The Role of Non-Discrimination in a World of Discriminatory Preferential Trade Agreements

Kamal Saggi
Vanderbilt University

Woan Foong Wong
University of Oregon

Halis Murat Yildiz
Ryerson University

Abstract. In a three-country model of endogenous trade agreements, we study the implications of the Most Favored Nation (MFN) clause when countries are free to form discriminatory preferential trade agreements (PTAs). Under current rules of the World Trade Organization (WTO), although non-member countries face discrimination at the hands of PTA members, they themselves are obligated to abide by MFN and treat PTA members in a non-discriminatory fashion. The non-discrimination constraint of MFN reduces the potency of a country's optimal tariffs and therefore its incentive for unilaterally opting out of trade liberalization. Thus, MFN can act as a catalyst for trade liberalization. However, when PTAs take the form of customs unions, the efficiency case for MFN as well as its pro-liberalization effect is weaker since one country finds itself *deliberately excluded* by the other two as opposed to staying out *voluntarily*.

Résumé.

JEL classification: F11, F12

We are grateful to Hiroyuki Kasahara as well as two anonymous referees for their helpful and constructive suggestions.
E-mail addresses: kamal.saggi@vanderbilt.edu (K. Saggi), wfwong@uoregon.edu (W.F. Wong), hyildiz@ryerson.ca (H.M. Yildiz).

Canadian Journal of Economics / *Revue canadienne d'économique* 20XX 00(0)
January 20XX. Printed in Canada / *Janvier 20XX. Imprimé au Canada*

ISSN: 0000-0000 / 20XX / pp. 1–?? / © Canadian Economics Association

1. Introduction

The General Agreement on Tariffs and Trade (GATT), now subsumed by the World Trade Organization (WTO), has governed multilateral trade liberalization since 1948. At the heart of the WTO system is the Most Favored Nation (MFN) principle, contained in Article I of the GATT. This non-discrimination principle applies to all WTO members and it requires that, when importing a given product from several countries, a WTO member must impose the same tariff – called its *applied MFN tariff* – on all foreign suppliers to its market. At the same time however, conflicting with this non-discrimination stance, GATT also allows discriminatory liberalization through the formation of Preferential Trade Agreements (PTAs) under Article XXIV. Specifically, this article allows member countries of a PTA to grant tariff cuts to each other that they do not have to extend to other WTO members that do not belong to the PTA, provided PTA members do not increase their tariffs on non-members. The increasing prevalence of PTAs has generated a large literature that focuses on understanding how the inherently discriminatory nature of PTAs impacts the degree of global trade liberalization that would have otherwise arisen if countries only used non-discriminatory MFN tariffs set either individually or through multilateral negotiations.

While PTAs embody discrimination against non-member countries, MFN requires countries not participating in a given PTA to treat PTA participants in a non-discriminatory fashion. This awkward asymmetry raises two substantive questions. One, is there a case for allowing PTA non-members to deny MFN treatment to PTA members? In other words, should non-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? We develop a model that allows us to directly address these novel questions that have been overlooked in the vast literature on PTAs.

Our formal approach follows Saggi and Yildiz (2010) who develop an equilibrium theory of PTAs in a modified version of the three-country competing exporters framework of Bagwell and Staiger (1999a).[1] 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 scenario* under which the non-member is obligated to follow the MFN principle of non-discrimination when setting its external tariffs. We next compare this benchmark scenario with an alternative setting, called *tariff discrimination.* Under the latter setting, the non-member country does not have to abide by MFN and is free to impose its optimal discriminatory tariffs on PTA members. Intuitively, a comparison of the two scenarios 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 PTA members.

Existing literature has shown that optimal MFN tariffs generally impose fewer distortions relative to optimally chosen discriminatory tariffs.[2] In general, since discriminatory tariffs tend to be biased against relatively efficient exporters, they cause socially harmful trade diversion. 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, MFN can act as a catalyst for further 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 with each other to a multilateral one involving all three countries). In other words, we show that

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

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

when a country is trading with two countries that are in a bilateral PTA with each other, the welfare case for requiring the non-member country to follow MFN is stronger if it has voluntarily chosen to stay out of the PTA relative to a scenario where it has been excluded by PTA members against its wishes.

Why would a country stay out of a PTA voluntarily? Our model shows that the answer to this question is driven by two factors, one having to do with how the formation of a PTA alters the external tariffs of member countries and the other with the degree of underlying economic asymmetry amongst countries. In our model, the restriction on the external tariffs of PTA members imposed by Article XXIV – i.e. that members should not raise their external tariffs when reducing their internal tariffs – is non-binding since members had an incentive to actually *reduce* their external tariffs. Such external liberalization on the part of PTA members is also a feature of our model and it generates an incentive for a country to stay out of a PTA between the other two countries since it can benefit from their external tariff reductions without having to reciprocate. It is worth noting here that in our partial equilibrium model, the formation of a PTA does not affect the non-member country's tariffs. By contrast, in the symmetric general equilibrium model of Syropoulos (1999), tariffs are strategic substitutes across countries and effects of a CU on global tariffs as well as welfare are a bit more nuanced since the tariff of the non-member country adjusts in response to changes in tariffs of CU members. However, just like in our model, though coordination provides members of a CU an incentive to raise their external tariffs, having to eliminate their internal tariffs (on one another) can blunt this incentive such that they end up reducing their common external tariff, although this is not always the case (like it is in our model). The second relevant factor related to endowment asymmetry works as follows: variation in economic size implies countries have differing economic incentives for retaining their optimal tariffs and therefore for not participating in a PTA. For example, when endowment asymmetry is sufficiently pronounced, the country whose volume of imports is much larger than its trading partners has a stronger incentive to retain its optimal tariffs (since they tend to be relatively more potent than those of its trading partners) so that it can choose not to join the PTA, preferring instead to free ride on

the external trade liberalization undertaken by the other two countries while it sets its own optimal tariffs.

Next, consider why two countries might have a joint incentive to exclude the third country from their bilateral PTA. Suppose two countries are in a bilateral PTA and consider how the move to free trade affects them. Since they already impose zero tariffs on each other, moving to global free trade costs them privileged access to each other's markets. On the other hand, they gain free access to the non-member's market. But, since countries are asymmetric in our model, the benefit of moving to free trade from a bilateral PTA can be dominated by the cost of giving up preferential access in each other's markets. While the economic incentive to exclude exists for both FTAs and CUs, this incentive goes unexercised in the more flexible FTA game since each of the excluding countries is tempted to become the hub country by entering into a separate FTA with the non-member country. Due to the common external tariff requirement of a CU, it is simply impossible for a CU member to position itself as a hub in a hub and spoke trading arrangement. As a result, a pair of countries are able to exercise their exclusion incentive only when PTA formation takes the form of CUs. The practical implication of this result is that the case for requiring MFN on the part of a non-member country is relatively weaker when the PTA between the other two countries is a CU as opposed to an FTA. We find this result because by making tariff discrimination infeasible, the non-discrimination constraint of MFN (*i*) reduces the benefit of being an outsider and thus lowers the unilateral incentive for opting out of trade liberalization while (*ii*) it increases the coalitional incentives of the two larger importers to exclude the third country from their bilateral CU.

We also examine the implications of allowing the PTA non-member to deny MFN treatment to PTA members when free trade is simply infeasible because of the underlying economic environment. Under such a case, we find that MFN adoption increases world welfare regardless of the nature of PTA. This result suggests that the MFN principle is not necessarily consistent with the objective of achieving global free trade although it helps avoid socially harmful trade diversion in a tariff ridden world.

It is worth noting here that Article XXIV of the World Trade Organization (WTO) allows countries to form PTAs under two conditions: (i) members should eliminate (substantially) all internal trade barriers and (ii) PTA members must not raise external trade barriers on non-members. While Article XXIV tries to limit the damage on non-member countries by requiring PTA members to not raise tariffs on outsiders, the fact remains that it contradicts the principle of non-discrimination that underlies the entire WTO system. It is worth pointing out that many earlier papers on PTA formation, including Yi (1996) and Krishna (1998), focused primarily on the first condition and imposed no restrictions on the setting of external tariffs of members. In other words, they allowed external tariffs to rise under CU formation. By contrast, Syropoulos (1999), Mrazova et al (2013) and Missios et al. (2016) study CU formation that considers both the internal and external tariff requirements of Article XXIV by explicitly modeling the fact that external tariffs are prevented from being raised. This paper offers a framework in which external tariffs are complements both under CU and FTA formation and thus makes it possible to examine tariff discrimination by countries, both as PTA members and a non-member.

Since Bhagwati (1991), a rich literature has emerged to address the question whether PTAs serve as *building* or *stumbling* blocs for multilateral trade liberalization. Taken together, this literature provides a fairly comprehensive evaluation of the effects of GATT Article XXIV on multilateral trade liberalization and global welfare. The first generation of theoretical models in this area focused primarily on the impact of PTA membership on the incentives for multilateral trade liberalization (see, for example, Krishna, 1998; Ornelas, 2005a, 2005b). The next wave of studies starting with Yi (1996) and subsequently Goyal and Joshi (2006), Aghion et al. (2007), Furusawa and Konishi (2007), and Seidman (2009) consider endogenous PTAs but ignore the possibility of MFN based trade liberalization. A series of papers published in the last decade or so – such as 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 them (granted to WTO members by GATT Article XXIV) is *necessary* for achieving global

free trade. Additionally, Saggi et al. (2019) have argued that the free internal trade requirement of Article XXIV makes it harder to achieve global free trade since in the presence of tariff complementarity it leads to lower external tariffs on the part of FTA members which in turn creates an incentive for free-riding on the part of the non-member. 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 contributes to the literature on the building versus stumbling bloc question by showing 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 for the case of CUs.

This paper proceeds as follows. Section 2 introduces our model while Section 3 details the two policy scenarios which are central to our analysis—the WTO-consistent and tariff discrimination scenarios. Section 4 characterizes the equilibrium outcomes under both scenarios. In Section 5, we consider the equilibrium outcomes for both when countries are allowed to coordinate their external tariffs, as would be the case under CUs. Section 6 highlights and discusses the robustness of our results to a number of extensions: an alternative equilibrium concept, allowing for the choice of trade agreement formation to be endogenous between an FTA or CU, as well as allowing for zero external tariffs and a more flexible endowment asymmetry structure. Section 7 concludes.

2. Tariffs and Trade

To explore the questions motivating the paper, we utilize an appropriately modified version of the partial equilibrium 'competing exporters' model of Bagwell and Staiger (1999a). The model considers three asymmetrically endowed countries (i, j, and k) and three (non-numeraire) goods I, J, and K.[3] Country i is endowed with zero units of good I and e_i units of the other two

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

goods. Country i imports good I from countries j and k who act as competing exporters to its market.

The demand for good z in country i is given by

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

where α measures the demand intercept and p_i^z the price of good z in country i. This demand function is derived from a utility function of the form $U(c^z) = u(c^z) + \nu$, where c^z denotes consumption of good z; ν denotes the numeraire good; and $u(c^z) = \sum_{z=I,J,K} \left[\alpha c^z - \frac{1}{2}(c^z)^2\right]$ is quadratic and additively separable in each of the three z goods.

Let t_{ij} denote 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:

$$p_i^I = p_j^I + t_{ij} \ = p_k^I + t_{ik} \tag{2}$$

Country i's imports of good I (denoted m_i^I) simply equal local demand since its endowment of good I equals zero:

$$m_i^I = d(p_i^I) = \alpha - p_i^I \tag{3}$$

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

$$x_j^I = e_j - [\alpha - p_j^I] \tag{4}$$

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

$$m_i^I = \sum_{j \neq i} x_j^I \tag{5}$$

Using equations (2) through (5) the equilibrium prices of good I in country i and country j are calculated as follows:

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

and

$$p_j^I = \frac{1}{3}\left(3\alpha - \sum_{j\neq i} e_j - 2t_{ij} + t_{ik}\right) \tag{7}$$

Observe from equation (6) that only a third of a given increase in the tariff t_{ij} passes through to local price with the rest falling on the shoulders of foreign exporters. Such incomplete pass through to local prices implies that tariffs lead to an improvement in the terms of trade of importing countries thereby providing a motive for their use.

From a welfare perspective, given the partial equilibrium nature of the model, it suffices to consider only the goods that might be subject to protection. A country's welfare is defined as the sum of consumer surplus, producer surplus, and tariff revenue over all such goods:[4]

$$w_i = \sum_z CS_i^z + \sum_z PS_i^z + TR_i \tag{8}$$

Using the equations above, one can calculate the 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:

$$ww = \sum_i w_i. \tag{9}$$

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: $\max\{e_i, e_j, e_k\} \leq \frac{5}{4}\min\{e_i, e_j, e_k\}$.[5]

4 We provide the exact formula of these welfare components in Appendix section 8.1.1

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

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

When countries do not enter into any type of trade agreement with each other (denoted as trade regime $\langle\Phi\rangle$), t_i^M denotes country i's optimal MFN tariff under this regime:

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

where $w_i(\Phi)$ is the welfare function for country i in the absence of any trade agreements (i.e. trade regime $\langle\Phi\rangle$).

Upon forming an FTA, member countries remove their internal tariffs on each other and each member imposes an individually optimal external tariff on the non-member. Under a single FTA, say between i and j, the optimal external tariff of an FTA member is as follows:

$$t_{ik}(ij) \equiv \text{Arg}\max w_i(ij) = \frac{5\,e_k - 4e_j}{11} \tag{10}$$

where $w_i(ij)$ is the welfare function for country i when countries i and j enter into an FTA with each other (trade regime $\langle ij\rangle$).

Comparing $t_{ik}(ij)$ and t_i^M reveals that $t_i^M > t_{ik}(ij)$. That is, our model exhibits "tariff complementarity": FTA formation induces member countries to lower their tariffs on the non-member.[6] Market separability implies that FTA formation leaves the FTA non-member's optimal MFN tariff unchanged: $t_{ki}\,(ij) = t_{kj}\,(ij) = t_k^M$.

Next, consider a bilateral CU between i and j (trade regime $\langle ij^u\rangle$). Like FTA members, CU members remove internal tariffs on each other ($t_{ij}\,(ij^u) = t_{ji}\,(ij^u) = 0$). But unlike FTA members, CU insiders can coordinate their

6 For tariff complementarity discussions, see Bagwell and Staiger (1998, 1999a, 1999b), Bond et al. (2004), Saggi and Yildiz (2009) and Estevedeordal et al. (2008). While only terms-of-trade considerations influence trade policy considerations here, this result is robust to including other trade policy motives like the presence of a production relocation externality (Suwanprasert (2020) and Ossa (2011))

external tariff to maximize the joint welfare of CU members. This results in the following optimal external tariff:

$$t_{ik}\left(ij^u\right) \equiv \underset{t_{ik}(ij^u)}{\arg\max}\; w_i(ij^u) + w_j(ij^u) \text{ subject to } t_{ij}\left(ij^u\right) = t_{ji}\left(ij^u\right) = 0$$
$$= \frac{2\,e_k - e_j}{5} \tag{11}$$

where $w_i(ij^u)$ and $w_j(ij^u)$ are the welfare functions for countries i and j respectively when they enter into a CU with each other (trade regime $\langle ij^u \rangle$).

We first note that, while CU members also practice tariff complementarity, $t_i^M > t_{ik}\left(ij^u\right)$, there exists a critical difference in the external tariff setting of FTA and CU members.[7] When setting external tariffs individually, each FTA member ignores the negative externality imposed on the export surplus of its FTA partner because of the reduction of its tariff on the non-member. By coordinating their external tariffs, CU members internalize this negative externality, i.e. $t_{ik}\left(ij^u\right) > t_{ik}\left(ij\right)$, and thereby benefit from tariff coordination. As a result, the degree of tariff complementarity for CU members is less than that for FTA members: $t_i^M - t_{ik}\left(ij^u\right) < t_i^M - t_{ik}(ij)$.

Here, one should note that, in the absence of an MFN clause, a non-member country is free to tariff discriminate across its trading partners. For a non-member country (say country i) under an FTA $\langle jk \rangle$, its optimal tariffs (t_{ij}^*, t_{ik}^*) solve $\arg\max w_i(jk)$ which gives:

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

7 In contrast, see Mrazova et al. (2013) for a setting where CU members are indeed bound by the GATT Article XXIV constraint such that they cannot raise their tariffs on non-members. Unlike the competing exporters model with a homogenous good, Mrazova et al. (2013) employs an oligopoly model where each country produces one specific variety and the other countries are competing importers in that particular variety. Thus, the relatively higher market power of a CU can lead to a possibility of tariff complementarity failing to hold. The same result would hold in a competing importers model as in Missios et al. (2016). However, in a competing importers setting, the MFN vs. discrimination question cannot be captured (because each country exports a unique good) and thus we opted for competing exporters framework as in Bagwell and Staiger (1999a).

Note that, when free to tariff discriminate, each country imposes a higher tariff on the larger exporter:

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

By increasing its volume of imports, a country's optimal tariff increases with the exporters' endowments. Relative to the case of MFN, countries impose higher discriminatory tariff on the imports from the country with larger endowment:

$$t^M_i - t^*_{ij}(\Phi) = \frac{2(e_k - e_j)}{8} < 0 \text{ iff } e_j > e_k$$

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

$$t^*_{ij}(jk) \leq t^M_i \leq t^*_{ik}(jk) \text{ where } e_j \leq e_k \tag{12}$$

We summarize the key messages of the above analysis in the following lemma which is well-established in the literature (see for example, Saggi (2004)):

Lemma 1. Suppose two countries form an FTA. If free to *tariff discriminate, the non-member country imposes a higher tariff on the FTA member from which it sources a larger volume of imports:* $t^*_{ik}(jk) \geq t^*_{ij}(jk)$ *iff* $e_j \leq e_k$. *If the non-member has to abide by MFN then its optimal MFN tariff is bounded by its optimal discriminatory tariffs:* $t^*_{ij}(jk) \leq t^M_i(jk) \leq t^*_{ik}(jk)$ *when* $e_j \leq e_k$.

One should note here that, the non-member country under a bilateral PTA has an incentive to discriminate using external tariffs and since tariff discrimination is biased against the country with a higher degree of comparative advantage, it leads to a socially harmful trade diversion. Thus, in a tariff ridden world, MFN adoption by each country improves world welfare by eliminating such trade diversion. Later, we emphasize the importance of

this result for scenarios where free trade is no longer an outcome and bilateral PTAs arise in equilibrium.

3. Endogenous Free Trade Agreements

We now derive and contrast outcomes under two well-defined policy scenarios – i.e. a WTO-consistent benchmark scenario and tariff discrimination – to address the questions motivating our analysis.

(a) *WTO-consistent scenario*: This benchmark 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 decided whether to 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 eliminate tariffs on each other while imposing their 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) *Tariff discrimination*: This scenario differs from the WTO-consistent benchmark in one way: at the second stage of the game, the non-member country is free to disregard MFN and impose its optimal discriminatory tariffs on FTA members. Thus, under this scenario, 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).

Following Saggi and Yildiz (2010), we now describe the process of FTA formation that constitutes 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. This announcement can be thought of as a proposal that each country makes to its trading partners. Denote country i's announcement by σ_i and its strategy set by S_i where

$$S_i = \{\{\phi, \phi\}, \{j, \phi\}, \{\phi, k\}, \{j, k\}\} \tag{13}$$

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 announcement profiles and the types of trade agreements that arise:

(*i*) Either no two announcements match or the only matching announcements are $\{\phi, \phi\}$. All of these announcement profiles result in no agreement $\langle\Phi\rangle$ under which all countries impose their optimal MFN tariffs on one another.

(*ii*) Two countries name each other *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 such an FTA, members impose individually optimal external tariffs on the non-member. The non-member instead imposes its optimal MFN tariff on members under the WTO scenario while imposing optimal discriminatory Nash tariffs under the tariff discrimination scenario.

(*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 $\langle ih\rangle$).

(*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 $\langle F \rangle$.

Since an FTA between two countries can form only if both sides want to do so, multiple announcement profiles can yield the same trade agreement. For example, the FTA trade regime $\langle ij \rangle$ is the outcome when countries i and j only announce each other, *regardless* of the nature of country k's announcement: if $\sigma_i = \{j,\phi\}$ and $\sigma_j = \{i,\phi\}$, then $\langle ij \rangle$ is the result of 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. Similarly, $\langle ij \rangle$ is also the outcome when countries i and j announce each other's name and either one or both of them also announce(s) country k but country k does not reciprocate. Thus, 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 $\sigma_k = \{\phi,\phi\}$. We consider the announcement $\sigma_i = \{j,\phi\}$, $\sigma_j = \{i,\phi\}$, and $\sigma_k = \{\phi,\phi\}$ as parsimonious and all other announcements as non-parsimonious. As in Saggi and Yildiz (2010), Saggi et al. (2013), Missios et al. (2016), and Saggi et al. (2019), we restrict the strategy set to the set of announcement profiles that are parsimonious.

When analyzing the above games, we refine the set of Nash equilibria by isolating the Nash equilibria that are *coalition proof*. In a classic paper, Bernheim et al. (1987) noted that "in an important class of "noncooperative" environments, it is natural to assume that players can freely discuss their strategies, but cannot make binding commitments." They further noted that any meaningful agreement in such cases had to be self-enforcing and that although "the best-response property of Nash equilibrium is a necessary condition for self-enforceability, it is not sufficient since it is in general possible

for coalitions to arrange plausible, mutually beneficial deviations from Nash agreements." A coalition proof Nash equilibrium (CPNE) is simply a Nash equilibrium that is immune to all *self-enforcing* coalitional deviations (i.e. those coalitional deviations that are not themselves susceptible to further deviations on the part of a subset of the original coalition).

4. Equilibrium Free Trade Agreements

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

Assumption 1. The pattern of endowment asymmetry is as follows: Medium-sized country m is a larger importer than small-sized country s while large-sized country l is a larger importer relative to both countries m and s: $e_s = \theta e \geq e_m = \frac{1}{2}\theta e + \frac{1}{2}e \geq e_l = e$ where $1 \leq \theta \leq 5/4$.[8]

Following this assumption, the reference to country size in this paper corresponds to the country's importer size. It is worth pointing out although country s has a weaker ability to manipulate its terms of trade via its import tariff relative to the other two, it is not 'small' in the traditional sense of the term wherein it would be a price-taker on world markets. Country s has the largest endowment so it is the largest exporter of two goods, but the smallest importer of the third good over which it imposes an import tariff. Similarly, country l has the smallest endowment and imports from countries with larger endowments so that it is the largest importer but the smallest exporter. Thus, since the only policy instrument in our framework is an import tariff, among all goods that are subject to import tariffs country l is the largest importer and therefore has the strongest ability to manipulate its terms of trade.

We proceed as follows. First, we study FTA formation in our WTO-consistent benchmark scenario and show that, while two countries may have

8 Countries having the same endowment in different goods is a simplifying assumption for expositional purpose. Since markets are strategically separated in goods, countries' external tariffs on different goods are independent. As long as we can identify countries with large, medium and smaller importing countries, the existing results would go through.

an incentive to form a bilateral trade agreement aimed at excluding the third country, this exclusion incentive is not self-enforcing. 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. Then, we argue that the ability of the non-member country under a bilateral FTA to tariff discriminate makes it harder to achieve global free trade due to larger free riding incentive on its part relative to the WTO-consistent scenario under which the non-member has to abide by MFN. Later, with our CU game we show that both the equilibrium and welfare implications of the MFN clause depend on whether the non-member voluntarily stays out of the PTA 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 and do not raise their external tariffs on non-members. Let country i's welfare under trade policy regime r be denoted by $w_i(r)$, where $r = \langle\Phi\rangle$,$\langle ij\rangle, \langle ih\rangle$, or $\langle F\rangle$ where it is understood that countries impose their optimal tariffs consistent with the policy regime r. Let $\Delta w_i(r-v) \equiv w_i(r) - w_i(v)$, 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 i.e. $\Delta w_i(r-v) = 0$ at $\theta = \theta_i(r-v)$.

We first state the following lemma that explains how differences in endowment 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 holds:*

(i) (***Attractiveness of bilateral FTA formation***) *Under no agreement, all countries have an incentive to form a bilateral FTA and each*

country prefers to form a bilateral FTA with the larger importer relative to the smaller one: $\Delta w_l(ml - sl) > 0$, $\Delta w_m(ml - sm) > 0$, *and* $\Delta w_s(sl - sm) > 0$ *for all* θ.

(*ii*) (Free riding incentive) *While neither the small nor the medium importer has an incentive to free ride and stay as an outsider under a bilateral FTA, the large importer does have such an incentive when the degree of endowment asymmetry is sufficiently large:* $\Delta w_i(F - jk) > 0$ *for all* θ *where* $i = s, m$ *and* $i \neq j, k$ *while* $\Delta w_l(F - sm) < 0$ *when* $\theta > \theta_l(F - sm)$.

(*iii*) (Exclusion incentive) *While the small importer does not participate in any joint deviation from free trade to exclude the third country via forming a bilateral FTA, medium and large importers do have such an incentive when the degree of endowment asymmetry is sufficiently large:* $\Delta w_s(F - sj) > 0$ *for all* θ *where* $j = m, l$ *while* $\Delta w_i(F - ij) < 0$ *when* $\theta > \theta_i(F - ij)$ *where* $i, j = m, l$.

(*iv*) (Attractiveness of being hub) *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* $i = s, m, l$.

(*v*) (Attractiveness of being spoke) *While the large importer prefers being a non-member under a bilateral FTA to being a spoke under a hub and spoke regime regardless of parameter values, the medium and smaller importers do so only when the degree of endowment asymmetry is sufficiently small:* $\Delta w_l(ih - sm) < 0$ *for all* θ *and* $i = s, m$ *and* $\Delta w_i(jh - jk) < 0$ *when* $\theta < \theta_i(jh - jk)$ *where* $i = s, m$ *and* $i \neq j, k$.

Part (i) of Lemma 2 governs the attractiveness of bilateral FTA formation starting from a regime of no trade agreements. Intuitively, the exchange of reciprocal market access makes bilateral FTA formation attractive. The latter part 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 2 argues that small and medium exporters have no incentive to unilaterally deviate from free trade to become an outsider under a bilateral FTA. However, when the large importer is sufficiently large, it has an incentive to free ride on the trade liberalization of the other two countries. Intuitively, despite the discrimination faced as an FTA outsider, it benefits from tariff complementarity which lowers the external tariffs faced when exporting to the FTA insiders and retains its ability to impose optimal tariffs.

The second and third parts of the Lemma together inform us that the small importing country has no unilateral or joint incentive to deviate from free trade. This is because its volume of export is large while its volume of imports is small relative to other countries. However, medium and large importers do have an incentive to jointly exclude the small importer from their free trade network when the asymmetry is sufficiently large.

The fourth part of the Lemma says that being a hub country is preferable for all countries relative to other trade policy regimes irrespective of their relative sizes. Intuitively, the hub country enjoys sole preferential access in the spoke countries while offering free trade in the domestic market. 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.

Finally, under a bilateral FTA, the large importer has no incentive to give up its ability to impose optimal tariffs to gain a free access in an export market in which the competing exporter already has a free access. This implies that

hub and spoke regimes in which small and medium importing countries are hub are not even Nash equilibrium.

An important message delivered by the above lemma is that the small importing country's preference does not matter for determining the equilibrium trade agreement while the larger importing country's choice is critical. We should also note from the first part of the Lemma 2 that countries do have an incentive to form a bilateral FTA when no such agreement exists. While members of an FTA discriminate against the non-member country, we know from the above tariff analysis that tariff complementarity partially benefits the non-member who also retains its ability to impose optimal external MFN tariffs. 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, regardless of the degree of asymmetry, the medium and the large importing countries benefit from the formation of a bilateral FTA to which they themselves do not belong:

$$\Delta w_i(jk-\Phi) > 0 \text{ for all } \theta,\ i=m,l \text{ and } i\neq j,k \tag{14}$$

Given its stronger preferences for better export access, the small importing country benefits from the formation of the bilateral FTA between larger importers only when the degree of endowment asymmetry is sufficiently small:

$$\Delta w_s(ml-\Phi) \geq 0 \text{ when } \theta \leq \theta_s(ml-\Phi)$$

Therefore, we can state the following:

Proposition 1. *Relative to no agreement* $\langle\Phi\rangle$, *bilateral FTAs* $\langle sm\rangle$ and $\langle sl\rangle$ *are Pareto-improving for all* θ *while the FTA* $\langle ml\rangle$ *is Pareto-improving iff* $\theta \leq \theta_s(ml-\Phi)$.

We are now ready to determine the CPNE of the FTA formation game under the WTO-consistent scenario. 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 parts (*ii*) and (*iii*) of the Lemma 2 that small 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 m and l. It is immediate from part (*iii*) of the above lemma that, taking country s' announcement fixed at $\{m, l\}$, countries m and l have an incentive to jointly deviate from their respective announcements $\{s, l\}$ and $\{s, m\}$ to $\{\phi, l\}$ and $\{\phi, m\}$ in order to exclude country s from a free trade network when country s is sufficiently small:

$$\begin{aligned} \Delta w_l(F - ml) < 0 \; when \; \theta > \theta_l(F - ml) \\ \Delta w_m(F - ml) < 0 \text{ when } \theta > \theta_m(F - ml) \end{aligned} \tag{15}$$

Since $\theta_m(F - ml) > \theta_l(F - ml)$, $\theta_m(F - ml)$ is the binding cutoff for this joint deviation. The above result establishes the existence of an *exclusion incentive*: when the endowment asymmetry is sufficiently pronounced (i.e. $\theta > \theta_m(F - ml)$) the two larger importers prefer a bilateral FTA between themselves over global free trade.

The key question is whether the joint exclusion incentive of the two larger importers is self-enforcing. The answer to this question turns out to be 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 importing country by jointly altering their announcements such that the announcement profile changes from Ω^F (which yields free trade) to $\Omega_1^{ml} = \{\sigma_l = \{\phi, m\}, \sigma_m = \{\phi, l\}, \sigma_s = \{m, l\}\}$ thereby altering the associated trade regime from free trade to the bilateral FTA $\langle ml \rangle$. 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 $\sigma_s = \{m, l\}$, each member of the initially deviating coalition (m 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 $\sigma_l = \{\phi, m\}$ to $\sigma_l = \{s, m\}$ which alters the trade regime from $\langle ml \rangle$ 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 m and l from Ω^F to Ω_1^{ml} 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 not-self enforcing.

Since all countries are better off under free trade relative to $\langle \Phi \rangle$, joint announcement deviations that convert the trade regime from $\langle F \rangle$ to $\langle \Phi \rangle$ can never arise. Based on the above discussion and the lemma, the only possible type of self-enforcing deviations from Ω^F that we need to consider are unilateral deviations from Ω^F and they are self-enforcing by definition. First, we establish that small and medium importing countries have no incentives to unilaterally deviate from Ω^F to announcements leading to hub and spoke regimes in which they are spokes. How about the large importing country? We find that, when the degree of asymmetry is sufficiently large, it has an incentive to unilaterally deviate from its announcements $\{s, m\}$ to $\{s, \phi\}$ or $\{\phi, m\}$ that leads to a hub and spoke regime under which countries s or m is a hub and it itself is a spoke:

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

where $\theta_l(F - sh) > \theta_l(F - mh)$. Then, it is immediate from the second part of the above lemma that only one unilateral deviation incentive remains to be examined: unilateral deviation of country l from $\{s, m\}$ to $\{\phi, \phi\}$ converting

free trade to $\langle sm \rangle$ and it happens when country l is sufficiently large importer relative to other countries:

$$\Delta w_l(F - sm) < 0 \text{ when } \theta > \theta_l(F - sm) \tag{16}$$

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

What if $\langle F \rangle$ is not a CPNE, as is the case when $\theta > \theta_l(F - sm)$? 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 (v) of Lemma 2 that the large importing country under $\langle sh \rangle$ and $\langle mh \rangle$ has an incentive to unilaterally deviate from its respective announcements $\{s, \phi\}$ and $\{\phi, m\}$ to $\{\phi, \phi\}$ and $\{\phi, \phi\}$, leading to a deviation from $\langle sh \rangle$ to $\langle sm \rangle$ and from $\langle mh \rangle$ to $\langle sm \rangle$. Since these unilateral deviations are self-enforcing, any announcement profile leading to $\langle sh \rangle$ and $\langle mh \rangle$ is not even a Nash equilibrium (thus cannot be a CPNE). Consider now the announcement profile that leads to $\langle lh \rangle$. Smaller importing spoke countries have an incentive to jointly deviate from their announcement to form the final FTA leading to global free trade and it is immediate from the above lemma that neither of these countries has incentive to unilaterally deviate further. Thus the initial deviation is self enforcing and the announcement profile that leads to $\langle lh \rangle$ is not a CPNE.

Next, we consider the announcement profile that leads to no agreement $\langle \Phi \rangle$. The first part of our Lemma 2 informs us that two countries have an incentive to deviate from their announcements to the ones that lead to a bilateral FTA and this joint announcement deviation is self-enforcing. As a result, the 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 small and medium importers, say $\langle sm \rangle$). Note from parts (ii) and (v) of Lemma 2 that country l has no incentive to unilaterally deviate from

its announcement $\{\phi, \phi\}$ to either include the small or medium importing country or both when $\theta > \theta_l(F - sm)$. Under such a case, the small and medium importing countries also have no incentive to break up their mutual FTA. As a result the announcement profile that yields $\langle sm \rangle$ is a CPNE when $\theta > \theta_l(F - sm)$ holds.

Next, we move to announcement profiles that yield $\langle sl \rangle$. We find that neither country has an incentive to unilaterally break up their agreements (see Proposition 1). Second, we know from part (i) of Lemma 2 that country s has no incentive to engage in the coalitional announcement deviation with country m that converts $\langle sl \rangle$ to $\langle sm \rangle$. Third, coalitional announcement deviation of countries m and l converting $\langle sl \rangle$ to $\langle ml \rangle$ 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. Fourth, note from the above discussion that the coalitional announcement deviation that replaces $\langle sl \rangle$ by $\langle F \rangle$ is self-enforcing only when $\theta \leq \theta_l(F - sm)$. Finally, we know from part (iv) of Lemma 2 that small and large importing countries have an incentive to become hub and it is immediate from part (v) of Lemma 2 that country m has an incentive to engage in any coalitional announcement deviations that replace $\langle sl \rangle$ by $\langle sh \rangle$ or $\langle sl \rangle$ by $\langle lh \rangle$ when the degree of endowment asymmetry is sufficiently large: $\theta > \theta_m(sh - sl) > \theta_m(lh - sl)$. These deviations are self-enforcing since neither country has an incentive to unilaterally deviate further. As a result, the announcement profile leading to $\langle sl \rangle$ is a CPNE whenever $\theta_l(F - sm) \leq \theta \leq \theta_m(lh - sl)$.

Finally, we consider the bilateral FTA between the two larger importing countries, i.e., $\langle ml \rangle$. First, as before, the coalitional announcement deviation from $\langle ml \rangle$ to $\langle F \rangle$ occurs $\theta \leq \theta_l(F - ml)$ and it is self-enforcing when $\theta \leq \theta_l(F - sm)$. Second, we can show that when $\theta > \theta_s(lh - ml)$ countries s and l have an incentive to jointly deviate from their respective announcements $\{\phi, \phi\}$ and $\{\phi, m\}$ to $\{l, \phi\}$ and $\{s, m\}$, leading to a deviation from $\langle ml \rangle$ to

$\langle lh \rangle$ and this deviation is self-enforcing. Since $\theta_s(lh - ml) < \theta_l(F - sm)$, these self-enforcing announcement deviations cover the entire parameter space and thus the announcement profile supporting $\langle ml \rangle$ is not a CPNE.

We summarize the main findings of the above analysis below:

Proposition 2. *The CPNE of the WTO-consistent game of trade liberalization where FTA members practice free internal trade and the non-member abides by MFN are as follows:*

(i) Free trade $\langle F \rangle$ *is the unique equilibrium when* $\theta < \theta_l(F - sm)$.[9]

(ii) Bilateral FTAs $\langle sl \rangle$ *and* $\langle sm \rangle$ *are both equilibria when* $\theta_l(F - sm) \leq \theta \leq \theta_m(lh - sl)$.

(iii) Bilateral FTA $\langle sm \rangle$ *is the unique equilibrium when* $\theta > \theta_m(lh - sl)$.

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. One important insight that emerges from the above discussion is that exclusion incentives go unexercised in equilibrium due to the lure of a hub and spoke trading arrangement to all parties. Moreover, the free riding incentive of the large importing country proves pivotal for the equilibrium condition of global free trade. To see this, recall from our previous discussion on optimal external tariffs that a country's optimal MFN tariff is increasing in the endowments of other countries. Thus, since country l has the smallest endowment, it sets a relatively higher external tariff by free riding on the agreement $\langle sm \rangle$, while the equation in (10) shows that, relative to other

9 We 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.

bilateral FTAs, the tariff complementarity is the largest under $\langle sm \rangle$ since member countries have incentives to reduce their external tariffs relatively more on the non-member with the smallest endowment. Balancing these asymmetries in the magnitude of tariffs for different sized importers, one can see that smaller importers gain relatively more from further trade liberalization while the relatively large importer gains more from free riding on the FTA with unconstrained ability to impose its optimal tariffs.

Consistent with this discussion, part (*ii*) says that if the degree of endowment asymmetry is sufficiently large, two asymmetric FTAs ($\langle sl \rangle$ or $\langle sm \rangle$) are the equilibrium outcomes – in both situations, one of the larger importing countries prefers being a non-member to participating in any bilateral or multilateral agreements. Under such a case, our theory offers no guidance regarding which of the trade regimes should be expected to arise in equilibrium. Note also from the above discussion that the bilateral FTA between the two larger importing countries $\langle ml \rangle$ fails to arise in equilibrium. Finally, when the degree of asymmetry is sufficiently large, part (*iii*) of Proposition 2 informs us that only the FTA between small and medium importing countries arises as a CPNE.

4.2. 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 if a country finds itself deliberately excluded from an FTA. To address these issues, we now analyze our *tariff discrimination* scenario under which not only the FTA members discriminate against the non-member but also the non-member 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.

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 who has a greater comparative advantage).

Let country i's welfare as a function of the underlying trade agreement r optimal discriminatory tariffs on the part of non-member be denoted by $w_i(\widetilde{r})$. As before, let $\Delta w_i(\widetilde{r} - \widetilde{v}) \equiv w_i(\widetilde{r}) - w_i(\widetilde{v})$. Here, it is important to note that only the external tariffs of the non-member country under a bilateral FTA are different relative to the benchmark WTO-consistent scenario. Therefore, the welfare levels change only under those regimes ($\left\langle \widetilde{ml} \right\rangle$, $\left\langle \widetilde{sl} \right\rangle$, and $\langle \widetilde{sm} \rangle$).

We next examine how Lemma 2 changes when the non-member is free to discriminate between the imports from member countries relative to the case where it abides by the MFN clause. It is immediate from our Lemma 1 that the non-member country and the member country with a weaker comparative advantage would prefer the case when the non-member country is able to discriminate relative to the case of MFN while the opposite is true for the member country with a greater comparative advantage:

$$\begin{aligned} &\Delta w_s(\widetilde{sm} - sm) \leq 0;\ \Delta w_s(\widetilde{sl} - sl) \leq 0 \text{ and } \Delta w_m(\widetilde{ml} - ml) \leq 0 \text{ for all } \theta, \\ &\text{and } \Delta w_i(\widetilde{jk} - jk) \geq 0 \text{ for all } \theta \text{ and } i = s, m, l \text{ and } i \neq j, k \\ &\text{while } \Delta w_m(\widetilde{sm} - sm) \geq 0; \Delta w_l(\widetilde{sl} - sl) \geq 0 \text{ and } \Delta w_l(\widetilde{ml} - ml) \geq 0 \text{ for all } \theta \end{aligned} \tag{17}$$

We first show that the first part of Lemma 2 continues to hold: countries still have an incentive to form a bilateral FTA and each country prefers to form a bilateral FTA with the larger importer relative to the smaller one. Moreover, medium and large importing countries still benefit from the formation of a bilateral FTA in which they are not member regardless of the

degree of asymmetry while the small importing country benefits only when the degree of asymmetry is sufficiently low. Here, one should note that since $\Delta w_s(\widetilde{ml} - ml) \geq 0$ obtains for all θ, the parameter range over which the FTA between medium and large importing countries is Pareto-improving expands. However, one should note that this positive result comes at the expense of lower world welfare under $\left\langle \widetilde{ml} \right\rangle$ relative to $\langle ml \rangle$. Thus, the following is immediate:

Proposition 3. *Relative to no agreement* $\langle \Phi \rangle$*, bilateral FTAs* $\langle \widetilde{sm} \rangle$ and $\left\langle \widetilde{sl} \right\rangle$ *are Pareto-improving for all* θ *while the FTA* $\left\langle \widetilde{ml} \right\rangle$ *is Pareto-improving iff* $\theta \leq \theta_s(\widetilde{ml} - \Phi)$ *where* $\theta_s(ml - \Phi) \leq \theta_s(\widetilde{ml} - \Phi)$.

Next, we examine how the second part of Lemma 2 adapts under the tariff discrimination scenario. We still obtain that small and medium importing countries do not have an incentive to free ride and stay an outsider under a bilateral FTA relative to free trade. As we know from prior discussion, under the WTO-consistent scenario the large importing country's choice is pivotal for global free trade to be the CPNE. We know from the inequalities in (17) that the large importing country has a stronger free riding incentive (stronger unilateral incentive to deviate from Ω^F) when it is free to discriminate as a non-member country relative to the MFN case: $\Delta w_l(F - \widetilde{sm}) < 0$ obtains *when* $\theta > \theta_l(F - \widetilde{sm})$ holds and $\theta_l(F - \widetilde{sm}) < \theta_l(F - sm)$.

We also find that, regardless of whether the non-member country imposes MFN or discriminatory tariffs, being the hub is very attractive due to the sole preferential access to spoke markets. Therefore, even when the exclusion incentive arises, it goes unexercised in the equilibrium as it was under the WTO-consistent scenario. Finally, since the non-member country benefits from being able to discriminate with the external tariffs, the countries' incentive to stay as an outsider rather than becoming a spoke strengthens under discrimination relative to MFN.

In the light of the above discussion, we first argue that $\langle\Phi\rangle$ is not a CPNE since any two countries have incentives to jointly deviate and form a bilateral FTA and this joint deviation is self-enforcing. Second, as before under MFN (even stronger under discrimination), we know a large importing spoke country under $\langle sh\rangle$ and $\langle mh\rangle$ has an incentive to unilaterally deviate from its respective announcements $\{s,\phi\}$ and $\{\phi,m\}$ to $\{\phi,\phi\}$ and $\{\phi,\phi\}$, leading to a deviation from $\langle sh\rangle$ to $\langle\widetilde{sm}\rangle$ and from $\langle mh\rangle$ to $\langle\widetilde{sm}\rangle$. Since unilateral deviations are self-enforcing, the announcement profiles leading to $\langle sh\rangle$ and $\langle mh\rangle$ are not a CPNE. Consider now the announcement profile that leads to $\langle lh\rangle$. As before, spoke countries s and m have an incentive to jointly deviate from their announcements to form the final FTA leading to global free trade and neither country has an incentive to unilaterally deviate further . Thus the initial deviation is self enforcing and the announcement profile that leads to $\langle lh\rangle$ is not a CPNE.

Third, the coalitional announcement deviation from $\left\langle\widetilde{ml}\right\rangle$ to $\langle F\rangle$ happens when $\theta\leq\theta_l(F-\widetilde{ml})$ and it is self-enforcing only when $\theta\leq\theta_l(F-\widetilde{sm})$. Second, we show that, when $\theta>\theta_s(lh-\widetilde{ml})$ holds, country s and the large country l have an incentive to jointly deviate from their respective announcements $\{\phi,\phi\}$ and $\{\phi,l\}$ to $\{\phi,m\}$ and $\{s,m\}$, leading to a deviation from $\left\langle\widetilde{ml}\right\rangle$ to $\langle lh\rangle$ and this deviation is self enforcing. Since $\theta_s(lh-\widetilde{ml})<\theta_l(F-\widetilde{sm})$, these self-enforcing announcement deviations cover the entire parameter space and thus the announcement profile leading to $\left\langle\widetilde{ml}\right\rangle$ is not a CPNE.

Based on the above discussion, the only possible announcement profiles that can be CPNE are the ones leading to $\langle F\rangle$, $\left\langle\widetilde{sl}\right\rangle$, and $\langle\widetilde{sm}\rangle$. As in the WTO-consistent scenario, we can immediately argue that the unilateral deviation of the large importing country from $\{s,m\}$ to $\{\phi,\phi\}$ is critical for the CPNE condition of global free trade:

$$\Delta w_l(F-\widetilde{sm})<0 \text{ when } \theta>\theta_l(F-\widetilde{sm})$$

and thus the announcement profile leading to $\langle F\rangle$ is CPNE whenever $\theta \leq \theta_l(F - \widetilde{sm})$. What if $\theta > \theta_l(F - \widetilde{sm})$ holds and global free trade fails to be a CPNE? The only remaining candidates for CPNE are the announcement profiles that lead to bilateral FTAs $\langle \widetilde{sm}\rangle$ and $\left\langle \widetilde{sl}\right\rangle$. We start with the announcement profiles that yields $\langle \widetilde{sm}\rangle$. First, note that country l has no incentive to engage in a deviation from its announcement $\{\phi, \phi\}$ to either include the small or medium countries or both when $\theta > \theta_l(F - \widetilde{sm})$. Moreover, the small and medium importing countries also have no incentive to unilaterally deviate from their announcements leading to $\langle \widetilde{sm}\rangle$ to the one that yields $\langle\Phi\rangle$. Therefore, $\langle \widetilde{sm}\rangle$ is a CPNE when $\theta \geq \theta_l(F - \widetilde{sm})$ holds.

Next we move to profiles that yield an FTA between the small and large importers, $\left\langle \widetilde{sl}\right\rangle$. We know from the above discussion that that neither country has an incentive to unilaterally deviate from its announcement to the one leading to $\langle\Phi\rangle$. We also know that the coalitional announcement deviation that converts $\left\langle \widetilde{sl}\right\rangle$ to $\left\langle \widetilde{ml}\right\rangle$ 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, country m has no incentive to engage in any coalitional announcement deviations that replace $\left\langle \widetilde{sl}\right\rangle$ by $\langle sh\rangle$ or $\left\langle \widetilde{sl}\right\rangle$ by $\langle lh\rangle$. Third, note from the above discussion that the coalitional announcement deviation that replaces $\left\langle \widetilde{sl}\right\rangle$ by $\langle F\rangle$ is self-enforcing only when $\theta \leq \theta_l(F - \widetilde{sm})$. As a result, the announcement profile leading to $\left\langle \widetilde{sl}\right\rangle$ is a CPNE whenever $\theta \geq \theta_l(F - \widetilde{sm})$.

The following proposition summarizes our findings under the tariff discrimination scenario:

Proposition 4. *The CPNE of the tariff discrimination game of trade liberalization where FTA members practice free internal trade and the non-member is free to tariff discriminate are as follows:*

(i) Free trade $\langle F\rangle$ *is the unique equilibrium when* $\theta \leq \theta_l(F - \widetilde{sm})$.

(ii) Both bilateral FTAs $\left\langle \widetilde{sl}\right\rangle$ *and* $\langle \widetilde{sm}\rangle$ *are equilibria when* $\theta > \theta_l(F - \widetilde{sm})$.

Insert Figure 2

Comparing the equilibrium outcome in the WTO-benchmark scenario (Proposition 2) and the equilibrium outcome in the tariff discrimination scenario (Proposition 3) yields the following result:

Proposition 5. *A comparison of the two scenarios (WTO-consistent benchmark versus tariff discrimination) yields:*

(i) When $\theta \leq \theta_l(F - \widetilde{sm})$, *free trade is the equilibrium outcome under both tariff discrimination and the WTO-consistent benchmark.*

(ii) When $\theta_l(F - \widetilde{sm}) < \theta \leq \theta_l(F - sm)$, *the WTO-consistent benchmark yields free trade whereas tariff discrimination yields bilateral FTAs* $\widetilde{\langle sm \rangle}$ *or* $\widetilde{\langle sl \rangle}$.

(iii) When $\theta > \theta_l(F - sm)$, *free trade is out of reach under both scenarios and world welfare is lower under tariff discrimination.*

Insert Figure 3

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 member countries can coordinate their external tariffs. Accordingly, in the second part of the paper, we discuss the case where the bilateral trade agreement is a customs union (CU) as opposed to an FTA.

5. Endogenous Customs Unions

Suppose the PTA under consideration is a CU under which member countries coordinate their external tariffs as opposed to an FTA. Similar to the FTA

analysis above, we study CU formation in our WTO-consistent benchmark and then we investigate the tariff discrimination scenario.

First, 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 Ω_i consists of four possible announcements:

$$\Omega_i = \{\{\phi\}, \{j\}, \{k\}, \{F\}\} \tag{18}$$

where the announcement $\{\phi\}$ by country i is in favor of the status quo (or no trade liberalization); $\{j\}$ is in favor of a CU with only country j; $\{k\}$ is in favor of a CU with only country k; and $\{F\}$ is in favor of global free trade. As it is clear from the strategy set, CU formation is more rigid relative to FTA formation due to the common external tariff determination. 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. This announcement stage determines the global policy regime. Next, given the policy regime, countries impose their optimal external tariffs. Finally, given trade agreements and tariffs, international trade and consumption take place.

We obtain the following mapping between various announcements profiles: (i) no agreement $\langle\Phi\rangle$ prevails when no two announcements match or when everyone announces $\{\phi\}$; (ii) the CU $\langle ij\rangle$ is formed if countries i and j announce each other's name $\sigma_i = \{j\}$ and $\sigma_j = \{i\}$; (iii) free trade $\langle F\rangle$ obtains if $\sigma_i = \{F\}$ for all $i, j, k = s, m, l$.

We first state the following lemma that summarizes CU formation incentives under both the WTO-consistent benchmark and under tariff discrimination:

Lemma 3. *Regardless of whether the non-member country abides by MFN or not, the following results hold:*

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

(ii) The small importing country has no incentive to unilaterally or jointly deviate from any agreement.

(iii) A country is worse off as a non-member facing a bilateral CU between the other two countries relative to all other trade regimes (i.e. no agreement, being a CU member, or free trade).

The above lemma informs us whether free trade obtains in equilibrium or not critically depends on the joint incentives of the medium and large importing countries to exclude the small importing country. It turns out that this exclusion incentive arises only when the degree of endowment asymmetry is sufficiently large:

$$\begin{aligned} &\Delta w_l(F - ml^u) < 0 \text{ when } \theta > \theta_l(F - ml^u); \\ &\Delta w_m(F - ml^u) < 0 \text{ when } \theta > \theta_m(F - ml^u) \text{ and} \\ &\quad \theta_l(F - ml^u) > \theta_m(F - ml^u) \end{aligned} \tag{19}$$

Here it is important to note that while the exclusion incentive does not arise 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. In other words, the joint announcement deviation leading to a

deviation from free trade to $\langle ml^u \rangle$ is self-enforcing since neither country has an incentive to unilaterally deviate further. In fact, the joint deviation incentive of the large importing country is pivotal for the stability of free trade. As a result, the announcement profile Ω^F leading to free trade is a CPNE only when $\theta \leq \theta_l(F - ml^u)$.

Next, consider the announcement profiles leading to no agreement $\langle \Phi \rangle$. Starting from $\langle \Phi \rangle$, countries m and l have an incentive to jointly alter their announcements to form $\langle ml^u \rangle$. Since this deviation is self-enforcing, $\langle \Phi \rangle$ cannot arise in equilibrium. Similarly, based on parts (i) and (iii) of the Lemma 3, we directly argue that countries m and l have an incentive to jointly alter their announcement profiles so that $\langle sm^u \rangle$ and $\langle sl^u \rangle$ are replaced by $\langle ml^u \rangle$ and these deviations are self-enforcing since neither country m 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 $\langle ml^u \rangle$. Based on the above discussion, it is immediate that $\langle ml^u \rangle$ is a CPNE whenever $\theta \geq \theta_m(F - ml^u)$.

We summarize our main findings in the following proposition:

Proposition 6. *Equilibrium agreements in the CU game under the WTO-consistent scenario are as follows::*

(i) Free trade $\langle F \rangle$ *is the unique equilibrium when* $\theta \leq \theta_m(F - ml^u)$.

(ii) Both free trade $\langle F \rangle$ *and the CU* $\langle ml^u \rangle$ *are equilibria when* $\theta_m(F - ml^u) \leq \theta \leq \theta_l(F - ml^u)$.

(iii) The CU $\langle ml^u \rangle$ *is the unique equilibrium if* $\theta > \theta_l(F - ml^u)$.

Insert Figure 4

We are now ready to examine the implications of requiring MFN on the part of non-member country that has been excluded from the CU. As established earlier, countries will set a higher tariff on the country from whom it imports more from when free to tariff discriminate (Lemma 1). In the case

of the stable CU $\langle ml^u \rangle$ in Proposition 3 above, this means that the excluded small importing country will set a higher tariff on the medium importing country relative to the large importing country. On one hand, faced with this tariff discrimination in the non-member's market, the medium importing country's incentive to exclude the small importing country weakens compared to the WTO benchmark scenario: $\theta_m(F - ml^u) < \theta_m(F - \widetilde{ml^u})$. On the other hand, since the large importing country faces a relative lower tariff under discrimination relative to MFN, this further strengthens its incentive to exclude the small importing country: $\theta_l(F - \widetilde{ml^u}) < \theta_l(F - ml^u)$. In fact, we show that these changes in the exclusion incentives of medium and large importing countries reverse the critical threshold ranking relative to the WTO-consistent scenario: $\theta_l(F - \widetilde{ml^u}) < \theta_m(F - \widetilde{ml^u})$. Therefore, it is the joint deviation incentive of the medium importing country that determines the CPNE condition for free trade: the announcement profile Ω^F leading to free trade is a CPNE when $\theta \leq \theta_m(F - \widetilde{ml^u})$ holds. Similar to the WTO-consistent scenario, we find that when $\theta \leq \theta_l(F - \widetilde{ml^u}) < \theta \leq \theta_m(F - \widetilde{ml^u})$ holds, both $\left\langle \widetilde{ml^u} \right\rangle$ and $\langle F \rangle$ are CPNE. Finally, when the degree of asymmetry is sufficiently large $\theta > \theta_m(F - \widetilde{ml^u})$ and global free trade fails to obtain due to exclusion incentives of the medium and large importing countries, $\left\langle \widetilde{ml^u} \right\rangle$ is the unique CPNE.

Combining the above results under tariff discrimination scenario with Proposition 6 has an interesting implication: when the threshold degrees of asymmetry are compared in the CU game under the WTO-consistent and tariff discrimination scenarios, we obtain $\theta_l(F - ml^u) < \theta_m(F - \widetilde{ml^u})$. We can now state the following result:

Proposition 7. *For* $\theta \leq \theta_l(F - ml^u)$ *free trade is the equilibrium outcome of the CU game under both tariff discrimination and the WTO-consistent scenario. When* $\theta_l(F - ml^u) < \theta \leq \theta_m(F - \widetilde{ml^u})$*, tariff discrimination yields free trade whereas the WTO-consistent scenario yields* $\left\langle \widetilde{ml^u} \right\rangle$*. Finally, when*

$\theta > \theta_m(F - \widetilde{ml^u})$, free trade is out of reach under both scenarios and world welfare is higher under the WTO-consistent scenario.

Insert Figure 5

As noted before in the FTA game, 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 such as the case in the CU game, the MFN requirement for the non-member country can hinder the prospect of global free trade via strengthening the exclusion incentives. As a result, whether MFN complements Article XXIV in achieving global free trade depends on the nature of the trade agreements – i.e. whether it is an FTA or a CU. Finally, irrespective of the nature of the PTA in question, MFN adoption of the non-member country is world welfare improving when free trade fails to obtain either due to free riding incentive or exclusion incentive. Therefore, while whether existence of Article I with Article XXIV increases the likelihood of global free trade depends on the nature of PTAs, it is necessarily world-welfare improving in a tariff-ridden world.

6. Extensions and Discussion

6.1. Alternative equilibrium concept: Strong Nash equilibrium

It is clear from our analysis so far that the self-enforceability requirement of the coalitional deviations in the concept of a CPNE plays a critical role in our model since the exclusion incentive goes unexercised in the FTA formation game (because each of the initially deviating countries has an incentive to further deviate and become the hub country in a hub and spoke regime). Since

a hub and spoke type of regime is simply not possible in the CU game due to the common external tariff requirement, the joint deviation of two countries to exclude the third one is self-enforcing and actually arises in equilibrium. While this is indeed a direct implication of the CPNE concept, we show that it does not play any crucial role for the equilibrium condition of global free trade in the FTA game since our findings suggest that, rather than the exclusion incentive, it is the free riding incentive that is critical for global free trade to arise.

In our one-shot trade agreement formation game, a reasonable alternative to our approach would be to use the notion of a strong Nash equilibrium (SNE) as the equilibrium refinement. Consistent with the arguments laid out by Bernheim et al. (1987), our choice of CPNE (rather than SNE) in this paper originates from the fact it is more appealing because the SNE concept is too strong and perhaps even internally inconsistent. This is because a SNE agreement must be immune to *any* deviations (unilateral or coalitional), *even those that are not self-enforcing* and further deviations would not be taken into consideration when establishing the SNE. In order to show that our results are robust to the choice of equilibrium concept and highlight the role of self-enforceability requirement of the CPNE in this context (as explained above) in a crystal clear way, we next examine which agreements arise as an SNE under MFN and discrimination.

To this end, since no hub and spoke arrangements can hold, and the medium and large importing countries have no incentives to unilaterally deviate from their announcement profiles under $\langle ml^u \rangle$ and $\left\langle \widetilde{ml^u} \right\rangle$, it is immediate to argue that our CPNE agreements under the CU game remain unchanged when we use SNE as our equilibrium concept. Therefore, we focus on the FTA game only. Based on the unilateral and coalitional deviations contained in the previous CPNE analysis, we can establish the following proposition (represented in Figure 6):

Proposition 8. *The Strong Nash equilibrium (SNE) agreements of the FTA game are as follows:*

(i) In the WTO-consistent scenario, free trade $\langle F \rangle$ *is the unique SNE when* $\theta < \theta_l(F - sm)$ *while the bilateral FTA* $\langle sm \rangle$ *is the unique SNE when* $\theta > \theta_l(F - sm)$*;*

(ii) In the tariff discrimination scenario, free trade $\langle F \rangle$ *is the unique SNE when* $\theta < \theta_l(F - \widetilde{sm})$ *while the bilateral FTA* $\langle \widetilde{sm} \rangle$ *is the unique SNE when* $\theta_l(F-\widetilde{sm}) < \theta \leq \theta_m(lh-\widetilde{sm})$*, where* $\theta_l(F-\widetilde{sm}) < \theta_l(F-sm) < \theta_m(lh-\widetilde{sm})$

Insert Figure 6

The results in the above proposition differ from the ones in Propositions 2, 4 and 5 in two important ways. First, the FTA between small and large importing countries $\langle sl \rangle$ ($\left\langle \widetilde{sl} \right\rangle$) does not arise as an SNE while it arises as a CPNE under both the WTO-consistent and tariff discrimination scenarios. To see the intuition behind this difference, note that countries m and l have a joint incentive to deviate from their announcement profile that leads to $\langle sl \rangle$ ($\left\langle \widetilde{sl} \right\rangle$) to the announcement profile leading to $\langle ml \rangle$ ($\left\langle \widetilde{ml} \right\rangle$). As a result, the announcement profile leading to $\langle sl \rangle$ ($\left\langle \widetilde{sl} \right\rangle$) is not an SNE. However, in finding CPNE, we examine further deviation incentives of either country m or country l, taking the announcement of country s as fixed and show that this joint deviation is not self-enforcing since either country m or country l has an incentive to further deviate to become the hub country. The second difference is that the announcement profile leading to $\langle \widetilde{sm} \rangle$ arises as a CPNE in the tariff discrimination scenario for the entire range of asymmetry over which global free trade is not an equilibrium outcome. However, the announcement profile leading to $\langle \widetilde{sm} \rangle$ is an SNE only when $\theta_l(F-\widetilde{sm}) < \theta \leq \theta_m(lh-\widetilde{sm})$. Note that when $\theta > \theta_m(lh-\widetilde{sm})$, all countries have an incentive to deviate coalitionally from their announcement profiles leading to $\langle \widetilde{sm} \rangle$ to the ones leading to $\langle lh \rangle$ and thus the announcement profile leading to $\langle \widetilde{sm} \rangle$ is not an SNE while it obtains as a CPNE since countries s and m have incentive to further deviate

jointly to the announcement profiles leading to global free trade and the initial coalitional deviation is not self-enforcing.

The important message Proposition 8 delivers is that our main results are robust to both equilibrium concepts. As before, the MFN constraint not only makes it easier to achieve global free trade but also delivers a welfare-superior outcome when global free trade cannot be reached.[10]

6.2. Endogenous choice between FTA and CU

So far, we have examined FTA formation and CU formation in isolation. Next, we investigate how robust our results are to a fully endogenous PTA formation game. In order to endogenize countries' choice between FTAs or CUs, we add an initial stage to our original FTA and CU formation game. In this new initial stage (stage 0), we allow the *majority* of countries to determine the type of PTAs they would like to pursue. Specifically, countries vote for the type of PTA game (FTA or CU) they prefer and majority vote determines which of the PTA formation games countries subsequently participate in. The rest of the game proceeds as described in the original model in Section 3. In such a game, our objective is to find out whether the MFN requirement for the non-member country changes the attainability of global free trade and the nature of equilibrium agreements (from an FTA to a CU or vice versa).[11]

It is also important to note from our previous discussion that, when free trade is not a CPNE, multiple CPNE ($\langle sl \rangle$, $\langle sm \rangle$ and $\left\langle \widetilde{sl} \right\rangle$, $\langle \widetilde{sm} \rangle$) arise in the FTA game under both the WTO-consistent and tariff discrimination scenarios. As indicated before, the theory offers no guidance which of the bilateral agreements ($\langle sl \rangle$ or $\langle sm \rangle$) arise in the equilibrium. Therefore, in order to proceed with the equilibrium trade agreements in our new game, we consider

10 It is important to note here that we use a static PTA formation game in this paper. Alternative way could be using a Dynamic PTA formation game in examining the implications of MFN constraint for the equilibrium path of agreements via using Markov perfect equilibrium concept.

11 In order to focus on the attainability of global free trade, when multiple CPNE involve global free trade, we assume that global free trade arises.

both possibilities in turn: (i) An FTA between small and medium importing countries ($\langle sm \rangle$ and $\langle \widetilde{sm} \rangle$) arise in WTO-consistent and tariff discrimination scenarios and (ii) An FTA between small and large countries ($\langle sl \rangle$ and $\left\langle \widetilde{sl} \right\rangle$) arise in WTO-consistent and tariff discrimination scenarios, To save space, we represent our main results in figures 8a and 8b for case (i) and 9a and 9b for case (ii) and provide the details in the appendix.

Insert Figures 7a, 7b, 8a and 8b here

The following proposition is immediate from comparisons of figures 7a and 7b for case (i) and figures 8a and 8b for case (ii):

Proposition 9. *When the choice between FTA and CU formation is endogenous, we find the following results:*

(i) Global free trade is more likely to be an equilibrium outcome under tariff discrimination relative to the WTO-consistent scenario.

(ii) When the degree of endowment asymmetry is sufficiently high and free trade is unattainable under both scenarios, FTA formation (rather than CU formation) is more likely to arise in the equilibrium under the WTO-consistent scenario relative to the tariff discrimination scenario.

The intuition behind the first part of the proposition can be explained as follows. For the equilibrium condition of global free trade we find that, rather than the free riding incentive of country l, the exclusion incentives of countries m and l via a bilateral CU $\langle ml^u \rangle$ and $\left\langle \widetilde{ml^u} \right\rangle$ are critical under both WTO-consistent and tariff discrimination scenarios. This is because of the fact that free riding incentive involves only country l while countries m and l (thus the majority) countries are involved in a coalitional deviation from global free trade to a bilateral CU and thus decide to vote for CU formation. We know from our previous discussion in the CU game that $\theta_l(F - ml^u) \leq \theta_m(F - \widetilde{ml^u})$ and thus when $\theta_l(F - ml^u) < \theta \leq \theta_m(F - \widetilde{ml^u})$ holds, global free trade

arises only if the excluded non-member country is not bound by the MFN constraint (tariff discrimination scenario). This result confirms our previous finding regarding the role of MFN requirement when a country is deliberately excluded.

When the degree of endowment asymmetry is sufficiently high, global free trade is not an equilibrium outcome under both scenarios and under such a case country l prefers free riding under $\langle sm \rangle$ to being a member under $\left\langle \widetilde{ml^u} \right\rangle$. Therefore, the second part of the above proposition argues that there are cases where the pair of countries choose CU formation and form $\left\langle \widetilde{ml^u} \right\rangle$ under tariff discrimination scenario but they choose FTA formation and $\langle sm \rangle$ under the WTO-consistent scenario. This result provides strong support for the MFN principle when global free trade is out of reach for two reasons: (i) country s (the country with the highest degree of comparative advantage) is less likely to be excluded under the WTO-consistent benchmark compared to the tariff discrimination scenario and (ii) WTO-consistent scenario leads to higher world welfare relative to tariff discrimination scenario when the bilateral PTAs between the same members arise as a CPNE. As a result, we conclude that our results are robust to fully endogenizing all countries' choice to forming either PTAs, FTA or CU.

6.3. Allowing for zero external tariffs

In this section, we allow for the possibility of PTA members to impose zero external tariffs when the non-negativity constraint of tariffs fails to hold due to a high degree of asymmetry. To be more specific, recall from our previous discussion that when $\theta > \frac{5}{4}$, country m's external tariff on imports from country l under $\langle sm \rangle$ turns negative and we constrained our attention to the range where $\theta < \frac{5}{4}$ in order to avoid this possibility. Here, we address the question what if we allow higher degree of asymmetry: $\theta > \frac{5}{4}$ and assume that a PTA member country imposes zero external tariff when its optimal external

tariff is negative.[12] It is important to note that global free trade is a CPNE only when the degree of endowment asymmetry is sufficiently small and thus our results concerning the attainment of global free trade stays unchanged. Therefore, since free trade is out of reach, our discussion revolves around the CPNE conditions for bilateral FTAs.

Before proceeding, we rule out the prohibitive tariffs by assuming $\theta < \frac{5}{3}$. This implies that our analysis in the CU game remains unchanged since the same constraint was valid in any case in our previous discussion. In the FTA game, our asymmetric structure $e_s = \theta e \geq e_m = \frac{1}{2}\theta e + \frac{1}{2}e \geq e_l = e$ implies two possible scenarios: (i) $t_m(sm) = t_m(\widetilde{sm}) = 0$ while optimal tariffs are positive in all other regimes when $\frac{3}{2} > \theta \geq \frac{5}{4}$; (ii) both $t_m(sm) = t_m(\widetilde{sm}) = 0$ and $t_s(sm) = t_s(\widetilde{sm}) = 0$ while positive optimal tariffs obtain in all other regimes when $\frac{5}{3} > \theta \geq \frac{3}{2}$. As we see from these two scenarios that, allowing for higher degree of asymmetry along with zero external tariff only changes the welfare levels under under $\langle sm \rangle$ (or $\langle \widetilde{sm} \rangle$ under discrimination).

Recall that, when $\theta < \frac{5}{4}$, $\langle sm \rangle$ is the unique CPNE under the WTO-consistent scenario while $\langle \widetilde{sm} \rangle$ or $\left\langle \widetilde{sl} \right\rangle$ is the CPNE under the tariff discrimination scenario when the degree of asymmetry is sufficiently large. Under the WTO-consistent scenario, country l has no incentive to engage in any trade liberalization efforts. As we increase the degree of asymmetry, an important question is whether $\langle sm \rangle$ continues to be a CPNE. The critical incentive here is the unilateral incentive of country m to break this agreement: $\langle sm \rangle$ is a CPNE under the WTO-consistent scenario when $\theta \leq \theta_m(sm - \Phi)$. Similarly, under the tariff discrimination scenario, the unilateral incentive of

12 The term "Open Regionalism" originated from Asia Pacific economic cooperation in the late 1980s and 1990s and was proposed as a form of PTA formation: the removal of barriers to and the encouragement of regional cooperation without discrimination against outsiders. In other words, member countries would not only eliminate internal tariffs but also practice free trade towards non-members. Our results suggest that it is indeed a possibility in a competing exporters framework with a sufficient degree of endowment asymmetry.

countries m and l to break the FTA determines the equilibrium condition for $\langle\widetilde{sm}\rangle$ and $\left\langle\widetilde{sl}\right\rangle$. The following result is represented in Figure 9:

Proposition 10. *Suppose* $e_s = \theta e \geq e_m = \frac{1}{2}\theta e + \frac{1}{2}e \geq e_l = e$, *where* $\theta < \frac{5}{3}$ *and countries impose zero external tariffs whenever parameters are such that their optimal tariffs are not positive. Then,*

(i) Under the WTO-consistent scenario, if $\theta < \theta_m(sm-\Phi)$ *the equilibrium outcome is the FTA* $\langle sm\rangle$.

(ii) Under tariff discrimination, the FTAs $\langle\widetilde{sm}\rangle$ *and* $\left\langle\widetilde{sl}\right\rangle$ *are equilibria of the FTA game when* $\theta \leq \theta_l(\widetilde{sl}-\Phi)$ *while the FTA* $\langle\widetilde{sm}\rangle$ *is the unique outcome when* $\theta > \theta_l(\widetilde{sl}-\Phi)$.

Insert Figure 9

The above proposition argues that, when we allow for higher degrees of asymmetry with zero external tariff constraint, requiring MFN on the part of the FTA non-member country is world welfare improving when $\theta < \theta_m(sm-\Phi)$. When the asymmetry is very large ($\theta > \theta_m(sm-\Phi)$), while bilateral FTAs arise in the equilibrium under the tariff discrimination scenario, there exists no CPNE in the WTO-consistent scenario. Thus, under such a case, we are not able to make any meaningful comparison.

6.4. Endowment asymmetry

In this section, we allow a more flexible structure for the endowment asymmetry. To this end, we assume that $e_l = e \leq e_m = \theta^m e \leq e_s = \theta^s e$, where $\theta^s < \frac{5}{4}$ holds in the FTA game and $\theta^s < \frac{5}{3}$ holds in the CU game. Before proceeding, let $\theta_i^s(r-v)$ denotes the critical threshold of asymmetry (as a function of θ^s) at which country i is indifferent between trade regime r and v. In order to examine whether MFN requirement facilitates the attainment of global free trade with the new structure of the endowment asymmetry, our primary focus is on the CPNE condition of global free trade in the FTA and

CU games under both WTO-consistent and tariff discrimination scenarios. We begin our analysis with the FTA game.

6.4.1. FTA game

In order to save space, we move directly to the following lemma that highlights the critical role of free riding incentives in the FTA formation game:

Lemma 4. *Suppose that* $e_l = e \leq e_m = \theta^m e \leq e_s = \theta^s e$, *where* $\theta^s < \frac{5}{4}$. *Then, the following results hold under the WTO-consistent scenario as well as tariff discrimination:*

(i) Country s has no incentive to engage in a unilateral or coalitional deviations from free trade.

(ii) Country m prefers being hub relative to being an FTA member.

The first part of the above lemma provides a confirmation for our previous results that the small importing country has strong preference for freer trade and thus has no incentive to unilaterally or jointly deviate from free trade. This is because its volume of export is large while its volume of import is small relative to other countries. Similarly, the second part of the above lemma confirms that, even when medium and large importers have incentives to jointly exclude the small importer from their free trade network, country m have an incentive to further deviate to become hub under $\langle mh \rangle$ or $\left\langle \widetilde{mh} \right\rangle$ and thus these exclusion incentives go unexercised as before. As a result, for global free trade to arise as a CPNE, the unilateral deviation incentives of medium and large importers are critical. Here, it is important to note that, since country l's import volume is larger relative to the one of country m, the unilateral deviation of country l from free trade determines the CPNE condition for free trade.

As before, when the degree of asymmetry is sufficiently large, country l has an incentive to unilaterally deviate from its announcements $\{s, m\}$ to $\{s, \phi\}$

or $\{\phi, m\}$ that leads to a hub and spoke regime under which countries s or m is a hub and it itself is a spoke:

$$\Delta w_l(F - sh) < 0 \text{ when } \theta > \theta_l^s(F - sh)$$
$$\Delta w_l(F - mh) < 0 \text{ when } \theta > \theta_l^s(F - mh)$$

where $\theta_l^s(F - sh) > \theta_l^s(F - mh)$. Only one unilateral deviation incentive remains to be examined: unilateral deviation of country l from $\{s, m\}$ to $\{\phi, \phi\}$ converting free trade to $\langle sm \rangle$ and $\langle \widetilde{sm} \rangle$ and it happens when country l is sufficiently large importer relative to other countries:

$$\Delta w_l(F - sm) < 0 \text{ when } \theta > \theta_l^s(F - sm) \text{ and}$$
$$\Delta w_l(F - \widetilde{sm}) < 0 \text{ when } \theta > \theta_l^s(F - \widetilde{sm}) \tag{20}$$

We next show that $\theta_l^s(F - \widetilde{sm}) < \theta_l^s(F - sm) < \theta_l^s(F - mh)$ from direct calculation and can establish the following proposition:

Proposition 11. *Suppose* $e_l = e \leq e_m = \theta^m e \leq e_s = \theta^s e$, where $\theta^s < \frac{5}{4}$. *When* $\theta_l^s(F - \widetilde{sm}) < \theta^s < \theta_l^s(F - sm)$, *global free trade fails to arise under tariff discrimination while it is the equilibrium outcome under the WTO-consistent scenario.*

The above result is represented in Figure 10 that generalizes our previous result: MFN imposes a constraint on country's optimal tariffs, reducing its incentive for unilaterally opting out of global free trade. Thus, by increasing the likelihood of each country voluntarily choosing to enter into international trade agreements, the MFN principle acts as a catalyst for further trade liberalization.

Insert Figure 10 here

Next, we execute a similar analysis for the CU game.

6.4.2. CU game

The following lemma highlights the critical role of the exclusion incentives for the equilibrium condition of global free trade:

Lemma 5. *Suppose that* $e_l = e \leq e_m = \theta^m e \leq e_s = \theta^s e$, where $\theta^s < \frac{5}{3}$. *Then, the following hold:*

(i) While countries s *and* m *have no incentive to deviate unilaterally from free trade, country* l *has such an incentive when the degree of asymmetry is sufficiently large:* $\Delta w_l(F - sm^u) < 0$ *when* $\theta^s > \theta_l^s(F - sm^u)$ *and* $\Delta w_l(F - \widetilde{sm^u}) < 0$ *when* $\theta^s > \theta_l^s(F - \widetilde{sm^u})$.

(ii) Country s *has no incentive to exclude country* l *from free trade:* $\Delta w_s(F - sm^u) > 0$ for all θ^s.

(iii) The incentives of countries m *and* l *to exclude country* s *are stronger relative to the incentives of countries* s *and* l *to exclude country* m.

The first part of the above lemma argues that, since CU members' common external tariff is higher relative to FTA members' individual tariffs, the free riding incentive under the CU game is weaker relative to FTA game. Only the large importing country has such an incentive when the degree of asymmetry in endowments is very large. As before, due to its unconstrained ability, the large importing country's free riding incentive is stronger under discrimination. The second and third parts of the lemma inform us that the critical exclusion incentive for free trade is the joint incentives of countries m and l to exclude country s. This is primarily due to stronger terms of trade motivations because of their relatively large import volumes from country s and relatively small export volumes to country s. In the WTO-consistent scenario, country l's incentive binds for such joint exclusion incentive:

$$\Delta w_m(F - ml^u) < 0 \text{ when } \theta^s > \theta_m^s(F - ml^u) \text{ and} \quad (21)$$
$$\Delta w_l(F - ml^u) < 0 \text{ when } \theta^s > \theta_l^s(F - ml^u) \text{ where } \theta_m^s(F - ml^u) < \theta_l^s(F - ml^u)$$

Since neither country m nor l has an incentive to further deviate from their announcements to the ones leading to no agreement, the initial deviation is self-enforcing. Moreover, we show that the exclusion incentive is stronger than the free riding incentive of country l: $\theta_l^s(F-ml^u) < \theta_l^s(F-sm^u)$. As a result, *the announcement profile leading to global free trade is a CPNE in WTO-consistent scenario only when* $\theta^s < \theta_l^s(F-ml^u)$.

Now consider the tariff discrimination scenario under which country l benefits from discrimination while country m is worse off under $\langle ml^u \rangle$ relative to MFN. This is a critical difference because, unlike the WTO-consistent scenario, it is the country m's incentive (rather than country l) that binds for the exclusion incentive:

$$\Delta w_m(F-\widetilde{ml^u}) < 0 \text{ when } \theta^s > \theta_m^s(F-\widetilde{ml^u}) \text{ and} \tag{22}$$
$$\Delta w_l(F-\widetilde{ml^u}) < 0 \text{ when } \theta^s > \theta_l^s(F-\widetilde{ml^u}) \text{ where } \theta_l^s(F-\widetilde{ml^u}) < \theta_m^s(F-\widetilde{ml^u})_m$$

As in the WTO-consistent scenario, neither country m nor l has an incentive to further deviate from their announcements to the ones leading to no agreement and this joint deviation is self-enforcing. However, unlike the WTO-consistent scenario, free riding incentive of country l can be stronger than the exclusion incentive when the degree of endowment asymmetry is sufficiently large because of the ability to discriminate under the tariff discrimination scenario. Therefore, we find that *the announcement profile leading to global free trade is a CPNE in tariff discrimination scenario when* $\theta_s \leq \min\{\theta_m^s(F-\widetilde{ml^u}), \theta_l^s(F-\widetilde{sm^u})\}$. The following result (represented in Figure 11) summarizes our discussion:

Proposition 12. *Suppose* $e_l = e \leq e_m = \theta^m e \leq e_s = \theta^s e$, where $\theta^s < \frac{5}{3}$. *When* $\theta_l^s(F-ml^u) < \theta^s < \min\{\theta_m^s(F-\widetilde{ml^u}), \theta_l^s(F-\widetilde{sm^u})\}$, *while global free trade fails to arise under the WTO-consistent scenario, it is the equilibrium outcome under tariff discrimination.*

The above two propositions confirm our previous findings under a more flexible asymmetric endowment structure: while MFN principle facilitates the prospect of global free trade in the FTA formation game by limiting the free riding incentives, it hampers the prospect of global free trade in the CU formation game in which one country is *deliberately excluded* by the other two countries.

Insert Figure 11

7. 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. To some degree, these two central GATT clauses clash with one another: although PTAs sanction discrimination against non-member countries, the MFN principle requires countries not participating in any PTAs themselves to refrain from discrimination amongst PTA participants. Motivated by this observation, we have developed a model that addresses two major questions. One, is there a case for allowing PTA non-members to tariff discriminate amongst PTA members? Two, does it matter whether a non-member contemplating such discrimination has *voluntarily* chosen to stay out of a PTA (knowing full-well that their non-participation will result in them facing discriminatory treatment) or finds itself *deliberately excluded* by PTA members? We show that holding the PTA non-member to MFN is desirable only when it chooses to voluntarily stay out of the trade agreement between the other two countries. Thus, there is no efficiency case for imposing MFN on countries that find themselves excluded from a Customs Union that they wish to participate in but are barred from doing so by other member countries. Our results are robust to different equilibrium concepts,

more general endowment asymmetry structure, and a fully endogenous PTA formation game that allows countries to choose between FTAs and CUs.

While we employ a static PTA formation model in this paper, a dynamic PTA formation model that uses the Markov Perfect equilibrium concept as in Seidman (2009), Lake and Yildiz (2016) and Lake et al. (2020), can be more appealing from the perspective of deriving the entire equilibrium path of agreements. We leave this for future research.

8. Appendix

In this Appendix, we detail all supporting calculations as well as proofs of lemmas and propositions.

8.1. Supporting calculations

First, we report the welfare levels as functions of an arbitrary tariff vectors. Using the welfare functions below and tariff levels reported in the text, we can calculate the formulae for optimum welfare levels under all possible regimes.

8.1.1. 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

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

8.2. Proofs of Lemmas and Propositions

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

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): **(Attractiveness of bilateral FTA formation)** Since $\Delta w_l(ml - sl) > 0$, $\Delta w_m(ml - sm) > 0$, and $\Delta w_s(sl - sm) > 0$ for all $1 \leq \theta \leq 5/4$, we conclude that countries have an incentive to form a bilateral FTA with the larger importer relative to the smaller one.

Part (*ii*): **(Free riding incentive)** We find that the small and medium countries do not have an incentive to free ride and stay an outsider under a bilateral FTA relative to free trade since $\Delta w_s(F-ml) > 0$ and $\Delta w_m(F-sl) > 0$ hold for all $1 \leq \theta \leq 5/4$. However, the large importer does have such incentive when the degree of endowment asymmetry is sufficiently large since $\Delta w_l(F - sm) < 0$ holds only when $\theta > \theta_l(F - sm) \cong 1.053$.

Part (*iii*): **(Exclusion incentives)** We find that the small country does not participate in any joint deviation from free trade to exclude the third country via a bilateral FTA, $\Delta w_s(F - sm) > 0$ and $\Delta w_s(F - sl) > 0$ hold for all $1 \leq \theta \leq 5/4$. However, the medium and large countries do have such incentives when the degree of endowment asymmetry is sufficiently large since for the medium country, $\Delta w_m(F-ml) < 0$ only when $\theta > \theta_m(F-ml) \cong 1.1304$ while for the large country, $\Delta w_l(F - ml) < 0$ only when $\theta > \theta_l(F - ml) \cong 1.1023$.

Part (*iv*): **(Attractiveness of being hub)** All countries prefer being the hub under a hub and spoke regime relative to all other trade policy regimes since $\Delta w_i(ih - \Phi) > 0$, $\Delta w_i(ih - F) > 0$ and $\Delta w_i(ih - ij) > 0$ for all $1 \leq \theta \leq 5/4$ and $i = s, m, l$.

Part (*v*): **(Attractiveness of being spoke)** We find that the large importer prefers being a non-member under a bilateral FTA to being a spoke under a

hub and spoke regime, since $\Delta w_l(sh - sm) < 0$ and $\Delta w_l(mh - sm) < 0$ for all $1 \leq \theta \leq 5/4$. The two smaller countries do so only when the degree of endowment asymmetry is sufficiently small: for the medium country, $\Delta w_m(lh - sl) < 0$ only when $\theta > \theta_m(lh - sl) \cong 1.204$ and $\Delta w_m(sh - sl) < 0$ only when $\theta > \theta_m(sh - sl) \cong 1.228$. For the small country, $\Delta w_s(lh - ml) > 0$ only when $\theta > \theta_s(lh - ml) \cong 1.0454$ and $\Delta w_s(mh - ml) > 0$ only when $\theta > \theta_s(lh - ml) \cong 1.0450$.

■

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 the following:

- $\Delta w_s(sl - \Phi) > 0$, $\Delta w_m(sl - \Phi) > 0$, and $\Delta w_l(sl - \Phi) > 0$ for all $1 \leq \theta \leq 5/4$. As such, we conclude that relative to no agreement $\langle \Phi \rangle$, bilateral FTA $\langle sl \rangle$ is Pareto-improving.

- $\Delta w_s(sm - \Phi) > 0$, $\Delta w_m(sm - \Phi) > 0$, and $\Delta w_l(sm - \Phi) > 0$ for all $1 \leq \theta \leq 5/4$. As such, we conclude that relative to no agreement $\langle \Phi \rangle$, bilateral FTA $\langle sm \rangle$ is Pareto-improving.

- $\Delta w_l(ml - \Phi) > 0$ and $\Delta w_m(ml - \Phi) > 0$ for all $1 \leq \theta \leq 5/4$ while $\Delta w_s(ml - \Phi) > 0$ holds only when $\theta < \theta_s(ml - \Phi) \cong 1.118$. As such, we conclude that relative to no agreement $\langle \Phi \rangle$, bilateral FTA $\langle ml \rangle$ is Pareto-improving only when $\theta < \theta_s(ml - \Phi)$.

■

Proof of Proposition 2. We prove Proposition 2 using the results from Lemma 2, the discussion in the main text and the following inequalities:

Part (i): First we consider the announcement profiles leading to global free trade $\langle F \rangle$. We show in Lemma 2 that the small country has no incentive to unilaterally or jointly deviate from global free trade. As such, a coalitional deviation away from $\langle F \rangle$ will involve the larger countries. We find that these

countries indeed have an incentive to jointly deviate to an FTA between themselves, excluding the small country, when the endowment asymmetry is above the following two thresholds respectively: - $\Delta w_l(F-ml) < 0$ when $\theta > \theta_l(F-ml) \cong 1.102$

- $\Delta w_m(F-ml) < 0$ when $\theta > \theta_m(F-ml) \cong 1.130$

Since $\theta_m(F-ml) > \theta_l(F-ml)$, $\theta_m(F-ml)$ is the binding cutoff for this joint deviation. From the discussion in the main text, we see that the only possible type of self-enforcing deviation from $\langle F \rangle$ is unilateral deviation. The large country is the only one with such incentives. When the degree of asymmetry is sufficiently large, it has an incentive to unilaterally deviate to a hub and spoke regime where the small or medium country is a hub:

- $\Delta w_l(F-sh) < 0$ when $\theta > \theta_l(F-sh) \cong 1.213$

- $\Delta w_l(F-mh) < 0$ when $\theta > \theta_l(F-mh) \cong 1.169$ where $\Delta w_l(F-sh) > \Delta w_l(F-mh)$. From part (v) of Lemma 2, we see that these regimes are not stable still the large importer has an incentive to deviate away from being a spoke to become a non-member under a bilateral FTA.

Parts (ii) and (iii): So the remaining unilateral deviation incentive is the deviation of the large country to being the outsider of an FTA between the two smaller countries, $\langle sm \rangle$. This happens when the large country is a sufficiently large importer relative to the other countries:

- $\Delta w_l(F-sm) < 0$ when $\theta > \theta_l(F-sm) \cong 1.0531$. where as discussed in the main text, this FTA is a CPNE since countries have no incentives to deviate from this when $\theta > \theta_l(F-sm)$.

We also consider trade regime $\langle sl \rangle$. From the discussion in the main text, we conclude that the announcement profile that leads to this trade regime is a CPNE whenever $\theta_l(F-sm) \leq \theta \leq \theta_m(lh-sl) \cong 1.204$.

■

Proof of Proposition 3. 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:

- $\Delta w_s(\widetilde{sl}-\Phi) > 0$; $\Delta w_l(\widetilde{sl}-\Phi) > 0$ and $\Delta w_m(\widetilde{sl}-\Phi) > 0$ for all $1 \leq \theta \leq 5/4$. As such, we conclude that relative to no agreement $\langle\Phi\rangle$, bilateral FTA $\left\langle\widetilde{sl}\right\rangle$ is Pareto-improving.

- $\Delta w_s(\widetilde{sm}-\Phi) > 0$; $\Delta w_m(\widetilde{sm}-\Phi) > 0$ and $\Delta w_l(\widetilde{sm}-\Phi) > 0$ for all $1 \leq \theta \leq 5/4$. As such, we conclude that relative to no agreement $\langle\Phi\rangle$, bilateral FTA $\langle\widetilde{sm}\rangle$ is Pareto-improving.

- $\Delta w_m(\widetilde{ml}-\Phi) > 0$, and $\Delta w_l(\widetilde{ml}-\Phi) > 0$ for all $1 \leq \theta \leq 5/4$ while $\Delta w_s(\widetilde{ml}-\Phi) > 0$ only when $\theta < \theta_s(\widetilde{ml}-\Phi) \cong 1.122$ where $\theta_s(ml-\Phi) \cong 1.118 < \theta_s(\widetilde{ml}-\Phi) \cong 1.122$. As such, we conclude that relative to no agreement $\langle\Phi\rangle$, bilateral FTA $\left\langle\widetilde{ml}\right\rangle$ is Pareto-improving only when $\theta < \theta_s(\widetilde{ml}-\Phi)$.

- Since $\theta_s(ml-\Phi) < \theta_s(\widetilde{ml}-\Phi)$, the parameter range over which the FTA between medium and large importing countries is Pareto-improving expands under tariff discrimination.

■

Proof of Proposition 4. We prove Proposition 4 using the results from Proposition 3, the discussion in the main text, and the following inequalities:

Part (i): First we consider the announcement profiles leading to global free trade $\langle F\rangle$. We continue to find that the small country has no incentive to unilaterally or jointly deviate from global free trade. As such, a coalitional deviation away from $\langle F\rangle$ will involve the larger countries. We find that these countries indeed have an incentive to jointly deviate to an FTA between themselves, excluding the small country, when the endowment asymmetry is above the following two thresholds respectively:

- $\Delta w_l(F-\widetilde{ml}) < 0$ when $\theta > \theta_l(F-\widetilde{ml}) \cong 1.076$;
- $\Delta w_m(F-\widetilde{ml}) < 0$ when $\theta > \theta_m(F-\widetilde{ml}) \cong 1.219$;

Since $\theta_m(F-\widetilde{ml}) > \theta_l(F-\widetilde{ml})$, $\theta_m(F-\widetilde{ml})$ is the binding cutoff for this joint deviation. From the discussion in the main text, we see that the only possible type of self-enforcing deviation from $\langle F\rangle$ is unilateral deviation. The large country is the only one with such incentives. When the degree of asymmetry is sufficiently large, it has an incentive to unilaterally deviate to a hub and spoke regime where the small or medium country is a hub:

- $\Delta w_l(F-sh) < 0$ when $\theta > \theta_l(F-sh) \cong 1.213$

- $\Delta w_l(F-mh) < 0$ when $\theta > \theta_l(F-mh) \cong 1.169$ where $\Delta w_l(F-sh) > \Delta w_l(F-mh)$. We again see that these regimes are not stable still the large importer has an incentive to deviate away from being a spoke to become a non-member under a bilateral FTA.

Part (*ii*): As we know under our discussion from Proposition 2, the large importing country's choice is pivotal for global free trade. So the remaining unilateral deviation incentive is the deviation of the large country to being the outsider of an FTA between the two smaller countries, $\langle\widetilde{sm}\rangle$. This happens when the large country is a sufficiently large importer relative to the other countries:

- $\Delta w_l(F-\widetilde{sm}) < 0$ when $\theta > \theta_l(F-\widetilde{sm}) \cong 1.0526$ where $\theta_l(F-\widetilde{sm}) \cong 1.0526 < \theta_l(F-sm) \cong 1.0531$

From the above, we see that the large country has a stronger free riding incentive (strong unilateral incentive to deviate away from global free trade) when it is free to discriminate as a non-member country relative to the MFN case.

Similar to the WTO-consistent scenario, the only possible remaining announcement profile that can be CPNE (other than $\langle F\rangle$ and $\langle\widetilde{sm}\rangle$) are the ones leading to $\left\langle\widetilde{sl}\right\rangle$. From the discussion in the main text, we see that $\left\langle\widetilde{sl}\right\rangle$ is a CPNE whenever $\theta > \theta_l(F-\widetilde{sm}) \cong 1.0526$.

■

Proof of Lemma 3. 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): Under the WTO-benchmark scenario: $\Delta w_l(ml^u - sl^u) > 0$, $\Delta w_m(ml^u - sm^u) > 0$, and $\Delta w_s(sl^u - sm^u) > 0$ for all $1 \leq \theta \leq 5/4$.

Under the tariff discrimination scenario: $\Delta w_l(\widetilde{ml^u} - \widetilde{sl^u}) > 0$, $\Delta w_m(\widetilde{ml^u} - \widetilde{sm^u}) > 0$, and $\Delta w_s(\widetilde{sl^u} - \widetilde{sm^u}) > 0$ for all $1 \leq \theta \leq 5/4$.

Part (*ii*): Under the WTO-benchmark scenario: For the small country, $\Delta w_s(sl^u - \Phi) > 0$, $\Delta w_s(sm^u - \Phi) > 0$, $\Delta w_s(F - sm^u) > 0$, $\Delta w_s(F - sl^u) > 0$, and $\Delta w_s(F - \Phi) > 0$ hold for all $1 \leq \theta \leq 5/4$.

Under the tariff discrimination scenario: For the small country, $\Delta w_s(\widetilde{sl^u} - \Phi) > 0$, $\Delta w_s(\widetilde{sm^u} - \Phi) > 0$, $\Delta w_s(F - \widetilde{sm^u}) > 0$, $\Delta w_s(F - \widetilde{sl^u}) > 0$, and $\Delta w_s(F - \Phi) > 0$ hold for all $1 \leq \theta \leq 5/4$.

Part (*iii*): Under the WTO-benchmark scenario: $\Delta w_s(ml^u - \Phi) < 0$, $\Delta w_m(sl^u - \Phi) < 0$, $\Delta w_l(sm^u - \Phi) < 0$, $\Delta w_s(ml^u - sm^u) < 0$, $\Delta w_s(ml^u - sl^u) < 0$, $\Delta w_m(sl^u - sm^u) < 0$, $\Delta w_m(sl^u - ml^u) < 0$, $\Delta w_l(sm^u - sl^u) < 0$, $\Delta w_l(sm^u - ml^u) < 0$, $\Delta w_s(F - ml^u) > 0$, $\Delta w_m(F - sl^u) > 0$, and $\Delta w_l(F - sm^u) > 0$ hold for all $1 \leq \theta \leq 5/4$.

Under the tariff discrimination scenario: $\Delta w_s(\widetilde{ml^u} - \Phi) < 0$, $\Delta w_m(\widetilde{sl^u} - \Phi) < 0$, $\Delta w_l(\widetilde{sm^u} - \Phi) < 0$, $\Delta w_s(\widetilde{ml^u} - \widetilde{sm^u}) < 0$, $\Delta w_s(\widetilde{ml^u} - \widetilde{sl^u}) < 0$, $\Delta w_m(\widetilde{sl^u} - \widetilde{sm^u}) < 0$, $\Delta w_m(\widetilde{sl^u} - \widetilde{ml^u}) < 0$, $\Delta w_l(\widetilde{sm^u} - \widetilde{sl^u}) < 0$, $\Delta w_l(\widetilde{sm^u} - \widetilde{ml^u}) < 0$, $\Delta w_s(F - \widetilde{ml^u}) > 0$, $\Delta w_m(F - \widetilde{sl^u}) > 0$, and $\Delta w_l(F - \widetilde{sm^u}) > 0$ hold for all $1 \leq \theta \leq 5/4$.

■

Proof of Proposition 6. We show this proof using Lemma 3 and the following inequalities: We have $\Delta w_l(F - ml^u) < 0$ when $\theta > \theta_l(F - ml^u) \cong 1.0312$, $\Delta w_m(F - ml^u) < 0$ when $\theta > \theta_m(F - ml^u) \cong 1.0304$, and $\Delta w_m(F - sl^u) < 0$ hold for all $1 \leq \theta \leq 5/4$.

■

Proof of Proposition 7. We can show this proof using Lemma 3 and the following inequalities:

$\Delta w_l(F - \widetilde{ml^u}) < 0$ when $\theta > \theta_l(F - \widetilde{ml^u}) \cong 1.0213$, $\Delta w_m(F - \widetilde{ml^u}) < 0$ when $\theta > \theta_m(F - \widetilde{ml^u}) \cong 1.0541$, and $\Delta w_m(F - \widetilde{sl^u}) < 0$ hold for all $1 \leq \theta \leq 5/4$.

■

Proof of Proposition 8. Using the above welfare formulae (as functions of arbitrary tariff vectors) and assuming $e_s = \theta e \geq e_m = \frac{1}{2}\theta e + \frac{1}{2}e \geq e_l = e$, where $\theta < \frac{5}{4}$, we show the following:

- The announcement profile leading to $\langle \Phi \rangle$ is not a SNE under WTO-consistent and tariff discrimination scenarios since countries m and l have an incentive to coalitionally deviate from their announcements to the ones leading to $\langle ml \rangle$ (or $\left\langle \widetilde{ml} \right\rangle$): $\Delta w_m(ml - \Phi) > 0$, $\Delta w_l(ml - \Phi) > 0$, $\Delta w_m(\widetilde{ml} - \Phi) > 0$ and $\Delta w_l(\widetilde{ml} - \Phi) > 0$ for all $\frac{5}{4} \leq \theta$

- The announcement profiles leading to $\langle sl \rangle$ and $\left\langle \widetilde{sl} \right\rangle$ are not an SNE under WTO-consistent and tariff discrimination scenarios since countries m and l have incentives to jointly deviate to announcement profiles leading to $\langle ml \rangle$ and $\left\langle \widetilde{ml} \right\rangle$): $\Delta w_m(ml - sl) > 0$, $\Delta w_l(ml - sl) > 0$, $\Delta w_m(\widetilde{ml} - \widetilde{sl}) > 0$ and $\Delta w_l(\widetilde{ml} - \widetilde{sl}) > 0$ obtain for all θ.

- The announcement profiles leading to $\langle ml \rangle$ and $\left\langle \widetilde{ml} \right\rangle$ are not an SNE under WTO-consistent and tariff discrimination scenarios since the following coalitional deviations cover the entire parameter space: (i) joint deviation of countries s and l from their announcement profiles to the ones leading to $\langle lh \rangle$ happens when $\theta > \theta_s(lh - \widetilde{ml}) \simeq 1.0461$ holds: $\Delta w_l(lh - ml) > 0$ and $\Delta w_l(lh - \widetilde{ml}) > 0$ obtain for all θ while $\Delta w_s(lh - ml) > 0$ when $\theta > \theta_s(lh - ml) \simeq 1.0454$ holds and $\Delta w_s(lh - \widetilde{ml}) > 0$ when $\theta > \theta_s(lh - \widetilde{ml}) \simeq 1.0461$ holds and (*ii*) the coalitional deviations of all countries from their announcement profiles $\langle ml \rangle$ to the ones leading to $\langle F \rangle$ happen when $\theta < \theta_l(F - ml) \simeq 1.1023$ holds: $\Delta w_s(F - ml) > 0$ obtains for all θ while

$\Delta w_m(F-ml) > 0$ when $\theta < \theta_m(F-ml) \simeq 1.1304$ holds and $\Delta w_l(F-ml) > 0$ when $\theta < \theta_l(F-ml) \simeq 1.1023$ holds. Similarly, the coalitional deviations of all countries from their announcement profile under $\left\langle \widetilde{ml} \right\rangle$ to the one leading to $\langle F \rangle$ happen when $\theta < \theta_l(F-\widetilde{ml}) \simeq 1.0758$ holds: $\Delta w_s(F-ml) > 0$ obtains for all θ while $\Delta w_m(F-\widetilde{ml}) > 0$ when $\theta < \theta_m(F-ml) \simeq 1.2195$ holds and $\Delta w_l(F-\widetilde{ml}) > 0$ when $\theta < \theta_l(F-\widetilde{ml}) \simeq 1.0758$ holds.

- The announcement profiles leading to $\langle sh \rangle$ and $\langle mh \rangle$ are not an SNE since country l has an incentive to unilaterally break its FTA with the hub country: $\Delta w_l(mh-sm) < 0$, $\Delta w_l(sh-sm) < 0$, $\Delta w_l(\widetilde{mh}-sm) < 0$ and $\Delta w_l(\widetilde{sh}-sm) < 0$ for all θ.

- The announcement profile leading to $\langle lh \rangle$ is not an SNE under both scenarios since countries s and m have incentives to jointly form the final FTA leading to $\langle F \rangle$: $\Delta w_s(F-lh) > 0$ and $\Delta w_m(F-lh) > 0$ for all θ.

- Country l has an incentive to unilaterally deviate from its announcement leading to $\langle F \rangle$ to the one leading to $\langle sm \rangle$ and $\langle \widetilde{sm} \rangle$ under both WTO-consistent and tariff discrimination scenarios when the degree of asymmetry is sufficiently large : $\Delta w_l(F-sm) < 0$ when $\theta > \theta_l(F-sm) \simeq 1.0531$ and $\Delta w_l(F-\widetilde{sm}) < 0$ when $\theta > \theta_l(F-\widetilde{sm}) \simeq 1.0526$. When $\theta \leq \theta_l(F-sm)$ and $\theta \leq \theta_l(F-\widetilde{sm})$ hold, there exist no other unilateral or coalitional announcement deviations and thus the announcement profile leading to $\langle F \rangle$ is an SNE.

- When $\theta > \theta_l(F-sm) \simeq 1.0531$ under the WTO-consistent scenario, there exist no unilateral or coalitional deviations from the announcement profile under $\langle sm \rangle$ and thus it is an SNE.

- Unlike the WTO-consistent scenario, when non-member countries are free to discriminate, all countries have incentives to jointly deviate to the announcement profile leading to $\langle lh \rangle$ when $\theta > \theta_m(lh-\widetilde{sm})$ holds. Therefore, the announcement profile leading to $\langle \widetilde{sm} \rangle$ is an SNE when $\theta_l(F-\widetilde{sm}) < \theta \leq \theta_m(lh-\widetilde{sm})$.

■

Proof of Proposition 9. Based on Propositions 6 and 7 we know that self-enforcing exclusion incentives of countries m and l happen in the WTO-consistent scenario when $\theta > \theta_l(F - ml^u) \cong 1.0312$ and in the tariff discrimination scenario when $\theta > \theta_m(F - \widetilde{ml^u}) \cong 1.0541$. Since exclusion incentive is a coalitional deviation involving the majority of countries, they determine the equilibrium condition for free trade.

In the WTO-consistent scenario, we also know from Propositions 2 and 6 that, when free trade fails to be an equilibrium, $\langle ml^u \rangle$ is a CPNE in the CU game while $\langle sl \rangle$ and $\langle sm \rangle$ arise a CPNE in the FTA game. Consider case (i) first that and assume that $\langle sm \rangle$ and $\langle \widetilde{sm} \rangle$ arise as the CPNE in the FTA game. We establish that the medium country prefers the CU $\langle ml^u \rangle$ over $\langle sm \rangle$ since $\Delta w_m(ml^u - sm) > 0$ hold for all $1 \leq \theta \leq 5/4$. The small country, however, prefers FTA formation to the CU formation since it is excluded in the CU: $\Delta w_s(ml^u - sm) < 0$. Then it boils down to the choice of the large country that prefers FTA formation (and thus $\langle sm \rangle$) to CU formation (and thus $\langle ml^u \rangle$) when the degree of endowment asymmetry is sufficiently high: $\Delta w_l(ml^u - sm) < 0$ when $\theta > \theta_l(ml^u - sm) \cong 1.0684$. Similarly, in the tariff discrimination scenario, we know from Propositions 4 and 7 that when global free trade is out of reach, $\langle ml^u \rangle$ is a CPNE in the CU game while $\langle sl \rangle$ and $\langle sm \rangle$ arise a CPNE in the FTA game. Focusing on case (i), we first show that the medium country prefers CU formation (thus $\left\langle \widetilde{ml^u} \right\rangle$) over FTA formation (thus $\langle \widetilde{sm} \rangle$) since $\Delta w_m(\widetilde{ml^u} - \widetilde{sm}) > 0$ holds for all θ. However, the opposite is true for the small country who prefers FTA formation (thus $\widetilde{sm}$) over CU formation (thus $\left\langle \widetilde{ml^u} \right\rangle$): $\Delta w_s(\widetilde{ml^u} - \widetilde{sm}) < 0$ holds for all θ. As in the WTO-consistent scenario, the choice of country l determines whether FTA formation or CU formation would be chosen and we show that large country prefers FTA formation (and thus $\langle \widetilde{sm} \rangle$) to CU formation (and thus $\left\langle \widetilde{ml^u} \right\rangle$) when the degree of endowment asymmetry is sufficiently high:

$\Delta w_l(\widetilde{ml^u} - \widetilde{sm}) < 0$ when $\theta > \theta_l(\widetilde{ml^u} - \widetilde{sm}) \cong 1.0954$. Thus, we obtain all the necessary equations for figures 8a and 8b.

Now consider figures 9a and 9b that correspond to case (ii) where $\langle sl \rangle$ and $\left\langle \widetilde{sl} \right\rangle$ arise as the CPNE in the FTA game under WTO-consistent scenario and tariff discrimination scenario, respectively. Under such a case, we show that both countries m and l prefer CU formation (and thus $\langle ml \rangle$ or $\left\langle \widetilde{ml^u} \right\rangle$) to FTA formation (thus $\langle sl \rangle$ or $\left\langle \widetilde{sl} \right\rangle$). In other words, if $\langle sl \rangle$ and $\left\langle \widetilde{sl} \right\rangle$ arise as the CPNE in the FTA game while the preference of country s is the opposite: $\Delta w_m(ml^u - sl) > 0$, $\Delta w_l(ml^u - sl) > 0$ and $\Delta w_s(ml^u - sl) < 0$ holds for all θ.

■

Proof of Proposition 10. Using the above welfare formulae (as functions of arbitrary tariff vectors) and assuming $e_s = \theta e \geq e_m = \frac{1}{2}\theta e + \frac{1}{2}e \geq e_l = e$, *where* $\theta < \frac{5}{3}$, we examine two cases:

(i) $t_m(sm) = t_m(\widetilde{sm}) = 0$ while positive optimal tariffs obtain in all other regimes when $\frac{3}{2} > \theta \geq \frac{5}{4}$

(ii) both $t_m(sm) = t_m(\widetilde{sm}) = 0$ and $t_s(sm) = t_s(\widetilde{sm}) = 0$ while positive optimal tariffs obtain in all other regimes when $\frac{5}{3} > \theta \geq \frac{3}{2}$.

We show the following:

- $\langle F \rangle$ is not a CPNE under both WTO-consistent and tariff discrimination scenarios since country l has an incentive to unilaterally deviate from its announcement leading to $\langle F \rangle$ to the one leading to $\langle sm \rangle$ (or $\langle \widetilde{sm} \rangle$): $\Delta w_l(F - sm) < 0$ and $\Delta w_l(F - \widetilde{sm}) < 0$ for all $\frac{5}{4} \leq \theta \leq \frac{5}{3}$.

- The announcement profile leading to $\langle \Phi \rangle$ is not a CPNE under both WTO-consistent and tariff discrimination scenarios since countries m and l have an incentive to coalitionally deviate from their announcements to the ones leading to $\langle ml \rangle$ (or $\left\langle \widetilde{ml} \right\rangle$): $\Delta w_m(ml - \Phi) > 0$, $\Delta w_l(ml - \Phi) > 0$, $\Delta w_m(\widetilde{ml} - \Phi) > 0$ and $\Delta w_l(\widetilde{ml} - \Phi) > 0$ for all $\frac{5}{4} \leq \theta \leq \frac{5}{3}$.

- The announcement profile leading to $\langle sh \rangle$ fails to be a CPNE under both WTO-consistent and tariff discrimination scenarios since country l has an incentive to unilaterally break its FTA with country s leading to $\langle sm \rangle$ (or $\langle \widetilde{sm} \rangle$): $\Delta w_l(sh - sm) < 0$ and $\Delta w_l(sh - \widetilde{sm}) < 0$ for all $\frac{5}{4} \leq \theta \leq \frac{5}{3}$.

- The announcement profile leading to $\langle mh \rangle$ is not a CPNE under both WTO-consistent and tariff discrimination scenarios since country l has an incentive to unilaterally break its FTA with country m leading to $\langle sm \rangle$ (or $\langle \widetilde{sm} \rangle$): $\Delta w_l(mh - sm) < 0$ and $\Delta w_l(mh - \widetilde{sm}) < 0$ for all $\frac{5}{4} \leq \theta \leq \frac{5}{3}$.

- The announcement profile leading to $\langle lh \rangle$ is not a CPNE under either scenario since countries s and m have incentives to jointly form the fila FTA leading to $\langle F \rangle$ and it is a self-enforcing deviation since neither country has an incentive to unilaterally deviate further : $\Delta w_s(F - lh) > 0$ and $\Delta w_m(F - lh) > 0$ for all $\frac{5}{4} \leq \theta \leq \frac{5}{3}$.

- The announcement profile leading to $\langle sl \rangle$ is not a CPNE under the WTO-consistent scenario since countries m and l have incentives to jointly deviate to the announcement profiles leading to a deviation to $\langle lh \rangle$ and it is a self-enforcing deviation.

- The announcement profile leading to $\left\langle \widetilde{sl} \right\rangle$ is a CPNE under the tariff discrimination scenario only when θ is not very large: $\Delta w_s(\widetilde{sl} - \Phi) \geq 0$ for all θ while $\Delta w_l(\widetilde{sl} - \Phi) \geq 0$ when $\theta \leq \theta_l(\widetilde{sl} - \Phi) \simeq 1.529$.

- The announcement profile leading to $\langle ml \rangle$ and $\left\langle \widetilde{ml} \right\rangle$ is not a CPNE since countries s and l have incentives to jointly deviate to the announcement profiles leading to a deviation to $\langle lh \rangle$ and it is a self-enforcing deviation.

- The announcement profile leading to $\langle sm \rangle$ is a CPNE under the WTO-consistent scenario only when θ is not very large: $\Delta w_s(sm - \Phi) \geq 0$ for all θ while $\Delta w_m(sm - \Phi) \geq 0$ when $\theta \leq (sm - \Phi) \simeq 1.463$.

- The announcement profile leading to $\langle \widetilde{sm} \rangle$ is a CPNE under the tariff discrimination scenario since there exists no unilateral or coalitional announcement deviation.

■

Proof of Lemma 4. Part (i): Under the WTO-consistent scenario, $\Delta w_s(F - ml) > 0$, $\Delta w_s(F - mh) > 0$, $\Delta w_s(F - lh) > 0$, $\Delta w_s(F - sm) > 0$, $\Delta w_s(F - sl) > 0$ and $\Delta w_s(F - \Phi) > 0$ for all θ^s. The same results obtain for the tariff discrimination scenario: $\Delta w_s(F - \widetilde{ml}) > 0$, $\Delta w_s(F - \widetilde{sm}) > 0$ and $\Delta w_s(F - \widetilde{sl}) > 0$ for all θ^s.

Part (ii): $\Delta w_m(hm - sm) > 0$, $\Delta w_m(hm - ml) > 0$, $\Delta w_m(hm - \widetilde{sm}) > 0$ and $\Delta w_m(hm - \widetilde{ml}) > 0$ for all θ^s.

■

Proof of Proposition 11. Country l has an incentive to unilaterally deviate from its announcement leading to $\langle F \rangle$ to the one leading to $\langle sh \rangle$ and $\langle mh \rangle$ under both WTO-consistent and tariff discrimination scenarios when the degree of asymmetry is sufficiently large: $\Delta w_l(F - sh) < 0$ when $\theta > \theta_l^s(F - sh)$ and $\Delta w_l(F - mh) < 0$ when $\theta > \theta_l^s(F - mh)$ where $\theta_l^s(F - sh) > \theta_l^s(F - mh)$ holds for all θ^m. Country l has an incentive to unilaterally deviate from its announcement leading to $\langle F \rangle$ to the one leading to $\langle sm \rangle$ and $\langle \widetilde{sm} \rangle$ under both WTO-consistent and tariff discrimination scenarios when the degree of asymmetry is sufficiently large: $\Delta w_l(F - sm) < 0$ when $\theta^s > \theta_l^s(F - sm)$ and $\Delta w_l(F - \widetilde{sm}) < 0$ when $\theta^s > \theta_l^s(F - \widetilde{sm})$. When $\theta^s \leq \theta_l^s(F - sm)$ and $\theta^s \leq \theta_l^s(F - \widetilde{sm})$ hold, there exist no other unilateral or self-enforcing coalitional announcement deviations and thus the announcement profile leading to $\langle F \rangle$ is a CPNE. Note also that $\theta_l^s(F - \widetilde{sm}) < \theta_l^s(F - sm) < \theta_l^s(F - mh)$ obtains for all θ^m.

■

Proof of Lemma 5. Part (i): $\Delta w_l(F - sm^u) < 0$ when $\theta^s > \theta_l^s(F - sm^u)$ and $\Delta w_l(F - \widetilde{sm^u}) < 0$ when $\theta^s > \theta_l^s(F - \widetilde{sm^u})$.

Part (ii): $\Delta w_s(F - sl^u) > 0$ and $\Delta w_s(F - \widetilde{sl^u}) > 0$ for all θ^s and θ^m.

Part (*iii*): In the WTO-consistent scenario, $\Delta w_m(F - ml^u) < 0$ when $\theta^s > \theta^s_m(F - ml^u)$ and $\Delta w_l(F - ml^u) < 0$ when $\theta^s > \theta^s_l(F - ml^u)$ while $\Delta w_s(F - sl^u) < 0$ when $\theta^s > \theta^s_s(F - sl^u)$ and $\Delta w_l(F - sl^u) < 0$ when $\theta^s > \theta^s_l(F - sl^u)$ where $\theta^s_m(F - ml^u) \leq \theta^s_l(F - ml^u)$, $\theta^s_l(F - sl^u) \leq \theta^s_s(F - sl^u)$ and $\theta^s_l(F - ml^u) \leq \theta^s_s(F - sl^u)$. In the tariff discrimination scenario, since $\Delta w_s(F - \widetilde{sl^u}) > 0$ for all θ^s and θ^m, we do not need to consider joint announcement deviations from free trade to the one leading to $\left\langle \widetilde{sl^u} \right\rangle$. Then, the following is sufficient: $\Delta w_m(F - \widetilde{ml^u}) < 0$ when $\theta^s > \theta^s_m(F - \widetilde{ml^u})$ and $\Delta w_l(F - \widetilde{ml^u}) < 0$ when $\theta^s > \theta^s_l(F - \widetilde{ml^u})$ hold, where $\theta^s_l(F - \widetilde{ml^u}) \leq \theta^s_m(F - \widetilde{ml^u})$.

■

Proof of Proposition 12. First, we show that countries s and m have no incentives to deviate from their announcements leading to $\langle F \rangle$ to the ones leading to $\langle \Phi \rangle$: $\Delta w_s(F - \Phi) > 0$ and $\Delta w_m(F - \Phi) > 0$ hold for all θ^s and θ^m. Combining the above expressions with the ones in the proof of Lemma 5, our proof is complete. ■

9. Figures

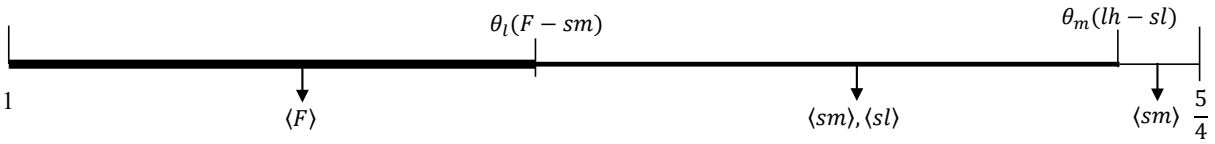

Figure 1: Equilibrium agreements under the benchmark WTO game with MFN (FTA)

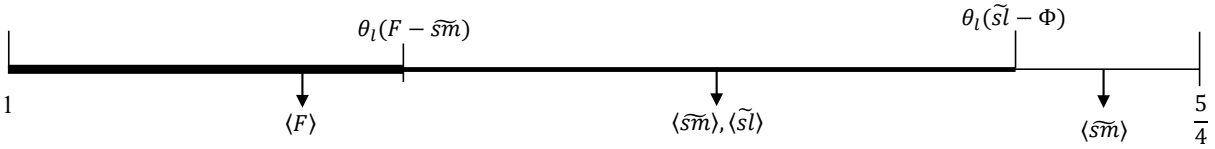

Figure 2: Equilibrium agreements with tariff discrimination (FTA)

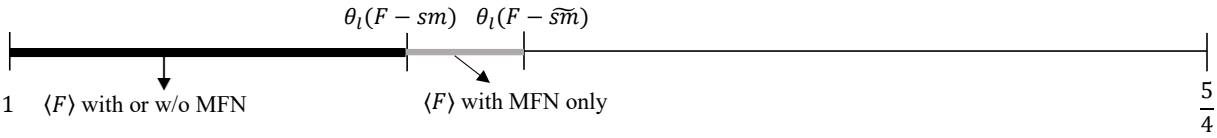

Figure 3: Free trade with MFN or with tariff discrimination (FTA)

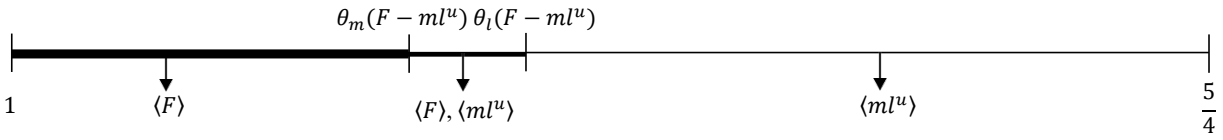

Figure 4: Equilibrium agreements under the benchmark WTO game (CU)

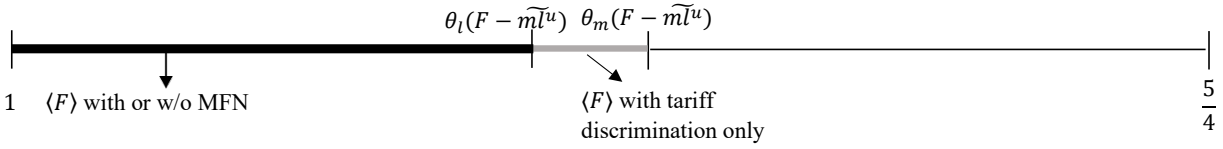

Figure 5: Free trade with MFN or with tariff discrimination (CU)

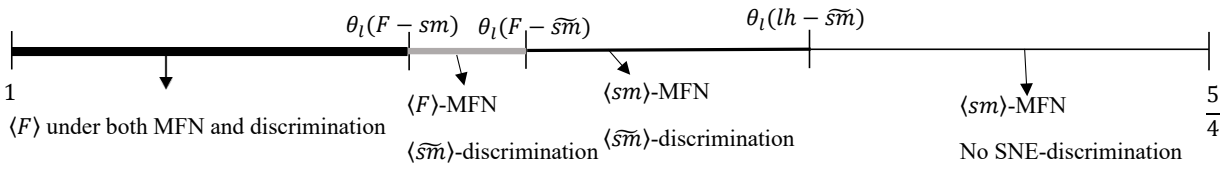

Figure 6: Strong Nash equilibrium with MFN and tariff discrimination

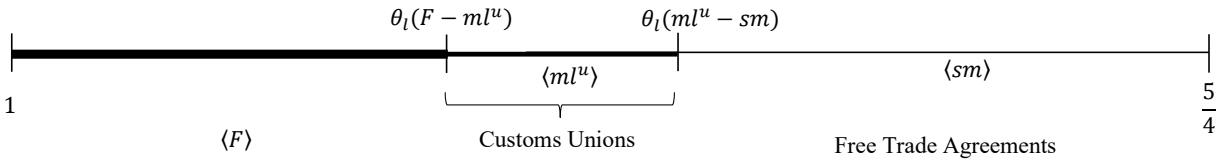

Figure 7a: Endogenous PTA formation game with MFN – Case (i): $\langle sm \rangle$ arises)

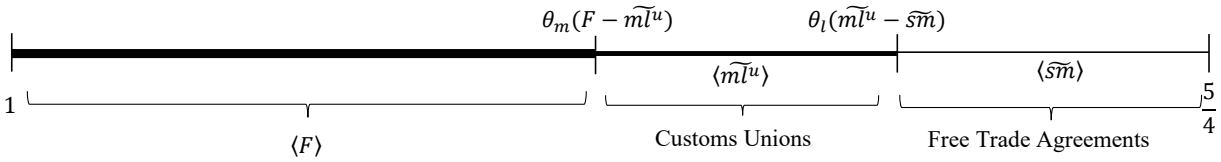

Figure 7b: Endogenous PTA formation game with tariff discrimination – Case (i): $\langle \widetilde{sm} \rangle$ arises)

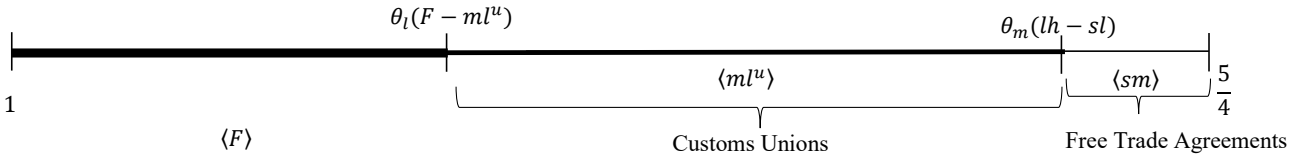

Figure 8a: Endogenous PTA formation game with MFN – Case (ii): $\langle sl \rangle$ arises

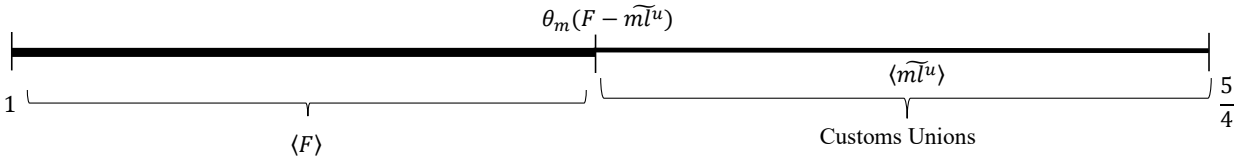

Figure 8b: Endogenous PTA formation game with tariff discrimination – Case (ii): $\langle \widetilde{sl} \rangle$ arises

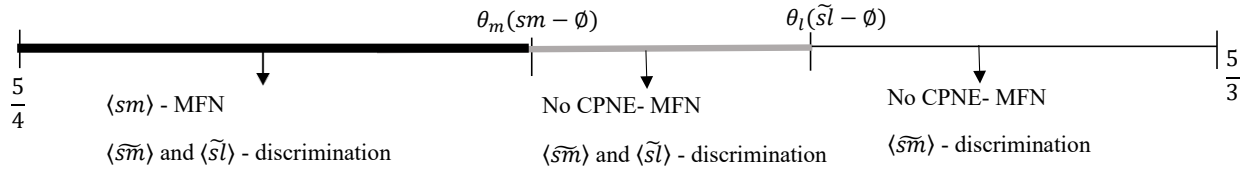

Figure 9: Higher Degree of Asymmetry with Zero External Tariffs: MFN vs. tariff discrimination

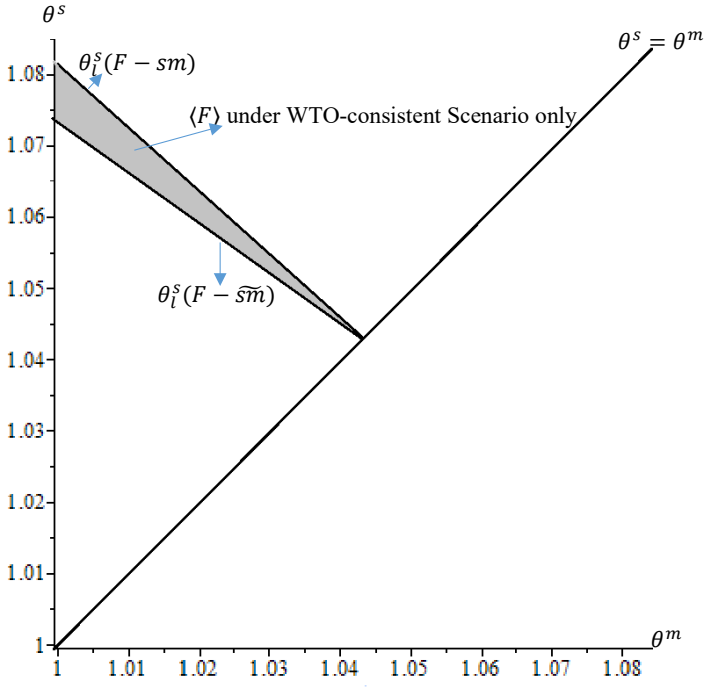

Figure 10: Global Free Trade under FTA game with Full Asymmetry: MFN vs. Discrimination

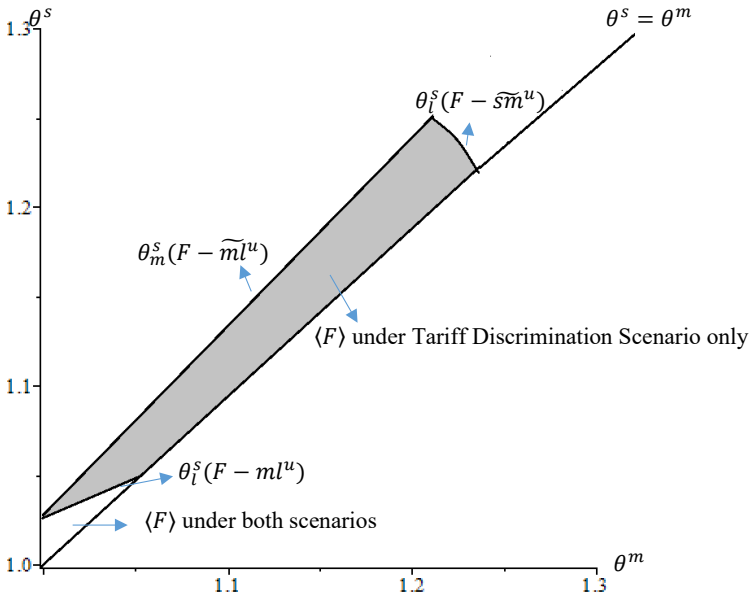

Figure 11: Global Free Trade under CU game with Full Asymmetry: MFN vs. Discrimination

References

Aghion, Philippe, Pol Antràs, and Elhanan Helpman 2007. "Negotiating Free Trade." *Journal of International Economics* 73: 1–30.

Bagwell, K. and Robert. W. Staiger, 1998. "Will preferential agreements undermine the multilateral trading system?" *The Economic Journal*, 108(449), pp.1162-1182.

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.

Bagwell, Kyle and Robert W. Staiger, 1999b. "An Economic Theory of GATT." *American Economic Review* 89(1): 215-48.

Bagwell, Kyle and Robert W. Staiger, 2010. "Backward Stealing and Forward Manipulation in the WTO." *Journal of International Economics* 82(1): 49-62.

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

Bond, E.W., Riezman, R.G. and Syropoulos, C., 2004. "A strategic and welfare theoretic analysis of free trade areas." *Journal of International Economics*, 64(1), pp.1-27.

Cadot, Olivier, Jaime De Melo, and Marcelo Olarreaga. 1999. "Regional integration and lobbying for tariffs against nonmembers." *International Economic Review* 40(3): 635-658.

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.

Estevadeordal, A., Freund, C. and Ornelas, E., 2008. "Does regionalism affect trade liberalization toward nonmembers?." *The Quarterly Journal of Economics*, 123(4), pp.1531-1575.

Furusawa,Taiji and Hideo Konishi, 2007. "Free Trade Networks." *Journal of International Economics* 72: 310–335.

Goyal, Sanjeev and Sumit Joshi, 2006. "Bilateralism and Free Trade." *International Economic Review* 47: 749–778.

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.

Krishna, Pravin,1998. "Regionalism and Multilateralism: a Political Economy Approach." *Quarterly Journal of Economics* 113: 227–51.

Lake, J. and Yildiz, H.M., 2016. On the different geographic characteristics of Free Trade Agreements and Customs Unions, *Journal of International Economics,* 103, 213-233.

Lake, J., Nken, M. and Yildiz, H.M., 2020.Tariff bindings and the dynamic formation of Preferential Trade Agreements, *Journal of International Economics,* 122.

Mrazova, Monika, David Vines, and Ben Zissimos. "Is the GATT/WTO's Article XXIV bad?." *Journal of International Economics* 89, no. 1 (2013): 216-232.

McCalman, Phil, 2002. "Multilateral trade negotiations and the most favored nation clause." *Journal of International Economics* 27: 151-76.

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.

Olarreaga, M. and Soloaga, I., 1998. "Endogenous tariff formation: the case of Mercosur." *World Bank Economic Review* 12(2): 297-320.

Ornelas, Emanuel, 2005a. "Trade Creating Free Trade Areas and the Undermining of Multilateralism." *European Economic Review* 49: 1717-35.
Ornelas, Emanuel, 2005b. "Rent Destruction and the Political Viability of Free Trade Agreements." *Quarterly Journal of Economics* 120: 1475-1506.
Ossa, Ralph, 2011. "A "new trade" theory of GATT/WTO negotiations. " *Journal of Political Economy*, 119(1): 122-152.
Saggi, Kamal, 2004. "Tariffs and the Most Favored Nation Clause." *Journal of International Economics* 63(2), 341-68.
Saggi, Kamal and Halis M. Yildiz, 2009. "Optimal tariffs of preferential trade agreements and the tariff complementarity effect." *Indian Growth and Development Review* 4 (17).
Saggi, Kamal and Halis M. Yildiz, 2010. "Bilateralism, Multilateralism, and the Quest for Global Free Trade." *Journal of International Economics* 81: 26-37.
Saggi, Kamal, Woan Foong Wong and Halis M. Yildiz, 2019. "Should the WTO require free trade agreements to eliminate internal tariffs?" *Journal of International Economics*, 118: 316-330.
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.
Seidmann, Daniel, 2009. "Preferential Trading Arrangements as Strategic Positioning." *Journal of International Economics* 79: 143–59.
Stoyanov, Andrey and Halis M. Yildiz, 2015. "Preferential versus Multilateral Trade Liberalization and the Role of Political Economy." *European Economic Review* 80: 140-164.
Syropoulos, C., 1999."Customs unions and comparative advantage." *Oxford Economic Papers*, 51(2): 239-266.
Suwanprasert, W., 2020. "The role of the most favored nation principle of the GATT/WTO in the New Trade model. " *Review of International Economics*, 28(3), pp.760-798.
Yi, Sang-Seung, 1996. "Endogenous formation of customs unions under imperfect competition: open regionalism is good." *Journal of International Economics* 41(1): 153-77.