when CU members are prevented from raising their tariffs on outsiders, a bilateral CU is still the unique stable equilibrium of the CU game. The only consequence of the tariff restriction is that it replaces the unconstrained CU $\langle ij^u \rangle$ by the constrained CU $\left\langle \overline{ij^u} \right\rangle$. Thus, the restriction on external tariffs of CU members fails to further the cause of global free trade, even as it makes the resulting CU more attractive from a global welfare perspective by softening the adverse impact of the CU on the non-member country.

In a recent paper, Mrázová, et al. (2013) compare CU formation with and without the constraint on external tariffs imposed by Article XXIV. In their multi-country oligopoly model, holding constant the structure of the CU, the constraint imposed by Article XXIV improves world welfare. But this positive effect is counterbalanced by a negative compositional effect since Article XXIV encourages the formation of more symmetric CUs and thereby causes a greater volume of trade to be subject to tariffs. In our model, the compositional effect is absent since countries are symmetric in terms of economic fundamentals. However, our primary focus is on how the endogenous choice between preferential and multilateral liberalization is affected by the nature of the trade agreement (FTA versus CU) under consideration and the effects of the Article XXIV tariff constraint are less central to our analysis.

4 Conclusion

At a time when multilateral trade liberalization seems to have come to a grinding halt, the question whether preferential trade agreements have a useful role to play in the multilateral trading system seems to have acquired an even greater degree of urgency. After all, other than unilateral trade liberalization, over the last decade or so PTAs appear to be the only game in town for countries interested in undertaking reciprocal trade liberalization. Of course, the concern that the pursuit of PTAs might undermine whatever appetite remains in the global trading system for multilateral trade liberalization remains a real one.

In this paper, we attempt to isolate the implications of the two commonly occurring PTAs for the prospects of multilateralism. The beginning point of our analysis is a well understood difference between FTAs and CUs: while CU members impose common external tariffs, each FTA member imposes external tariffs of its own choosing. This difference in tariff setting behavior between the two PTAs implies that while FTA members have less market power than CU members, they enjoy a greater degree of flexibility in the sense that they are free to enter into further agreements with outsiders whereas CU members can only do so with consent of existing members.

An important result of this paper is that the formation of either type of PTA induces the non-member to lower its tariffs on PTA members: the external trade diversion – i.e. the reduction in the volume of exports flowing from PTA members to the non-members – reduces the ability of the non-member to manipulate its terms of trade thereby making it optimal for it to lower its tariffs. Thus, PTA formation not only benefits members because of the internal trade liberalization that they undertake but also because of the external trade liberalization induced abroad.

The central result of the paper is that the more flexible nature of FTAs (which in turn emanates from independent external tariff setting on the part of members) helps in the attainment of global free trade. Specifically, the ability of an existing FTA member to form an independent FTA with the non-member makes global free trade the only stable equilibrium of our FTA game. By contrast, the only stable equilibrium of the CU game is a bilateral CU. This result captures the intuition that, once formed, CUs constrain their members in a way that FTAs do not. Finally, we also find that the tariff restriction imposed by Article XXIV of the GATT fails to further the cause of global free trade although it makes the resulting CU more attractive from a global welfare. Thus, our model shows that the two key differences between FTAs and CUs – i.e. the pooling of market power under a CU and the inability of a CU member to enter into an independent trade agreements with

outsiders – both impose welfare costs on the world trading system but it is only the latter that prevents the obtainment of free trade.

5 Appendix

5.1 Supporting calculations

Here we report the key formulae that are necessary for proving our results. For an arbitrary tariff vector $\mathbf{t} = (t_{ij}, t_{ik}, t_{ji}, t_{jk}, t_{ki}, t_{kj})$, we can write country i's welfare as

$$w_i = CS_i + PS_i + TR_i,$$

where consumer surplus in country i equals

$$CS_i = \frac{1}{2}\left[(\alpha - p_i^I)^2 + (\alpha - p_i^J - t_{ij})^2 + (\alpha - p_i^K - t_{ik})^2\right],$$

whereas its producer surplus equals

$$PS_i = \frac{(\lambda+1)}{2}(p_i^I)^2 + \frac{1}{2}\left[(p_i^J + t_{ij})^2 + (p_i^K + t_{ik})^2\right]$$

and the tariff revenue is given by

$$TR_i = \frac{t_{ij}[\alpha\lambda + 4t_{kj} - 2t_{ij}(\lambda+4)]}{\lambda+6} + \frac{t_{ik}[\alpha\lambda + 4t_{jk} - 2t_{ik}(\lambda+4)]}{\lambda+6}$$

and prices are given by

$$p_i^I = \frac{3\alpha - 2\sum_{z\neq i} t_{zi}}{\lambda+6};\ p_i^J = \frac{3\alpha - 2\sum_{z\neq j} t_{zj}}{\lambda+6};\ \text{and}\ p_i^K = \frac{3\alpha - 2\sum_{z\neq k} t_{zk}}{\lambda+6}$$

Using the above formulae and the optimal tariff levels reported in the text, we can easily calculate welfare levels under all possible trade agreements. To save space, we do not include the algebraic details underlying these straightforward calculations.

5.2 Proof of Proposition 2

In our model, it is straightforward that no agreement $\langle\Phi\rangle$ is a Nash equilibrium since no country has an incentive to announce another's name if the latter does not announce its

name in return. Which of the other three policy regimes – i.e. a bilateral FTA $\langle ij \rangle$, global free trade $\langle F \rangle$, and a hub and spoke agreement $\langle ih \rangle$ – can emerge as Nash equilibria? Note from inequality (25) that neither member country (i or j) has an incentive to unilaterally change their announcement that would only lead to no agreement $\langle \Phi \rangle$. Thus, a bilateral FTA is a Nash equilibrium. Now consider the hub and spoke regime $\langle ih \rangle$. From inequalities (25), (26) and (28), the following is immediate: (i) hub country i has no incentive to unilaterally change its announcement from $\{j, k\}$ to $\{j, \phi\}$ or $\{\phi, k\}$ or $\{\phi, \phi\}$ since it would lead to a deviation from the hub and spoke regime $\langle ih \rangle$ where i is the hub country to $\langle ij \rangle$ or $\langle ik \rangle$ or $\langle \Phi \rangle$, respectively; (ii) neither spoke country (say j) has an incentive to unilaterally deviate from its announcement $\{i, \phi\}$ to $\{\phi, \phi\}$ or $\{i, k\}$ since the former deviation would lead to deviation from $\langle ih \rangle$ where j is a spoke country to $\langle ik \rangle$ where j is a non-member country while the latter deviation would not lead to any change in the agreement. As a result, a hub and spoke regime $\langle ih \rangle$ is also a Nash equilibrium. Finally, consider global free trade $\langle F \rangle$. Since world welfare is the highest under $\langle F \rangle$, each country prefers $\langle F \rangle$ to $\langle \Phi \rangle$ and this result together with (25) implies that a country (say k) has no incentive to unilaterally deviate from its announcement $\{i, j\}$ to $\{\phi, \phi\}$ that would result a deviation from $\langle F \rangle$ to $\langle ij \rangle$:

$$\Delta w_k(F - ij) > 0 \tag{48}$$

Moreover, we know from (29) that a country (say k) has no incentive to unilaterally change its announcement from $\{i, j\}$ to $\{i, \phi\}$ or $\{\phi, j\}$ since this would mean a deviation from $\langle F \rangle$ to $\langle ih \rangle$ or $\langle jh \rangle$ under both of which country k is a spoke country. Thus, free trade is also a Nash equilibrium.

To deal with the multiplicity of equilibria described above and to capture the process of FTA formation in a more realistic fashion, we now isolate Nash equilibria that are coalition proof (i.e. Nash equilibria that are immune to self-enforcing coalitional deviations). We begin by considering whether global free trade $\langle F \rangle$ is a coalition proof Nash equilibrium. As indicated above, since world welfare is the highest under $\langle F \rangle$, each country prefers $\langle F \rangle$ to $\langle \Phi \rangle$ and thus we can immediately rule out any coalitional announcement deviations that would lead to a deviation from $\langle F \rangle$ to $\langle \Phi \rangle$. Similarly, we know from (29) that no two countries (say j and k) have incentives to jointly alter their announcements from $\{i, k\}$ to $\{i, \phi\}$ and $\{i, j\}$ to $\{i, \phi\}$, respectively since it would lead to a deviation from $\langle F \rangle$ to $\langle ih \rangle$ where both are spokes (and spokes are worse of relative to free trade). Finally, taking the announcement of their complement (country k) fixed, consider the coalitional announcement deviation of two countries (say i and j) from their announcements $\{j, k\}$ and $\{i, k\}$ to $\{j, \phi\}$ and $\{i, \phi\}$ respectively. This would lead to a coalitional deviation from free trade $\langle F \rangle$ to a bilateral FTA $\langle ij \rangle$. Using the welfare levels reported above, it is straightforward to show that two countries indeed have a joint incentive to exclude the

third country from free trade:

$$\Delta w_m(F - ij) < 0 \tag{49}$$

This implies that, taking the announcement of their complement (country k) fixed: $\sigma_k = \{i, j\}$, the above coalitional deviation in announcements would occur. The question then becomes whether it is a self-enforcing deviation in nature. It would be self enforcing, if no proper subset of initially deviating countries (neither i nor j) has an incentive to further alter their announcements from $\{j, \phi\}$ and $\{i, \phi\}$ respectively, taking country k's announcement fixed: $\sigma_k = \{i, j\}$. However, given that $\sigma_k = \{i, j\}$, we know from (26) that country i has an incentive to further deviate from its announcement $\{j, \phi\}$ to $\{j, k\}$ that would lead to a deviation from $\langle ij \rangle$ to $\langle ih \rangle$ where it becomes the hub. Similarly, country j has an incentive to further deviate from its announcement $\{i, \phi\}$ to $\{i, k\}$ so as to itself become the hub. Thus, the initial coalitional announcement deviation that leads to a deviation from free trade $\langle F \rangle$ to a bilateral FTA $\langle ij \rangle$ is not self-enforcing. As a result, we have shown that free trade is a coalition-proof Nash equilibrium.

Next, we consider whether the other agreements are coalition-proof Nash equilibrium. First, consider no agreement $\langle \Phi \rangle$. Note from (25) that any two countries (say i and j) have an incentive to coalitionally change their announcements from $\{\phi, \phi\}$ and $\{\phi, \phi\}$ to $\{j, \phi\}$ and $\{i, \phi\}$ respectively, taking country k's announcement fixed: $\sigma_k = \{\phi, \phi\}$. This initial deviation is self-enforcing since no proper subset of the initially deviating countries (neither i nor j) has an incentive to further alter their announcements unilaterally (i.e. $\langle ij \rangle$ is a Nash equilibrium). Therefore, no agreement $\langle \Phi \rangle$ is not a coalition-proof Nash equilibrium.

We next examine whether a bilateral FTA $\langle ij \rangle$ is a coalition-proof Nash equilibrium. It is immediate from the inequalities in (26) and (28) that countries i and k have incentives to coalitionally change their announcements from $\{j, \phi\}$ and $\{\phi, \phi\}$ to $\{j, k\}$ and $\{i, \phi\}$ respectively, taking country j's announcement fixed: $\sigma_j = \{i, \phi\}$. This announcement deviation would convert FTA $\langle ij \rangle$ to the hub and spoke regime $\langle ih \rangle$ where i is the hub and j and k are spokes. This initial coalitional deviation is self-enforcing since no proper subset of the initially deviating countries (neither i nor k) has an incentive to further alter their announcements unilaterally (i.e. $\langle ih \rangle$ is a Nash equilibrium). Therefore, a bilateral FTA is not a coalition-proof Nash equilibrium.

Finally, we examine whether a hub and spoke regime (say $\langle ih \rangle$) is a coalition-proof Nash equilibrium. Note from the inequality in (29) that countries j and k have incentives to coalitionally change their announcements from $\{i, \phi\}$ and $\{i, \phi\}$ to $\{i, k\}$ and $\{i, j\}$ respectively, taking country i's announcement fixed: $\sigma_i = \{j, k\}$. This announcement deviation would convert the hub and spoke regime $\langle ih \rangle$ to free trade $\langle F \rangle$. This initial coalitional deviation is self-enforcing since no proper subset of the initially deviating countries (neither j nor k) has an incentive to further alter their announcements unilaterally (i.e. $\langle F \rangle$ is a Nash

equilibrium). Therefore, a hub and spoke regime is not a coalition-proof Nash equilibrium.

References

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

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

[3] Bagwell, Kyle and Robert. W. Staiger, 1998. "Regionalism and Multilateral Tariff Cooperation." In John Piggott and Allan Woodland, eds, *International Trade Policy and the Pacific Rim*, London: MacMillan.

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

[5] 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.

[6] Bond, Eric W., Raymond G. Riezman, and Constantinos Syropoulos, 2004. "A Strategic and Welfare Theoretic Analysis of Free Trade Areas." *Journal of International Economics* 64, 1-27.

[7] Broda, Christian, Nuno Limao and David E. Weinstein, 2008. "Optimal Tariffs and Market Power: The Evidence," *American Economic Review*, 98(5), 2032-65.

[8] Coates, Daniel and Rodney Ludema, 2001. "A Theory of Trade Policy Leadership." *Journal of Development Economics* 65, 1–29.

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

[10] Estevadeordal, Antoni, Caroline Freund and Emanuel Ornelas, 2008. "Does Regionalism Affect Trade Liberalization toward Non Members?" *Quarterly Journal of Economics* 123(4), 1531-1575.

[11] Feenstra, Robert, 2004. *Advanced International Trade*. Princeton University Press.

[12] Freund, Caroline and Emanuel Ornelas, 2010. "Regional Trade Agreements," *Annual Review of Economics* 2, 139-167.

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

[14] Horn, Henrik, Giovanni Maggi, and Robert W. Staiger, 2010. "Trade agreements as endogenously incomplete contracts." *American Economic Review* 100(1), 394-419.

[15] Kennan, John and Raymond Riezman, 1990. "Optimal Tariff Equilibria with Customs Unions?" *Canadian Journal of Economics* 23, 70-83.

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

[17] Krishna, Pravin and Devashish Mitra, 2005. "Reciprocated Unilateralism in Trade Policy." *Journal of International Economics* 65, 461–487.

[18] Mrázová, Monika, David Vines, and Ben Zissimos, 2013. "Is the GATT/WTO's Article XXIV Bad?" *Journal of International Economics* 89(1), 216-232.

[19] Olarreaga, Marcelo, Isidro Soloaga, and Alan Winters, 1999. "What's Behind MERCOSUR's Common External Tariff?" Policy Research Working Paper Series 2231, World Bank.

[20] Ornelas, Emanuel, 2005a. "Trade Creating Free Trade Areas and the Undermining of Multilateralism." *European Economic Review* 49, 1717-1735.

[21] Ornelas, Emanuel, 2005b. "Rent Destruction and the Political Viability of Free Trade Agreements." *Quarterly Journal of Economics* 120, 1475-1506.

[22] Saggi, Kamal and Halis M. Yildiz, 2010. "Bilateralism, Multilateralism, and the Quest for Global Free Trade." *Journal of International Economics* 81, 26-37.

[23] Saggi, Kamal, Alan Woodland and Halis M. Yildiz, 2013. "On the Relationship between Preferential and Multilateral Trade Liberalization: The Case of Customs Unions." *American Economic Journal: Microeconomics*, 5(1), 63-99.