

Preferential Trade Agreements and Rules of the Multilateral Trading System

Kamal Saggi*, Woan Foong Wong[†], Halis Murat Yildiz[‡]

Abstract

In a three-country model of endogenous trade agreements, we study the effects of major WTO rules governing the conduct of free trade agreements (FTAs). We show that FTA members retain positive internal tariffs even if they seek to maximize their joint welfare. Requiring FTAs to eliminate internal tariffs – as stipulated by current WTO rules – makes the non-member better off although it simultaneously reduces the likelihood of achieving global free trade by encouraging free-riding on its part. While the WTO's non-discrimination constraint is not necessarily conducive to reaching global free trade, it raises welfare in a tariff-ridden world.

Keywords: Free Trade Agreements, Tariffs, Customs Unions, World Trade Organization, Coalition proof Nash equilibrium, Welfare. *JEL Classifications:* F11, F12.

*Department of Economics, Vanderbilt University, Nashville, TN 37235-1828. E-mail: kamal.saggi@vanderbilt.edu.

[†]Department of Economics, University of Oregon, Eugene, Oregon, 97403. E-mail: wfwong@uoregon.edu.

[‡]Department of Economics, Ryerson University, 350 Victoria Street, Toronto, ON, Canada M5B 2K3. E-mail: hyildiz@ryerson.ca.

1 Introduction

Under current rules of the World Trade Organization (WTO), countries entering into a preferential trade agreement (PTA) are required to eliminate tariffs on “substantially all trade” with each other.¹ This paper develops a model of endogenous trade agreements to investigate the welfare implications of this *free internal trade requirement* facing PTAs at the WTO as well as the affect it has on the likelihood of achieving global free trade.² We focus on free trade agreements (FTAs), the most commonly occurring type of PTA in today’s global economy. Although the WTO system sanctions discrimination in the specific form of PTAs, it also requires all member countries to grant most favored nation (MFN) status to one another which generally forbids discrimination on their part. This clash between PTAs and MFN raises two interesting questions. One, is there a case for allowing PTA non-members to deny MFN treatment to PTA members? In other words, should WTO members be permitted to engage in tariff discrimination when they find themselves facing such discrimination at the hands of PTA members? Two, does the answer to this question depend upon whether non-members have *voluntarily* chosen to stay out of a PTA (knowing full-well that their non-participation will result in them facing discriminatory treatment) or have been *deliberately excluded* by PTA members? Our model allows us to directly address these novel questions that have been overlooked in the vast literature on PTAs.

Our conceptual approach to the formation of trade agreements follows Saggi and Yildiz (2010) who develop an equilibrium theory of FTAs in a modified version of the three-country competing exporters framework of Bagwell and Staiger (1999a).³ Assuming FTA members impose zero tariffs on one another, they compare the relative merits of bilateralism and multilateralism as alternative routes to global trade liberalization. In the present paper, like Saggi and Yildiz (2010), we begin with a *WTO-consistent benchmark* scenario under which FTA members are required to eliminate tariffs on each other and the non-member is obligated to follow the MFN principle of non-discrimination when setting its tariffs on FTA

¹This condition and other related provisions governing PTAs are specified in Article XXIV of the General Agreement on Tariffs and Trade (GATT), the key multilateral agreement governing international trade in goods amongst WTO members.

²In the existing literature, Article XXIV has often been invoked as a justification for the *assumption* that PTA members impose zero tariffs on each other. Though reasonable, this approach masks the incentives underlying the tariff-setting behavior of PTA members and, by design, fails to shed light on the consequences of requiring them to fully liberalize internal trade.

³Saggi et. al (2013) build on Saggi and Yildiz (2010) by considering trade agreements that take the form of customs unions as opposed to FTAs.

members. We next compare this WTO-consistent benchmark with two alternative settings. Under our first alternative scenario – called *unconstrained preferential liberalization* – FTA members have the freedom to implement jointly optimal internal tariffs as opposed to having to eliminate them as a precondition for forming the FTA.⁴ Under the second scenario – called *tariff discrimination* – the non-member country does not have to abide by MFN and is free to impose its optimal discriminatory tariffs on FTA members. Intuitively, the tariff discrimination scenario helps determine whether there is a sound rationale for requiring a country to practice non-discrimination when it itself faces discrimination at the hands of FTA members.

A comparison of the WTO-consistent scenario with the unconstrained preferential liberalization scenario delivers several interesting results. First, we show that if FTA members choose internal tariffs to maximize their joint welfare, they indeed have an incentive to impose positive tariffs on one another. The intuition for this surprising result rests on the interplay between two mechanisms: the *lack of external tariff coordination* between FTA members and the *complementarity of imports tariffs*. Each FTA member individually set its external tariff on the non-member and this lack of coordination leads each FTA member to ignore the benefits that its external tariff confers on its partner – if one FTA member raises its tariff on the non-member country, exports of the other member to its market increase but each FTA member ignores this effect on its partner’s export surplus while setting its external tariff. However, each FTA member ignores this effect on its partner’s export surplus while setting its own external tariff. Thus, the individually optimal external tariffs of FTA members are too low from the perspective of their joint welfare. The existence of tariff complementarity and the lack of external tariff coordination together imply that, while coordinating their internal tariffs, FTA members deliberately choose to set positive internal tariffs on each other: doing so commits each of them to a higher external tariff on the non-member country thereby bringing the individually optimal external tariffs of FTA members closer to jointly optimal ones. To confirm the role that external tariff coordination plays in generating positive internal tariffs within an FTA, we also consider a setting where FTA members can coordinate their external as well as internal tariffs, as they might be able to do under a customs union (CU). Under such a case, members indeed

⁴While GATT Article XXIV requires FTA members to impose zero internal tariffs on each other, FTA members do not always abide by this restriction. An analysis of PTAs involving 85 countries and 90 percent of world trade in 2007 found that roughly two-thirds of tariff lines with MFN rates greater than 15 percent were not reduced through PTAs (see Bagwell et. al, 2016 and WTO, 2011). Our model sheds light on the consequences of such non-compliance on the part of PTA members regarding the free internal trade requirement of GATT Article XXIV.

find it optimal to engage in free internal trade. This result suggests that the free internal trade requirement of Article XXIV is likely to be more binding for FTAs relative to CUs.⁵ Although we find general evidence that FTAs have more excluded sectors than CUs, there is a lack of comprehensive empirical evidence on internal tariffs and excluded sectors in FTAs and CUs.⁶ Freund and Ornelas (2010) highlighted the wide range of implementation rates of PTAs as a vital research area that has received little attention.⁷

The second major insight delivered by our analysis is that requiring FTA members to eliminate internal tariffs *benefits* the non-member since it leads to *lower* external tariffs on the part of FTA members. This result, driven by tariff complementarity, is noteworthy since part of the original intent behind the design of Article XXIV is plausibly to minimize any potential negative effects of FTAs on non-member countries. Ostensibly, this objective was met by prohibiting FTA members from raising their external tariffs on outsiders. However, in our model, it turns out that FTA members have no incentive to increase their external tariffs on the non-member country anyway.⁸ Thus, the Article XXIV stipulation that FTA members cannot raise tariffs on outsiders may actually do little to protect the interests of outsiders. The idea that the requirement of free internal trade amongst FTA members could imply *lower tariffs for outsiders* was probably unforeseen at the time the relevant GATT rules were crafted. Instead, it seems more likely that the requirement of zero internal tariffs was designed to promote trade creation amongst FTA members. Our analysis demonstrates that, somewhat surprisingly, it is the Article XXIV requirement of free internal trade within an FTA that ends up protecting the non-member as opposed to the restriction imposed on the external tariffs of FTA members.

⁵This result is in line with Kennan and Riezman (1990), Yi (1996), Bagwell and Staiger (1998), Cadot, de Melo, and Olarreaga (1999), Freund (2000), and Ornelas (2007).

⁶Liu (2010) studies how the influence of special interest groups relative to voters affects the choice between partial-scope (formed under the Enabling Clause of GATT) and full-fledged trade agreements.

⁷Using product exclusions from 15 FTAs signed by the US, EU, Japan, and Canada, Damuri (2012) shows that 7 percent of tariff lines are excluded, either temporarily or permanently. Agriculture and food products are the most protected products while manufactured products are the least protected. These product exclusions are also different across FTAs with different partners, highlighting the discriminatory feature of FTAs. Product exclusion is correlated with the regime of trade protection proxied by MFN tariff rates. Studying the bilateral trade agreements of countries in ASEAN, APEC, and South Asia, Menon (2009) also finds that the most commonly excluded sector is agriculture. In the example of Japan's trade agreement with Mexico, 13 percent of Mexico's exports to Japan are excluded from the trade agreement. In comparison, CUs like the European Union are fully implemented (Freund and Ornelas, 2010). Mercusor also only excluded sugar and automobiles (Olarreaga and Soloaga, 1998).

⁸This result also arises in Richardson (1993, 1995). Since an FTA shifts imports away from non-member countries, Richardson (1993) shows that in a model with endogenous protection FTA countries have an incentive to lower external tariffs to shift these imports back if the diverted imports reduce its welfare.

Our third major result pertaining to the free internal trade requirement of Article XXIV is that having such a requirement makes it harder to achieve global free trade. The logic for this result is as follows. By lowering the external tariffs of FTA members, the free internal trade requirement of Article XXIV makes it *less* attractive for the non-member to enter into trade agreements with them – by staying out, it remains free to impose its optimal import tariffs while facing relatively lower tariffs in the markets of FTA countries due to the disciplining force of the free internal trade requirement.⁹ Thus, the free internal trade requirement of Article XXIV might facilitate some degree of free-riding in the WTO system since it allows non-member countries to benefit from reductions in external tariffs of FTA members that result from their internal trade liberalization *without* having to offer any tariff cuts of their own. Thus, our overall message is somewhat nuanced: when circumstances are such that achieving global free trade is not possible, the free internal trade requirement of Article XXIV increases world welfare by reducing tariffs world-wide but, at the same time, it also reduces the likelihood of reaching global free trade.

From existing literature we know that optimal MFN tariffs generally impose fewer distortions than optimally chosen discriminatory tariffs.¹⁰ A comparison of the WTO-consistent scenario with the tariff discrimination scenario brings to light a hitherto ignored benefit of MFN: by making tariff discrimination infeasible, MFN reduces the potency of a country's optimal tariffs and therefore its incentive for unilaterally opting out of trade liberalization with other countries. Thus, by increasing the likelihood of each country voluntarily choosing to enter into international trade agreements, the MFN principle can act as a catalyst for trade liberalization. However, we also show that this pro-liberalization effect of MFN is weaker when one country is *deliberately excluded* by the other two (who prefer a bilateral trade agreement to a multilateral one). In other words, we show that the welfare case for requiring a country to follow MFN as a non-member trading with countries that are in a bilateral FTA with each other is stronger if it has voluntarily chosen to not enter into trade agreements with its trading partners relative to a scenario where it has been excluded from their bilateral FTA against its wishes. In our model, a pair of countries have an incentive to exclude the third country only when they can coordinate their external tariffs which is in the case of CUs. The practical implication of this result is that the case

⁹The the role of the free-rider problem caused by MFN during multilateral trade negotiations has been examined by Johnson (1965), Caplin and Krishna (1988), and Ludema and Mayda (2009, 2013). In a recent paper, Wong (2017) shows that the free rider problem removes global free trade as a stable outcome in multilateral trade negotiations.

¹⁰See Choi (1995), Bagwell and Staiger (1999b), Horn and Mavroidis (2001), McCalman (2002), Saggi (2004), and Bagwell and Staiger (2010) for analyses of the various legal and economic aspects of MFN.

for requiring MFN on the part of countries trading with CUs from which they have been deliberately excluded is relatively weaker than the case for MFN on the part of countries that have chosen not to join FTAs in order to benefit from the reductions in the external tariffs of FTA members while themselves retaining the freedom to utilize their optimal discriminatory tariffs.

Since Bhagwati (1991), the literature has paid significant attention to whether PTAs serve as *building* or *stumbling* blocs for multilateral trade liberalization. Early theoretical research on this issue generally took PTAs to be exogenously given and focused on how PTA membership affects the incentives that countries have for participating in multilateral trade liberalization (see, for example, Krishna, 1998; Ornelas, 2005a, 2005b). More recent studies, such as Goyal and Joshi (2006), Aghion et al. (2007), Furusawa and Konishi (2007), and Seidman (2009) consider endogenous PTAs but ignore the possibility of trade liberalization on an MFN basis. Under this approach, PTAs are seen as building blocs so long as their pursuit eventually leads to global free trade. However, Saggi and Yildiz (2010), Saggi et al. (2013), Missios et al. (2016) and Stoyanov and Yildiz (2015) have argued that PTAs ought to be seen as building blocks only if the freedom to pursue PTAs (granted to WTO members by GATT Article XXIV) is necessary for achieving global free trade. An attractive feature of this line of research is that it treats both preferential and multilateral liberalization as being endogenous. The present paper follows this approach and furthers the literature on the building versus stumbling bloc question by showing that the free internal trade requirement of Article XXIV makes it harder to achieve global free trade, i.e., it reduces the likelihood that PTAs act as building blocs. Moreover, we show that whether or not requiring MFN on the part of the non-member country is conducive for the cause of global free trade depends upon the nature of the PTA in question: MFN facilitates free trade when PTAs take the form of FTAs whereas it hinders it if they take the form of CUs.

2 Tariffs and trade

Our underlying trade model is an appropriately adapted version of the partial equilibrium ‘competing exporters’ framework developed by Bagwell and Staiger (1999a) to analyze the effects of PTAs. There are three asymmetrically endowed countries: i , j , and k and three (non-numeraire) goods: I , J , and K .¹¹ Each country’s market is served by two competing

¹¹All countries have large enough endowments of the freely traded numeraire good that they consume in positive quantities.

exporters and I denotes the good that corresponds to the upper case value of i . Country i is endowed with zero units of good I and e_i units of the other two goods.

The demand for good z in country i is given by

$$d(p_i^z) = \alpha - p_i^z \text{ where } z = I, J, \text{ or } K \quad (1)$$

As is well known, the above demand functions can be derived from a utility function of the form $U(c^z) = u(c^z) + w$ where c^z denotes consumption of good z ; w denotes the numeraire good; and $u(c^z)$ is quadratic and additively separable in each of the three goods. Country i must import good I in order to consume it and it can import it from either trading partner.

Let t_{ij} be the tariff imposed by country i on its imports of good I from country j . Ruling out prohibitive tariffs yields the following no-arbitrage conditions:

$$p_i^I = p_j^I + t_{ij} = p_k^I + t_{ik} \quad (2)$$

Let m_i^I be country i 's imports of good I . Since country i has no endowment of good I , we have

$$m_i^I = d(p_i^I) = \alpha - p_i^I \quad (3)$$

Each country's exports of a good must equal its endowment of that good minus its local consumption:

$$x_j^I = e_j - [\alpha - p_j^I] \quad (4)$$

Market clearing for good I requires that country i 's imports equal the total exports of the other two countries:

$$m_i^I = \sum_{j \neq i} x_j^I \quad (5)$$

Equations (2) through (5) imply that the equilibrium price of good I in country i equals:

$$p_i^I = \frac{1}{3} \left(3\alpha - \sum_{j \neq i} e_j + \sum_{j \neq i} t_{ij} \right) \quad (6)$$

A country's terms of trade motive for import tariffs is evident from equation (6): only a third of a given increase in either of its tariffs is passed on to domestic consumers in the form of a price increase, with the rest of the burden falling on the shoulders of foreign exporters.

From a welfare perspective, given the partial equilibrium nature of the model, it suffices to consider only protected goods. A country's welfare is defined as the sum of consumer

surplus, producer surplus, and tariff revenue over all such goods:

$$w_i = \sum_z CS_i^z + \sum_z PS_i^z + TR_i \quad (7)$$

Using equations (2) through (6) one can easily obtain welfare of country i as a function of endowment levels and tariffs. Let aggregate world welfare be defined as the sum of each country's welfare:

$$ww = \sum_i w_i. \quad (8)$$

Before proceeding further, we note that in order to guarantee non-negative exports and positive tariffs under all trade policy regimes, we impose the following parameter restriction throughout the paper: $\max\{e_i, e_j, e_k\} \leq \frac{5}{4} \min\{e_i, e_j, e_k\}$.¹²

We are now ready to report the key properties of the different types of optimal tariffs that arise under the various trade policy regimes that can arise in our model.

Suppose countries do not enter into any type of trade agreement with each other. Then, in the absence of an MFN clause, each country is free to tariff discriminate across its trading partners. Let country i 's optimal discriminatory tariff pair be given by $(t_{ij}^*, t_{ik}^*) \equiv \arg \max w_i(t_{ij}, t_{ik})$.

When forced to abide by MFN, country i must set $t_{ij}^* = t_{ik}^*$. Let t_i^M denote country i 's optimal MFN tariff where

$$t_i^M \equiv \arg \max w_i(t_{ij}, t_{ik}) \text{ such that } t_{ij} = t_{ik} \quad (9)$$

Using derivations reported in the appendix, it is straightforward to show that, when free to tariff discriminate, each country imposes a higher tariff on the larger exporter:

$$t_{ik}^* \geq t_{ij}^* \text{ iff } e_k \geq e_j \quad (10)$$

and that each country's optimal MFN tariff is bound by its discriminatory tariffs:

$$t_{ij}^* \leq t_i^M \leq t_{ik}^* \text{ where } e_j \leq e_k \quad (11)$$

Now let us consider how the formation of an FTA between two countries, say i and j , affects the non-member country. It is useful to begin with exogenously given internal and external tariffs and consider how variations in these tariffs affect the non-member. Let the

¹²Calculations supporting this restriction and all of the results reported in the paper are contained in the appendix.

pair of internal tariffs set by FTA members i and j on each other be denoted by (τ_{ij}, τ_{ji}) . Our first point is simply that, all else equal, the non-member loses if the internal tariffs within the FTA decline (we call this as the *discrimination effect*):

$$\frac{\partial w_k}{\partial \tau_{ij}} > 0 \text{ and } \frac{\partial w_k}{\partial \tau_{ji}} > 0 \quad (12)$$

Consider now the relationship between internal and external tariffs of an FTA between countries i and j . We assume that FTA members first choose their internal tariffs (τ_{ij}, τ_{ji}) to maximize their joint welfare and then, given internal tariffs, each FTA member chooses its external tariff to maximize its own welfare. Thus, as a member of a bilateral FTA with country j , country i chooses t_{ik} to $\max w_i(t_{ik}; \tau_{ij})$.¹³ The optimal external tariff of FTA member i as a function of its internal tariff on FTA member j is given by

$$t_{ik}^*(\tau_{ij}) \equiv \arg \max_{t_{ik}} w_i(t_{ik}; \tau_{ij})$$

Using the first order condition for the above problem, we can show the following:

$$\frac{dt_{ik}^*(\tau_{ij})}{d\tau_{ij}} > 0 \quad (13)$$

i.e. the individually optimal external tariff of an FTA member country is increasing in its internal tariff on the other member country. In other words, there is *tariff complementarity* between the internal and external tariffs of FTA member countries. This tariff complementarity implies that the deeper the degree of internal trade liberalization in an FTA, the lower the tariffs that FTA members impose on the non-member.

The above tariff analysis shows that the preferential trade liberalization undertaken by FTA members has two conflicting effects on the non-member country. On the one hand, the non-member loses from the discrimination that is inherent to FTAs. On the other hand, the internal liberalization within an FTA induces each member to lower its tariff on the non-member. Furthermore, when external tariffs are chosen by FTA members to maximize their respective welfare, the tariff complementarity effect *outweighs* the discrimination effect so that the larger the degree of internal trade liberalization between FTA members, the higher the non-member's welfare, i.e., at $t_{ik} = t_{ik}^*(\tau_{ij})$ and $t_{jk} = t_{jk}^*(\tau_{ji})$ we have:

$$\frac{\partial w_k}{\partial \tau_{ij}} < 0 \text{ and } \frac{\partial w_k}{\partial \tau_{ji}} < 0$$

¹³Due to the structure of the model, a country's individually tariff is independent of the tariffs of its trading partners (since these apply to different goods). In other words, country i 's choice of t_{ik} only depends upon τ_{ij} and is independent of all other tariffs.

Now consider tariff setting within an FTA. While setting their internal tariffs, FTA members jointly solve

$$\max_{\tau_{ij}, \tau_{ji}} [w_i(\tau_{ij}, \tau_{ji}, t_{ik}^*(\tau_{ij}), t_{jk}^*(\tau_{ji})) + w_j(\tau_{ij}, \tau_{ji}, t_{ik}^*(\tau_{ij}), t_{jk}^*(\tau_{ji}))]$$

In other words, while setting their internal tariffs, FTA member account for the fact that each of them chooses an individually optimal external tariff subsequently. The first order condition for τ_{ij} when evaluated at is given by

$$\frac{\partial w_i}{\partial \tau_{ij}} + \frac{\partial w_i}{\partial t_{ik}} \frac{dt_{ik}^*(\tau_{ij})}{d\tau_{ij}} + \frac{\partial w_j}{\partial \tau_{ij}} + \frac{\partial w_j}{\partial t_{ik}} \frac{dt_{ik}^*(\tau_{ij})}{d\tau_{ij}} = 0$$

which is the same as

$$\frac{\partial(w_i + w_j)}{\partial \tau_{ij}} + \frac{dt_{ik}^*}{d\tau_{ij}} \left[\frac{\partial(w_i + w_j)}{\partial t_{ik}} \right] = 0 \quad (14)$$

Note that

$$\frac{\partial(w_i + w_j)}{\partial \tau_{ij}} < 0$$

i.e., all else equal, an increase in country i 's internal tariff lowers the joint welfare of FTA members but, as noted above in (13), due to tariff complementarity we have $\frac{dt_{ik}^*}{d\tau_{ij}} > 0$. Furthermore, at the individually optimal external tariff chosen by country i the following must hold:

$$\frac{\partial w_i}{\partial t_{ik}} = 0$$

But since $\frac{\partial w_j}{\partial t_{ik}} > 0$, it immediately follows from (14) that at the individually optimal external tariff chosen by country i we must have

$$\frac{\partial(w_i + w_j)}{\partial t_{ik}} > 0$$

Intuitively, since country i does not take into account the effect of its tariff on its partner country, it is jointly welfare improving for the two FTA members to raise their external tariffs above their individually optimal tariffs. As a result, though positive internal tariffs hurts FTA members by lowering internal trade, they also benefit them by committing them to higher tariffs on the non-member. As a result, FTA members find it jointly optimal to impose positive internal tariffs on each other. Let the optimal internal tariffs set by countries i and j on each other be denoted by $(\tau_{ij}^*, \tau_{ji}^*)$.

We summarize the key messages of the above analysis in the following lemma:

Lemma 1: (i) *In the absence of trade agreements, each country imposes a higher tariff on the partner from which it imports more under free trade: $t_{ik}^* \geq t_{ij}^*$ iff $e_j \leq e_k$; (ii) a*

country's optimal MFN tariff is bound by its optimal discriminatory tariffs: $t_{ij}^* \leq t_i^M \leq t_{ik}^*$ where $e_j \leq e_k$; (iii) the larger the degree of internal trade liberalization undertaken by FTA members, the higher the welfare of the non-member country; and (iv) FTA members impose strictly positive internal tariffs on each other, i.e. $\tau_{ij}^* > 0$ and $\tau_{ji}^* > 0$.

The intuition behind the first two results reported in Lemma 1 is already well-established in the literature – see, for example, Saggi (2004). As noted above, the result in part (iii) obtains because the tariff complementarity effect of an FTA dominates its discrimination effect. The intuition behind part (iv) is rather subtle and is as follows. Due to the lack of external tariff coordination in an FTA, each FTA member does not take into account the fact that an increase in its external tariff benefits its FTA partner whose exports compete with those of the non-member. Thus, the individually optimal external tariffs of FTA members are too low from the perspective of maximizing their joint welfare. But the coordination of internal tariffs that occurs prior to the setting of external tariffs provides FTA members with a partial remedy to this problem. Due to the existence of tariff complementarity, FTA members deliberately choose to set positive internal tariffs on each other: doing so commits each of them to imposing a higher external tariff on the non-member country thereby bringing their individually optimal external tariffs closer to jointly optimal ones.

To confirm the role that tariff coordination plays in generating positive internal tariffs within an FTA, suppose FTA members could coordinate both internal and external tariffs, as they might be able to do under a customs union (CU). Then, members solve¹⁴

$$\max_{\tau_{ij}, \tau_{ji}, t_{ik}, t_{jk}} [w_i(\tau_{ij}, \tau_{ji}, t_{ik}, t_{jk}) + w_j(\tau_{ij}, \tau_{ji}, t_{ik}, t_{jk})]$$

Since tariffs of different countries apply to different goods, it suffices to focus on the choice of τ_{ij} and t_{ik} . Differentiating the objective function with respect to τ_{ij} we have

$$\frac{\partial(w_i + w_j)}{\partial \tau_{ij}} < 0$$

If external tariffs can also be coordinated, an FTA becomes equivalent to a CU in our model and its members find it optimal to engage in free internal trade since their joint welfare is strictly decreasing in each of the internal tariffs. The optimal external tariff t_{ik}^u of the CU between i and j defined, as usual, by $\frac{\partial(w_i + w_j)}{\partial t_{ik}} = 0$. It is straightforward to show that CU members impose higher external tariffs than FTA members: $t_{zk}^u > t_{zk}^*$ where $z = i, j$. Thus,

¹⁴When both external and internal tariffs are coordinated, the tariff problem compresses to a single stage.

due to the dual coordination of internal and external tariffs, a CU between two countries yields (i) deeper internal trade liberalization and (ii) higher external tariffs relative to an FTA between them.

3 Endogenous trade agreements

The three policy scenarios that we contrast are formalized as follows:

(a) *WTO-consistent scenario*: This scenario is captured by a three stage game of trade liberalization under which countries abide by both Article I and Article XXIV of GATT. In the first stage, countries enter into FTAs with one another (the process of FTA formation is described in greater detail below). In the second stage, given the trade policy regime that results from the first stage, countries choose their tariffs. If an FTA is formed, its members practice free internal trade while imposing individually optimal external tariffs on the non-member who, in accordance with MFN, imposes non-discriminatory tariffs on the two member countries. At the third stage of the game, given trade agreements and tariffs, international trade and consumption take place.

(b) *Unconstrained preferential liberalization scenario*: This scenario is formalized as a four stage game that proceeds as follows. The first stage of the game remains the same as the first stage of the WTO-consistent scenario. At the second stage, given the policy regime, FTA members set their internal tariffs to maximize their joint welfare. As opposed to the WTO-consistent scenario described in (a), the internal tariffs of an FTA do not have equal zero. Next, all countries simultaneously choose their individually optimal external tariffs. At the last stage of the game, international trade and consumption occur.

(c) *Tariff discrimination scenario*: This scenario differs from the WTO-consistent benchmark in only one way: at the second stage of the game, the non-member country is free to impose discriminatory tariffs on FTA members as opposed to having to treat them in an MFN manner. Thus, all countries engage in some type of tariff discrimination: FTA members discriminate against the non-member by imposing higher tariffs on it than they do on each other while the non-member discriminates between them by imposing a higher tariff on the country from whom it imports more (see Lemma 1).

We now describe the process of FTA formation that occurs during the first stage of the game and is common to all three scenarios.

The process of FTA formation: At the first stage of the game, each country announces whether or not it wants to sign an FTA with each of the other two countries. Denote

country i 's announcement by σ_i and its strategy set by S_i where

$$S_i = \{\{\phi, \phi\}, \{j, \phi\}, \{\phi, k\}, \{j, k\}\} \quad (15)$$

In S_i , $\{\phi, \phi\}$ denotes an announcement in favor of no FTAs, $\{j, \phi\}$ an announcement 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. Since a trade agreement requires consent from both sides, we posit the following mapping between various announcements profiles and the types of trade agreements that countries can form:

(i) No two announcements match or the only matching announcements are $\{\phi, \phi\}$. All of these announcement profiles yield no agreement $\langle \Phi \rangle$. Under the WTO consistent and unconstrained preferential liberalization scenarios, all countries impose their optimal MFN tariffs on one another. Under the tariff discrimination scenario, all countries impose their optimal discriminatory Nash tariffs on one another.

(ii) Two countries announce each others' name and there is no other matching announcement: i.e., $j \in \sigma_i$ and $i \in \sigma_j$ while $i \notin \sigma_k$ and/or $k \notin \sigma_i$ and $j \notin \sigma_k$ and/or $k \notin \sigma_j$. All of these announcements yield an FTA between countries i and j denoted by $\langle ij \rangle$ under which members impose their jointly optimal internal tariffs on each other and their individually optimal external tariffs on the non-member.

(iii) Country i announces in favor of signing an FTA with countries j and k while countries j and/or k announce only in favor of signing an FTA with country i : i.e. $\sigma_i = \{j, k\}$; $i \in \sigma_j$; and $i \in \sigma_k$ while $k \notin \sigma_j$ and/or $j \notin \sigma_k$. This set of announcements yields a pair of independent FTAs (i.e. a hub and spoke trading regime) with i as the common member denoted by $\langle ij, ik \rangle$ (or simply $\langle ih \rangle$). Under a hub and spoke agreement $\langle ih \rangle$, country i sets jointly optimal internal tariffs with the two other spokes while the spokes solve the same tariff problems as they do under a bilateral FTA with country i .

(iv) All countries announce each others' names, i.e., the announcement profile is $\Omega^F \equiv \{\sigma_i = \{j, k\}, \sigma_j = \{i, k\}, \sigma_k = \{i, j\}\}$. This announcement profile yields global free trade $\langle F \rangle$.

Note that since an FTA between two countries can arise only if it is mutually acceptable to both sides, multiple announcement profiles can map into the same agreement. For example, the FTA $\langle ij \rangle$ obtains when (i) countries i and j call only each other, *regardless* of the nature of country k 's announcement: if $\sigma_i = \{j, \phi\}$ and $\sigma_j = \{i, \phi\}$, then $\langle ij \rangle$ obtains for all four possible announcements on the part of country k , i.e., for $\sigma_k = \{\phi, \phi\}, \{i, \phi\}, \{\phi, j\}$ and $\{i, j\}$ so that country k 's announcement has no bearing upon the outcome when neither of the other two countries' announce its name; (ii) countries i and j announce each

other's name and either one or both of them also announce country k but country k does not reciprocate: i.e. all of the following types of announcements map into the FTA $\langle ij \rangle$: (a) $\sigma_i = \{j, k\}$ and $\sigma_j = \{i, \phi\}$ but $i \notin \sigma_k$ or (b) $\sigma_i = \{j, \phi\}$ and $\sigma_j = \{i, k\}$ but $j \notin \sigma_k$ or (c) $\sigma_i = \{j, k\}$ and $\sigma_j = \{i, k\}$ but $\sigma_k = \{\phi, \phi\}$.

When analyzing the above games, we only consider those Nash equilibria that are *coalition-proof*. 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 coalition proof Nash equilibrium (CPNE) is a Nash equilibrium that is immune to all *self-enforcing* coalitional deviations.

4 Equilibrium agreements

In order to simplify exposition, we make the following assumption:

Assumption 1: *Countries l and l' are larger importers than country s : $e_s = \theta e \geq e_l = e_{l'} = e$ where $1 \leq \theta \leq 5/4$.*¹⁵

It is worth pointing out here that, in our model, all countries have the ability to manipulate their terms of trade via import tariffs. Country s has a weaker ability to manipulate its terms of trade but its not a 'small' country in the traditional sense of the term wherein it would be a price-taker on world markets.

We proceed as follows. First, we study FTA formation in our WTO-consistent benchmark scenario and show that, in this scenario, no two countries have an incentive to form a bilateral trade agreement aimed at excluding the third country. Instead, it is the strength of the free-riding incentive of the non-member country that proves pivotal in determining whether or not global free trade obtains in equilibrium. Next, we derive equilibrium trade agreements under unrestricted preferential liberalization scenario where FTA members are free to impose positive internal tariffs on each other. In equilibrium, FTA members utilize this freedom and they end up imposing higher external tariffs relative to the WTO-consistent benchmark where they are forced to eliminate internal tariffs. This in turn reduces the free-riding incentive of the non-member country and therefore furthers the

¹⁵The qualitative nature of our results is robust to a scenario where all three countries are asymmetric, such as when $e_s = \theta_s e \geq e_m = \theta_m e \geq e_l = e$. But since the key insights can be illustrated more easily in the simpler case where the two larger countries are symmetric, in what follows we proceed with this assumption.

cause of global free trade. On the other hand, when global free trade is infeasible, the free internal trade requirement of Article XXIV raises global welfare by lowering internal *and* external tariffs of FTA countries. Finally, we examine the tariff discrimination scenario under which the non-member country is free to tariff discriminate between FTA members. We show that the welfare benefits of the MFN clause depend on whether the non-member voluntarily stays out of the FTA between the other two countries or has been deliberately excluded by them.

4.1 WTO-consistent agreements

In this section, we derive equilibrium trade agreements under our benchmark scenario where countries follow both Articles I and XXIV of GATT – i.e. the non-member country follows MFN and FTA members engage in free internal trade. Let country i 's welfare as a function of the underlying trade policy regime r be denoted by $w_i(r)$, where $r = \langle \Phi \rangle, \langle ij \rangle, \langle ih \rangle$, or $\langle F \rangle$ and it is understood that all countries impose optimal tariffs consistent with regime r . For example, if $r = \langle ij \rangle$ then countries i and j impose the optimal internal tariffs τ_{ij}^* and τ_{ji}^* on each other respectively while imposing the tariffs $t_{ik}^*(\tau_{ij}^*)$ and $t_{jk}^*(\tau_{ji}^*)$ on country k . Let $\Delta w_i(r - v)$ denote the difference between country i 's welfare under trade agreements r and v : $\Delta w_i(r - v) \equiv w_i(r) - w_i(v)$, where $r, v = \langle \Phi \rangle, \langle ij \rangle, \langle ih \rangle$, or $\langle F \rangle$. Furthermore, let $\theta_i(r - v)$ denote the critical threshold of asymmetry at which country i is indifferent between regimes r and v .

We first state the following lemma that explains how differences in market power across countries lead them to have asymmetric preferences over various trade regimes:

Lemma 2: *In the WTO-consistent approach to the formation of trade agreements, the following hold:*

(i) *Each country prefers to form a bilateral FTA with the larger importer relative to the smaller one: $\Delta w_l(ll' - sl) > 0$ for all θ .*

(ii) *The smaller importer (s) has an incentive to form an additional bilateral FTA under any trade regime except for when it is a non-member facing an FTA between the other two countries.*

(iii) *Each larger importer prefers being a non-member under a bilateral FTA to being a spoke under a hub and spoke regime while the smaller importer does so only when the degree of endowment asymmetry is sufficiently small: $\Delta w_l(lh - sl) < 0$ and $\Delta w_l(sh - sl) < 0$ for all θ and $\Delta w_s(lh - ll') < 0$ when $\theta < \theta_s(lh - ll')$.*

(iv) All countries prefer being the hub under a hub and spoke regime relative to all other trade policy regimes: $\Delta w_i(ih - \Phi) > 0$; $\Delta w_i(ih - F) > 0$ and $\Delta w_i(ih - ij) > 0$ for all $i = s, l, l'$.

Part (i) of Lemma 1 follows from two reinforcing effects. The larger a country's trading partner's import volume, the larger the increase in export surplus it enjoys from the elimination of its partner's optimal tariff and the smaller the loss it suffers from its own trade liberalization since its tariff reduction applies to a smaller volume of imports. Thus, a country prefers to form a bilateral FTA with the larger importer amongst its two trading partners. The second part of Lemma 1 argues that the smaller importer (i.e. country s) has an incentive to form an additional FTA under any given regime except when the existing regime is $\langle ll' \rangle$ and the endowment asymmetry is sufficiently large (see part (iii)). This implies that, generally speaking, the larger importing country's choice is critical in determining whether or not an FTA between two asymmetric countries arises. Finally, part (iv) says that being a hub country is better for all countries irrespective of their size relative to all other trade policy regimes. Note in particular that, relative to free trade, the hub country enjoys privileged access in both spoke countries while its domestic surplus is no different. Moreover, this privileged access in export markets is so desirable that a hub country has no incentive to unilaterally revoke either or both of its FTAs.

While members of an FTA discriminate against the non-member country, we know from the above tariff analysis that the internal trade liberalization undertaken by FTA partners actually benefits the non-member. This raises the possibility that, starting from no agreement $\langle \Phi \rangle$, the formation of an FTA makes all countries better off (i.e. is Pareto improving relative to $\langle \Phi \rangle$). Indeed, we can show that the smaller country benefits from the formation of an FTA between large countries only when the degree of endowment asymmetry is sufficiently small:

$$\Delta w_s(ll' - \Phi) > 0 \text{ when } \theta < \theta_s(ll' - \Phi) \quad (16)$$

Second, while the larger non-member (country l') always benefits from the formation of $\langle sl \rangle$, the larger member country benefits from the formation of $\langle sl \rangle$ only when the degree of asymmetry is sufficiently small:

$$\Delta w_l(sl - \Phi) > 0 \text{ when } \theta < \theta_l(sl - \Phi) \quad (17)$$

Therefore, we find the following:

Proposition 1: *Relative to no agreement $\langle \Phi \rangle$ wherein all countries impose their optimal Nash tariffs on each other, the FTA $\langle ll' \rangle$ is Pareto-improving iff $\theta < \theta_s(ll' - \Phi)$ while the the FTA $\langle sl \rangle$ is Pareto-improving iff $\theta < \theta_l(sl - \Phi)$.*

Armed with the underlying incentives identified by Lemma 2, we are now ready to determine the CPNE of the WTO-consistent game of trade agreement formation. We proceed by considering each of the announcement profiles that yield the various trade policy regimes in turn. First, consider the announcement profile leading to global free trade $\langle F \rangle$. First note from part (ii) of the Lemma 2 that smaller importer (i.e. country s) has no incentive to participate in any deviation (unilateral or coalitional). Thus, if there exists a coalitional deviation, it must involve countries l and l' . Taking country s ' announcement fixed at $\{l, l'\}$, countries l and l' have an incentive to jointly deviate from their respective announcements $\{s, l'\}$ and $\{s, l\}$ to $\{\phi, l'\}$ and $\{\phi, l\}$ in order to exclude country s from a free trade network when country s is sufficiently small:

$$\Delta w_l(F - ll') < 0 \text{ when } \theta > \theta_l(F - ll') \quad (18)$$

The above result establishes the existence of an *exclusion incentive*: when the endowment asymmetry is sufficiently pronounced (i.e. $\theta > \theta_l(F - ll')$) the two larger importers prefer a bilateral FTA between themselves to global free trade. Furthermore, since world welfare under free trade is higher than that under a bilateral FTA, it follows that the non-member country is better off under free trade relative to the bilateral FTA $\langle ij \rangle$.

The key question is whether the joint exclusion incentive of the two larger importers is self-enforcing or not. The answer to this question turns out to be in the negative. To see why, suppose each country announces in favor of an FTA with both its trading partners. Starting with these announcements the two larger importers have an incentive to exclude the smaller country by jointly altering their announcements such that the announcement profile changes from Ω^F (which yields free trade) to $\Omega_1^{ll'} = \{\sigma_l = \{\phi, l'\}, \sigma_{l'} = \{\phi, l\}, \sigma_s = \{l, l'\}\}$ thereby altering the associated trade regime from free trade to the bilateral FTA $\langle ll' \rangle$. However, from part (iv) of Lemma 2 we know that each country's most preferred trading arrangement is a hub and spoke regime with itself as the hub. It follows then that, holding constant the announcement of the excluded country at $\sigma_s = \{l, l'\}$, each member of the deviating coalition (l or l') has an incentive to alter its announcement so as to include country s . For example, country l has an incentive to alter its announcement from $\sigma_l = \{\phi, l'\}$ to $\sigma_l = \{s, l'\}$ which alters the trade regime from $\langle ll' \rangle$ to $\langle lh \rangle$. Since the welfare of a hub is higher than that of a member country in a single FTA – see part (iv) of Lemma 2 –

the original coalitional deviation of countries l and l' from Ω^F to $\Omega_1^{ll'}$ is *not* self-enforcing. Thus, in a nutshell, the lure of a hub and spoke trading arrangement makes any joint deviation from Ω^F to an announcement profile that supports a bilateral FTA between any two countries not-self enforcing.

Consider now announcement deviations that convert the trade regime from $\langle F \rangle$ to $\langle \Phi \rangle$. It is easy to see that since all countries are better off under free trade relative to $\langle \Phi \rangle$, no two countries have an incentive to deviate from Ω^F to an announcement profile that yields $\langle \Phi \rangle$. For example, holding $\sigma_s = \{l, l'\}$, countries l and l' have no incentive to jointly deviate from their respective announcements $\{s, l'\}$ and $\{s, l\}$ to $\{\phi, \phi\}$ and $\{\phi, \phi\}$. Based on the above discussion, the only possible type of self-enforcing deviation from Ω^F that we need to consider is a unilateral deviation from Ω^F by one of the large importers. To this end, we find that there exists no incentive of a large country (say l) to unilaterally deviate from its announcements $\{s, l'\}$ to any announcement that leads to a hub and spoke regime under which country s is a hub and it itself is a spoke:

$$\Delta w_l(F - sh) = \Delta w_{l'}(F - sh) \geq 0 \text{ for all } \theta \quad (19)$$

Then two unilateral deviation incentives remain to be examined: (i) country l unilaterally deviates from $\{s, l'\}$ to $\{\phi, l'\}$:

$$\Delta w_l(F - l'h) = \Delta w_{l'}(F - lh) < 0 \text{ when } \theta > \theta_l(F - l'h) \quad (20)$$

and (ii) country l unilaterally deviates from $\{s, l'\}$ to $\{\phi, \phi\}$:

$$\Delta w_l(F - sl') = \Delta w_{l'}(F - sl) < 0 \text{ when } \theta > \theta_l(F - sl') \quad (21)$$

We find that $\theta_l(F - sl') < \theta_l(F - l'h)$ and thus the announcement profile leading to $\langle F \rangle$ is CPNE whenever $\theta \leq \theta_l(F - sl')$.

What if $\langle F \rangle$ is not a CPNE, as is the case when $\theta > \theta_l(F - sl')$? We can quickly rule out the various announcement profiles leading to the hub and spoke regimes as candidates for CPNE. To see why, recall from part (iii) of Lemma 2 that a larger spoke country (say l) under $\langle sh \rangle$ and $\langle l'h \rangle$ has an incentive to unilaterally deviate from its respective announcements $\{s, \phi\}$ and $\{\phi, l'\}$ to $\{\phi, \phi\}$ and $\{\phi, \phi\}$, leading to a deviation from $\langle sh \rangle$ to $\langle sl' \rangle$ and from $\langle l'h \rangle$ to $\langle sl' \rangle$. Since these unilateral deviations are self-enforcing, any announcement profile leading to a hub and spoke regime cannot be a CPNE.

Next, we consider the various announcement profiles that lead to no agreement $\langle \Phi \rangle$. Since countries l and l' have an incentive to jointly deviate from their respective announcements $\{\phi, \phi\}$ and $\{\phi, \phi\}$ to $\{\phi, l'\}$ and $\{\phi, l\}$ in order to form $\langle ll' \rangle$, this joint deviation is self-enforcing. As a result, any announcement profile that yields $\langle \Phi \rangle$ cannot be a CPNE.

The only remaining candidates for CPNE are the announcement profiles that lead to bilateral FTAs. We start with those profiles that yield an FTA between the smaller importer and one of the larger ones, say $\langle sl \rangle$). We find that, when $\theta > \theta_l(sl - \Phi)$, country l has an incentive to unilaterally deviate from its announcement $\{s, \phi\}$ to $\{\phi, \phi\}$ thereby converting the trade policy regime from the bilateral FTA $\langle sl \rangle$ to no agreement $\langle \Phi \rangle$. Second, we know from part (iv) of Lemma 2 that the coalitional announcement deviation leading that converts $\langle sl \rangle$ to $\langle ll' \rangle$ is *not* self-enforcing since the common member country (i.e. country l) has an incentive to further deviate to become the hub country, taking the announcement of its complement as fixed. Third, note from the above discussion that the coalitional announcement deviation that replaces $\langle sl \rangle$ by $\langle F \rangle$ is self-enforcing only when $\theta \leq \theta_l(F - sl')$. Finally, it is immediate from the part (iii) of Lemma 2 that country l' has no incentive to engage in any coalitional announcement deviations that replace $\langle sl \rangle$ by $\langle sh \rangle$ or $\langle sl \rangle$ by $\langle lh \rangle$. As a result, the announcement profile leading to $\langle sl \rangle$ is a CPNE whenever $\theta_l(F - sl') \leq \theta \leq \theta_l(sl - \Phi)$.

Finally, we consider the bilateral FTA between the two larger countries, i.e., $\langle ll' \rangle$. First, as before, the coalitional announcement deviation from $\langle ll' \rangle$ to $\langle F \rangle$ occurs $\theta \leq \theta_l(F - ll')$ and it is self-enforcing when $\theta \leq \theta_l(F - sl')$. Second, we can show that when $\theta > \theta_s(lh - ll')$ country s and either of the larger countries (say l) have an incentive to jointly deviate from their respective announcements $\{\phi, \phi\}$ and $\{\phi, l'\}$ to $\{l, \phi\}$ and $\{s, l'\}$, leading to a deviation from $\langle ll' \rangle$ to $\langle lh \rangle$ and this deviation is self-enforcing. Since $\theta_s(lh - ll') < \theta_l(F - sl')$, these self-enforcing announcement deviations cover the entire parameter space and thus the announcement profile supporting $\langle ll' \rangle$ is not a CPNE.

We summarize the main findings of the above analysis below:

Proposition 2: *The equilibria of the WTO-consistent game of trade liberalization where FTA members have to practice free internal trade and the non-member has to abide by MFN are as follows:*

- (i) Free trade $\langle F \rangle$ is the equilibrium agreement when $\theta \leq \theta_l(F - sl')$.¹⁶
- (ii) An asymmetric bilateral FTA $\langle sl \rangle$ (or $\langle sl' \rangle$) is the equilibrium when $\theta_l(F - sl') \leq \theta \leq \theta_l(sl - \Phi)$.
- (iii) There exists no equilibrium if $\theta > \theta_l(sl - \Phi)$.

¹⁶We should note here that, technically speaking, the equilibrium is the announcement profile Ω^F that yields free trade as the agreement. In what follows, for expositional ease, we state our results directly in terms of various trade agreements that emerge as equilibrium outcomes as opposed to the announcement profiles that support them.

Insert Figure 1

The above proposition relates the degree of underlying asymmetry to the nature of equilibrium agreements. Part (i) simply says that if the degree of endowment asymmetry is sufficiently small, free trade is the equilibrium outcome. It is important to reiterate that while the exclusion incentives of larger importing countries go unexercised in equilibrium, each large importing country's incentive to unilaterally deviate from free trade proves critical for determining the viability of free trade. Part (ii) says that if the degree of endowment asymmetry is sufficiently large, only an asymmetric FTA ($\langle sl \rangle$ or $\langle sl' \rangle$) is an equilibrium – in such a situation, one of the larger importing countries prefers being a non-member to participating in any bilateral or multilateral agreements. Note from the above discussion that the bilateral FTA between the two larger countries $\langle ll' \rangle$ fails to arise in equilibrium. Finally, part (iii) of Proposition 1 says that there exists no CPNE if the degree of endowment asymmetry is very large. In such a situation, our theory offers no guidance regarding which of the trade regimes should be expected to arise in equilibrium.¹⁷

What if Article XXIV allows for positive internal tariffs? Next we allow this possibility.

4.2 Equilibrium agreements with internal tariffs

Here, we consider the scenario of *unconstrained preferential liberalization* wherein FTA member countries jointly choose their internal tariffs before setting their individually optimal external tariffs. Recall that, due to the existence of tariff complementarity in our model, the deeper the internal trade liberalization in an FTA, the lower the external tariffs of member countries. As a result, when allowed, member countries set positive internal tariffs on trade from each other and due to incomplete internal trade liberalization tariff complementarity is smaller relative to the case with free internal trade requirement.

Under a hub and spoke agreement $\langle ih \rangle$, country i has a trade agreement with both countries j and k and its internal tariffs are chosen to maximize the joint welfare of all three countries which leads to zero internal tariffs on its part: $\tau_{ij}^*(ih) = \tau_{ik}^*(ih) = 0$, while the spoke countries' tariffs solve the same problem as they do under a bilateral trade agreement so that $t_{jk}^*(ih) = t_{jk}^*(ij)$.

Let country i 's welfare as a function of the underlying trade agreement r with positive internal tariffs be denoted by $w_i(\hat{r})$ and let $\Delta w_i(\hat{r} - \hat{v})$ denote the difference between country

¹⁷When we compare this parameter space under different scenarios, we do not take any stand regarding the trade regimes that can arise.

i 's welfare under trade agreements r and v with positive internal tariffs: $\Delta w_i(\widehat{r} - \widehat{v}) \equiv w_i(\widehat{r}) - w_i(\widehat{v})$. The following lemma explains the preferences of asymmetric countries over trade regimes when member countries are able to impose internal tariffs before setting their external tariffs:

Lemma 3: *When member countries of an FTA choose their internal tariffs jointly before setting their individually optimal external tariffs, the following hold:*

(i) *Starting from no agreement $\langle \Phi \rangle$, all countries have an incentive to form a bilateral FTA: $\Delta w_i(\widehat{ij} - \Phi) > 0$ for all θ and $i, j = s, l, l'$.*

(ii) *A large importer prefers a bilateral FTA with the other larger importer relative to the smaller one: $\Delta w_l(\widehat{ll'} - \widehat{sl}) > 0$ for all θ .*

(iii) *The smaller importer has an incentive to form a bilateral FTA under any trade regime.*

(iv) *Each larger importer prefers being a non-member under a bilateral FTA to being a spoke under a hub and spoke regime provided endowments are sufficiently asymmetric across countries: $\Delta w_l(\widehat{l'h} - \widehat{sl'}) < 0$ when $\theta > \theta_l(\widehat{l'h} - \widehat{sl'})$ and $\Delta w_l(\widehat{sh} - \widehat{sl'}) < 0$ when $\theta > \theta_l(\widehat{sh} - \widehat{sl'})$.*

(v) *All countries prefer being the hub country under a hub and spoke regime relative to no agreement as well as to being a member under a bilateral FTA: $\Delta w_i(\widehat{ih} - \Phi) > 0$ and $\Delta w_i(\widehat{ih} - \widehat{ij}) > 0$ for all $i = s, l, l'$.*

The intuition behind part (i) of Lemma 3 is that when member countries under a bilateral FTA can coordinate internal tariffs before setting their individually optimal external tariffs, they partially internalize the effects of their external tariffs on one another and this increases the incentive of larger importing countries to form a bilateral FTA. We find that, relative to the WTO-consistent benchmark case, the incentives for forming FTAs are generally stronger under constrained preferential liberalization since FTA members are less constrained and can therefore achieve higher levels of welfare under FTAs. Furthermore, due to joint determination of internal tariffs, a country's preference to form a bilateral FTA with the larger of its two trading partners is even stronger. Parts (iii) and (iv) of Lemma 3 differ from part (ii) and part (iii) of Lemma 2 in an important way: while coordinating their internal tariffs, FTA members deliberately choose to set positive internal tariffs: doing so leads each member to impose a higher external tariff on the non-member country relative to our WTO-consistent benchmark case. This in turn decreases the incentive of countries to stay outside FTAs, whether as a non-member facing a bilateral FTA or as a

spoke under a hub and spoke regime. Finally, part (v) of Lemma 3 says that being a hub country is better for all countries (irrespective of their size) relative to no agreement and to being a member of a bilateral FTA.

An interesting question is whether bilateral FTA formation is more or less likely to be Pareto-improving over no agreement when FTA members are free to impose positive internal tariffs on each other. Since tariff complementarity is weaker when FTA members are not constrained by Article XXIV, the non-member country's relative situation is worse off under the unconstrained liberalization scenario relative to the WTO-consistent benchmark scenario. As indicated above, since member countries always benefit from forming an FTA relative to no agreement, the Pareto-improvement condition of a bilateral FTA with internal tariffs relies only on the welfare preference of the non-member country. We first find that, starting from no agreement, a larger country always benefits from the formation of an FTA between the other two countries:

$$\Delta w_V(\widehat{sl} - \Phi) > 0 \text{ for all } \theta \quad (22)$$

Second, as under the WTO-consistent benchmark case, the small country benefits from the formation of an FTA between the two larger countries only when the degree of asymmetry is sufficiently small:

$$\Delta w_s(\widehat{ll'} - \Phi) > 0 \text{ when } \theta < \theta_s(\widehat{ll'} - \Phi) \quad (23)$$

A comparison of conditions in (16) and (23) yields

$$\theta_s(\widehat{ll'} - \Phi) < \theta_s(ll' - \Phi)$$

implying that the formation of $\langle \widehat{ll'} \rangle$ is *less likely* to be Pareto improving relative to $\langle ll' \rangle$. We have:

Proposition 3: (i) *Relative to no agreement $\langle \Phi \rangle$, an unconstrained FTA between two asymmetric countries $\langle \widehat{sl} \rangle$ is necessarily Pareto-improving whereas the unconstrained FTA between the two larger importers $\langle \widehat{ll'} \rangle$ is Pareto-improving only when $\theta < \theta_s(\widehat{ll'} - \Phi)$.*

(ii) *The freedom to set positive internal tariffs in a coordinated fashion makes the FTA between two asymmetric partners more likely to be Pareto-improving while the opposite is true for the FTA between the two larger importers.*

We are now ready to derive equilibria under the game of unrestricted preferential liberalization. First note, it is immediate from part (i) of Lemma 3 that any two countries

have an incentive to jointly deviate from their respective announcements under $\langle \Phi \rangle$ to announcement profiles leading to a bilateral FTA. Since this deviation is self-enforcing, $\langle \Phi \rangle$ is not an equilibrium.

Next, we consider the announcement profiles leading to $\widehat{\langle ll' \rangle}$. It is immediate from part (iii) and part (v) of the Lemma 3 that, taking the announcement profile of a large country (say l') as given, country s and either of the large member countries (say l) have incentives to jointly deviate from their respective announcements $\{\phi, \phi\}$ and $\{\phi, l'\}$ to $\{l, \phi\}$ and $\{s, l'\}$, leading to a deviation from $\widehat{\langle ll' \rangle}$ to $\widehat{\langle lh \rangle}$ and this deviation is self enforcing. As a result, the announcement profile leading to $\widehat{\langle ll' \rangle}$ is never a CPNE.

Next, we consider the announcement profile leading to global free trade $\langle F \rangle$. As in the benchmark case, note from part (iii) of Lemma 2 that any deviation (unilateral or coalitional) from $\langle F \rangle$ does not involve country s . Thus, if there exists a coalitional deviation, it must be by countries l and l' . Similar to the benchmark WTO case, when countries have the ability to set positive internal tariffs, large countries still have incentives to exclude the small country. In other words, taking country s ' announcement fixed $\{l, l'\}$, countries l and l' have incentives to jointly deviate from their respective announcements $\{s, l'\}$ and $\{s, l\}$ to $\{\phi, l'\}$ and $\{\phi, l\}$ in order to exclude country s from a free trade network when country s is sufficiently small:

$$\Delta w_l(F - \widehat{ll'}) < 0 \text{ when } \theta > \theta_l(F - \widehat{ll'}) \quad (24)$$

The following result is based on the comparison of the exclusion incentives contained in (18) and (24):

Lemma 4: *The larger importers have a stronger incentive to exclude the smaller country from their mutual trade agreement under the unconstrained preferential liberalization scenario under which they impose positive internal tariffs on each other relative to the WTO-consistent benchmark where they are required to fully liberalize internal trade: $\theta_l(F - \widehat{ll'}) < \theta_l(F - ll')$.*

We next argue that, as under the benchmark WTO case, the flexible nature of FTAs ensures that the exclusion incentive goes unexercised even when countries are able to impose positive internal tariffs on each other. To see why, suppose each country announces in favor of an FTA with both its trading partners. Part (v) of Lemma 2 informs us that a hub and spoke regime $\widehat{\langle lh \rangle}$ is a preferred regime for the hub country relative to being a member under $\widehat{\langle ll' \rangle}$. It follows then that, holding constant the announcement of the excluded small

country at $\sigma_s = \{l, l'\}$, each member of the deviating coalition (l or l') has an incentive to alter its announcement to form a separate FTA with the excluded country. As a result, the original coalitional deviation of countries l and l' is not self-enforcing and thus the lure of a hub and spoke trading arrangement ends up undermining the exclusion incentives as before.

Next, taking country s ' announcement as fixed $\{l, l'\}$, countries l and l' have no incentives to jointly deviate from their respective announcements of $\{s, l'\}$ and $\{s, l\}$ to $\{\phi, \phi\}$ and $\{\phi, \phi\}$, leading to a deviation from $\langle F \rangle$ to $\langle \Phi \rangle$. As before, the only possible self-enforcing deviation is the unilateral deviation of the either large importer from free trade. To this end, we find that, when the degree of asymmetry is sufficiently large, a large country (say l) has incentive to unilaterally deviate from its announcements $\{s, l'\}$ to an announcement leading to a hub and spoke regime where the small country or the other large country is a hub and it itself is a spoke:

$$\Delta w_l(F - \widehat{sh}) < 0 \text{ when } \theta > \theta_l(F - \widehat{sh}) \quad (25)$$

and

$$\Delta w_l(F - \widehat{l'h}) < 0 \text{ when } \theta > \theta_l(F - \widehat{l'h}) \quad (26)$$

where $\theta_l(F - sh) < \theta_l(F - l'h)$. Then, the unilateral deviation incentive that remains to be examined is the unilateral deviation of a large country (say l) from $\{s, l'\}$ to $\{\phi, \phi\}$:

$$\Delta w_l(F - \widehat{sl'}) = \Delta w_{l'}(F - \widehat{sl}) < 0 \text{ when } \theta > \theta_l(F - \widehat{sl'}) \quad (27)$$

We find that $\theta_l(F - \widehat{sl'}) < \theta_l(F - \widehat{sh})$ holds and thus the announcement profile leading to $\langle F \rangle$ is CPNE when $\theta \leq \theta_l(F - \widehat{sl'})$.

We next examine the hub and spoke regimes. From part (iv) of Lemma 3, we know that a large spoke country (say l) under \widehat{sh} and $\widehat{l'h}$ has an incentive to unilaterally deviate from its respective announcements $\{s, \phi\}$ and $\{\phi, l'\}$ to $\{\phi, \phi\}$ and $\{\phi, \phi\}$, leading to a deviation from \widehat{sh} to $\widehat{sl'}$ and from $\widehat{l'h}$ to $\widehat{sl'}$ when small country is sufficiently small and $\theta_l(\widehat{l'h} - \widehat{sl'}) < \theta_l(\widehat{sh} - \widehat{sl'})$. Moreover, when $\theta < \theta_l(\widehat{l'h} - \widehat{sl'})$, the joint announcement deviations of small and large countries leading to deviations from hub and spoke regimes to free trade are self-enforcing. Thus, the announcement profiles leading to any hub and spoke regime is never a CPNE.

The only remaining candidate for CPNE is the announcement profile leading to \widehat{sl} . We know from part (i) of Lemma 3 that no country has an incentive to unilaterally deviate from its announcement leading to a deviation from \widehat{sl} to $\langle \Phi \rangle$. Second, we know from

part $\langle vl \rangle$ that the coalitional announcement deviation leading to a deviation from $\langle sl \rangle$ and $\langle ll' \rangle$ is not self-enforcing since the common member country (l here) always has an incentive to further deviate to become the hub country, taking the announcement of its complement fixed. Third, note from the above discussion that the coalitional announcement deviation leading a deviation from $\langle sl \rangle$ to $\langle F \rangle$ is self-enforcing only when $\theta \leq \theta_l(F - \widehat{sl'})$. When $\theta > \theta_l(F - \widehat{sl'})$ holds, the non-member country l' has no incentive to engage in any coalitional announcement deviations that lead to a deviation from $\langle sl \rangle$ to $\langle sh \rangle$ or from $\langle sl \rangle$ to $\langle lh \rangle$. As a result, we argue that the announcement profile leading to $\langle sl \rangle$ is a CPNE when $\theta \geq \theta_l(F - \widehat{sl'})$.

The following result is immediate:

Proposition 4: *The equilibria of the game of unrestricted preferential liberalization under which FTA member countries coordinate their internal tariffs before setting their individually optimal external tariffs are as follows: if $\theta \leq \theta_l(F - \widehat{sl'})$, free trade is the equilibrium outcome; otherwise, it is the asymmetric FTA $\langle sl \rangle$ (or $\langle sl' \rangle$).*

Insert Figure 2

A comparison of Propositions 2 and 4 yields the following result:

Proposition 5: *(i) For $\theta \leq \theta_l(F - \widehat{sl'})$, the equilibrium outcome is free trade whether or not FTA members are required to practice free internal trade whereas over $\theta_l(F - \widehat{sl'}) < \theta \leq \theta_l(F - \widehat{sl'})$ it is the equilibrium only if FTA members are free to set positive internal tariffs on each other.*

(ii) When free trade is out of reach, i.e. when $\theta > \theta_l(F - \widehat{sl'})$, the free internal trade requirement of the WTO increases world welfare by yielding (weakly) lower global tariffs.

Insert Figure 3

The above proposition argues that, when the degree of endowment asymmetry is sufficiently small, free trade arises regardless of whether or not FTA members are required to engage in free internal trade. However, when the degree of endowment asymmetry is moderate, free trade arises *only when* FTA members are free to impose positive internal tariffs on each other. In other words, the free internal trade requirement of GATT's Article

XXIV hinders the cause of global free trade. To understand this result, we should first note that the viability of free trade is determined by the unilateral deviation incentive of one of the larger importers regardless of whether FTA members are required to engage in free internal trade or not. Due to the presence of tariff complementarity, the freedom to set positive internal tariffs leads FTA members to impose higher external tariffs which in turn makes it less attractive for one of the larger importers to opt out of global free trade – i.e. its incentive to free ride on the external trade liberalization of FTA members without having to offer any trade liberalization of its own is reduced. Finally, when free trade is out of reach, the free internal trade requirement of Article XXIV acts as a disciplining device in a tariff-ridden world and it helps protect the interest of non-member country by leading FTA members to adopt lower external tariffs. Thus, our overall message is as follows: when circumstances are such that achieving complete global free trade is not possible, the free internal trade requirement of Article XXIV increases world welfare by reducing both internal and external tariffs of FTAs but, at the same time, it also reduces the likelihood of reaching global free trade. From a practical perspective, given the multitude of ways in which countries can prevent the obtainment of global free trade, it would seem that the beneficial effects of Article XXIV’s free internal trade requirement for FTAs are likely to be of greater real-world relevance than their negative effect on the prospects of achieving global free trade.

4.3 Equilibrium agreements under tariff discrimination

While Article XXIV of GATT sanctions tariff discrimination in the form of FTAs, the MFN principle requires WTO members to treat all other members (with whom they do not have FTAs) in a non-discriminatory fashion. A natural question is whether there is a case for imposing the MFN constraint on the tariff-setting behavior of countries trading with FTAs whose members actively discriminate against them. This concern would appear to be even more acute for the case of a country that finds itself deliberately excluded from an FTA. To address these issues, we now analyze our *tariff discrimination* scenario under which a country trading with member countries of an FTA is free to set its optimal discriminatory tariffs on them as opposed to having to treat them in a non-discriminatory fashion. To isolate the role of MFN, we assume that, under the tariff discrimination scenario, FTA countries have to abide by GATT Article XXIV that obligates them to eliminate internal tariffs.

The structure of a country’s optimal discriminatory tariffs and their relationship to its

optimal MFN tariff is described in Lemma 1. As this lemma indicates, when free to tariff discriminate across its trading partners, a country sets a higher tariff on the country from whom it sources a larger volume of imports (i.e. the one whose endowment is larger). Using these optimal tariffs, we can calculate each country's welfare under various trade regimes.

Let country i 's welfare a function of the underlying trade agreement r with free internal trade between FTA members and optimal discriminatory tariffs on the part of non-member(s) be denoted by $w_i(\tilde{r})$. As before, let $\Delta w_i(\tilde{r} - \tilde{v}) \equiv w_i(\tilde{r}) - w_i(\tilde{v})$. Here, it is important to note that only the tariffs of larger importers under no agreement $\langle \tilde{\Phi} \rangle$, and the external tariffs of the non-member country under $\langle \tilde{sl} \rangle$ are different relative to the benchmark WTO case. Therefore, the welfare levels change only under those regimes.

To avoid redundancy, we directly state the following result:

Proposition 6: *The following hold under tariff discrimination:*

(i) *While the FTA between the two larger importers $\langle \tilde{ll}' \rangle$ is always Pareto-improving over no agreement $\langle \tilde{\Phi} \rangle$, the asymmetric FTA $\langle \tilde{sl} \rangle$ is Pareto-improving only when $\theta < \theta_l(\tilde{sl} - \tilde{\Phi})$.*

(ii) *While the symmetric FTA between the two larger importers $\langle \tilde{ll}' \rangle$ is more likely to be Pareto-improving over no agreement relative to the WTO-consistent benchmark case, the opposite is true for the asymmetric FTA $\langle \tilde{sl} \rangle$.*

Following the analysis under the previous two scenarios, we can derive equilibrium outcomes under tariff discrimination and compare them with those obtained under the WTO-consistent benchmark case. For brevity, we directly state the main result (proof is in the appendix):

Proposition 7: *For $\theta \leq \theta_l(F - \tilde{sl}')$ free trade is the equilibrium outcome under both tariff discrimination and the WTO-consistent benchmark. When $\theta_l(F - \tilde{sl}') < \theta \leq \theta_l(F - sl')$, the WTO-consistent benchmark yields free trade whereas tariff discrimination yields an asymmetric FTA (i.e. $\langle \tilde{sl} \rangle$ or $\langle \tilde{sl}' \rangle$). Finally, when $\theta > \theta_l(F - sl')$, free trade is out of reach under both scenarios but world welfare is lower under tariff discrimination.*

Insert Figure 4

The above analysis provides strong support for the MFN principle. Not only does the MFN constraint make it easier to achieve global free trade, it also delivers a welfare-superior outcome when global free trade cannot be reached due to the high degree of asymmetry

in the underlying economic environment. Yet, the above analysis cannot shed light on the effects of MFN when a bilateral agreement emerges because members deliberately exclude the third country since such an incentive on the part of members only arises when they can coordinate their external tariffs. Accordingly, in the final part of the paper, we discuss the case where the bilateral trade agreement is a customs union (CU) as opposed to an FTA.

4.4 Agreements with external tariff coordination

Suppose the PTA under consideration is a CU under which member countries coordinate both internal and external tariffs as opposed to an FTA wherein they only coordinate internal tariffs.

Now consider the endogenous formation of CUs. 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 S_{iu} consists of four possible announcements:

$$S_{iu} = \{\{\phi, \phi\}, \{j^u, \phi\}, \{\phi, k^u\}, \{j^u, k^u\}\} \quad (28)$$

where $\{\phi, \phi\}$ denotes an announcement in favor of no CU with either trading partners, $\{j^u, \phi\}$ in favor of a CU with only country j ; $\{\phi, k^u\}$ in favor of a CU with only country k ; and $\{j^u, k^u\}$ in favor of a CU that includes both its trading partners.

The mapping between various announcements profiles and the CUs that can arise is as follows: (i) when no two announcements match or the only matching announcements are $\{\phi, \phi\}$ we obtain no agreement $\langle \Phi \rangle$; (ii) a CU between countries i and j denoted by $\langle ij^u \rangle$ is formed if two countries announce each others' names and there is no other matching announcement: i.e. $\langle ij^u \rangle$ is formed if $j^u \in \sigma_i$, $i^u \in \sigma_j$ and both (a) $k^u \notin \sigma_i$ and/or $i^u \notin \sigma_k$ and (b) $k^u \notin \sigma_j$ and/or $j^u \notin \sigma_k$ hold; (iv) if all countries' announce each other's names, global trade agreement $\langle F \rangle$ obtains. Note that a hub and spoke type trading regime cannot arise under the CU game due to the fact that CU members coordinate their external tariffs.

Under no agreement $\langle \Phi \rangle$, the optimal MFN and discriminatory tariffs in the CU game are the same as that in the FTA game. However, as indicated earlier, CU members find it optimal to engage in free internal trade since their joint welfare is strictly decreasing in each of the internal tariffs. Thus, the *free internal trade requirement of Article XXIV does not bind for CU members*. Furthermore, recall that CU members impose higher external tariffs than FTA members: $t_{zk}^u > t_{zk}^*$ where $z = i, j$ and i and j are the two member countries. Since CU members engage in (i) more internal trade liberalization and (ii) less external

liberalization relative to FTA members, one can expect that the free-riding incentive is weaker whereas the exclusion incentive is more pronounced under the CU game relative to the FTA game.

Next, we examine the role of MFN under Assumption 1. We first state the following lemma:

Lemma 5A: *Regardless of whether the non-member country abides by MFN or not, the following hold:*

(i) *Each country prefers to form a bilateral CU with the larger importer relative to the smaller one.*

(ii) *The smaller importer (s) has no incentive to unilaterally or jointly deviate from any agreement.*

(iii) *A country is worse off as a non-member than as a CU member as well as relative to no agreement or free trade.*

It is immediate from the third part of the lemma that, a *bilateral CU fails to be Pareto-improving over no agreement*. By contrast, Propositions 1 and 3 provide conditions under which a bilateral FTA is necessarily Pareto-improving. Furthermore, the above lemma informs us whether free trade obtains in equilibrium or not critically depends on the incentive that the two larger importers have to exclude the smaller country. It turns out this exclusion incentive arises only when the two larger importers are sufficiently larger than the third country: $\theta > \theta_i(F - ll^u)$. Here it is important to note that, since the member countries under $\langle ll^u \rangle$ are *symmetric*, the freedom to tariff discriminate is irrelevant from the viewpoint of the non-member country. Thus, free trade arises in equilibrium whenever $\theta \leq \theta_i(F - ll^u)$ whether or not countries are free to violate MFN.

But what if parameters are such that free trade does not arise in equilibrium? Does MFN matter then? We next show that it does not. To this end, consider first the announcement profiles leading to no agreement $\langle \Phi \rangle$. Starting from $\langle \Phi \rangle$, countries l and l' have an incentive to jointly alter their announcements to form $\langle ll^u \rangle$. Since this deviation is self-enforcing, $\langle \Phi \rangle$ cannot arise in equilibrium. Similarly, based on parts (i) and (iii) of the Lemma 5A, we directly argue that countries l and l' have an incentive to jointly alter their announcement profiles so that $\langle sl^u \rangle$ is replaced by $\langle ll^u \rangle$ and this deviation is self-enforcing since neither country l nor l' has an incentive to unilaterally deviate further since doing so leads to $\langle \Phi \rangle$. As a result, the only remaining candidate for equilibrium is the announcement profile leading to $\langle ll^u \rangle$ under which the two member countries are symmetric. Based on the

above discussion, it is immediate that $\langle ll^u \rangle$ is an equilibrium whenever $\theta \geq \theta_l(F - ll^u)$ and this result obtains regardless of whether the non-member country is free to tariff discriminate or not. The following proposition summarizes our findings.

Proposition 8: *Regardless of whether the non-member country has to follow MFN or not, free trade obtains in equilibrium whenever $\theta \leq \theta_l(F - ll^u)$ whereas the custom union between the two larger countries CU $\langle ll^u \rangle$ arises otherwise.*

The analysis above has shown that, given Assumption 1, imposing the MFN requirement has no consequences for equilibrium outcomes (or welfare) of the CU game. But what if the pattern of endowment asymmetry differs from that given under Assumption 1? To answer this question and to highlight the role of MFN, we now consider a slightly different structure of endowment asymmetry and show that MFN indeed matters when the CU involves two asymmetric countries:

Assumption 2: *Country l is a larger importer than countries s and s' : $e_s = e_{s'} = \theta e \geq e_l = e$ where $1 \leq \theta \leq 5/4$.¹⁸*

We first state the following lemma that explains the preferences of asymmetric countries over trade regimes:

Lemma 5B: *Suppose Assumption 2 describes the underlying endowment structure of the world economy. Then, the following hold:*

(i) *Each country prefers to form a bilateral CU with a larger importer relative to a smaller one: $\Delta w_i(ij^u - ik^u) > 0$ iff $e_k > e_j$.*

(ii) *The smaller importers (s and s') neither have a unilateral incentive to break up any agreement nor a joint incentive (with the other small country) to deviate from free trade.*

(iii) *The non-member country is worse off relative to no agreement as well as relative to being the member of a bilateral CU: $\Delta w_k(ij^u - \Phi) < 0$ and $\Delta w_k(ij^u - ik^u) < 0$.*

It is worth noting that, even under Assumption 2, a bilateral CU is not Pareto-improving over no agreement.

We are now ready to derive equilibrium outcomes under Assumption 2. First, let us consider whether free trade as a candidate for an equilibrium outcome. Note from parts

¹⁸As in the FTA game, the qualitative nature of our results is robust to a scenario where all three countries are fully asymmetric, such as when $e_s = \theta_s e \geq e_m = \theta_m e \geq e_l = e$.

(*ii*) and (*iii*) that the incentive for free riding on the trade liberalization of others is rather limited in the CU game and the only country that would engage in a unilateral deviation from free trade is country l and this happens only when the degree of asymmetry is sufficiently large: $\theta > \theta_l(F - ss'^u)$. Similarly, it is immediate from part (*iii*) that, if there exists a coalitional deviation from free trade, it must involve one of the smaller countries (say s) and the large importer l . Countries s and l have an incentive to jointly deviate from their respective announcements $\{s'^u, l^u\}$ and $\{s^u, s'^u\}$ to $\{\phi, l^u\}$ and $\{s^u, \phi\}$ in order to exclude country s' when $\theta > \theta_l(F - sl^u)$ and since neither has an incentive to further deviate from their announcements, their initial joint deviation is self-enforcing. Note also that since $\theta_l(F - sl^u) < \theta_l(F - ss'^u)$, the announcement profile leading to supporting free trade is an equilibrium whenever $\theta \leq \theta_l(F - sl^u)$. Here it is important to note that while the exclusion incentive goes unexercised in the FTA formation game due to the relatively flexible nature of FTAs, such is not the case in the CU game. In the FTA game, if two countries (i and j) jointly exclude the third country from free trade by forming a bilateral FTA then each member has an incentive to sign an independent FTA with the excluded country thereby making itself a hub. The ability to act on this incentive acts as a deterrent for the other initially deviating country (say country j) 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 excluded country without the consent of its CU partner.

What if $\theta > \theta_l(F - sl^u)$ so that free trade fails to arise in equilibrium? It is easy to dismiss the announcement profiles leading to no agreement $\langle \Phi \rangle$ as candidates for an equilibrium. This is because countries s and s' have an incentive to jointly deviate from their respective announcements $\{\phi, \phi\}$ and $\{\phi, \phi\}$ to $\{\phi, s'^u\}$ and $\{s^u, \phi\}$ to form the CU $\langle ss'^u \rangle$. Furthermore, since this deviation is self-enforcing, no agreement cannot be an equilibrium. Similarly, based on parts (*i*) and (*iii*) of Lemma 5B, we can argue that countries s and l always have incentives to jointly deviate from their announcement profiles to the ones that lead to a deviation from $\langle ss'^u \rangle$ to $\langle sl^u \rangle$ and this deviation is self-enforcing since either country s or l has no unilateral incentive to further deviate further to any announcement that leads to $\langle \Phi \rangle$. As a result, the only remaining candidate for equilibria are announcement profiles leading to $\langle sl^u \rangle$. Based on the above discussion, it is immediate that $\langle sl^u \rangle$ is an equilibrium whenever $\theta \geq \theta_s(F - sl^u)$.

Next, we examine the implications of requiring MFN on the part of non-member countries that are trading with CU members. As indicated in Proposition 1, the larger member

country of $\langle sl^u \rangle$ is better off under tariff discrimination relative to MFN while the small member is worse off. As a result, the exclusion incentive of the larger country is stronger under tariff discrimination relative to MFN while that of the smaller country is weaker. In fact, when the non-member is free to tariff discriminate, the smaller country has no incentive exclude the other smaller country: $\Delta w_s(F - \widetilde{sl^u}) > 0$. Thus, it is the unilateral deviation incentive of the larger country that determines the equilibrium condition for free trade: it turns out that when tariff discrimination is possible, free trade obtains in equilibrium iff $\theta \leq \theta_l(F - \widetilde{ss'^u})$.¹⁹ We can now state:

Proposition 9: *Given Assumption 2, the following hold:*

- (i) *Free trade obtains in equilibrium regardless of whether the non-member country is required to follow MFN whenever $\theta \leq \theta_l(F - sl^u)$.*
- (ii) *When $\theta_l(F - sl^u) < \theta \leq \theta_l(F - \widetilde{ss'^u})$, free trade obtains in equilibrium only if the non-member country is free to tariff discriminate.*
- (iii) *When $\theta > \theta_l(F - \widetilde{ss'^u})$ free trade is infeasible both under tariff discrimination and MFN but imposing MFN increases world welfare by eliminating discrimination against the larger country.*

Insert Figure 5

As noted before, by making tariff discrimination infeasible, the non-discrimination constraint of MFN reduces the benefit of being an outsider and thus lowers the unilateral incentive for opting out of trade liberalization. Therefore, in the FTA game, by increasing the likelihood of each country voluntarily choosing to enter into international trade agreements, the MFN principle can act as a catalyst for the cause of global free trade. However, when one country is *deliberately excluded* by the other two because they prefer a bilateral trade agreement to a multilateral one (as can happen in the CU game), the pro-liberalization effect of MFN is weaker. As a result, whether MFN complements Article XXIV in achieving global free trade depends on the nature of the trade agreement – i.e. whether it is an FTA or a CU. Finally, irrespective of the nature of the PTA in question, MFN is welfare improving when free trade is infeasible (i.e. fails to obtain whether or not an excluded country has to follow MFN). Thus, while whether WTO requirements specified in Article I and Article XXIV increase the likelihood of global free trade depends on the nature of PTAs, they are necessarily welfare improving in a tariff-ridden world.

¹⁹Note that since member countries are symmetric under $\langle ss'^u \rangle$, $\theta_l(F - ss'^u) = \theta_l(F - \widetilde{ss'^u})$.

5 Conclusion

In this paper, we have provided an analysis of the interaction between two core GATT rules governing trade liberalization at the WTO: Article 1 (MFN) that obligates members to adopt non-discriminatory trade policies toward one another and Article XXIV that lays down conditions that countries entering into PTAs are required to follow. The two most important conditions imposed by Article XXIV on PTA members are: (a) members should eliminate trade restrictions on *substantially all trade* between themselves and (b) refrain from raising trade restrictions on non-member countries. In our competing exporters model, due to the existence of tariff complementarity, the second requirement of Article XXIV turns out to be non-binding and the fate of the outside countries ends up depending upon whether or not PTA members are required to fully liberalize their internal trade.

To draw out the implications of requiring FTA members to eliminate tariffs on one another, we derive and contrast optimal tariffs and equilibrium trade agreements under two scenarios: under the WTO-consistent scenario, members are required to engage in free internal trade whereas under unrestricted preferential liberalization scenario members are free to impose non-zero internal tariffs on one another. Under both scenarios, the non-member is required to follow MFN. A comparison of these scenarios delivers several new insights. First, we show that whether members of a PTA have an incentive to maintain positive *internal* tariffs on one another or not depends upon how they set their *external* tariffs. If PTA members set external tariffs independently, as they do in an FTA, they benefit from not eliminating tariffs on one another since doing so commits them to higher external tariffs. On the other hand, when external tariffs are coordinated – as they are under a CU – PTA members find it optimal to eliminate internal tariffs so that the restriction on internal tariffs imposed by Article XXIV becomes moot.

Our second major result is rather surprising: requiring FTA members to eliminate internal tariffs *benefits* the non-member since it leads to *lower* external tariffs on the part of FTA members. In other words, it is the Article XXIV requirement of free internal trade amongst FTAs that ends up protecting the interest of the non-member as opposed to the restriction on external tariffs imposed on FTA members. Indeed, we show that the free internal trade requirement can make it more likely that an FTA between two countries is Pareto-improving relative to a scenario where no trade agreements exist.

Since our analysis derives equilibrium agreements in a game in which all countries are free to form trade agreements with one another, we are able to speak to the consequences of the free internal trade requirement of Article XXIV for the likelihood of achieving global

free trade. Our major finding here is that this requirement on internal tariffs of FTAs makes it *harder* to achieve global free trade by limiting the negative impact of an FTA on the non-member country: due to tariff complementarity, lower internal tariffs within an FTA also imply lower external tariffs. By not entering into a trade agreement with FTA members, the non-member country remains free to impose its optimal import tariffs on them while itself facing relatively lower tariffs in their markets. Thus, it is possible that the free internal trade requirement of Article XXIV facilitates some degree of free-riding in the WTO system by making it possible for non-member countries to benefit from reductions in external tariffs of FTA members without having to reciprocate with tariff cuts of their own. Thus, while the free internal trade requirement of Article XXIV reduces the likelihood of obtaining global free trade, it also increases welfare by lowering tariffs world-wide when global free trade is simply out of reach. Finally, we show that the case for requiring MFN on the part of countries trading with CUs from which they have been deliberately excluded is relatively weaker than the case for MFN on the part of those countries that have themselves chosen not to join FTAs.

Finally, while we have examined the implications of the free internal trade requirement facing PTAs for both FTAs and CUs, our approach has abstracted from the endogenous choice between these two types of PTAs. This is an important question that we leave for future research.

6 Appendix

In this Appendix we provide all supporting calculations and proofs.

6.1 Supporting calculations

We begin by reporting welfare levels as functions of an arbitrary tariff vectors. Then, we report the optimal tariffs under each trade regime. Using the welfare and tariff levels reported below, we can easily obtain the formulae for optimum welfare levels under all possible regimes. Lemmas 1, 2, 3, 4, 5 and the various inequalities reported in the main text follow from a direct application of the relevant formulae.

6.1.1 Welfare levels

We report welfare levels for country i under a trade regime r as a function of an arbitrary tariff vector $\mathbf{t}(r)$ where $\mathbf{t}(r) = (t_{ij}(r), t_{ik}(r))$:

$$w_i(r) = \sum_z CS_i^z(r) + \sum_z PS_i^z(r) + TR_i(r)$$

where

$$\sum_z CS_i^z(r) = \frac{1}{2} \left[\left(\frac{e_j + e_k - t_{ij}(r) - t_{ik}(r)}{3} \right)^2 + \left(\frac{e_i + e_k + 2t_{ji}(r) - t_{jk}(r)}{3} \right)^2 + \left(\frac{e_i + e_j + 2t_{ki}(r) - t_{kj}(r)}{3} \right)^2 \right]$$

$$\sum_z PS_i^z(r) = \frac{e_i[6\alpha - 2e_i - e_j - e_k + t_{jk}(r) + t_{kj}(r) - 2t_{ji}(r) - 2t_{ki}(r)]}{3}$$

and

$$TR_i(r) = \frac{t_{ij}(r)[2e_j - e_k + t_{ik}(r) - 2t_{ij}(r)]}{3} + \frac{t_{ik}(r)[2e_k - e_j + t_{ij}(r) - 2t_{ik}(r)]}{3}.$$

6.1.2 Optimal Tariffs

Next, we report the optimal tariffs under each regime and provide supporting calculations for our tariff discussion in the text. Country i 's optimal MFN tariff is

$$t_i^\phi \equiv \text{Arg max } w_i(\Phi) = \frac{e_j + e_k}{8} \quad (29)$$

First, consider a scenario where countries are free to discriminate. Under no agreement $\langle \Phi \rangle$ (or under an FTA for a non-member country), (t_{ij}^*, t_{ik}^*) solve $\arg \max w_i(\Phi)$:

$$t_{ij}(\Phi) = \frac{3e_j - e_k}{8} \quad \text{and} \quad t_{ik}(\Phi) = \frac{3e_k - e_j}{8}$$

The following is immediate:

$$t_{ij}^*(\Phi) - t_{ik}^*(\Phi) = \frac{(e_j - e_k)}{2} > 0 \text{ iff } e_j > e_k$$

Relative to the case of MFN, countries impose higher discriminatory tariff on the imports from the country with larger endowment:

$$t_i^\phi - t_{ij}^*(\Phi) = \frac{2(e_k - e_j)}{8} < 0 \text{ iff } e_j > e_k$$

Next, we examine the FTA member tariffs. First we show that, holding everything else constant, non-member country loses as internal tariffs of an FTA decline:

$$\frac{\partial w_k(ij)}{\partial \tau_{ij}} = \frac{2(e_k - \tau_{ij}) - (e_j - \tau_{ij})}{9} > 0$$

Suppose now that external tariffs are optimally chosen. Then, we find the following optimum external tariff as a function of internal tariff between member countries:

$$t_{ik}(ij) = \frac{5e_k - 4e_j + 7\tau_{ij}}{11}$$

Note that the tariff complementarity holds:

$$\frac{\partial t_{ik}(ij)}{\partial \tau_{ij}} = \frac{7}{11} > 0$$

We next show that, when external tariffs are optimally chosen by FTA members, we obtain:

$$\frac{\partial w_k(ij)}{\partial \tau_{ij}} = -\frac{4e_k - t_{ij} - e_j}{121} < 0$$

If countries could coordinate internal tariffs before setting their individually optimum external tariffs, FTA members can partially internalize the effects of their tariffs on one another: $(\tau_{ij}, \tau_{ji}) \equiv \arg \max [w_i(ij) + w_j(ij)]$:

$$\tau_{ij} = \frac{3e_j - e_k}{63} > 0$$

Then the optimal external tariff is as follows:

$$t_{ik}(ij) = \frac{4e_k - 3e_j}{9} \tag{30}$$

Under free internal trade, the optimum external tariff under an FTA (and the optimal spoke's tariff under a hub and spoke regime) is immediate:

$$t_{ik}(ij) = t_{ik}(jh) = \frac{5e_k - 4e_j}{11}$$

Under a CU, we found the following optimum external tariffs as a function of the internal tariffs:

$$t_{ik}(ij^u) = \frac{2e_k - e_j}{5} + \frac{\tau_{ij}}{2} \tag{31}$$

Note that, while it is weaker relative to an FTA game, the tariff complementarity still holds:

$$\frac{\partial t_{ik}(ij^u)}{\partial \tau_{ij}} = \frac{1}{2} > 0$$

We find that it is optimum for CU members to eliminate internal tariffs:

$$\frac{\partial[w_i(ij^u) + w_j(ij^u)]}{\partial\tau_{ij}} = -\frac{\tau_{ij}}{2} < 0$$

As a result, the following jointly optimal external tariffs under $\langle ij^u \rangle$ obtain:

$$t_{ik}(ij^u) = \frac{2e_k - e_j}{5} \quad (32)$$

Note that we obtain higher external tariffs under a CU relative to an FTA: $t_{ik}(ij^u) > t_{ik}(ij)$.

6.2 Proofs of Lemmas and Propositions

Note that the proof of Lemma 1 is immediate from the optimal tariff discussion above.

Proof of Lemma 2

Using the above welfare formulae (as functions of an arbitrary tariff vectors) and plugging the above optimum tariffs into them, it is straightforward to show the following inequalities:

Part (i): $\Delta w_l(ll' - sl) > 0$ for all $1 \leq \theta \leq 5/4$.

Part (ii): $\Delta w_s(sl - \Phi) > 0$, $\Delta w_s(sh - sl) > 0$, and $\Delta w_s(F - lh) > 0$ hold for all $1 \leq \theta \leq 5/4$ while $\Delta w_s(lh - ll') > 0$ only when $\theta > \theta_s(lh - ll') \cong 1.03$.

Part (iii): $\Delta w_l(lh - sl) < 0$ and $\Delta w_l(lh - sl) < 0$ for all $1 \leq \theta \leq 5/4$ while $\Delta w_s(lh - ll') > 0$ only when $\theta > \theta_s(lh - ll') \cong 1.03$.

Part (iv): $\Delta w_i(ih - \Phi) > 0$, $\Delta w_i(F - ih) < 0$ and $\Delta w_i(ih - ij) > 0$ for all for all $1 \leq \theta \leq 5/4$ and $i = s, l, l'$.

Proof of Proposition 1

Using the above welfare formulae (as functions of an arbitrary tariff vectors) and plugging the above optimum tariffs into them, it is straightforward to show that $\Delta w_s(sl - \Phi) > 0$ holds for all $1 \leq \theta \leq 5/4$ while $\Delta w_l(sl - \Phi) > 0$ only when $\theta < \theta_l(sl - \Phi) \cong 1.24$. Similarly, we obtain $\Delta w_l(ll' - \Phi) > 0$ for all $1 \leq \theta \leq 5/4$ while $\Delta w_s(ll' - \Phi) > 0$ only when $\theta < \theta_s(ll' - \Phi) \cong 1.09$.

Proof of Proposition 2

Using the results from Lemma 2, the discussion in the main text and the following inequalities, it is straightforward to prove Proposition 2:

- $\Delta w_l(F - ll') < 0$ when $\theta > \theta_l(F - ll') \cong 1.085$;
- $\Delta w_l(F - l'h) = \Delta w_{l'}(F - lh) < 0$ when $\theta > \theta_l(F - l'h) \cong 1.18$;
- $\Delta w_l(F - sl') = \Delta w_{l'}(F - sl) < 0$ when $\theta > \theta_l(F - sl') \cong 1.181$;
- $\Delta w_s(lh - ll') > 0$ when $\theta > \theta_s(lh - ll') \cong 1.03$.

Proof of Lemma 3

Using the above welfare formulae (as functions of an arbitrary tariff vectors) and plugging the above optimum tariffs (without free internal trade requirement) into them, it is straightforward to show the following inequalities:

Part (i): $\Delta w_s(\widehat{sl} - \Phi) > 0$, $\Delta w_l(\widehat{sl} - \Phi) > 0$, $\Delta w_l(\widehat{ll}' - \Phi) > 0$ for all $1 \leq \theta \leq 5/4$.

Part (ii): $\Delta w_l(\widehat{ll}' - \widehat{sl}) > 0$ for all $1 \leq \theta \leq 5/4$.

Part (iii): $\Delta w_s(\widehat{sl} - \Phi) > 0$, $\Delta w_s(\widehat{sh} - \widehat{sl}) > 0$, $\Delta w_s(F - \widehat{lh}) > 0$ and $\Delta w_s(lh - ll') > 0$ hold for all $1 \leq \theta \leq 5/4$.

Part (iv): $\Delta w_l(\widehat{l'h} - \widehat{sl'}) < 0$ when $\theta > \theta_l(\widehat{l'h} - \widehat{sl'}) \cong 1.029$ and $\Delta w_l(\widehat{sh} - \widehat{sl'}) < 0$ when $\theta > \theta_l(\widehat{l'h} - \widehat{sl'}) \cong 1.037$.

Part (iv): $\Delta w_i(\widehat{ih} - \Phi) > 0$ and $\Delta w_i(\widehat{ih} - \widehat{ij}) > 0$ for all for all $1 \leq \theta \leq 5/4$ and $i = s, l, l'$.

Proof of Proposition 3

Along with the first part of Lemma 3, using the above welfare formulae (as functions of an arbitrary tariff vectors) and plugging the above optimum tariffs (without free internal trade requirement) into them, it is straightforward to show that $\Delta w_{l'}(\widehat{sl} - \Phi) > 0$ holds for all $1 \leq \theta \leq 5/4$ while $\Delta w_s(\widehat{ll}' - \Phi) > 0$ only when $\theta < \theta_s(\widehat{ll}' - \Phi) \cong 1.076$.

Proof of Lemma 4

Using the above welfare formulae (as functions of an arbitrary tariff vectors) and plugging the above optimum tariffs (with and without free internal trade requirement) into them, it is straightforward to show that $\theta_l(F - \widehat{ll}') \cong 1.082 < \theta_l(F - ll') \cong 1.085$.

Proof of Proposition 4

Using the results from Lemmas 3 and 4, the discussion in the main text and the following inequalities, it is straightforward to prove Proposition 4:

- $\Delta w_l(F - \widehat{ll}') < 0$ when $\theta > \theta_l(F - \widehat{ll}') \cong 1.085$;

- $\Delta w_l(F - \widehat{l'h}) = \Delta w_{l'}(F - \widehat{l'h}) < 0$ when $\theta > \theta_l(F - \widehat{l'h}) \cong 1.138$;
- $\Delta w_l(F - \widehat{s'h}) = \Delta w_l(F - \widehat{s'h}) < 0$ when $\theta > \theta_l(F - \widehat{s'h}) \cong 1.130$;
- $\Delta w_l(F - \widehat{s'l'}) = \Delta w_{l'}(F - \widehat{s'l'}) < 0$ when $\theta > \theta_l(F - \widehat{s'l'}) \cong 1.037$;

Proof of Proposition 5

The proof is immediate from the proofs of Propositions 2 and 4.

Proof of Proposition 6

Using the above welfare formulae (as functions of an arbitrary tariff vectors) and plugging the above optimum tariffs under discrimination into them, it is straightforward to show the following inequalities:

- $\Delta w_s(\widetilde{s'l} - \widetilde{\Phi}) > 0$ and $\Delta w_{l'}(\widetilde{s'l} - \widetilde{\Phi}) > 0$ for all $1 \leq \theta \leq 5/4$ while $\Delta w_l(\widetilde{s'l} - \widetilde{\Phi}) > 0$ holds only when $\theta < \theta_l(\widetilde{s'l} - \widetilde{\Phi}) \cong 1.159$.
- $\Delta w_s(l'l' - \widetilde{\Phi}) > 0$ and $\Delta w_l(l'l' - \widetilde{\Phi}) = \Delta w_{l'}(l'l' - \widetilde{\Phi}) > 0$ for all $1 \leq \theta \leq 5/4$.
- $\theta_l(\widetilde{s'l} - \widetilde{\Phi}) \cong 1.159 < \theta_l(s'l - \Phi) \cong 1.24$.

Proof of Proposition 7

Using the above welfare formulae (as functions of an arbitrary tariff vectors) and plugging the above optimum tariffs under discrimination into them, we can make the following discussion. First, we show when the announcement profiles leading to free trade is a CPNE. We can argue the following three results that follows directly the benchmark case. First, countries l and l' always have incentives to jointly deviate from their respective announcements $\{\phi, \phi\}$ and $\{\phi, \phi\}$ to $\{\phi, l'\}$ and $\{\phi, l\}$ and form $\langle ll' \rangle$. Since this deviation is self-enforcing, $\langle \Phi \rangle$ is never a CPNE. Second, we know from Lemma 2 that a large spoke country (say l) under $\langle sh \rangle$ and $\langle l'h \rangle$ always has an incentive to unilaterally deviate from its respective announcements $\{s, \phi\}$ and $\{\phi, l'\}$ to $\{\phi, \phi\}$ and $\{\phi, \phi\}$, leading to a deviation from $\langle sh \rangle$ to $\langle \widetilde{s'l'} \rangle$ and from $\langle l'h \rangle$ to $\langle \widetilde{s'l'} \rangle$. Since unilateral deviations are self-enforcing, the announcement profiles leading to any hub and spoke regime is never a CPNE. Third, the coalitional announcement deviation from $\langle ll' \rangle$ to $\langle F \rangle$ happens when $\theta \leq \theta_l(F - ll') \cong 1.19$ and it is self-enforcing only when $\theta \leq \theta_l(F - \widetilde{s'l'}) \cong 1.075$. Second, we show that, when $\theta > \theta_s(lh - ll') \cong 1.033$ holds, country s and either of the large member countries (say l) have incentives to jointly deviate from their respective announcements $\{\phi, \phi\}$ and $\{\phi, l'\}$ to $\{l, \phi\}$ and $\{s, l'\}$, leading to a deviation from $\langle ll' \rangle$ to $\langle lh \rangle$ and this deviation is self-enforcing. Since $\theta_s(lh - ll') < \theta_l(F - \widetilde{s'l'})$, these self-enforcing announcement deviations cover

the entire parameter space and thus the announcement profile leading to $\langle ll' \rangle$ is never a CPNE.

Based on the above discussion, the only possible announcement profiles that can be CPNE are the ones leading to $\langle F \rangle$ and $\langle \widetilde{sl} \rangle$. As in the benchmark case, we can immediately argue that the unilateral deviation of the either large importer from free trade announcements is critical for the CPNE condition. To this end, we find that country l unilaterally deviates from $\{s, l'\}$ to $\{\phi, \phi\}$:

$$\Delta w_l(F - \widetilde{sl}') = \Delta w_{l'}(F - \widetilde{sl}) < 0 \text{ when } \theta > \theta_l(F - \widetilde{sl}') \cong 1.075 \quad (33)$$

We find that $\theta_l(F - \widetilde{sl}') < \theta_l(F - l'h) \cong 1.138$ holds and thus the announcement profile leading to $\langle F \rangle$ is CPNE when $\theta \leq \theta_l(F - \widetilde{sl}')$.

Finally consider $\langle \widetilde{sl} \rangle$. We find that, when $\theta > \theta_l(\widetilde{sl} - \widetilde{\Phi}) \cong 1.159$, country l has an incentive to unilaterally deviate from its announcement $\{s, \phi\}$ to $\{\phi, \phi\}$ leading to a deviation from $\langle sl \rangle$ to $\langle \Phi \rangle$ and as indicated in Proposition 6, we have $\theta_l(\widetilde{sl} - \widetilde{\Phi}) < \theta_l(sl - \Phi)$. Second, we know from part (iv) of Lemma 2 that the coalitional announcement deviation leading to a deviation from $\langle \widetilde{sl} \rangle$ to $\langle ll' \rangle$ is not self-enforcing since the common member country (l here) always has an incentive to further deviate to become the hub country, taking the announcement of its complement fixed. Third, note from the above discussion that the coalitional announcement deviation leading a deviation from $\langle \widetilde{sl} \rangle$ to $\langle F \rangle$ is self-enforcing only when $\theta \leq \theta_l(F - \widetilde{sl}')$. Finally, country l' has no incentive to engage in any coalitional announcement deviations that lead to a deviation from $\langle \widetilde{sl} \rangle$ to $\langle sh \rangle$ or from $\langle \widetilde{sl} \rangle$ to $\langle lh \rangle$. As a result, we argue that the announcement profile leading to $\langle sl \rangle$ is a CPNE when $\theta_l(F - \widetilde{sl}') \leq \theta \leq \theta_l(\widetilde{sl} - \widetilde{\Phi})$.

Proof of Lemma 5A

Using the above welfare formulae (as functions of an arbitrary tariff vectors) and plugging the above optimum tariffs under CUs into them, it is straightforward to show the following inequalities:

Part (i): $\Delta w_l(ll'^u - sl^u) > 0$ for all $1 \leq \theta \leq 5/4$.

Part (ii): $\Delta w_s(sl^u - \Phi) > 0$, $\Delta w_s(F - sl^u) > 0$, and $\Delta w_s(F - ll'^u) > 0$ hold for all $1 \leq \theta \leq 5/4$.

Part (iii): $\Delta w_s(ll'^u - \Phi) < 0$, $\Delta w_l(sl'^u - \Phi) < 0$, $\Delta w_s(F - ll'^u) > 0$ and $\Delta w_l(F - sl'^u) > 0$ for all $1 \leq \theta \leq 5/4$.

Proof of Proposition 8

Using Lemma 5A and the following inequality, the proof is complete: $\Delta w_l(F - ll^u) < 0$ when $\theta > \theta_l(F - ll^u) \cong 1.023$.

Proofs of Lemma 5B

Part (i): $\Delta w_s(sl^u - ss^u) > 0$ for all $1 \leq \theta \leq 5/4$.

Part (ii): $\Delta w_s(sl^u - \Phi) > 0$, $\Delta w_s(ss^u - \Phi) > 0$, $\Delta w_s(F - ss^u) > 0$ and $\Delta w_s(F - s'l^u) > 0$ hold for all $1 \leq \theta \leq 5/4$.

Part (iii) $\Delta w_s(s'l^u - \Phi) < 0$, $\Delta w_l(ss^u - \Phi) < 0$, $\Delta w_s(s'l^u - sl^u) < 0$, $\Delta w_s(s'l^u - ss^u) < 0$, and $\Delta w_l(ss^u - sl^u) < 0$ hold for all $1 \leq \theta \leq 5/4$.

Proof of Proposition 9

Using Lemma 5B and the following inequalities, the proof is complete. We have $\Delta w_l(F - ss^u) < 0$ when $\theta > \theta_l(F - ss^u) \cong 1.231$, $\Delta w_s(F - sl^u) < 0$ when $\theta > \theta_s(F - sl^u) \cong 1.049$ and $\Delta w_l(F - sl^u) < 0$ when $\theta > \theta_l(F - sl^u) \cong 1.045$. Under discrimination, $\Delta w_s(F - \widetilde{sl^u}) > 0$ holds for all $1 \leq \theta \leq 5/4$. Moreover, $\Delta w_l(F - \widetilde{ss^u}) < 0$ when $\theta > \theta_l(F - \widetilde{ss^u}) \cong 1.231$.

References

- [1] Aghion, Philippe, Pol Antràs, and Elhanan Helpman 2007. “Negotiating Free Trade.” *Journal of International Economics* 73: 1–30.
- [2] Bagwell, Kyle and Robert W. Staiger, 1999a. “Regionalism and Multilateral Tariff Cooperation.” In John Piggott and Allan Woodland, eds, *International Trade Policy and the Pacific Rim*, London: MacMillan.
- [3] Bagwell, Kyle and Robert W. Staiger, 1999b. “An Economic Theory of GATT.” *American Economic Review* 89(1): 215–48.
- [4] Bagwell, Kyle and Robert W. Staiger, 2010. “Backward Stealing and Forward Manipulation in the WTO.” *Journal of International Economics* 82(1): 49–62.
- [5] Bagwell, Kyle, Chad P. Bown, and Robert W. Staiger, 2016. “Is the WTO Passé?” *Journal of Economic Literature* 54(4): 1125–1231.
- [6] Bernheim, Douglas B., Bezalel Peleg and Michael Whinston, 1987. “Coalition-proof Nash Equilibria I. Concepts.” *Journal of Economic Theory* 42: 1–12.

- [7] Cadot, Olivier, Jaime De Melo, and Marcelo Olarreaga. 1999. "Regional integration and lobbying for tariffs against nonmembers." *International Economic Review* 40(3): 635-658.
- [8] Caplin, Andrew, and Kala Krishna, 1998. "Tariffs and the Most Favored Nation Clause: A Game Theoretic Approach." *Seoul Journal of Economics* 1, 267-289.
- [9] Choi, Jay P., 1995. "Optimal tariffs and the choice of technology: discriminatory tariffs vs. the most favored nation clause." *Journal of International Economics* 38: 143-160.
- [10] Damuri, Yose Rizal, 2012. "How Preferential Are Preferential Trade Agreements? Analysis of Product Exclusions in PTAs." Graduate Institute Centre for Trade and Economic Integration Working Paper, 2012-3.
- [11] Freund, Caroline 2000. "Multilateralism and the endogenous formation of preferential trade agreements." *Journal of International Economics* 52(2): 359-376.
- [12] Freund, Caroline L., and Emanuel Ornelas, 2010. "Regional Trade Agreements." *Annual Review of Economics* 2: 139-166.
- [13] Furusawa, Taiji and Hideo Konishi, 2007. "Free Trade Networks." *Journal of International Economics* 72: 310-335.
- [14] Goyal, Sanjeev and Sumit Joshi, 2006. "Bilateralism and Free Trade." *International Economic Review* 47: 749-778.
- [15] Horn, Henrik and Petros Mavroidis, 2001. "Economic and legal aspects of the Most-Favored-Nation clause." *European Journal of Political Economy* 17(2) 233-279.
- [16] Johnson, H.G., 1965. "An Economic Theory of Protectionism, Tariff Bargaining, and the Formation of Customs Unions." *The Journal of Political Economy* 72: 256-83.
- [17] Kennan, John and Raymond Riezman, 1990. "Optimal Tariff Equilibria with Customs Unions." *Canadian Journal of Economics* 23(1): 70-83.
- [18] Krishna, Pravin, 1998. "Regionalism and Multilateralism: a Political Economy Approach." *Quarterly Journal of Economics* 113: 227-51.
- [19] Liu, Xuepeng. 2010. "Testing Conflicting Political Economy Theories: Full-fledged versus Partial-scope Regional Trade Agreements." *Southern Economic Journal* 77(1): 78-103.

- [20] Ludema, R.D. and Mayda, A.M., 2009. "Do countries Free Ride on MFN?" *Journal of international Economics* 77(2), pp.137-50.
- [21] Ludema, R.D. and Mayda, A.M., 2013. "Do Terms-of-Trade Effects Matter for Trade Agreements? Theory and Evidence from WTO Countries." *The Quarterly Journal of Economics* 128(4): 1837-93.
- [22] McCalman, Phil, 2002. "Multilateral trade negotiations and the most favored nation clause." *Journal of International Economics* 27: 151-76.
- [23] Missios, Paul, Kamal Saggi and Halis M. Yildiz, 2016. "External Trade Diversion, Exclusion Incentives and the Nature of Preferential Trade Agreements." *Journal of International Economics* 99: 105-19.
- [24] Olarreaga, M. and Soloaga, I., 1998. "Endogenous tariff formation: the case of Mercosur." *World Bank Economic Review* 12(2): 297-320.
- [25] Ornelas, Emanuel, 2005a. "Trade Creating Free Trade Areas and the Undermining of Multilateralism." *European Economic Review* 49: 1717-35.
- [26] Ornelas, Emanuel, 2005b. "Rent Destruction and the Political Viability of Free Trade Agreements." *Quarterly Journal of Economics* 120: 1475-1506.
- [27] Ornelas, Emanuel, 2007. "Exchanging market access at the outsiders' expense: the case of customs unions." *Canadian Journal of Economics* 40(1): 207-24.
- [28] Richardson, Martin, 1993. "Endogenous Protection and Trade Diversion." *Journal of International Economics* 34(3-4): 309-24.
- [29] Richardson, Martin, 1995. "Tariff Revenue Competition in a Free Trade Area." *European Economic Review* 39(7): 1429-37.
- [30] Saggi, Kamal, 2004. "Tariffs and the Most Favored Nation Clause." *Journal of International Economics* 63(2), 341-68.
- [31] Saggi, Kamal and Halis M. Yildiz, 2010. "Bilateralism, Multilateralism, and the Quest for Global Free Trade." *Journal of International Economics* 81: 26-37.
- [32] 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.

- [33] Seidmann, Daniel, 2009. “Preferential Trading Arrangements as Strategic Positioning.” *Journal of International Economics* 79: 143–59.
- [34] Stoyanov, Andrey and Halis M. Yildiz, 2015. “Preferential versus Multilateral Trade Liberalization and the Role of Political Economy.” *European Economic Review* 80: 140-164.
- [35] Wong, Woan Foong, 2016. “The Free Rider Effect and Market Power in Trade Agreements.” Working Paper.
- [36] WTO. 2011. “World Trade Report 2011: The WTO and Preferential Trade Agreements: From Co-Existence to Coherence.” Geneva: WTO.
- [37] Yi, Sang-Seung, 1996. “Endogenous formation of customs unions under imperfect competition: open regionalism is good.” *Journal of International Economics* 41(1): 153-77.

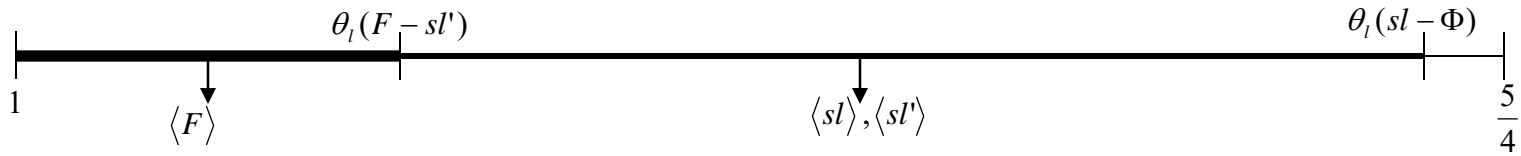


Figure 1: Equilibrium agreements under the benchmark WTO game (FTA)

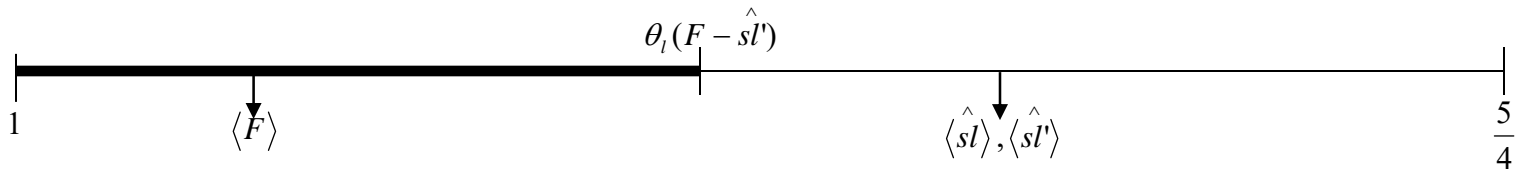


Figure 2: Equilibrium agreements without free internal trade requirement (FTA)

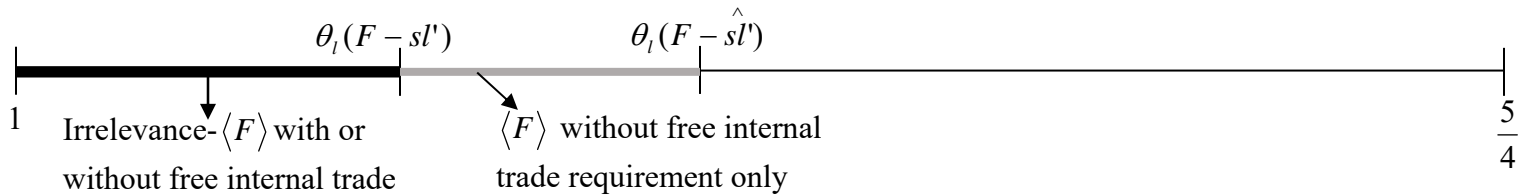


Figure 3: Free trade with/without free internal trade requirement (FTA)

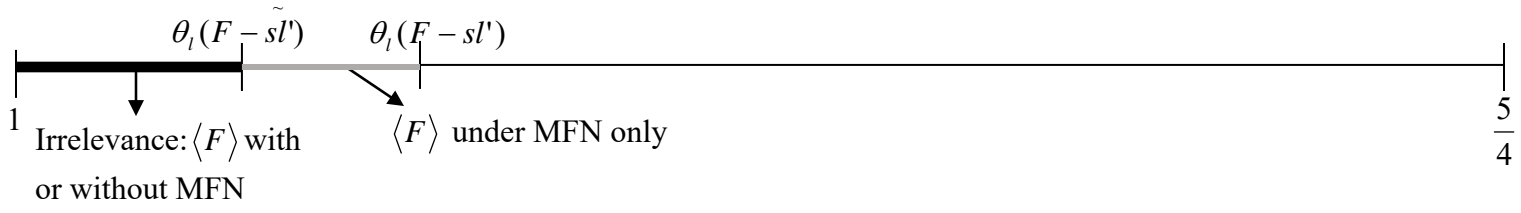


Figure 4: Free trade with/without MFN requirement (FTA)

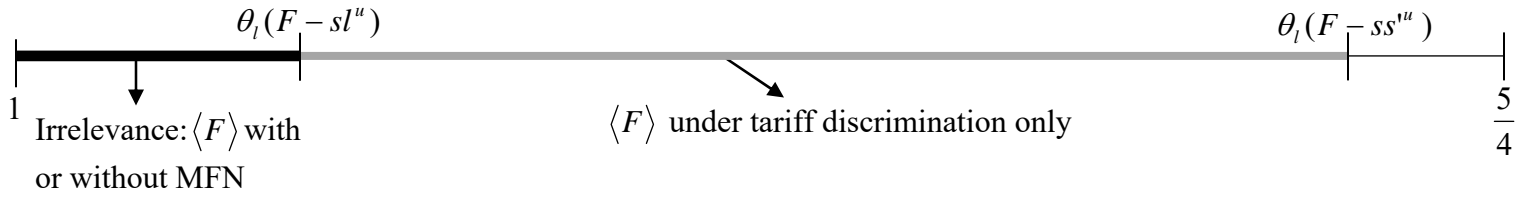


Figure 5: Free trade with/without MFN requirement (CU): two small and one large countries