

# **The Farsighted Stability of Global Trade Policy Arrangements**

**Stefan Berens and Lasha Chochua**

# THE FARSIGHTED STABILITY OF GLOBAL TRADE POLICY ARRANGEMENTS

STEFAN BERENS AND LASHA CHOCHUA

**ABSTRACT.** In this paper, we study and compare the stability of trade policy arrangements in two different regulatory scenarios, one with and one without Preferential Trade Agreements (PTAs), i.e. current vs. modified WTO rules. Unlike the existing literature, our paper considers an extensive choice set of trade constellations, containing both available PTAs, Customs Unions (CUs) and Free Trade Agreements (FTAs), as well as Multilateral Trade Agreements (MTAs), while assuming unlimited farsightedness of the negotiating parties. With symmetric countries and under both the current and the modified WTO rules, the Global Free Trade (GFT) regime emerges as the unique stable outcome. In the case of asymmetry, the results are driven by the relative size of the countries. If the world is in the vicinity of symmetry and two out of three countries are close to identical while relatively smaller than the other one, the area where the GFT regime is stable increases when prohibiting PTAs. However, when two similar countries are relatively larger, the availability of PTAs is conducive to the stability of the GFT regime. Finally, if the world is further away from symmetry, full trade liberalization is not attainable at all and an area where the Most-Favoured-Nation (MFN) regime is stable appears in the scenario without PTAs. Thus, the direction of the effect of PTAs on trade liberalization depends on the degree of asymmetry among countries.

*Keywords:* Trade Policy Arrangements, Stability, Unlimited Farsightedness  
*JEL Classification:* F13, F55

## 1. INTRODUCTION

Following the General Agreement on Tariffs and Trade (GATT) of 1947, an increasing number of signatory countries liberalized their trade policies primarily via two channels: bilateral and multilateral negotiations. To the present day, there have been eight rounds of multilateral trade negotiations with the current ninth one, the Doha Round, still ongoing. At the same time, parallel to the arrangements observed on the multilateral level, the world has seen an ever-increasing number of Preferential Trade Agreements (PTAs) mainly in the wake of bilateral negotiations. Currently, about forty percent of all countries/territories are a member of more than five PTAs while about a quarter participates in more than ten.<sup>1</sup>

The World Trade Organization (WTO), successor of the GATT in 1995, provides the rule set for the trade liberalization process of a significant number of countries.<sup>2</sup>

---

Bielefeld Graduate School of Economics and Management (BiGSEM), Bielefeld University.

The following people supported and influenced this paper through various fruitful discussions: Herbert Dawid, Gerald Willmann, and Michael Chwe. Also, the authors are grateful for the (financial) support of the BiGSEM.

<sup>1</sup>Source: <http://www.wto.org>

<sup>2</sup>All members of the WTO account for 96.4 percent of world trade, 96.7 percent of world GDP, and 90.1 percent of world population as of 2007 (Source: <http://www.wto.org>).

Its Article I acts as the foundation for any multilateral trade liberalization by formulating the so-called Most-Favoured-Nation (MFN) principle: Any concession granted to one member needs to be extended to all other members of the WTO.<sup>3</sup> In this paper, trade policy arrangements that are consistent with the MFN principle are referred to as Multilateral Trade Agreements (MTAs).<sup>4</sup> Contrary to the core MFN principle, Article XXIV Paragraph 5 explicitly allows countries to form PTAs, specifically Customs Unions (CUs) and Free Trade Agreements (FTAs), that do not need to extend the concessions granted within the arrangement to other countries.<sup>5</sup> However, Article XXIV Paragraph 5 Subparagraph (a), (b), and (c) each require that these are without (negative) influence on other trade relations.

The (direction of the) influence of Article XXIV Paragraph 5 on the development of trade policy arrangements is a controversial topic and the focus of many papers.<sup>6</sup> Likewise, the primary purpose of this paper is the analysis of the stability of different trade policy arrangements in two scenarios, that is with PTAs (current WTO rules) and without PTAs (modified WTO rules). In particular, it is our intent to examine whether PTAs act as ‘building blocks’ or ‘stumbling blocks’ on the path towards global free trade (Bhagwati (1993)).

The existing literature usually considers a limited selection of trade agreements or assumes limited farsightedness of the negotiating countries. It certainly allows for a cleaner description of the model and interpretation of its results, but ultimately raises the question about whether or not these restrictions significantly influence the analysis and to what degree these frameworks capture reality. In our opinion, certain empirical observations favor an extensive choice set and full farsightedness. During the past rounds of multilateral trade negotiations, many countries were simultaneously involved in other trade liberalization processes.<sup>7</sup> Moreover, such trade negotiations are usually complicated processes with significant effect on the countries’ economies and accompanied by elaborate studies about feasibility and future developments.<sup>8</sup> Taking these assessments into account, the contribution of our paper is an answer to the question concerning the influence on the analysis.

First of all, our paper considers an extensive set of trade agreements, containing PTAs, i.e., CUs and FTAs, as well as MTAs. Next, endogenizing the formation of trade agreements, each country ranks them based on a three-country two-good general equilibrium model of international trade.<sup>9</sup> The stability of all trade agreements is then examined using these rankings together with the concept of ‘consistent sets’

<sup>3</sup>Article I states that ‘any [...] favour [...] granted by any contracting party to any product originating in or destined for any other country shall be accorded immediately and unconditionally to the like product originating in or destined for [...] all other contracting parties’ (GATT, 1947).

<sup>4</sup>Furthermore, we interchangeably use the terms trade policy arrangements, trade agreements, trade constellations, and trade relations.

<sup>5</sup>Article XXIV Paragraph 5 states that ‘[...] this agreement shall not prevent [...] the formation of a customs union or of a free-trade area [...]’ (GATT, 1947).

<sup>6</sup>The next part of this paper contains further information on the related literature.

<sup>7</sup>Maggi (2014) showcases the importance of an extensive set of trade constellations.

<sup>8</sup>Aumann and Myerson (1988) provides a (brief) description of the criticism against the use of limited farsightedness in general: ‘When a player considers forming a link with another one, he does not simply ask himself whether he may expect to be better off with this link than without it, given the previously existing structure. Rather, he looks ahead and asks himself, “Suppose we form this new link, will other players be motivated to form further new links that were not worthwhile for them before? Where will it all lead? Is the end result good or bad for me?”’

<sup>9</sup>A model similar to that of Saggi and Yildiz (2010), which itself is a modification of the one in Bagwell and Staiger (1997). The modified one is also used in Saggi, Woodland and Yildiz (2013).

as stable sets - a notion proposed by Chwe (1994). As a result, our paper expands the set of trade agreements under consideration and also extends the farsightedness of the negotiating parties in comparison to the literature. In fact, to the best of our knowledge, no other paper considers a choice set as extensive as ours.

In the end, our analysis shows that the effect of PTAs on trade liberalization depends on the size distribution of the countries. As long as the countries are close to symmetric, Global Free Trade (GFT) emerges as the unique stable outcome under both the existing and the hypothetical institutional arrangement. However, when two countries are considerably smaller, a modified WTO without PTAs would facilitate the formation of GFT. By contrast, if two countries are relatively larger, this modified WTO would actually obstruct the development towards GFT. Once the world is further away from symmetry, full trade liberalization is not attainable at all and abolishing the exception for PTAs might result in the worst possible state from the perspective of overall world welfare, the non-cooperative MFN regime.

The findings of our paper notably deviate from those of the existing literature. Compared to the paper of Saggi, Woodland and Yildiz (2013), the composition of the stable set of trade policy arrangements differs on a substantial part of the parameter space under consideration (while coinciding on the remainder). Beyond that, the comparison with the work of Lake (2017) yields not only a difference in terms of stability but also with respect to the driving force(s).

The remainder of this paper is organized as follows. Section 2 focuses on the related literature, Section 3 specifies the model, Section 4 analyzes the findings while further details are discussed in Section 5, and Section 6 concludes our paper.

## 2. RELATED LITERATURE

An ever increasing body of literature studies the different aspects of international trade agreements. It is not our goal to completely review this stream of literature.<sup>10</sup> The emphasis of this part of our paper is on the methodology of the related papers. Further details, in particular a comparison of the model predictions, can be found in Section 5. In the following, the focus is on the so-called ‘rules-to-make-rules’ literature (Maggi (2014)) that tries to determine the role of PTAs in the global trade liberalization process.

A number of relevant papers are the work of Saggi, Yildiz and various co-authors. Saggi and Yildiz (2010) considers a three-country trade model where the degree and nature of trade liberalization, bilateral and multilateral, are endogenously determined. Using Coalition-Proof Nash Equilibria, the authors study the stability of FTAs and MTAs while varying the extent of asymmetry among the countries with respect to their size. In a subsequent paper Saggi, Woodland and Yildiz (2013) study the complementary case by focusing on the combination of CUs and MTAs while leaving everything else fixed (in terms of their framework). By contrast, the paper of Missions, Saggi and Yildiz (2016) analyzes the effect of both forms of PTAs, i.e. CUs and FTAs, on attaining global free trade, but excludes MTAs. In a sense, this completes their ‘2 out of 3’ pattern of trade agreements under consideration.

Another related paper (in terms of farsightedness) is the work of Lake (2017), who uses a dynamic approach to understand whether FTAs facilitate or impede the formation of GFT. The approach uses a three-country dynamic model where a

---

<sup>10</sup>The reader may want to consult the papers of Maggi (2014), Grossmann (2016), and Bagwell and Staiger (2016) for a detailed review of the related literature.

fixed protocol specifies for each period the exact nature (and order) of negotiations. Then, on the basis of Markov Perfect Equilibria in pure strategies, the author analyzes the effect of country asymmetries on global trade liberalization.

Furthermore, a variety of research focuses purely on analyzing the effect of FTAs. Goyal and Joshi (2006) consider several countries with a homogeneous good in their model and study different degrees of asymmetry across countries. They employ the notion of Pairwise Stability by Jackson and Wolinsky (1996) as the solution method. Furusawa and Konishi (2007) use similar methods but introduce heterogeneity with respect to goods. In a separate section, they also briefly discuss a setting with CUs, but overall focus on FTAs. Another related paper to Goyal and Joshi (2006) is that of Zhang et al. (2013) in which the concept of Pairwise Stability is replaced with Pairwise Farsighted Stability by Herings, Mauleon and Vannetelbosch (2009), thereby comparing myopia with farsightedness in an otherwise fixed framework. Also connected to this is the paper of Zhang et al. (2014), which uses the work of Goyal and Joshi (2006) as a benchmark and analyzes the evolutionary effect of the number of countries in a dynamic framework featuring random perturbations. Now, while all of the aforementioned papers employ (different) network-theoretic concepts, there is also Aghion et al. (2007), which features standard cooperative game theory. In the three-country model presented there, a single country takes on the role of negotiation leader and decides to either engage in sequential bilateral or single multilateral bargaining with the other countries.

The stability concept of our approach is that of Chwe (1994). It is (in parts) a response to the criticism of the von-Neumann-Morgenstern stable set (solution)<sup>11</sup>. The approach aims to achieve two goals, namely to include unlimited consideration of the future by the participants while simultaneously avoiding emptiness of the stable set that plagues other (more) restrictive solution concepts. It is also closely related to the stability concept found in Herings, Mauleon and Vannetelbosch (2009) and its extension (HMV (2014)). In fact, as is noted by the authors, their criterion constitutes a stricter version, but in specific cases (like our model) they coincide.

### 3. MODEL

**3.1. Setting.** Let  $N = \{a, b, c\}$  denote the set of all (three) countries in the world. Furthermore, let  $X$  denote the set of all trade agreements between these countries, see Section 3.3 for an explicit list. Then, the welfare function of each country induces a collection of preferences on  $X$  denoted by  $\{\prec_i\}_{i \in N}$ , see Section 3.2 for a description of the employed trade model that determine the welfare functions. Moreover, the non-empty subsets  $S$  of  $N$  specify the coalitions of countries, i.e. the grand coalition, coalitions of two, and single coalitions. Naturally, the preferences of the individual countries induce those of the coalitions, namely for  $x_1, x_2 \in X$  and  $S \subseteq N$ ,  $S \neq \emptyset$ :  $x_1 \prec_S x_2$  if and only if  $x_1 \prec_i x_2$  for all  $i \in S$ . Further, the actual ability of coalitions to change the status quo of trade agreements is captured via the collection  $\{\rightarrow_S\}_{S \subseteq N, S \neq \emptyset}$  of effectiveness relations defined on  $X$ , see Section 3.5 for the resulting overall network structure. In combination, the preferences together with the effectiveness relations will allow us to analyze the (potential) stability of different trade agreements, see Section 3.4 for a formal definition of the employed concept of stability. Finally, to determine the stable and unstable trade agreements

<sup>11</sup>Consult von Neumann and Morgenstern (1944) for a description of this (solution) concept and Harsanyi (1974) for its criticism.

an algorithm numerically evaluates a grid of the parameter space, see Section 3.6 for details.

**3.2. Underlying Trade Model.** In order to study the stability of different constellations of trade agreements, our framework utilizes a three-country trade model with competition via exports. It will determine the welfare of each country and thereby induce preferences and rankings over all regimes. The model itself follows the one used by Saggi and Yildiz (2010).

Recall that  $N = \{a, b, c\}$  denotes the set of countries. Further, let  $G = \{A, B, C\}$  denote three (corresponding) non-numeraire goods. Now, each country  $i$  is endowed with zero units of good  $I$  (corresponding capital letter) and  $e_i$  units of the others. Ultimately, it will end up importing  $I$  and exporting  $J$  and  $K$  with  $J, K \neq I$ . To guarantee the ‘competing exporters’-structure, a general condition needs to be applied to the degree of asymmetry with respect to the endowments of the countries. For  $i$  and  $j$  in  $N$  with  $i \neq j$ , in order for the exports from  $i$  to  $j$  to be non-negative the condition  $3e_j \leq 5e_i$  needs to be satisfied. Thus, the general condition reads as follows:

$$\frac{3}{5} \max\{e_j, e_k\} \leq e_i \leq \frac{5}{3} \min\{e_j, e_k\} \quad \forall i, j, k \in N$$

The preferences of individuals in each country are furthermore assumed to be identical. The demand for any non-numeraire good  $L \in G$  in country  $i \in N$  is given by the function  $d(p_i^L) = \alpha - p_i^L$  with  $p_i^L$  the price of good  $L$  in country  $i$  and the (universal) reservation price  $\alpha$ .<sup>12</sup> Each country also (possibly) imposes tariffs on the goods imported by them. Let  $t_{ij}$  denote the tariff imposed by country  $i$  on the import from country  $j$ . All prices and tariffs of a specific good  $I \in G$  are connected via the following no-arbitrage condition

$$(1) \quad p_i^I = p_j^I + t_{ij} = p_k^I + t_{ik}$$

where  $i, j, k \in N$  are pairwise distinct. In this model, the resulting prices together with the corresponding endowments are the only factors influencing imports and exports. In particular, the level of imports  $m_i^I$  of good  $I$  to country  $i$  is completely determined by the demand function (depending on the price),  $m_i^I = d(p_i^I) = \alpha - p_i^I$ . The exports  $x_j^I$  of good  $I$  from country  $j$  are the combination of the demand function (or prices) and the corresponding endowment,  $x_j^I = e_j - d(p_j^I) = e_j + p_j^I - \alpha$ . Now, a market-clearing condition for any good  $I$  requires that country  $i$ ’s import is equal to the total export of the countries  $j$  and  $k$  (again  $i, j, k \in N$  pairwise distinct):

$$(2) \quad m_i^I = x_j^I + x_k^I$$

Ultimately, the objective function of country  $i$  is its welfare<sup>13</sup>, denoted  $W_i$ , which includes Consumer Surplus (CS), Producer Surplus (PS), and Tariff Revenue (TR):

$$W_i = \sum_{L \in G} CS_i^L + \sum_{L \in G \setminus \{I\}} PS_i^L + TR_i$$

<sup>12</sup>The demand function is derived from a utility function that is additively separable and also quadratic in each non-numeraire good.

<sup>13</sup>In certain cases (depending on the trade agreement) the objective function of a country includes the welfare of other countries as well. See Section 3.3 for the details.

Now, CS is composed of three parts itself, namely one for each good. The consumer surplus  $CS_i^I$  with respect to the foreign good  $I$  is  $CS_i^I = \frac{1}{2}(\alpha - p_i^I)m_i^I$  and  $CS_i^L = \frac{1}{2}(\alpha - p_i^L)(e_i - x_i^L)$  for a domestic good  $L$ . Also, PS splits into two. The producer surplus  $PS_i^L$  for a domestic good  $L$  is given by  $PS_i^L = x_i^L(p_i^L - t_{li}) + (\alpha - p_i^L)p_i^L$ . Finally, the tariff revenue  $TR_i$  is given by  $TR_i = x_j^I t_{ij} + x_k^I t_{ik}$ .

3.2.1. *Equilibrium.* Let us start by using no-arbitrage (1) and market-clearing (2) to compute the equilibrium prices:

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

Using these equilibrium values, it is possible to calculate imports, exports, and also the welfare of each country up to the value of the tariffs (Appendix B.1). Note, that the maximization of welfare with respect to tariffs is going to be restricted depending on the trade agreement under consideration, see Section 3.3. For example in the case of MFN, country  $i$  maximizes  $W_i$  under the restriction that  $t_{ij} = t_{ik}$ . Therefore, country  $i$  aims to maximize its welfare  $W_i$  over  $(t_{ij}, t_{ik}) \in T_i$  given  $(t_{ji}, t_{jk}) \in T_j$  and  $(t_{ki}, t_{kj}) \in T_j$ , where  $T_l$  is the set of possible tariff pairs for country  $l$  in a fixed trade agreement.

The full equilibrium of this model is computed as follows. Fix a trade agreement and thereby the restrictions on the tariffs. Compute the best-response functions for each country (with respect to the tariffs) and determine the optimal choices. While Section 3.3 contains all information on the trade agreements that is necessary to compute the equilibria, the actual results are presented in Appendix B.2. Finally, an overview of the (resulting) overall welfare can be found in Appendix B.3.

**3.3. Trade Policy Arrangements.** All trade relations in our model are one of four types: MFN, CU, FTA, and MTA. Each type, except for MFN, naturally induces different combinations of insiders and outsiders. Namely, three combinations of two members and one of three (each for CU, FTA, and MTA).<sup>14</sup> Additionally, the case of FTA contains the possibility of a special hub structure with two FTAs at the same time - adding another three combinations. In total, our model allows for 16 different trade constellations.<sup>15</sup> For each of these trade agreements the tariffs are bounded from below and above by zero and the MFN-tariff respectively, which is discussed in more detail in Appendix B.2. The corresponding set of tariffs for country  $i$ , i.e.  $[0, t_i^{MFN}]$ , is denoted by  $T_i$ . Any additional restrictions on tariffs, specific to trade agreements, are listed here:

In the baseline case, i.e. MFN, countries do not liberalize their trade relations at all, but the non-discrimination principle still applies. Each country unilaterally

<sup>14</sup>Note that in our model Global Free Trade is essentially listed in three different variations, via CUs, FTAs, and MTAs. The actual welfare is necessarily equal across all three variations, but not their position in the network (Section 3.5). In particular, for our concept of stability it is important which group of countries can create or destroy specific trade agreements (Appendix C.1). Occasionally, all three variants together are going to be referred to as 'GFT' (when applicable).

<sup>15</sup>The framework does not contain combinations of different classes of trade agreement due to the possibly conflicting restrictions on tariffs that the different classes entail. In order to circumvent potential conflicts one would need to fix an (arbitrary) ordering in terms of priority (or importance) of trade agreements, which would reduce the explanatory power more than the inclusion of other combinations of trade agreements would increase it (in our opinion).

chooses its (optimal) tariffs accordingly. Therefore, the optimization problem of country  $i$  is  $\max_{(t_{ij}, t_{ik}) \in T_i^{MFN}} W_i$  with  $T_i^{MFN} = \{(t_{ij}, t_{ik}) \in \mathbb{R}_{\geq 0}^2 \mid t_{ij} = t_{ik}\}$ . Note, that in this reference scenario each tariff is chosen from  $\mathbb{R}_{\geq 0}$  instead of  $T_i$ .

In case country  $i$  and  $j$  form CU(i,j), each of them removes any trade restriction on the other country and then jointly imposes an optimal tariff on country  $k$ . Thus, the optimization problem of country  $i$  and  $j$  is  $\max_{(t_{ij}, t_{ik}) \in T_i^{CU}, (t_{ji}, t_{jk}) \in T_j^{CU}} W_i + W_j$  with  $T_i^{CU} = \{(t_{ij}, t_{ik}) \in T_i^2 \mid t_{ij} = 0\}$  and  $T_j^{CU}$  analogous. Finally, country  $k$  simply follows and applies the principle of MFN (as before). However, as soon as all three countries enter a single CU together, the (common) optimization problem is trivial, because the only possible tariff of each country towards any other country is zero, and the scenario is denoted by CUGFT.

In case country  $i$  and  $j$  form FTA(i,j), each of them removes any trade restriction on the other country and then unilaterally imposes an optimal tariff on country  $k$ . Thus, the (representative) optimization problem of country  $i$  is  $\max_{(t_{ij}, t_{ik}) \in T_i^{FTA}} W_i$  with  $T_i^{FTA} = \{(t_{ij}, t_{ik}) \in T_i^2 \mid t_{ij} = 0\} (= T_i^{CU})$ . The optimization problem of country  $k$  is identical to that of the third country in case of a CU. Further, in case country  $i$  forms an FTA both with  $j$  and  $k$ , that is FTAHub(i), then both tariffs of country  $i$  are set to zero by nature of its trade relation with both other countries. Each of the other two countries operates as before: Country  $j$  ( $k$  analogous) faces  $\max_{(t_{ji}, t_{jk}) \in T_j^{FTA}} W_j$  where  $T_j^{FTA} = \{(t_{ji}, t_{jk}) \in T_j^2 \mid t_{ji} = 0\}$ . Thus, in terms of decision problem, it does not matter for a country whether its partner also forms another trade agreement with the other country. Finally, if all three countries in pairs of two countries form FTAs, then the optimization problem is identical to the case of CUGFT, denoted FTAGFT, but the actual trade agreement is different in terms of structure and network position, see Section 3.5.

In case country  $i$  and  $j$  form MTA(i,j), then both jointly change their tariffs with respect to each other and also for the third country (at the same time). Thus, the optimization problem of country  $i$  and  $j$  is  $\max_{(t_{ij}, t_{ik}) \in T_i^{MTA}, (t_{ji}, t_{jk}) \in T_j^{MTA}} W_i + W_j$  with  $T_i^{MTA} = \{(t_{ij}, t_{ik}) \in T_i^2 \mid t_{ij} = t_{ik}\}$  and  $T_j$  analogous. As seen before, the optimization problem of country  $k$  is identical to that of the third country in case of a CU. Again, as soon as all three countries enter a single MTA together, the optimization problem is identical to the case of CUGFT, denoted MTAGFT, but also different in terms of network position, see Section 3.5.

**3.4. Stability Concept.** As concept of stability our framework makes use of the approach of Chwe (1994).<sup>16</sup> Consider the tuple  $\Gamma = (N, X, \{\prec_i\}_{i \in N}, \{\rightarrow_S\}_{S \subseteq N, S \neq \emptyset})$  that correspondingly describes the evolution of the status quo of trade agreements driven by the combination of preferences and effectiveness relations:

Let  $x \in X$  be the status quo of trade agreements at the start. Next, each coalition  $S \subseteq N$ ,  $S \neq \emptyset$  (including individuals) is able to make  $y \in X$  the new status quo as long as  $x \rightarrow_S y$ . Continue with such  $y$  as the new status quo. If a status quo  $z \in X$  is reached without any coalition moving away, then the state is actually realized and each country receives their corresponding welfare.<sup>17</sup> In consequence, any coalition only favors following through on their ability to move,  $x \rightarrow_S y$ , when

<sup>16</sup>Consult the paper of Chwe (1994) for the proofs of the propositions that are presented here.

<sup>17</sup>Technically, the model is without any true sense of time. Any start (or end) as well as any sequence of actions should be interpreted as a thought-experiment. Furthermore, a path created in this fashion is generally not unique.



preferring the final welfare over the current one,  $x \prec_S z$ . Formally, this comparison of states by (chains of) coalitions is captured in the definition of direct and indirect dominance:

**Definition 1** (Dominance). *Let  $x_1, x_2 \in X$ . Then,*

- i)  $x_1$  is directly dominated by  $x_2$ , write  $x_1 < x_2$ , if there exists  $S \subseteq N$ ,  $S \neq \emptyset$ , such that  $x_1 \rightarrow_S x_2$  and  $x_1 \prec_S x_2$ .
- ii)  $x_1$  is indirectly dominated by  $x_2$ , write  $x_1 \ll x_2$ , if there exist sequences  $y_0, y_1, \dots, y_m \in X$  (with  $y_0 = x_1$  and  $y_m = x_2$ ) and  $S_0, S_1, \dots, S_{m-1} \subseteq N$ , such that  $S_i \neq \emptyset$ ,  $y_i \rightarrow_{S_i} y_{i+1}$ , and  $y_i \prec_{S_i} y_m$  for  $i = 0, 1, \dots, m-1$ .

Note, that if  $x_1 < x_2$  for some  $x_1, x_2 \in X$ , then automatically  $x_1 \ll x_2$ .

Using this definition, the concept of ‘consistent set’ describes a (sub-)set that exhibits internal stability in the form of a lack of incentive to deviate:

**Definition 2** (Consistent Set). *A set  $Y \subseteq X$  is consistent if  $y \in Y$  if and only if for all  $x \in X$  and all  $S \subseteq N$ ,  $S \neq \emptyset$ , with  $y \rightarrow_S x$  there exists  $z \in Y$  where  $x = z$  or  $x \ll z$  such that  $y \not\prec_S z$ .*

In general, a consistent set is not necessarily unique, but the following proposition allows us to talk about the unique ‘largest consistent set’, i.e. the (consistent) set that contains all consistent sets:

**Proposition 1.** *There uniquely exists a  $Y \subseteq X$  such that  $Y$  is consistent and  $Y' \subseteq X$  consistent implies  $Y' \subseteq Y$ . The set  $Y$  is called the largest consistent set or simply LCS.*

*Or put differently, it is the unique fixed point of the correspondence  $f: 2^X \rightarrow 2^X$  defined by*

$$Y \mapsto f(Y) = \{y \in X \mid \forall x \in X, \forall S \subseteq N, S \neq \emptyset, \text{ with } y \rightarrow_S x: \\ \exists z \in Y \text{ s.t. } (x = z \text{ or } x \ll z) \wedge y \not\prec_S z\}.$$

Now, similar to the internal stability captured in the definition of consistent sets, a form of external stability is captured via an incentive to gravitate towards the consistent set:

**Definition 3** (External Stability). *Let  $Y \subseteq X$  be the largest consistent set. Then, it satisfies the external stability condition if for all  $x \in X \setminus Y$  there exists  $y \in Y$  such that  $x \ll y$ .*

The following result characterizes one setting in which this condition is satisfied:

**Proposition 2.** *Let  $X$  be finite and the underlying preferences irreflexive. Then, the LCS is non-empty and satisfies the external stability property.*

Finally, let us state a couple of comments on the application and interpretation of this stability concept with respect to our model:

3.4.1. *Application.* First of all, applying Proposition 1 to our model is trivial, because it is stated without any (additional) requirements on the involved objects. Furthermore, the application of Proposition 2 is straight forward as well: First, the set of outcomes  $X$  is clearly finite in our setting as we are only considering a finite number of different trade agreements. Second, any strict preference is automatically irreflexive and our preferences are induced by strict welfare comparisons. Thus, while the definition of the (largest) consistent set in general only guarantees internal stability, our setting actually implies external stability as well:

**Corollary 1.** *In our setting, the (unique) LCS is non-empty and satisfies the external stability property (in addition to the internal stability).*

Now, the LCS is going to be the focus point of our analysis. Any trade agreement is considered to be ‘(potentially) stable’ if it is in the LCS, ‘unstable’ otherwise. The nomenclature is a tribute to the fact that the LCS as a stability concept is ‘weak: not so good at picking out, but ruling out with confidence’, because ultimately it ‘does not try to say what will happen but what can possibly happen’ (Chwe (1994)).

**3.5. Network Structure.** The complete network structure consists of a collection of transition matrices  $\{A_S\}_{S \subseteq N, S \neq \emptyset}$  induced by  $\{\rightarrow_S\}_{S \subseteq N, S \neq \emptyset}$ . Let  $S \subseteq N, S \neq \emptyset$  be any coalition, then the entry  $(A_S)_{x_i, x_j}$  is 1 if  $x_i \rightarrow_S x_j$  and 0 otherwise. Thus, the matrix for  $\{a, b, c\}$ , the full coalition, is simply given by  $(A_{\{a, b, c\}})_{x_i, x_j} = 1$  for all  $x_i, x_j \in X$ . Further, each of the transition matrices induces a directed graph with the trade agreements as vertices and the effectiveness relations as edges. Therefore, the corresponding directed graph of the full coalition is a complete directed graph with loops.

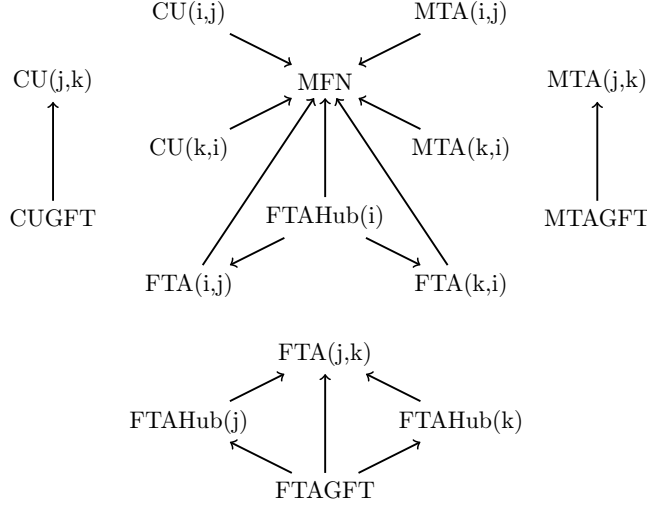
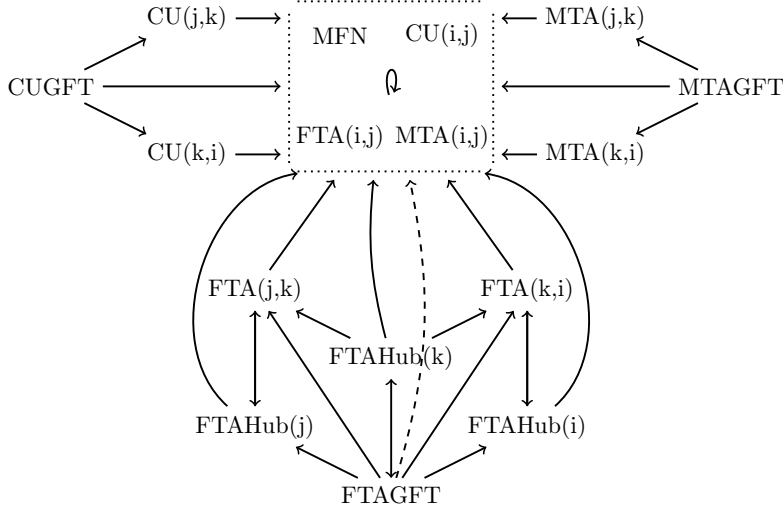
It is noteworthy to point out that the relation (or transition)  $x \rightarrow_S x$  holds for all trade agreements  $x$  and all coalitions  $S$ , but is ultimately irrelevant for the analysis with respect to the stability. The reason for this is the fact that our model contains no sense of time - essentially stalling negotiations serves no purpose.<sup>18</sup> Therefore, these transitions are ignored from now on or, put differently, the framework only considers a form of equivalence classes, namely modulo loops. Furthermore, whenever coalition  $S$  is able to destroy one trade agreement, say  $x_1$ , and subsequently create another one, say  $x_2$ , then it is able to move directly, i.e.  $x_1 \rightarrow_S x_2$ . Finally, for the remaining coalitions (of two and one country) only the transition graphs are presented here. The corresponding transition tables can be found in Appendix B.4.

Let us now consider the transition graph for a single country coalition  $i \in N$  with  $j, k \in N \setminus \{i\}, j \neq k$ , denoting the other two countries. In this case, MFN is connected to a number of other different elements, but not to the three variants of Global Free Trade, CUGFT, FTAGFT and MTAGFT. Now, each of those forms a separate group of connected trade agreements. Thus, the overall transition graph, see Figure 1 (modulo loops), consists of four sub-graphs.

Finally, consider the transition graph for a coalition of two countries  $i, j \in N, i \neq j$  with  $k \in N \setminus \{i, j\}$  denoting the other country. In this case, MFN, CU( $i, j$ ), FTA( $i, j$ ), and MTA( $i, j$ ) are all interconnected. Also, any element connected to one of these is automatically connected to all of them. Thus, in the transition graph, see Figure 2 (again, modulo loops), this group of four corresponds to a complete directed sub-graph pictured as one ‘(super) node’ (dotted box).

---

<sup>18</sup>While staying in one trade constellation, the overall strategic situation remains the same. Specifically, for each country and each coalition the welfare of each trade agreements only depends on the parameters of the underlying trade model. Similarly, the network structure stays constant. Additionally, the number of (potential) movements in a chain of trade agreements is unlimited.

FIGURE 1. The transition graph for coalition  $\{i\}$ ,  $i \in N$ .FIGURE 2. The transition graph for coalition  $\{i, j\}$ ,  $i, j \in N$ ,  $i \neq j$ .

**3.6. Algorithm and Parameters.** The (additional) explanatory power from the introduction of an extensive set of trade agreements and unlimited farsightedness comes at the cost of a complex computational problem. This problem is solved numerically with the help of an algorithm - the pseudocode of which can be found in Appendix A.<sup>19</sup> The parameter space therefore needs to be specified and discretized:

First, recall that the endowments satisfy  $\frac{3}{5} \max\{e_j, e_k\} \leq e_i \leq \frac{5}{3} \min\{e_j, e_k\}$  for all  $i, j, k \in N$  in order to guarantee the ‘competing exporters’-structure, see Section 3.2. Now, without loss of generality, normalize one endowment to one,

<sup>19</sup>The authors are grateful to Michael Chwe for the provision of an exemplary algorithm.

namely  $e_b = 1$ . Consequentially, for  $i, j \in N \setminus \{b\}$ :  $e_{\min} := \frac{3}{5} \leq \frac{3}{5} \max\{1, e_j\} \leq e_i$  and  $e_i \leq \frac{5}{3} \min\{1, e_j\} \leq \frac{5}{3} =: e_{\max}$ . Furthermore, the resulting parameter space, Figure 3, can be split into six right-angled triangles, which are mirror images of one another (in terms of relative endowments). Thus, again without loss of generality, focus on one of them, namely the marked triangle, and then cover it with a grid for the actual computation.<sup>20</sup>

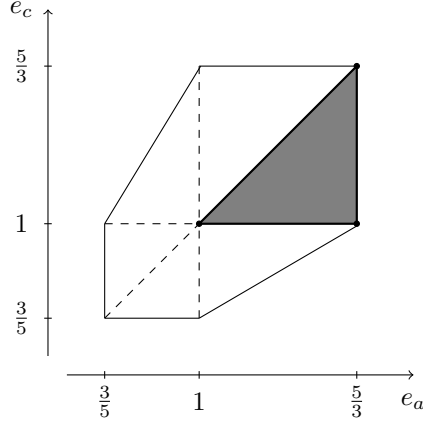


FIGURE 3. The parameter space of the endowments with  $e_b = 1$

Additionally, to produce plausible results, e.g positive prices, the factor  $\alpha$  needs to be chosen above a minimal value for each tuple of endowments,  $\alpha_{\min}(e_a, e_b, e_c)$ . Above this minimal value, the results remain unchanged.<sup>21</sup> Thus, by taking the maximum over all these minimal values,  $\alpha_{\max \min} = \max_{e_a, e_b, e_c} \{\alpha_{\min}(e_a, e_b, e_c)\}$ , adding an epsilon,  $\alpha = \alpha_{\max \min} + \epsilon$ , and using it for all endowments makes sure that all results are plausible and comparable at the same time.<sup>22</sup>

#### 4. ANALYSIS

Let us now present the resulting structure of stability among trade agreements according to our framework. Figure 4 depicts the parameter space of endowments under consideration for this - it is the (marked) triangle from before. The analysis starts with the three extreme points, then turns to the connecting intervals, and finishes with the entire interior. In each of these cases, two scenarios are examined. The first scenario corresponds to the current WTO institutional arrangement while the second one assumes modified WTO rules without Article XXIV Paragraph 5, which would prevent the formation of PTAs (specifically CUs and FTAs).

<sup>20</sup>The distance is set to 0.0013360053440215 - due to 500 points per dimension of the grid.

<sup>21</sup>The factor  $\alpha$  always enters the welfare of country  $i$  as  $2\alpha e_i$  (see Appendix B.1). Therefore, any changes above the minimal value leave the welfare levels and therefore the rankings unaffected.

<sup>22</sup>In our computation  $\epsilon$  is simply fixed to 0.01, which yields  $\alpha = 1.3988888888888888$ .

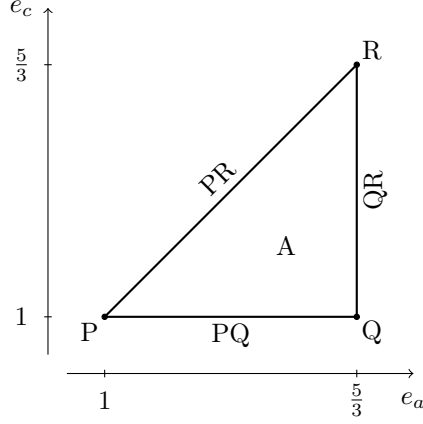


FIGURE 4. Overview of the different points, intervals and areas of interest depending on the (partially normalized) endowment tuple

The remainder of this analysis is structured as follows. First, Section 4.1 considers the symmetric case, see point P in Figure 4, where all countries are identical. Second, Section 4.2 features the two extreme asymmetric cases, points Q and R, with countries that are small, small, and large (Q) or small, large, and large (R). Next, Section 4.3 discusses the three related intervals, sides PQ, QR, and PR, where the countries are small, small, and varying (PQ), small, large, and varying (QR), or small with two varying equally (PR). Finally, Section 4.4 describes the inner area, area A, with three distinct countries.

**4.1. Symmetric Case.** First, let us consider the symmetric case, where symmetry refers to identical endowments for all countries, i.e.  $e_a = e_b = e_c = 1 = e_{\min}$ , and corresponds to point P in the triangle of Figure 4. As the countries do not differ from one another, the only thing that matters for welfare is whether a country is an insider or an outsider in a specific trade agreement. In the following, we present the ranking of preferences from the perspective of country  $a$ , which represents that of all other countries as well, for fixed  $i, j \in N \setminus \{a\}$  with  $i \neq j$ :

$$CU(i, j) \prec_a MFN \prec_a MTA(a, i) \prec_a FTAH_{ub}(i) \prec_a FTA(i, j) \prec_a FTA(a, i) \\ \prec_a CU(a, i) \prec_a MTA(i, j) \prec_a GFT \prec_a FTAH_{ub}(a)$$

The case where two countries form a CU is the least favorable trade constellation for the third country. Under such circumstances, the outsider faces the second-highest tariffs (with MFN-tariffs the highest), while the insiders cancel the tariffs among themselves. The exports of country  $a$  to the other countries,  $i$  and  $j$ , are the lowest under  $CU(i, j)$  compared to all alternative trade agreements. The same applies to the total imports. In other words, the ‘trade diversion’ effect is the strongest for country  $a$  in case of  $CU(i, j)$ . In general, the MFN regime favors country  $a$  when compared to  $CU(i, j)$ . The tariff revenue remains the same, while the consumer surplus is lower and the producer surplus is higher - the increase offsets the decrease. The MFN regime slackens the ‘trade diversion’ effect present in the case of  $CU(i, j)$  by virtue of increased export values of country  $a$ .

Among the group of bilateral trade agreements where the country is an insider, the MTAs result in the lowest welfare (for this country).  $MTA(a,i)$  itself generates a higher welfare for country  $a$  in comparison with the MFN regime on the grounds of increased consumer and producer surplus. The  $FTAHub(i)$  constellation results in even further gains in welfare for country  $a$  through higher export values and producer surplus accordingly (the tariff revenue and also the consumer surplus are lower under  $FTAHub(i)$  compared to  $MTA(a,i)$  though). However, country  $a$  does not have an incentive to remain in this constellation. The unilateral deviation from  $FTAHub(i)$  to  $FTA(i,j)$  comes with a decrease of consumer and producer surplus but enough increase in tariff revenue to ultimately ensure higher welfare under the latter constellation. Nonetheless, among FTAs being an outsider is less desirable than being an insider for any country. The drop in tariff revenue is offset by an expansion of the consumer and producer surplus, resulting in higher welfare for country  $a$  in case of  $FTA(a,i)$  compared to  $FTA(i,j)$ . As an insider, country  $a$  prefers  $CU(a,i)$  over  $FTA(a,i)$  though. More precisely, in spite of the decline in the consumer surplus, the actual welfare goes up through an expansion of tariff revenue and producer surplus.

The formation of  $MTA(i,j)$  guarantees the highest welfare for country  $a$  compared to any other bilateral trade agreement. The driving factor is the MFN-principle, which implies that in case of  $MTA(i,j)$  the insiders need to apply the same tariff to both each other and the outsider - a form of free-rider problem. At the same time, country  $a$  attains the highest possible tariff revenue.

Each country obtains the second-highest welfare level when the world reaches global free trade. Under full trade liberalization, the producer surplus is also the second-highest among all trade agreements (effectively driving the ranking). It is only surpassed by that of  $FTAHub(a)$ . The latter constellation brings about the highest possible welfare for country  $a$ . But note that such a trade agreement disproportionately favors the hub country over the other countries.

Countries' strong preference rankings are the crucial ingredient for computing the LCS. In fact, for each country all three variants of global free trade are ranked as second-best option while each first-best option, a hub structure, is ranked considerably lower for the other countries. Intuitively, global free trade seems like a stable compromise. The following proposition and its proof reinforce this:

**Proposition 3.** *Under symmetry and with the current institutional arrangement of the WTO, the LCS contains three elements: CUGFT, FTAGFT, and MTAGFT. In other words, (the trinity of) global free trade is the unique stable outcome.*

*Proof.* Based on the definition of indirect dominance and the transition graphs, see Section 3.4 and 3.5, the preference rankings from earlier allow us to derive the indirect dominance matrix. If the entry in the matrix is equal to one (resp. zero), then the trade arrangement corresponding to the row of the entry is (resp. isn't) indirectly dominated by the one corresponding to the column of the entry. For example,  $FTAHub(a)$  is indirectly dominated by  $CUGFT$  as there exists a (finite) sequence of outcomes and coalitions such that all coalitions in the sequence prefer the final outcome over the current one:

$$FTAHub(a) \rightarrow_{\{b,c\}} CU(b,c) \rightarrow_{\{a,b,c\}} CUGFT$$

Checking for all possible sequences yields the following indirect dominance matrix:<sup>23</sup>

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1 <i>MFN</i>	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2 <i>CU(a, b)</i>	0	0	0	0	1	0	0	0	1	1	0	1	0	1	1	1
3 <i>CU(b, c)</i>	0	0	0	0	1	0	0	0	0	1	1	1	1	0	1	1
4 <i>CU(c, a)</i>	0	0	0	0	1	0	0	0	1	0	1	1	1	1	0	1
5 <i>CUGFT</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6 <i>FTA(a, b)</i>	0	1	1	1	1	0	0	0	1	1	0	1	0	1	1	1
7 <i>FTA(b, c)</i>	0	1	1	1	1	0	0	0	0	1	1	1	1	0	1	1
8 <i>FTA(c, a)</i>	0	1	1	1	1	0	0	0	1	0	1	1	1	1	0	1
9 <i>FTAHub(a)</i>	0	1	1	1	1	1	1	0	0	0	1	0	0	0	0	1
10 <i>FTAHub(b)</i>	0	1	1	1	1	1	1	0	0	0	1	0	0	0	0	1
11 <i>FTAHub(c)</i>	0	1	1	1	1	1	1	0	0	0	1	0	0	0	0	1
12 <i>FTAGFT</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13 <i>MTA(a, b)</i>	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
14 <i>MTA(b, c)</i>	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
15 <i>MTA(c, a)</i>	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
16 <i>MTAGFT</i>	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0

Note that intuitively any outcome is stable if all deviations from it are deterred. Also, a deviation from the outcome is hindered if there is a stable outcome which might be reached and some member of the deviating coalition does not prefer it over the initial outcome. In the following procedure, start with the full set and then keep removing elements that are unstable until the remaining ones are stable

Take  $x \in \{MFN, FTA(i, j), CU(i, j), MTA(i, j)\}$ , where  $i, j \in N$  with  $i \neq j$ , and then consider the joint deviation  $x \rightarrow_{\{a,b,c\}} FTAGFT$ . The *FTAGFT* regime is not indirectly dominated by any other outcome (see the matrix above) and also  $x \prec_{\{a,b,c\}} FTAGFT$  for each of those  $x$ . Thus the deviation  $x \rightarrow_{\{a,b,c\}} FTAGFT$  cannot be deterred and therefore no such  $x$  can be part of the stable set.

Consider *FTAHub*( $i$ ),  $i \in N$ , and the deviation *FTAHub*( $i$ )  $\rightarrow_{\{j,k\}} FTAGFT$ ,  $j, k \in N \setminus \{i\}$  with  $j \neq k$ . Using *FTAHub*( $i$ )  $\prec_{\{j,k\}} FTAGFT$  together with the logic from before eliminates *FTAHub*( $i$ ) for each  $i \in N$ .

Focus on the set of remaining elements  $Y = \{CUGFT, FTAGFT, MTAGFT\}$ . Start with any element  $y$  in  $Y$ . If there is a deviation to any element  $x \in X \setminus Y$ , then there always exists an indirect dominance path (see indirect dominance matrix)  $x \ll y'$  coming back to an element  $y' \in Y$ . In addition, for any  $y_1, y_2 \in Y$ ,  $y_1 \neq y_2$ , there does not exist a coalition  $S \subseteq N$ ,  $S \neq \emptyset$ , for which  $y_1 \prec_S y_2$ . Thus, the set  $Y$  satisfies the (internal) stability condition while being maximal, i.e.  $Y = LCS$ .  $\square$

In the symmetric case, under the current institutional arrangement of the WTO, the global free trade variations appear as the only stable constellation according to our framework. But what would happen without Article XXIV Paragraph 5? In this case, countries would not have the option to liberalize trade through the formation of CUs or FTAs - leaving MTAs as the only possibility. The representative preference ranking of country  $a$  would look as follows:

$$MFN \prec_a MTA(a, i) \prec_a MTA(i, j) \prec_a MTAGFT$$

<sup>23</sup>Appendix A contains the pseudocode for this procedure.

Each country achieves the peak welfare under MTAGFT. Thus, it is reasonable to conjecture stability of MTAGFT. The following proposition proves this intuition:

**Proposition 4.** *Under symmetry and with the modified institutional arrangement of the WTO (no PTAs), the LCS contains one element: MTAGFT. In other words, global free trade is the unique stable outcome.*

*Proof.* The indirect dominance matrix is derived as before:

$$\begin{array}{c}
 1 \text{ MFN} \\
 2 \text{ MTA}(a, b) \\
 3 \text{ MTA}(b, c) \\
 4 \text{ MTA}(c, a) \\
 5 \text{ MTAGFT}
 \end{array}
 \begin{array}{c}
 1 \quad 2 \quad 3 \quad 4 \quad 5 \\
 \left( \begin{array}{ccccc}
 0 & 1 & 1 & 1 & 1 \\
 0 & 0 & 1 & 1 & 1 \\
 0 & 1 & 0 & 1 & 1 \\
 0 & 1 & 1 & 0 & 1 \\
 0 & 0 & 0 & 0 & 0
 \end{array} \right)
 \end{array}$$

Let us start with the full set of trade agreements again (limited to the setting). If the grand coalition moves from *MFN* to *MTAGFT*, then the only possibility is to stay there, as *MTAGFT* is not indirectly dominated by any other outcome. Moreover,  $MFN \prec_{a,b,c} MTAGFT$ . Thus, *MFN* cannot be stable. Furthermore, if the grand coalition moves from any bilateral *MTA* regime to *GFTMTA*, by the same argument, it is clear that no bilateral *MTA* can be stable. Finally, any deviation from *MTAGFT* will come back to itself due to the indirect dominance. Consequentially, the set  $Y = \{MTAGFT\}$  is consistent and also the largest one.  $\square$

If symmetry among all countries holds, then Article XXIV Paragraph 5 does not change anything in terms of stability and corresponding welfare, both individual and overall. The only stable trade constellation is (the trinity of) global free trade.

**4.2. Asymmetric Case - Vertices of the Triangle.** It is natural to start the analysis of the asymmetric case by considering its two extreme scenarios, which correspond to the points Q and R in the triangle of Figure 4. In the following, Section 4.2.1 discusses the case of countries that are small, small, and large (Q) while Section 4.2.2 focuses on countries that are small, large, and large (R).

**4.2.1. The case of two small and one large country.** In this scenario, fix  $e_a = e_{max}$  and  $e_b = e_c = e_{min}$  (point Q). Let us start with the ranking of preferences for country *a* and another country  $i \in N \setminus \{a\}$  - representing also  $j \in N \setminus \{a, i\}$ :

$$\begin{aligned}
 CU(i, j), FTA(i, j) \prec_a FTAH_{ub}(i) \prec_a MFN, MTA(i, j) \prec_a FTA(a, i) \\
 \prec_a MTA(a, i) \prec_a CU(a, i) \prec_a GFT, FTAH_{ub}(a)
 \end{aligned}$$

$$\begin{aligned}
 MTA(a, i) \prec_i GFT, FTAH_{ub}(a) \prec_i CU(a, i) \prec_i FTA(a, i) \prec_i CU(a, j) \\
 \prec_i MFN, MTA(i, j) \prec_i FTAH_{ub}(i) \prec_i FTA(a, j) \prec_i MTA(a, j) \\
 \prec_i FTAH_{ub}(j) \prec_i CU(i, j), FTA(i, j)
 \end{aligned}$$

One immediately notices that small and large countries have different rankings. A large country profoundly dislikes the scenarios where it is an outsider; while the small countries, by contrast, dislike any trade arrangements with the large country. Note that in certain cases countries actually do not differentiate between different



trade constellations.<sup>24</sup> For example,  $CU(i,j)$  and  $FTA(i,j)$  result in same welfare for all countries. In this case, under the given pattern of endowments, the optimal tariffs of the small countries for  $CU$  and  $FTA$  are above the MFN-tariff. However, the Sub-paragraphs of Article XXIV Article 5 rule this out and therefore the tariffs are capped at the MFN-level. A similar argument applies to the case of  $FTAHub(a)$ . Here, the optimal tariffs of the small countries would be negative. By restricting tariffs from below by zero implies that  $FTAHub(a)$  corresponds to GFT, or rather a Pseudo-GFT. Finally, the MTA between the small countries actually coincides with the MFN regime because of identical optimal tariffs for both cases.

Next, let us analyze the preferences of the large country  $a$ . As mentioned above, being the outsider produces the least favorable constellations for a large country. The worst scenarios are those where the small countries form a PTA. In such cases, the export, and hence the producer surplus, is the lowest in the large country. Now, compared to these PTAs among the small countries, both the tariff revenue and the consumer surplus are lower under  $FTAHub(i)$ , but the comparably strong growth of the producer surplus produces an increase in welfare. Further increases in producer surplus are possible via the MFN regime or  $MTA(a,i)$ . As soon as the large country forms an FTA, its welfare increases due to trade relations that benefit its exports within the constellation.  $MTA(a,i)$  leads to an even higher welfare, but there the driving factor are the tariff revenues. Among all bilateral trade agreements, where the large country is an insider, the optimal outcome is  $CU(a,i)$ . The highest welfare for the large country occurs when trade is fully liberalized though. In that scenario, it is able to completely reap the trade benefits - the producer surplus peaks when compared to the other trade agreements.

Now, let us consider the preferences of a small country  $i$ , the countries  $i$  and  $j$  are indistinguishable from each other in this respect. Contrary to the preferences of a large country, it is in its best interest for a small country to avoid forming any trade arrangement with the large country. The smallest welfare of a small country occurs in the case of  $MTA(a,i)$ , as the constellation generates one of the least desirable combinations of tariff revenue and producer surplus. Under  $FTAHub(a)$  or GFT, the small country achieves higher welfare through an increase in producer surplus and despite a decrease in tariff revenue and consumer surplus.  $CU(a,i)$  and  $FTA(a,i)$  each lead to further welfare improvements for the small country. In both cases, the driving factor is a higher consumer surplus. Next, regardless of the lower consumer and producer surplus under  $CU(a,j)$  (compared to  $FTA(a,i)$ ), its higher tariff revenue actually results in a higher welfare. The tariff revenue stays at its peak under  $MTA(i,j)$  or the MFN regimes as well, but with higher welfare. Specifically, an increase in consumer surplus offsets the decrease in producer surplus. By comparison,  $FTAHub(i)$  actually increases the producer surplus and thereby also the total welfare.  $FTA(a,j)$  then generates its peak tariff revenue from before and through this a higher welfare for the small country as an outsider. Under  $MTA(a,j)$  the tariff revenue stays the same and the lower producer surplus gets (more than) compensated by the higher consumer surplus. Compared to this,  $FTAHub(j)$  decreases the tariff revenue but increases both the consumer and the producer surplus enough to increase the welfare. Finally, a small country attains the best result by forming a PTA with the other small country, either through  $FTA(i,j)$

---

<sup>24</sup>Additional details on this can be found in Appendix B.2 and Appendix B.5.1.

or  $CU(i,j)$ . While keeping relatively high tariff revenues, the small countries manage to have a high producer surplus as well.

Using the preference rankings to derive the LCS yields the following proposition:

**Proposition 5.** *With the endowments given by  $e_a = e_{max}$  and  $e_b = e_c = e_{min}$ , and under the current institutional arrangement of the WTO, the stable constellations are the PTAs between the two small countries, that is  $CU(b,c)$  and  $FTA(b,c)$ .*

*Proof.* Let us start by giving the indirect dominance matrix:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1 <i>MFN</i>	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
2 $CU(a,b)$	1	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0
3 $CU(b,c)$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4 $CU(c,a)$	1	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0
5 <i>CUGFT</i>	1	1	1	1	0	0	1	0	0	0	0	0	0	1	0	0
6 $FTA(a,b)$	1	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0
7 $FTA(b,c)$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8 $FTA(c,a)$	1	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0
9 $FTAHub(a)$	1	0	1	0	0	1	1	1	0	1	1	0	0	1	0	0
10 $FTAHub(b)$	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
11 $FTAHub(c)$	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
12 <i>FTAGFT</i>	1	0	1	0	0	1	1	1	0	1	1	0	0	1	0	0
13 $MTA(a,b)$	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0
14 $MTA(b,c)$	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
15 $MTA(c,a)$	1	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0
16 <i>MTAGFT</i>	1	1	1	1	0	0	1	0	0	0	0	0	1	1	1	0

Recall that  $X$  denotes the full set and let  $Y = \{CU(b,c), FTA(b,c)\}$  be the candidate for the LCS. Take any element  $x$  from the set  $X \setminus Y$  and consider the deviation  $x \rightarrow_{\{b,c\}} CU(b,c)$ . Note that  $CU(b,c)$  is not indirectly dominated by any other element from  $X$  and furthermore  $x \prec_{\{b,c\}} CU(b,c)$  for all  $x \in X \setminus Y$ . Thus, the deviation  $x \rightarrow_{\{b,c\}} CU(b,c)$  can not be deterred for all  $x \in X \setminus Y$ . Therefore, no such  $x$  can be part of the stable set.<sup>25</sup>

As each outcome in  $X \setminus Y$  is indirectly dominated by  $y \in Y$  (see the matrix), for any coalition and any deviation away from  $y \in Y$  there always exists a path of indirect dominance back to  $Y$ . Moreover, no coalition is actually better off when coming back to  $Y$ , as  $x \not\prec_S y$  for all  $x, y \in Y$ ,  $x \neq y$ , and  $S \subseteq N$ ,  $S \neq \emptyset$ . Therefore, the set  $Y$  satisfies the (internal) stability condition while being maximal, i.e.  $Y = LCS$ .  $\square$

Even though global free trade is the most desirable regime for the large country, the two small countries do not have any incentive to form such a constellation and the large country can not enforce it. As a consequence, country  $a$  ends up with the worst trade agreement (from its perspective). Thus, in this scenario the size advantage of the large country does not translate into a favorable stable regime.

<sup>25</sup>It might appear that this proof deviates from the general approach of eliminating element by element from the full set until the remainder forms the stable set. However, in this proof it is purely a coincidence that in one step all elements but the stable ones can be eliminated with one argument (or rather deviation).

Moreover, this specific case showcases the relevance of the restrictions on PTAs (remember that insiders are not allowed to raise tariffs on outsiders). The constraint makes the small countries be indifferent between the two forms of PTAs.

Now we turn to the hypothetical scenario without Article XXIV Paragraph 5. Here, the ranking of preferences for the countries, with country  $a$  the large one and country  $b$  and  $c$  small (represented by  $i$  and  $j$ ), are as follows:

$$MTA(i, j), MFN \prec_a MTA(a, i) \prec_a GFT$$

$$MTA(a, i) \prec_i MTAGFT \prec_i MFN, MTA(i, j) \prec_i MTA(a, j)$$

As a result, the best outcome for a small country  $i$  is the  $MTA(i, j)$  regime, as the PTAs are not available anymore. The next proposition presents the new LCS as a consequence of these changes:

**Proposition 6.** *With the endowments given by  $e_a = e_{max}$  and  $e_b = e_c = e_{min}$ , and under a modified institutional arrangement of the WTO, the stable constellations are MFN and  $MTA(b, c)$ .*

*Proof.* The indirect dominance matrix is given as follows:

$$\begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 \\ 1 & MFN \\ 2 & MTA(a, b) \\ 3 & MTA(b, c) \\ 4 & MTA(c, a) \\ 5 & MTAGFT \end{matrix} \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix} \end{array}$$

Start with the full set again. If we consider the deviations  $MTA(c, a) \rightarrow_c MFN$  and  $MTA(a, b) \rightarrow_b MFN$ , then no further deviations are expected as  $MFN$  is not indirectly dominated by any other outcome. In addition,  $MTA(c, a) \prec_c MFN$  and  $MTA(a, b) \prec_b MFN$ , so  $MTA(c, a)$  and  $MTA(a, b)$  cannot be part of the stable set. The same argument works in the case of  $MTAGFT$  and the deviation  $MTAGFT \rightarrow_{b,c} MFN$ , as  $MTAGFT \prec_{b,c} MFN$ . So, the global free trade regime cannot be stable as well.

Let  $Y = \{MFN, MTA(b, c)\}$ . Following any deviation from the elements in  $Y$ , there is always an indirect dominance path coming back to  $Y$  ( $MFN$  in this case). In addition, for any  $x, y \in Y$  with  $x \neq y$  there does not exist coalition  $S$  for which  $x \prec_S y$ . Thus, the set  $Y$  is consistent and the largest one as well.  $\square$

In summary, when there are two small and one large country, the GFT regime is unstable under the current and hypothetical institutional set-up of the WTO. At best, world trade can be partially liberalized. Additionally, the small countries profit when they can form a PTA instead of an MTA, as the limiting MFN principle can be avoided that way.

**4.2.2. The case of one small and two large countries.** In this scenario, fix  $e_b = e_{min}$  and  $e_a = e_c = e_{max}$  (point R). Let us start with the ranking of preferences for country  $b$  and another country  $i \in N \setminus \{b\}$  - representing also  $j \in N \setminus \{b, i\}$ :

$$\begin{aligned} GFT \prec_b CU(i, b) \prec_b FTAH_{ub}(i) \prec_b FTA(i, b) \prec_b FTAH_{ub}(b) \prec_b MTA(i, b) \\ \prec_b MFN \prec_b CU(i, j) \prec_b MTA(i, j) \prec_b FTA(i, j) \end{aligned}$$

$$\begin{aligned}
CU(j, b) \prec_i FTA(j, b) \prec_i MFN \prec_i MTA(i, b), MTA(j, b) \prec_i FTA(i, j) \\
\prec_i MTA(i, j) \prec_i CU(i, j) \prec_i FTAHub(b) \prec_i FTAHub(j) \prec_i GFT \\
\prec_i FTAHub(i) \prec_i FTA(i, b) \prec_i CU(i, b)
\end{aligned}$$

Under the given pattern of endowments, the preference rankings of the countries are considerably different from the previous cases. For the small country, the MFN regime generates higher welfare than any other trade agreement where it is part of. As for a large country, being an outsider is on the lower end of the ranking, while being an insider in a PTA with a small country is on the other end.

Let us take a closer look at the preference ranking of the small country. First, GFT actually generates the lowest total welfare - driven by no tariff revenue and not enough compensation via consumer and producer surplus. As mentioned before, any trade arrangement involving the small country results in lower welfare compared with other constellations (but higher welfare than GFT). The lowest among those are the CU with any of the large countries, which through increased tariff revenue (and despite a decrease in consumer surplus) yield higher welfare in comparison with the GFT regime. Even though FTAHub(i) reduces those gains in tariff revenue again, by virtue of a growing consumer surplus it still raises the total welfare. Further improvement in the welfare of the small country is possible if the world moves from FTAHub(i) to FTA(i,b); the sole reason is a higher consumer surplus. Under FTAHub(b), the export volumes to the large countries are at its peak and it generates substantially higher producer surplus. As a consequence, it results in the small country preferring to form a hub structure (as the hub node) over an FTA with one of the large countries. Replacing the FTA with an MTA with similar structure is the most desirable configuration for the small country among the constellations where it participates. Under MTA(i,b) the producer surplus is actually the smallest compared to all other alternatives, but high tariff revenue and consumer surplus determine its position in the ranking. The MFN regime surpasses all configurations mentioned above. When there are two large countries, the tariff revenue becomes an important factor in the welfare of the small country. Any further improvements with respect to the welfare of the small country depend on the large countries liberalizing trade among themselves - the small country essentially free-rides in these cases (exhausting its tariff revenue to the fullest). The driving factor among these three is the export volume. Consequentially, CU(i,j) is the worst option, followed by MTA(i,j), and FTA(i,j) is the (overall) best outcome.

The following discusses the preferences of the two large countries. The least favorable scenario occurs when the other large country forms a CU together with the small country. Its position in the ranking is driven by the lowest export volumes and producer surplus. Now, FTA(j,b) produces higher welfare compared to the previous constellation due to growth in producer surplus (based on rising exports to the small country) which makes up for the drop in consumer surplus. A similar development makes the MFN regime an even better constellation (here the exports to the large country increase). All tariffs (and thus prices) are identical under both MTA(i,b) and MTA(j,b), as a consequence they generate the same welfare. On the grounds of increased exports, the welfare tops that of the MFN regime. Among the class of bilateral trade agreements between the large countries, the ranking goes as follows: FTA(i,j) followed by MTA(i,j) only surpassed by CU(i,j). In comparison with MTA(i,b) and MTA(j,b), the greater consumer and producer

surplus of  $FTA(i,j)$  guarantees an increase of total welfare. An MTA between the two large countries produces more tariff revenues and actually results in a more desirable outcome. Moving from  $MTA(i,j)$  to  $CU(i,j)$  decreases tariff revenue and also consumer surplus but the gain in producer surplus through increased exports to the other large country makes more than up for this.  $FTAHub(b)$  and even more so  $FTAHub(j)$  further improve the welfare via growth of the tariff revenue and consumer surplus (the case of  $FTAHub(b)$ ), and increased exports to the other large country (for  $FTAHub(j)$ ). Now, the GFT regime allows the large country to raise the exports to the small country while retaining the same level of exports to the other large country. As a consequence, the welfare of GFT surpasses that of the previous mentioned constellations. However, when the large country is part of a hub structure as the hub node itself, then its exports to the small country increase such that the welfare exceeds that of full trade liberalization. Furthermore, the FTA with the small country constitutes the second-best outcome for the large country on the grounds of high tariff revenue accompanied by similar consumer surplus. Finally,  $CU(i,b)$  is the most desirable constellation driven by the high exports to the small country.

Let us compute the LCS under these preference rankings in the next proposition:

**Proposition 7.** *With the endowments given by  $e_b = e_{min}$  and  $e_a = e_c = e_{max}$ , and under the current institutional arrangement of the WTO, the stable constellation is the CU between the two large countries, that is  $CU(c,a)$ .*

*Proof.* The indirect dominance matrix is given as follows:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1 <i>MFN</i>	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0
2 $CU(a,b)$	1	0	0	1	0	1	1	1	0	1	0	0	1	1	1	0
3 $CU(b,c)$	1	0	0	1	0	1	1	1	0	1	0	0	1	1	1	0
4 $CU(c,a)$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5 <i>CUGFT</i>	0	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0
6 $FTA(a,b)$	1	0	0	1	0	0	0	1	0	1	0	0	1	1	1	0
7 $FTA(b,c)$	1	0	0	1	0	0	0	1	0	1	0	0	1	1	1	0
8 $FTA(c,a)$	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
9 $FTAHub(a)$	0	0	0	1	0	1	1	1	0	0	0	0	0	0	1	0
10 $FTAHub(b)$	1	0	0	1	0	0	0	1	1	0	1	1	1	1	1	0
11 $FTAHub(c)$	0	0	0	1	0	1	1	1	0	0	0	0	0	0	1	0
12 <i>FTAGFT</i>	0	1	1	1	0	1	1	1	1	0	1	0	0	0	1	0
13 $MTA(a,b)$	1	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0
14 $MTA(b,c)$	1	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0
15 $MTA(c,a)$	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
16 <i>MTAGFT</i>	0	1	1	1	0	1	1	0	0	0	0	0	0	0	1	0

First, take  $x \in \{CU(i,b), FTA(i,b), MTA(i,b), FTAHub(i), FTAHub(b)\}$ , with  $i \in \{a,c\}$ . Country  $b$  can destroy such trade agreements and, depending on the initial constellation, either  $FTA(c,a)$  or the *MFN* regime remains. Then, further deviations are possible, namely  $MTA(c,a)$  and  $CU(c,a)$ . However, each of the aforementioned trade agreements is indirectly dominated by  $CU(c,a)$  and simultaneously country  $b$  is better off compared to the initial situation. Consequently, such deviations can not be avoided and no such  $x$  can be part of the stable set.

Now, consider  $x \in \{MFN, FTA(c, a), MTA(c, a)\}$  for which  $x \rightarrow_{\{a,c\}} CU(a, c)$  presents a deviation that can not be deterred. As in the previous paragraph,  $CU(c, a)$  is not indirectly dominated any element and also  $x \prec_{\{a,c\}} CU(a, c)$ . Thus, no such  $x$  can be the part of the stable set as well.

At last, let  $x \in \{CUGFT, FTAGFT, MTAGFT\}$  and consider the deviations where country  $b$  leaves the agreements.  $CU(c, a)$ ,  $FTA(c, a)$ , or  $MTA(c, a)$  can be the result. We have shown that the last two outcomes can not be stable. As for  $CU(a, c)$ , we have that for all  $x$  considered  $x \prec_{\{b\}} CU(a, c)$ . As a result, we conclude that no such  $x$  can be in the consistent set.

$CU(a, c)$  indirectly dominates each outcome, all deviations from it are deterred. So, the set containing  $CU(a, c)$  is consistent and the largest one as well.  $\square$

The small country manages to block many desirable outcomes for large countries. Country  $b$  can unilaterally deviate from any trade agreement with higher welfare than  $CU(i, j)$  for the large countries. Thus, the majority of countries cannot impose their will on the other country. What the large countries can achieve is the best trade agreement that they can reach without the participation of the small country, in this case one among themselves.

A similar story unfolds in the scenario without Article XXIV Paragraph 5. There, the countries' preference rankings are as follows, with country  $b$  the small one and country  $a$  and  $c$  large (represented by  $i$  and  $j$ ):

$$MTAGFT \prec_b MTA(i, b) \prec_b MFN \prec_b MTA(i, j)$$

$$MFN \prec_i MTA(i, b), MTA(j, b) \prec_i MTA(i, j) \prec_i MTAGFT$$

As the logic of the corresponding preference rankings of the countries is similar to before, let us directly present the proposition:

**Proposition 8.** *With the endowments given by  $e_b = e_{min}$  and  $e_a = e_c = e_{max}$ , and under a modified institutional arrangement of the WTO, the stable constellation is the MTA between the two large countries, that is  $MTA(c, a)$ .*

*Proof.* In this case, the indirect dominance matrix has the following form:

$$\begin{array}{l} 1 \text{ } MFN \\ 2 \text{ } MTA(a, b) \\ 3 \text{ } MTA(b, c) \\ 4 \text{ } MTA(c, a) \\ 5 \text{ } MTAGFT \end{array} \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Assume,  $x \in \{MTA(a, b), MTA(b, c), MTAGFT\}$  and consider the deviations, where country  $b$  dismantles any above mentioned constellation. Two possibilities: Either  $MFN$  or  $MTA(c, a)$  remain. From  $MFN$  either no coalition moves away or, as it is indirectly dominated by  $MTA(c, a)$  (see the indirect dominance matrix), the latter might be approached. In either case,  $b$  is better off. Thus, no such  $x$  can be part of the stable set.

Now, analyze the case of the  $MFN$  regime. Take the following deviation:  $MFN \rightarrow_{\{a,c\}} MTA(c, a)$ . As  $MTA(c, a)$  is not indirectly dominated by any other trade agreement and  $MFN \prec_{\{a,c\}} MTA(c, a)$ , the  $MFN$  regime can not be stable as well.

As  $MTA(c, a)$  indirectly dominates each trade agreement, all deviations from it are deterred. So, the set consisting of  $MTA(c, a)$  is consistent and the largest one as well.  $\square$

Thus, similar to the other asymmetric case, one small and two large countries allow for partial but not full liberalization of world trade irrespective of the actual scenario (current vs. modified WTO rules). In terms of overall welfare, the world is better off in the hypothetical scenario without Article XXIV Paragraph 5 though. Individually, the small country is in a better position in case of  $MTA(i, j)$  compared to  $CU(i, j)$ , as it exploits the MFN obligation of the large countries. By contrast, the large countries are better off in the other case. Therefore, while none of the two institutional arrangement facilitate global free trade, they influence the welfare for the stable set (both overall and individual).

**4.3. Asymmetric Case - Edges of the Triangle.** Let us now turn to the cases where the endowments of countries vary along one dimension - corresponding to the sides PQ, QR, and PR in the triangle of Figure 4. Specifically, Section 4.3.1 presents the scenario where the countries are small, small, and varying (PQ), Section 4.3.2 discusses the setting where the countries are small, large, and varying (QR), and Section 4.3.3 describes the case of a small country with two varying equally (PR).

While in the previous cases it was still possible to solve the problems analytically, the following require the use of a numerical approach. The analysis presented here consists of graphics picturing the composition of the stable sets and accompanying descriptions that explore the underlying mechanics. The exact numerical values for these (sub-)intervals can be found in Section C.2.

**4.3.1. The case of one small, one large, and one varying country.** First, let us consider the case  $e_b = e_{min}$ ,  $e_a = e_{max}$ , and  $e_c \in (e_{min}, e_{max})$  (side QR). Under the given pattern of the endowments, a number of trade agreements can be completely ruled out (with respect to the LCS). The MFN and GFT regimes for example are never part of the stable set. Additionally, none of the PTAs between the small and the large country appear as a stable outcome. The same holds for the hub structures where either the small or the large country is the hub node. As for the actual composition of the LCS, see Figure 5 for a graphical representation.

The general observation is that when the varying country is close in size to the small country, then the PTAs between these smaller countries appear as elements in the stable set. When the country becomes larger the trade constellation between the larger countries replaces these. Additionally, there are two small, separated, regions in the middle of the interval where  $FTA_{Hub}(c)$  is stable.

In order to get an intuitive understanding of the results, let us identify specific trade agreements that go from stable to unstable (or the other way around) for certain endowment tuples. Then, explore the underlying mechanics to understand why the changes happen.

Start with the PTAs between country  $b$  and  $c$ , the small and the varying one. Interestingly, the only factor driving their stability are the preferences of country  $b$  (with fixed minimal endowments). Once the MFN regime becomes more desirable than  $CU(b, c)$  for country  $b$ , the constellation  $CU(b, c)$  drops from the stable set. Now, an identical story holds for the case of  $FTA(b, c)$ . Thus, for both constellations it only requires a single change in the preference ranking of country  $b$  to influence the stable set.

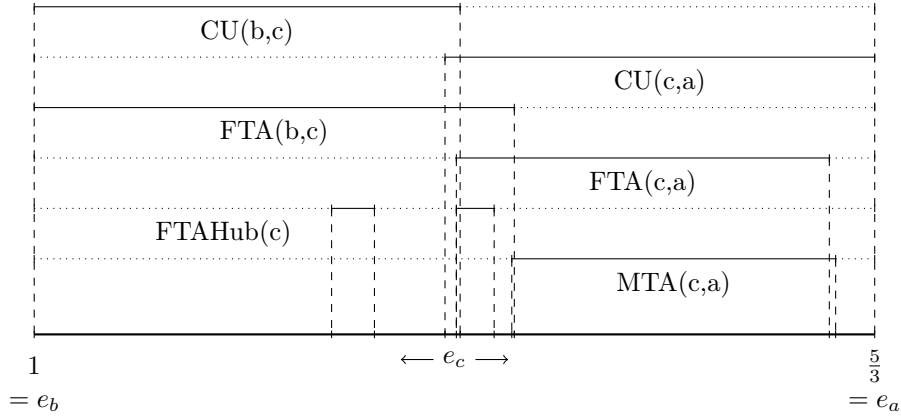


FIGURE 5. Characterization of the case of small, varying, and large country

The PTAs and MTAs between country  $a$  and  $c$  start to appear in the LCS when country  $c$  is becoming relatively large and closer to country  $a$  in size. At first both countries actually prefer to form a CU with country  $b$ , that is when country  $c$  is relatively small (and  $CU(b,c)$  actually is an element in the stable set). However, once it is preferable for country  $b$  to be the outsider instead of the insider in a CU,  $CU(c,a)$  emerges as a stable outcome (even though  $CU(b,c)$  still remains stable). Moreover, as soon as country  $c$  prefers  $FTA(c,a)$  respectively  $MTA(c,a)$  over the MFN regime, each of them becomes part of the LCS as well. For the interval where all PTAs and MTAs between country  $a$  and  $c$  are stable, both countries have fixed preference relations over these outcomes:

$$\begin{aligned} FTA(c,a) \prec_a CU(c,a) \prec_a MTA(c,a) \\ MTA(c,a) \prec_c FTA(c,a) \prec_c CU(c,a) \end{aligned}$$

However, as soon as country  $c$  also prefers  $MTA(c,a)$  over  $FTA(c,a)$ , the joint FTA drops out of the LCS. Similarly, as soon as country  $a$  prefers  $CU(c,a)$  over  $MTA(c,a)$ , this also applies to the joint MTA - leaving  $CU(c,a)$  as the only stable outcome.

$FTAHub(c)$  is stable in the two small, separated, regions in (or near) the middle of the interval. In the first region, the stability is driven by the fact that country  $b$  starts to value  $FTAHub(c)$  more than  $FTA(b,c)$  and gets in unison with country  $a$  in this respect. Once the preferences of country  $b$  over these outcomes get reversed,  $FTAHub(c)$  drops out of LCS again. In the second region, the stability of the same hub structure is largely determined by the change in the preferences of country  $c$ . Now, as soon as it starts to value  $FTA(c,a)$  over the MFN regime, which also puts  $FTA(c,a)$  in the LCS, both FTAs with  $c$  as a partner are stable and consequentially the corresponding hub structure is stable as well. As soon as the free-riding incentives of country  $b$  increase (valuing the MFN regime more than  $FTAHub(c)$ ), this hub structure is not part of the stable set anymore.

The hypothetical institutional arrangement without Article XXIV Paragraph 5 does not promote the appearance of GFT as part of the stable set. GFTMTA, but also  $MTA(a,b)$  and  $MTA(b,c)$  never emerge as stable outcomes. Varying the size



of country  $c$  generates either the MFN regime or  $\text{MTA}(c,a)$  as the stable element. Figure 6 presents these findings.<sup>26</sup>

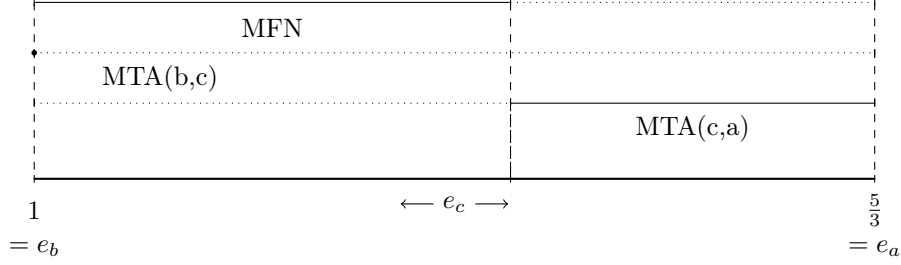


FIGURE 6. Characterization of the case of small, varying, and large country

Over the whole interval, country  $b$  does not have any incentive to form an MTA with any of the other countries. This is one reason why the MFN regime is stable over the specific range of the interval. The other reason is that country  $c$  prefers to not have a trade agreement with country  $a$  as long as its own size is not too large. Once country  $c$  gets sufficiently large though,  $\text{MTA}(c,a)$  presents a better option than the MFN regime. As a consequence,  $\text{MTA}(c,a)$  replaces the MFN regime as the stable set.

As a sidenote, while in the first scenario (with PTAs), the LCS near and at each respective extreme point corresponded to each other (continuity), the situation is different in the second scenario (without PTAs). When country  $c$  and  $b$  are equal in size,  $\text{MTA}(b,c)$  appears in the LCS even though it is not there before. Here, both the MFN regime and  $\text{MTA}(b,c)$  generate the same welfares for all countries (see also the discussion on point Q in Section 4.2.1).

Finally, under this given pattern of endowments, the GFT regime does not appear as part of the stable set independent of the scenario (with and without PTAs). However, the choice of rules does determine whether partial trade liberalization takes place or not. The possibility of forming PTAs reduces the incentive of the small(est) country to free ride. Otherwise, the MFN regime is the unique stable outcome when there is one small, one large, and one comparably small country.

**4.3.2. The case of two small, and one varying country.** Second, let us showcase the scenario with  $e_b = e_c = e_{min}$  and  $e_a \in (e_{min}, e_{max})$  (side PQ). In contrast to the previous case, it is not possible to rule out many of the trade agreements. Only MFN,  $\text{FTAHub}(a)$ , and  $\text{MTA}(b,c)$  never appear in the LCS. The stable set is then presented in Figure 7.<sup>27</sup>

In the immediate vicinity around symmetry, the GFT regime is the only element of the LCS (or rather the group of the three variants forms the stable set), but both  $\text{FTAHub}(b)$  and  $\text{FTAHub}(c)$  emerge as stable outcomes when moving away from the extreme point. On the whole interval a number of different PTAs and MTAs,

<sup>26</sup>In addition to the aforementioned elements, it also pictures  $\text{MTA}(b,c)$  as a single point, see the dot, but this appears only for completion sake because that point corresponds to one of the extreme cases (point Q) discussed earlier.

<sup>27</sup>The dot marks a single point again.

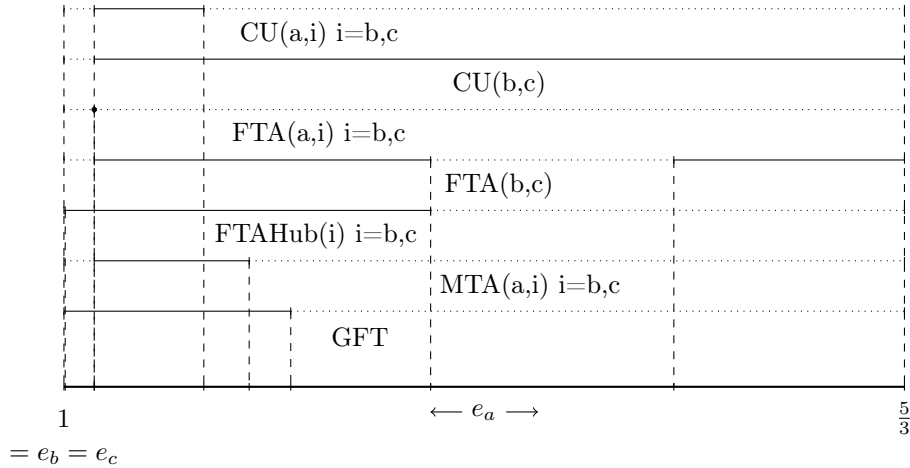


FIGURE 7. Characterization of the case of small, small, and varying country

mostly between a small and the larger country, appear. Near the other point, only PTAs among the small countries are still stable.

First, the spike in the number of stable constellations close to symmetry actually follows a change in the preferences of the varying country with respect to  $CU(b,c)$  and the  $GFT$  regimes - it starts preferring the first over the latter. Furthermore,  $FTA(a,b)$  and  $FTA(c,a)$  become unstable because the small countries start to like the  $MFN$  regime more than the  $GFT$  variants (or rather these are only stable for that instance where it is not the case). When country  $a$  gets sufficiently large, country  $b$  prefers  $FTAHub(c)$  over  $CU(a,b)$  and  $c$  prefers  $FTAHub(b)$  over  $CU(c,a)$ . As a consequence, both of these  $CUs$  drop out from the  $LCS$ . Similarly, when country  $b$  and  $c$  start preferring  $FTAHub(c)$  and  $FTAHub(b)$  over  $GFT$ , the latter stops being stable. A similar argument also applies to the  $MTAs$ . When the size of country  $a$  increases even more, both country  $b$  and  $c$  favor  $CU(b,c)$  over their respective hub structure, which results in  $FTA(b,c)$ ,  $FTAHub(b)$ , and  $FTAHub(c)$  becoming unstable. When the endowment of country  $a$  gets close to maximum, the small countries are constrained by the  $MFN$ -tariffs and do not differentiate between  $CU(b,c)$  and  $FTA(b,c)$  anymore, which makes  $FTA(b,c)$  stable again.

In the scenario without Article XXIV Paragraph 5, the interval of  $GFT$  increases significantly. Moreover, over two-thirds of this interval the  $GFT$  regime is the unique element in the  $LCS$ . Additionally, all possible combinations of  $MTA$  appear at some point (mostly close to symmetry). Figure 8 demonstrates the results.<sup>28</sup>

Around symmetry,  $GFTMTA$  is the only element in the stable set. As soon as the small countries start to prefer  $MTAs$  with country  $a$  over  $GFTMTA$ , all three  $MTAs$  appear in the  $LCS$ . When the size of country  $a$  increases, the  $MTAs$  drop out from the  $LCS$ , because the small countries rank the one with the large country as the worst trade agreements (switching last place with the  $MFN$  regime), which actually also influences the stability of the  $MTA$  among themselves. Furthermore,

<sup>28</sup>As before, in addition to the mentioned trade agreements, the graphic also contains  $MFN$  as a single point, see the dot, at an extreme point (again point  $Q$ ).

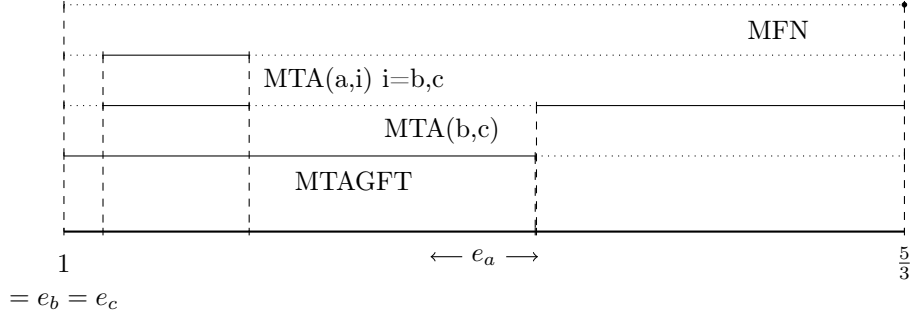


FIGURE 8. Characterization of the case of small, small, and varying country

the GFT regime becomes unstable when the small countries start to prefer their joint MTA over GFTMTA.

Similar to the previous case, the LCS changes at one extreme point. Namely, when the endowment of country  $a$  reaches the maximum, the MFN regime appears in the LCS, as MFN and MTA(b,c) generate identical welfare for all countries (again, see also the discussion on point Q in Section 4.2.1).

Under this pattern of endowments, the first scenario does not allow for a sharp prediction via the LCS (unlike the previous case). Especially around symmetry, where almost all trade agreements are part of the stable set. In the second scenario, the effect of the PTAs on the stability of the GFT regime is significant though - essentially the abolishment of Article XXIV Paragraph 5 would facilitate the formation of GFT as long as there are two small countries and the third country is not substantially larger.

**4.3.3. The case of one small, and two varying countries.** Finally, let us turn to the case where  $e_b = e_{min}$  and  $e_a = e_c \in (e_{min}, e_{max})$  (side PR). In this scenario, depending on the size of the larger countries, any trade agreement can be part of the stable set. The exact composition of the LCS is the basis for Figure 9.

In the interval around symmetry, the GFT variants are stable and stay the unique elements of the stable set for longer (compared to the previous case). Also, a collection of different trade agreements is stable relatively close to symmetry. However, near the other extreme point, the CUs between the varying countries is the unique stable outcome. Also, MFN is stable in two small, separated, regions.

Again, the peak in stability near symmetry comes from a shift in the preferences of the varying countries with respect to CUs. At that point, both of these start to prefer a CU with the small country over the different forms of GFT. The occurrence of the MFN regime actually follows a preference of the small country of MFN over GFT (the first region) and then FTAHub(a) and FTAHub(c) (the second region). As countries  $a$  and  $c$  are getting bigger, first MTA(a,b) and MTA(b,c) drop out from the LCS when they rank the lowest according to the preferences of country  $b$ . The three variants of GFT become unstable once the small country prefers CU(c,a). Next, CU(a,b) and CU(b,c) follow as the small country starts to prefer to be in the MFN regime over a CU with any of the larger countries. As soon as CU(c,a) becomes the more desirable trade agreement for country  $b$  when comparing it to

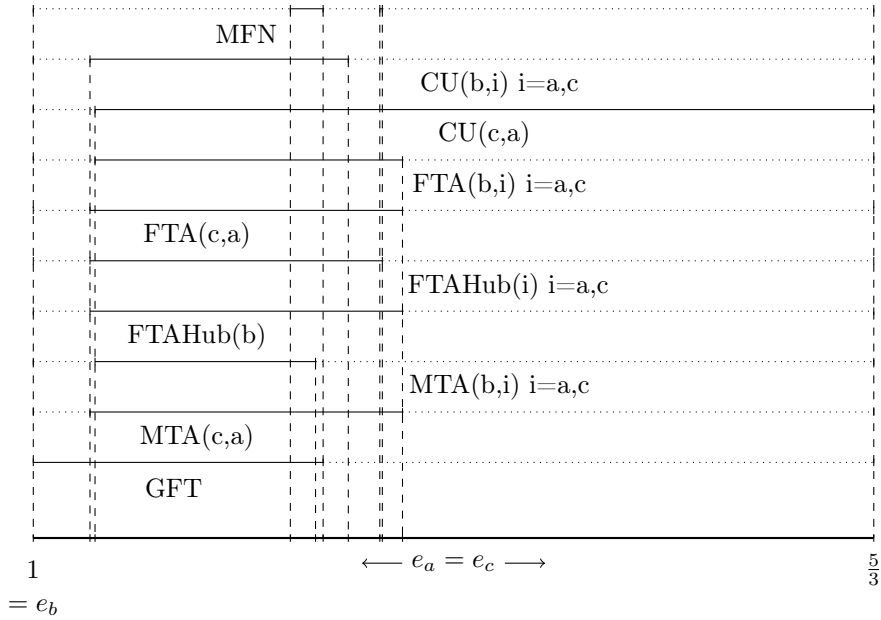


FIGURE 9. Characterization of the case of small, varying, and varying country

any FTA where  $b$  participates or any hub structure with a large country as hub, all aforementioned constellations drop out from the LCS.

Contrary to the previous case, switching off Article XXIV Paragraph 5, actually decreases the interval where the GFT regime is part of the stable set. However, this effect is considerably smaller. A similar observation holds for the range where the GFT regime is the unique stable outcome. The exact composition can be seen in Figure 10.

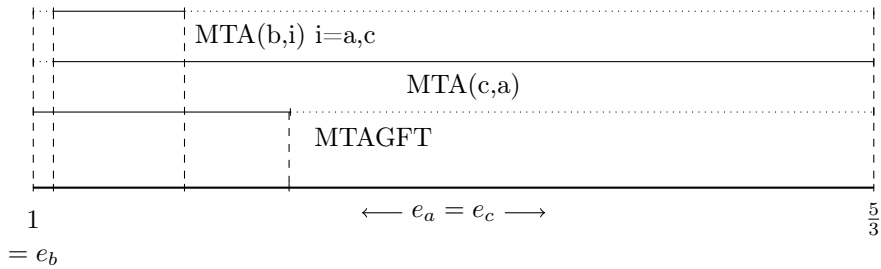


FIGURE 10. Characterization of the case of small, varying, and varying country

The main driving force behind the stability are alterations in the preferences of the small country over the interval. More precisely, it is important where exactly the small country places the MFN regime in its ranking of preferences compared to the other trade agreements. As soon as country  $b$  prefers MFN over another

constellation, the latter drops out from the LCS. After a certain point, the MTA between the large countries remains the only stable outcome.

Similar to the previous pattern of endowments, this case makes a clear analysis in the first scenario difficult, especially around symmetry where, as before, almost all constellations are stable. The effect of Article XXIV Paragraph 5 actually works in the other direction on the stability of the GFT regime when compared to the previous case though.

**4.4. Asymmetric Case - Interior of the Triangle.** In the following (and final) part of the analysis, the focus lies on the interior of the triangle of Figure 4. Here, unlike in the previous discussions, both CU and FTA appear together under the label of PTA. However, a variation of the graphics of this analysis that actually distinguishes between the two can be found in Appendix C.1. For the purpose of a general overview, this level of abstraction suffices though - in fact, the members of a specific trade agreement are suppressed for clarity as well, i.e. who is insider and outsider.

First, we consider the existing institutional set-up, where PTAs are available to the countries. Figure 11 shows the (simplified) stable sets. In a small region close to symmetry, region one, the trinity of GFT regimes is the unique stable element. In both a neighboring and another distant area, region two, PTAs become stable as well. The connecting area, region three, adds MTAs as another stable element. In a tiny area near the diagonal, region four, no form of trade agreement can actually be excluded from the stable set. Further along the diagonal and in the asymmetric corners, region five, PTAs are the only stable trade constellation. In between, region six, MTAs are also stable. In another tiny area, also close to the diagonal, region seven, MFN enters the stable set as well. In general, with a certain degree of asymmetry among countries at most partial trade liberalization can be expected.

Next, we consider a modified institutional set-up, where PTAs are not available. Figure 12 depicts the corresponding stable sets. In an area near symmetry as well as in a sizeable area away from it, region one, GFT is again the unique stable element. Connected to these are two areas, region two, where MTAs become stable as well. Moving towards the asymmetric corners, region three, yields MTAs as the only stable element. In between, region four, only MFN remains in the stable set.

The comparison of the graphics allows us to deduce two compelling statements. The first noteworthy result is the extent of MFN in each scenario. In the modified institutional arrangement without PTAs the area where MFN is (uniquely) stable increases substantially (note that this effect is present away from symmetry). Under (significant) asymmetry, it seems that PTAs allow countries to move towards their international efficiency frontier (cf. Bagwell et al. (2016)).

The second interesting result is the difference in the extent of stability of GFT in the two regulatory scenarios. First, recall that once the degree of asymmetry surpasses a certain threshold, none of the GFT regimes remains in the stable set, independent of the institutional set-up. Around symmetry the opposite holds in that the GFT regimes are always stable there (in both scenarios). In between, the effect of PTAs on the stability of GFT depends on the structure of asymmetry. See Figure 13 for the different areas of stability depending on the regulatory scenario. Note that region one corresponds to the aforementioned stability around symmetry. In the case of two relatively larger countries (but not too large), the abolishment of PTAs results in a reduction of the area where GFT is stable, see region two.

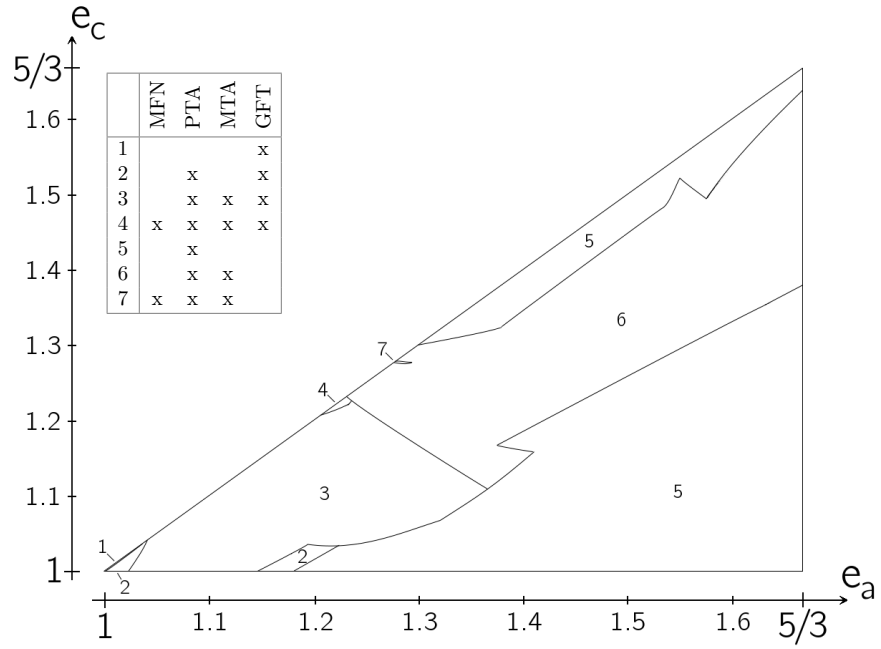


FIGURE 11. Simplified Overall Stability with PTAs

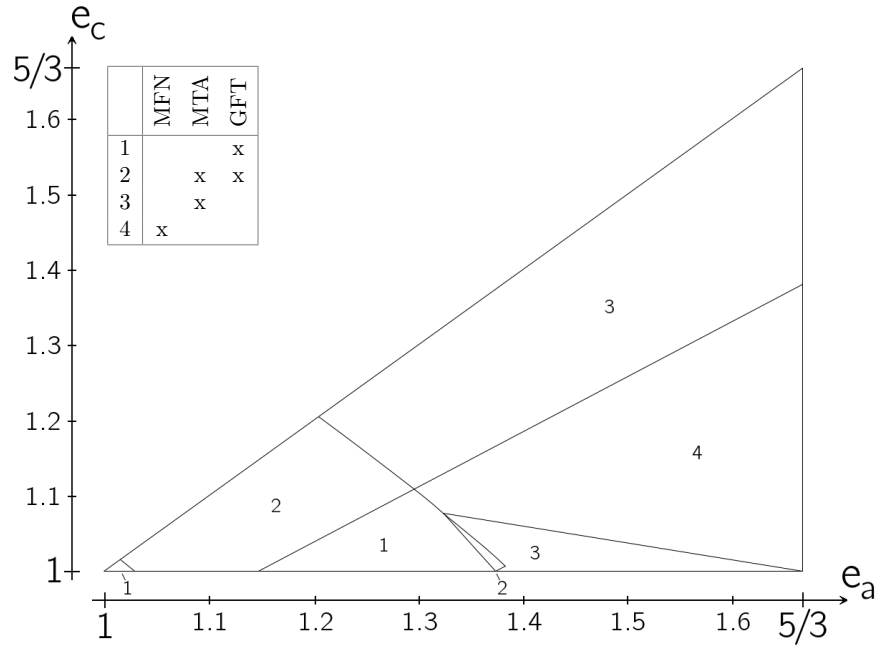


FIGURE 12. Simplified Overall Stability without PTAs

In this instance, PTAs act as ‘building blocks’ on the road to GFT. But when two countries are relatively smaller (but not too small), the same regulatory action yields the exact opposite effect, see region three. Here, PTAs are ‘stumbling blocks’. Thus, whether PTAs are ‘building blocks’ or ‘stumbling blocks’ in the vicinity of symmetry depends on the relative size of the majority of the countries.

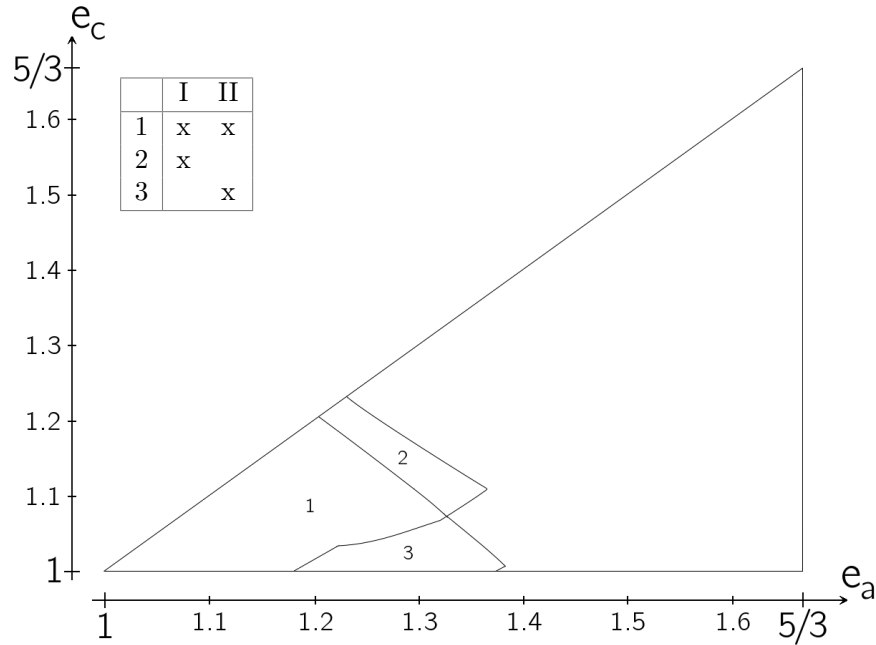


FIGURE 13. The different areas of stability of the GFT regime in the scenario with (I) and without (II) PTAs

In a nutshell: If the world is in the vicinity of symmetry and two out of three countries are close to identical while relatively smaller than the other one, the area where the GFT regime is stable increases when prohibiting PTAs. However, when two similar countries are relatively larger, the availability of PTAs is conducive to the stability of the GFT regime. Finally, if the world is further away from symmetry, full trade liberalization is not attainable at all and an area where the MFN regime is stable appears in the scenario without PTAs.

## 5. DISCUSSION

In this section, let us first compare the findings of our paper with those of several similar studies and underline the differences in the modeling strategies, especially with respect to the explanatory power of each approach. Second, this section links our predictions to different empirical observations, thereby validating our approach.

Let us start with the paper of Saggi, Woodland and Yildiz (2013).<sup>29</sup> First note that the underlying trade model in our paper is similar to theirs, which allows a direct comparison of the findings in certain scenarios (found in the next paragraph). The first distinction is the set of trade agreements under consideration. While in our model countries can be involved in multilateral trade liberalization via MTAs or they may choose to carry out their favored form of preferential trade liberalization through CUs or FTAs, Saggi, Woodland and Yildiz (2013) focuses on two out of these three possibilities, namely CUs and MTAs. In our opinion, the expanded set of trade arrangements in our model allows us to fully capture the trade-offs among the alternatives and make the model realistic. The second significant difference is the concept of stability. While our framework uses the notion of LCS, the paper of Saggi, Woodland and Yildiz (2013) utilizes Coalition-Proof Nash Equilibria. As Bernheim, Peleg and Whinston (1987) note, their notion of self-enforceability, which is critical for Coalition-Proof Nash Equilibria, is too restrictive in one crucial aspect, mainly: ‘When a deviation occurs, only members of the deviating coalition may contemplate deviations from the deviation. This rules out the possibility that some member of the deviating coalition might form a pact to deviate further with someone not included in this coalition.’ Importantly, this limitation does not affect the concept of LCS. It is not a pure academic difference. The historic development of the two (disjoint) trade constellations in Europe in the 1960s, the European Economic Community (EEC) and the European Free Trade Association (EFTA), can not be captured by a model using Coalition-Proof Nash Equilibria, because it excludes those strategies that the UK actually followed during that time.<sup>30</sup> While being a member of EFTA the UK applied for EEC membership in 1961 and thereby undermined the stability of the EFTA. Furthermore, ‘the more ambitious Kennedy Round between 1964 and 1967 coincided with negotiations to expand the EEC to include Britain, Ireland, Denmark, Greece, and Norway - and was motivated in part by US concerns about being excluded from an ever-broader and more unified European market.’ (World Trade Report 2011). Thus, unlike the Coalition-Proof Nash Equilibrium, the LCS allows interactions among members and non-members of coalitions simultaneously, thereby accommodating these historic developments. Additionally, the (conjectured) motivation of the US reinforces the importance of the interaction among different modes and forms of trade liberalization.

In specific cases it is actually possible to directly compare the composition of the stable sets of our paper with those of Saggi, Woodland and Yildiz (2013). In fact, their ‘multilateralism game’ fits our scenario of a modified institutional arrangement without PTAs. Compare Figures 2 and 5 of Saggi, Woodland, and Yildiz (2013) with Figures 10 and 8 in our paper correspondingly. In the case with one small country and the other two varying, both approaches predict the same stable sets near the endpoints of the interval. In our paper GFT stays part of the stable set even when MTAs, either between the large countries or a small and a large country, become stable as well - which is in contrast to Saggi, Woodland, and Yildiz (2013). A similar observation follows for the case of two small and one varying country, i.e.

---

<sup>29</sup>As mentioned in Section 2, this paper analyzes the case of CUs and MTA while their other papers (Saggi and Yildiz (2010) resp. Missions, Saggi and Yildiz (2016)) focus on different combinations of trade agreements (FTAs and MTAs resp. CUs and FTAs) but use a similar framework. As a consequence, the comparison of methodology applies to these papers as well.

<sup>30</sup>See Baldwin and Gylfason (1995).



near the endpoints of the interval results coincide while the appearance of MTAs does not prevent GFT from staying in the stable set. Furthermore, in our model there exists an interval where GFT becomes the unique stable element once more. It seems that one effect of the unlimited farsightedness is the proliferation of GFT.

Another relevant paper is that of Lake (2017). Apart from the stability concept, the approach of Lake also differs with respect to the choice set. There, the focus lies on FTAs. In this respect, a direct comparison of the findings is difficult. Moreover, compared to the previous paper, the ‘multilateralism game’ is further simplified in Lake (2017), as the only possible regime there is the three country constellation that results in GFT. Furthermore, as the underlying trade model, the paper employs the political economy oligopolistic model. However, according to the paper, the findings are robust with respect to various underlying trade models, including the competition via exports model. Additionally, Lake himself compares his results to those of Saggi and Yildiz (2010). Due to the similarity of the ‘multilateralism game’ in Saggi and Yildiz (2010) and Saggi, Woodland, and Yildiz (2013), it is only logical to compare our results with those of Lake (2017) as with the previous paper.

According to Lake (2017), specifically Figure 3, the exact role of FTAs under asymmetry depends on the nature of asymmetry (similar to our findings). However, the direction of the effect of PTAs on trade liberalization is the opposite. There, in case of two larger and one small country, FTAs act as ‘(strong) stumbling blocs’, and with two smaller and one larger country, as ‘(strong) building blocs’. Furthermore, there it seems that the determining factor are the preferences of the larger countries, while in our case the findings are driven by the preferences of the smaller countries. Thus, the aforementioned differences in the choice set and stability concept appear to shift the power to influence the negotiations among the countries, which then produces a different outcome. Specifically, in the case of the ‘multilateralism game’, which corresponds to our scenario without PTA, there are essentially two areas, that is one where GFT emerges as unique equilibrium and one with MFN instead. In the parameter space triangle the first makes up the upper left part of the triangle, while the second makes up the opposing lower right.<sup>31</sup> Therefore, for two larger countries the GFT regime remains the unique equilibrium for the whole interval, whereas in our model it only stays stable in the vicinity of symmetry (only partially unique) and then the MTA between the larger countries is the unique stable outcome, as seen in Figure 10. Furthermore, for two smaller countries first GFT then MFN is the unique equilibrium, while in the beginning GFT is also stable in our case (although only partially unique) the MTA between the small countries takes it place as the unique stable element near the end, depicted in Figure 8. Finally, in the case of three different countries, it starts with MFN and ends with GFT as the unique equilibrium, which corresponds to our findings for the first part but then in the second part the MTA between the medium and large country is the unique stable element, visible in Figure 6.

Another aspect of Lake (2017) necessitates a remark, namely the assumption that a once created trade agreement remains binding from then on. Lake argues that ‘the binding nature of trade agreements is pervasive in the literature and realistic’. However, the latest developments in the world cast doubt on this plausibility. The USA, for example, pulled out of the negotiations for the Trans-Pacific Partnership

---

<sup>31</sup>The first corresponds to the regions denoted WBB and SSB in Figure 3 of Lake (2017), while the second matches the regions SBB and WSB.

at the final stage and currently negotiates with South Korea to amend the so-called KORUS FTA. The developments around ‘Brexit’ are another argument for modeling non-binding trade agreements. Using the LCS as stability concept allowed us to accommodate such deviations.

A final remark on the relation of our research with empirical observations. As the analysis has shown, a growing degree of asymmetry among countries produced a significant area of stability for the MFN regime when PTAs are prohibited, see region four in Figure 12. If one would interpret the expansion of the WTO rule set to an increasing number of countries as an amplification of asymmetry among its member states, then the potential of PTAs to prevent the MFN regime might be one of the driving factors of the prevalence of PTAs in recent history. However, the World Trade Report (2011) casts doubt on this motive. According to the report: ‘Approximately 66 per cent of tariff lines with MFN rates above 15 percentage points have not been reduced in PTAs.’ Note that a reduction of the tariff peaks for 34 percent of the tariff lines is still a significant effect considering the fact that the majority of tariff peaks occurs in agricultural and labor-intensive manufacturing sectors, which are politically sensitive and countries usually try to exclude them from the trade liberalization via PTAs. Furthermore, according to the same report, over time more and more PTAs have included provisions regarding technical barriers to trade - a category of the Non-Tariff Barriers (NTBs). The paper of Kee, Nicita, and Olarreaga (2009) estimates that restrictiveness measures that include NTBs are on average about 87 percent higher than the measures based on tariffs alone. Thus, in order to evaluate the aforementioned motive properly, the effects of NTBs should be included in the analysis as well. Moreover, contrary to the conclusion of the report that ‘preference margins are small and market access is unlikely in many cases to be an important reason for creating new PTAs’, Keck and Lendle (2012) show that the preferential utilization rates are often high even in the case of small preference margins and they increase both with the preference margin and the export volume. All in all, it is our opinion that there is (partial) evidence for the motive of avoidance of MFN via PTAs which coincides with the predictions of our model about the trade-off between trade liberalization and the MFN regime in the asymmetric part of the parameter space.

As this overview showed, a number of core attributes of the LCS, specifically the farsightedness, the non-binding nature of agreements, and the possibility of interactions between members and non-members of coalitions, capture important mechanisms present in the world and influence the composition of the stable set significantly (when compared to other stability concepts).

## 6. CONCLUSION

Under the rules of the WTO (previously GATT), a group of countries can engage in both multilateral and preferential trade liberalization. The formation of global trade agreements is a complex game and the rules of the game influence the nature of the exact outcomes. WTO’s Article I aims at creating the global free trade system, while Article XXIV Paragraph 5 allows countries to seemingly circumvent the liberalization process. In this paper, our focus lies on the stability of trade policy arrangements under two different regulatory scenarios (with and without PTAs) assuming unlimited farsightedness of the participants in the trade negotiations and considering an extensive set of trade agreements - moving our model closer to reality.

Unfortunately, the answer to the question whether PTAs are ‘building blocks’ or ‘stumbling blocks’ on the path towards global free trade is not as straightforward as one would like it to be. In the end, the results presented here are mixed and depend on the size distribution of the countries. Under symmetry, GFT is the unique stable trade constellation in both regulatory scenarios. But as soon as one moves away from symmetry, GFT might not be reached at all. In between, the effect of switching off Article XXIV Paragraph 5 depends on the exact asymmetry. In case two countries are relatively smaller, prohibiting PTAs increases the area of stability of the GFT regimes. When two countries are relatively larger, it reduces the area. Once the world is further away from symmetry, abolishing the exception for PTAs might result in the worst possible state from the perspective of overall world welfare, the non-cooperative MFN regime. Therefore, under such circumstances, PTAs act as a mechanism that prevents the MFN regime.

Our research also raises a couple of questions in need of further investigation. First, it would be interesting to study the robustness of the findings with respect to the underlying trade model. While the model of competition via exports remains popular in the related literature, economists also extensively use both oligopoly and competition via imports model. Fortunately, the framework presented here does allow for a different underlying trade model such as the ones mentioned above. Another potential area of inquiry might be an extension of the framework to increase the number of countries. Nowadays, in addition to bilateral negotiations, so-called plurilateral negotiations play an important role in the development of preferential trade liberalization. Recent examples are the Trans-Pacific Partnership (TTP) and the Regional Comprehensive Economic Partnership (RCEP). Including more than three countries in a model would allow us to investigate the strategic interactions among countries whilst taking these negotiations into account. The introduction of political economy considerations to the underlying trade model is another area of interest<sup>32</sup>, as it might allow us to understand the nature of tariff peaks occurring after PTAs come into effect. It is our opinion that modifications or extensions of our framework (as mentioned here) are directions worthy of further research.

As a final remark, it is perhaps important, going forward, to move the debate of ‘building blocks’ vs. ‘stumbling blocks’ to a level of detail that goes beyond this binary choice.

#### REFERENCES

- [1] General agreement on tariffs and trade, 1947. Source: <http://www.wto.org>.
- [2] World trade report 2011, 2011. Source: <http://www.wto.org>.
- [3] Philippe Aghion, Pol Antràs, and Elhanan Helpman. Negotiating free trade. *Journal of International Economics*, 73(1):1–30, 2007.
- [4] Robert Aumann and Roger Myerson. Endogenous formation of links between players and coalitions: an application of the shapley value. *The Shapley Value*, pages 175–191, 1988.
- [5] Kyle Bagwell, Chad P. Bown, and Robert W. Staiger. Is the wto passé? *Journal of Economic Literature*, 54(4):1125–1231, December 2016.
- [6] Kyle Bagwell and Robert W. Staiger. Regionalism and multilateral tariff cooperation. Working Paper 5921, National Bureau of Economic Research, February 1997.
- [7] Kyle Bagwell and Robert W Staiger. The design of trade agreements. *Handbook of Commercial Policy*, 1:435–529, 2016.
- [8] Richard E. Baldwin and Thorvaldur Gylfason. A domino theory of regionalism. In *Expanding Membership of the European Union*, pages 25–53. Cambridge University Press, 1995.

---

<sup>32</sup>See for example Facchini et al. (2013).

- [9] B. Douglas Bernheim, Bezalel Peleg, and Michael Whinston. Coalition-proof nash equilibria i. concepts. *Journal of Economic Theory*, 42(1):1–12, 1987.
- [10] Jagdish Bhagwati. Regionalism and multilateralism: an overview. *New dimensions in regional integration*, 22:51, 1993.
- [11] Michael Suk-Young Chwe. Farsighted coalitional stability. *Journal of Economic Theory*, 63(2):299 – 325, 1994.
- [12] Giovanni Facchini, Peri Silva, and Gerald Willmann. The customs union issue: Why do we observe so few of them? *Journal of International Economics*, 90(1):136 – 147, 2013.
- [13] Taiji Furusawa and Hideo Konishi. Free trade networks. *Journal of International Economics*, 72(2):310–335, 2007.
- [14] Sanjeev Goyal and Sumit Joshi. Bilateralism and free trade. *International Economic Review*, 47(3):749–778, 2006.
- [15] Gene M Grossman. The purpose of trade agreements. *Handbook of Commercial Policy*, 1:379–434, 2016.
- [16] John Harsanyi. An equilibrium-point interpretation of stable sets and a proposed alternative definition. *Management Science*, 20(11):1472–1495, 1974.
- [17] P. Jean-Jacques Herings, Ana Mauleon, and Vincent Vannetelbosch. Farsightedly stable networks. *Games and Economic Behavior*, 67(2):526–541, 2009.
- [18] P.J.J. Herings, A. Mauleon, and V. Vannetelbosch. Stability of networks under level-k farsightedness. Research Memorandum 030, Maastricht University, Graduate School of Business and Economics (GSBE), January 2014.
- [19] Matthew O Jackson and Asher Wolinsky. A strategic model of social and economic networks. *Journal of economic theory*, 71(1):44–74, 1996.
- [20] Alexander Keck and Andreas Lendle. New evidence on preference utilization. WTO Staff Working Papers ERSD-2012-12, World Trade Organization (WTO), Economic Research and Statistics Division, September 2012.
- [21] Hiau Looi Kee, Alessandro Nicita, and Marcelo Olarreaga. Estimating trade restrictiveness indices. *The Economic Journal*, 119(534):172–199, 2009.
- [22] James Lake. Free trade agreements as dynamic farsighted networks. *Economic Inquiry*, 55(1):31–50, 2017.
- [23] Giovanni Maggi. International trade agreements. In *Handbook of International Economics*, volume 4, chapter 6, pages 317–390. 2014.
- [24] Paul Missios, Kamal Saggi, and Halis Murat Yildiz. External trade diversion, exclusion incentives and the nature of preferential trade agreements. *Journal of International Economics*, 99:105–119, 2016.
- [25] Kamal Saggi, Alan Woodland, and Halis Murat Yildiz. On the relationship between preferential and multilateral trade liberalization: the case of customs unions. *American Economic Journal: Microeconomics*, 5(1):63–99, 2013.
- [26] Kamal Saggi and Halis Murat Yildiz. Bilateralism, multilateralism, and the quest for global free trade. *Journal of International Economics*, 81(1):26–37, 2010.
- [27] John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.
- [28] Jin Zhang, Zhiwei Cui, and Lei Zu. The evolution of free trade networks. *Journal of Economic Dynamics and Control*, 38:72–86, 2014.
- [29] Jin Zhang, Licun Xue, and Lei Zu. Farsighted free trade networks. *International Journal of Game Theory*, 42(2):375–398, 2013.

## APPENDIX A. PSEUDOCODE

Note, that a couple of functions and variables are directly baked into the program without any further explanation in the pseudocode below - for example the matrix that determines the general network structure (for each player and all coalitions). The origin and characterization of these can be found in their respective parts in the main paper. The network structure  $A$  and the preference relations  $B$  both enter as a collection of  $|X| \times |X|$ -matrices,  $\{A_S\}_{S \subseteq N, S \neq \emptyset}$  and  $\{B_S\}_{S \subseteq N, S \neq \emptyset}$  resp., where  $(A_S)_{i,j} = \mathbb{1}_{\{i \rightarrow_S j\}}(i, j)$  and  $(B_S)_{i,j} = \mathbb{1}_{\{i \prec_S j\}}(i, j)$  for  $(i, j) \in \bar{X} \times X$ .

---

**Algorithm** Largest Consistent Set

---

**Input:** Countries  $N$ , Outcomes  $X$ , Network Structure  $A$ , Preference Relations  $B$ **Output:** Largest Consistent Set  $\{Y\}$ 

```

1: procedure PARAMETERSPACELCS( $N, X, A, B$ )
2:    $E = eMaxArea$  ▷ See Section 3.6
3:    $\alpha = \alpha MinValue(E)$  ▷ See Section 3.6
4:   for  $e \in E$  do
5:      $Y = GeneralLCS(N, X, A, B)$ 
6:   return  $\{Y\}$ 

7: function GENERALLCS( $N, X, A, B$ )
8:   for  $S \subseteq N$  do
9:      $C_S = \min\{A_S, B_S\}$ 
10:   $D^0 = \max_{S \subseteq N} \{C_S\}$  ▷ : Direct Dominance
11:   $n = 0$ 
12:  repeat
13:     $n = n + 1$ 
14:    for  $S \subseteq N$  do
15:       $A_S^n = (\mathbb{1}_{\{(A_S \cdot D^{n-1})_{i,j} \neq 0\}}(i, j))_{(i,j) \in X \times X}$ 
16:       $D_S^n = \min\{A_S^n, B_S\}$ 
17:       $D^n = \max_{S \subseteq N} \{D_S^n\}$  ▷ : Indirect Dominance
18:    until  $D^n = D^{n-1}$ 
19:     $D = \mathbb{1}_X + D^n$ 
20:     $Y^0 = (1)_{x \in X}$ 
21:     $m = 0$ 
22:    repeat
23:       $m = m + 1$ 
24:      for  $x \in X$  do
25:        if  $Y_x^{m-1} = 0$  then
26:           $Y_x^m = 0$ 
27:        else
28:           $y = \max_{k \in X, S \subseteq N} \left\{ (A_S)_{x,k} \left( 1 - \max_{z \in X} \{Y_z^{m-1}(D)_{k,z} (1 - (B_S)_{x,z})\} \right) \right\}$ 
29:           $Y_x^m = Y_x^{m-1} - y$ 
30:        until  $Y^m = Y^{m-1}$ 
31:     $Y = Y^m$ 
32:  return  $Y$ 

```

---

## APPENDIX B. MODEL

**B.1. Individual Welfare.** The following table lists the individual welfare for each (representative) trade agreement, depending on endowments and tariffs, multiplied with the factor 18. Note that for MFN, CUGFT, FTAGFT, and MTAGFT the welfare  $W_i$  resembles  $W_j$  and  $W_k$ . In case of CU(i,j), FTA(i,j), and MTA(i,j) the welfare  $W_i$  is similar to  $W_j$ . For FTAHub(i) the welfare  $W_j$  resembles  $W_k$ .

Trade Agreement	Individual Welfare
MFN	
$\hookrightarrow W_i$	$-10e_i^2 + 2e_j^2 + 2e_k^2 - 8t_i^2 + t_j^2 + t_k^2 + 4e_i(9\alpha - e_j - e_k - t_j - t_k) + 2e_j(e_k + t_i + t_k) + 2e_k(t_i + t_j)$
CU(i,j)	
$\hookrightarrow W_i$	$-10e_i^2 + 2e_j^2 + 2e_k^2 - 11t_{ik}^2 + t_{jk}^2 + t_k^2 + 4e_i(9\alpha - e_j - e_k + t_{jk} - t_k) + 2e_j(e_k - 4t_{ik} + t_k) + 2e_k(5t_{ik} - t_{jk})$
$\hookrightarrow W_k$	$2e_i^2 + 2e_j^2 - 10e_k^2 + 4t_{ik}^2 + 4t_{jk}^2 - 8t_k^2 + 2e_i(e_j - 2e_k + 2t_{jk} + t_k) + 2e_j(-2e_k + 2t_{ik} + t_k) + 4e_k(9\alpha - 2t_{ik} - 2t_{jk})$
CUGFT	
$\hookrightarrow W_i$	$-10e_i^2 + 2e_j^2 + 2e_k^2 + 4e_i(9\alpha - e_j - e_k) + 2e_j e_k$
FTA(i,j)	
$\hookrightarrow W_i$	$-10e_i^2 + 2e_j^2 + 2e_k^2 - 11t_{ik}^2 + t_{jk}^2 + t_k^2 + 4e_i(9\alpha - e_j - e_k + t_{jk} - t_k) + 2e_j(e_k - 4t_{ik} + t_k) + 2e_k(5t_{ik} - t_{jk})$
$\hookrightarrow W_k$	$2e_i^2 + 2e_j^2 - 10e_k^2 + 4t_{ik}^2 + 4t_{jk}^2 - 8t_k^2 + 2e_i(e_j - 2e_k + 2t_{jk} + t_k) + 2e_j(-2e_k + 2t_{ik} + t_k) + 4e_k(9\alpha - 2t_{ik} - 2t_{jk})$
FTAHub(i)	
$\hookrightarrow W_i$	$-10e_i^2 + 2e_j^2 + 2e_k^2 + t_{jk}^2 + t_{kj}^2 + 4e_i(9\alpha - e_j - e_k + t_{jk} + t_{kj}) + 2e_j(e_k - t_{kj}) - 2e_k t_{jk}$
$\hookrightarrow W_j$	$2e_i^2 + 2e_j^2 - 10e_k^2 + 4t_{jk}^2 - 11t_{kj}^2 + 2e_i(e_j - 2e_k + 2t_{jk} - 4t_{kj}) + e_j(-4e_k + 10t_{kj}) + 4e_k(9\alpha - 2t_{jk})$
FTAGFT	
$\hookrightarrow W_i$	$-10e_i^2 + 2e_j^2 + 2e_k^2 + 4e_i(9\alpha - e_j - e_k) + 2e_j e_k$
MTA(i,j)	
$\hookrightarrow W_i$	$-10e_i^2 + 2e_j^2 + 2e_k^2 - 8t_i^2 + t_j^2 + t_k^2 + 4e_i(9\alpha - e_j - e_k - t_j - t_k) + 2e_j(e_k + t_i + t_k) + 2e_k(t_i + t_j)$
$\hookrightarrow W_k$	$2e_i^2 + 2e_j^2 - 10e_k^2 + t_i^2 + t_j^2 - 8t_k^2 + 2e_i(e_j - 2e_k + t_j + t_k) + 2e_j(-2e_k + t_i + t_k) + 4e_k(9\alpha - t_i - t_j)$
MTAGFT	
$\hookrightarrow W_i$	$-10e_i^2 + 2e_j^2 + 2e_k^2 + 4e_i(9\alpha - e_j - e_k) + 2e_j e_k$

TABLE 14. The individual welfare for each trade agreement depending on endowments and tariffs

**B.2. Tariffs.** The following describes the tariffs that the countries choose for each trade agreement. In addition to the specific restrictions mentioned in Section 3.3, all tariffs are bounded both from below and above by zero and the MFN-tariff respectively. As per WTO rule, the formation of any PTA does not allow additional tariffs towards others - which results in the upper bound of the MFN-tariff. Also, any form of subsidies is excluded here - which results in the lower bound of zero. Now, the following determines and describes the optimal tariffs for each scenario and the cases where capping occurs:

**B.2.1. MFN.** In this case, the optimal tariff of country  $i$ , given by  $t_i^* = \frac{1}{8}(e_j + e_k)$ , is always greater than zero as the endowments themselves are greater than zero. Additionally,  $t_i^*$  is going to play the role of the maximal tariff for country  $i$  for all the other agreements, then denoted  $t_i^{MFN}$ .

**B.2.2. CU.** Consider the scenario CU(i,j), then the optimal tariff of country  $i$  towards country  $k$ , given by  $t_{ik}^* = \frac{1}{5}(2e_k - e_j)$ , is always greater than zero but not always less than the MFN-tariff (and the one towards country  $j$ ,  $t_{ij}^*$ , is always zero):

i) Lower Bound. By assumption on the endowments  $e_k \geq \frac{3}{5}e_j$  and thus  $e_k > \frac{1}{2}e_j$ , which guarantees  $t_{ik}^* > 0$ .

ii) Upper Bound. By assumption on the endowments  $e_k \leq \frac{5}{3}e_j$  however  $t_{ik}^* \leq t_i^{MFN}$  requires  $e_k \leq \frac{13}{11}e_j$ , which leaves the interval  $\frac{13}{11}e_j < e_k \leq \frac{5}{3}e_j$  to require capping. For this interval, the (maximal) MFN-tariff is optimal as the derivative of the joint welfare with respect to  $t_{ik}$  is always greater than zero on the interval  $[0, t_i^{MFN}]$ :

$$\frac{\partial(W_i + W_j)}{\partial t_{ik}} = \frac{1}{9}(-10t_{ik} - 2e_j + 4e_k) \geq \frac{1}{36}(-13e_j + 11e_k) > 0$$

**B.2.3. FTA.** Consider the scenario FTA(i,j), then the optimal tariff of country  $i$  towards country  $k$ , given by  $t_{ik}^* = \frac{1}{11}(5e_k - 4e_j)$ , is neither always greater than zero nor always less than the MFN-tariff (but the one towards country  $j$ ,  $t_{ij}^*$ , is zero):

i) Lower Bound. By assumption on the endowments  $e_k \geq \frac{3}{5}e_j$  however  $t_{ik}^* \geq 0$  requires  $e_k \geq \frac{4}{5}e_j$ , which leaves the interval  $\frac{3}{5}e_j \leq e_k < \frac{4}{5}e_j$  to require capping. For this interval, the (minimal) zero-tariff is optimal as the derivative of the welfare with respect to  $t_{ik}$  is always lesser than zero on the interval  $[0, t_i^{MFN}]$ :

$$\frac{\partial W_i}{\partial t_{ik}} = \frac{1}{9}(-11t_{ik} - 4e_j + 5e_k) \leq \frac{1}{9}(5e_k - 4e_j) < 0$$

ii) Upper Bound. By assumption on the endowments  $e_k \leq \frac{5}{3}e_j$  however  $t_{ik}^* \leq t_i^{MFN}$  requires  $e_k \leq \frac{43}{29}e_j$ , which leaves the interval  $\frac{43}{29}e_j < e_k \leq \frac{5}{3}e_j$  to require capping. For this interval, the (maximal) MFN-tariff is optimal as the derivative of the welfare with respect to  $t_{ik}$  is always greater than zero on the interval  $[0, t_i^{MFN}]$ :

$$\frac{\partial W_i}{\partial t_{ik}} = \frac{1}{9}(-11t_{ik} - 4e_j + 5e_k) \geq \frac{1}{72}(-43e_j + 29e_k) > 0$$

**B.2.4. MTA.** Consider the scenario MTA(i,j), then the optimal tariff of country  $i$ , given by  $t_i^* = \frac{1}{7}(2e_k - e_j)$ , is greater than zero and less or equal to the MFN-tariff as per assumption on the endowments  $\frac{3}{5}e_j \leq e_k \leq \frac{5}{3}e_j$ .

B.2.5. *Notes.* The analysis considered country  $i$  and an agreement with country  $j$ , but it naturally extends to all other combinations. Also, the perspective of the third country needs no further analysis as it always chooses the MFN-tariff. Furthermore, the case of FTAHub(i) is simply a combination of FTA(i,j) and FTA(i,k). Finally, the three variants of GFT require no additional analysis as every country always chooses the zero-tariff. Information on another form of GFT, Pseudo-GFT, that technically exists but turns out to be negligible, can be found in Appendix B.5.1.

B.3. **Overall Welfare.** The following table lists the overall welfare for each (representative) trade agreement, depending purely on endowments, computed modulo  $2\alpha (\sum_{n \in N} e_n)$ , which is the common term associated with the factor  $\alpha$ . Also, the notation  $l_c$  and  $l^c$  is used to indicate that country  $l$  is capped in terms of tariffs from below or above respectively. Note that one specific comparison of trade agreements is presented in more detail in Appendix B.5.2.

Trade Agreement	Overall Welfare
MFN	
↳ no cap	$\frac{11}{32}(-e_i^2 - e_i e_j - e_i e_k - e_j^2 - e_j e_k - e_k^2)$
CU(i,j)	
↳ no cap	$\frac{1}{1600}(-563e_i^2 - 550e_i e_j - 448e_i e_k - 563e_j^2 - 448e_j e_k - 704e_k^2)$
↳ $i^c$	$\frac{1}{1600}(-563e_i^2 - 550e_i e_j - 448e_i e_k - 550e_j^2 - 550e_j e_k - 627e_k^2)$
↳ $i^c, j^c$	$\frac{11}{32}(-e_i^2 - e_i e_j - e_i e_k - e_j^2 - e_j e_k - e_k^2)$
CUGFT	
↳ no cap	$\frac{1}{3}(-e_i^2 - e_i e_j - e_i e_k - e_j^2 - e_j e_k - e_k^2)$
FTA(i,j)	
↳ no cap	$\frac{1}{7744}(-2963e_i^2 - 2662e_i e_j - 1728e_i e_k - 2963e_j^2 - 1728e_j e_k - 3648e_k^2)$
↳ $i_c$	$\frac{1}{23232}(-8889e_i^2 - 7986e_i e_j - 5184e_i e_k - 7865e_j^2 - 7744e_j e_k - 9344e_k^2)$
↳ $i_c, i_c$	$\frac{1}{192}(-65e_i^2 - 66e_i e_j - 64e_i e_k - 65e_j^2 - 64e_j e_k - 64e_k^2)$
↳ $i^c$	$\frac{1}{7744}(-2963e_i^2 - 2662e_i e_j - 1728e_i e_k - 2662e_j^2 - 2662e_j e_k - 3155e_k^2)$
↳ $i^c, j^c$	$\frac{11}{32}(-e_i^2 - e_i e_j - e_i e_k - e_j^2 - e_j e_k - e_k^2)$
FTAHub(i)	
↳ no cap	$\frac{1}{363}(-153e_i^2 - 81e_i e_j - 81e_i e_k - 146e_j^2 - 121e_j e_k - 146e_k^2)$
↳ $j_c$	$\frac{1}{363}(-137e_i^2 - 81e_i e_j - 121e_i e_k - 146e_j^2 - 121e_j e_k - 121e_k^2)$
↳ $j_c, k_c$	$\frac{1}{3}(-e_i^2 - e_i e_j - e_i e_k - e_j^2 - e_j e_k - e_k^2)$
↳ $j^c$	$\frac{1}{23232}(-8889e_i^2 - 5184e_i e_j - 7986e_i e_k - 9344e_j^2 - 7744e_j e_k - 7865e_k^2)$
↳ $j^c, k^c$	$\frac{1}{192}(-66e_i^2 - 66e_i e_j - 66e_i e_k - 65e_j^2 - 64e_j e_k - 65e_k^2)$
FTAGFT	
↳ no cap	$\frac{1}{3}(-e_i^2 - e_i e_j - e_i e_k - e_j^2 - e_j e_k - e_k^2)$
MTA(i,j)	
↳ no cap	$\frac{1}{3136}(-1083e_i^2 - 1078e_i e_j - 960e_i e_k - 1083e_j^2 - 960e_j e_k - 1216e_k^2)$
MTAGFT	
↳ no cap	$\frac{1}{3}(-e_i^2 - e_i e_j - e_i e_k - e_j^2 - e_j e_k - e_k^2)$

TABLE 15. The overall welfare for each trade agreement depending on endowments



**B.4. Transition Tables.** The following lists the network structure of Section 3.5, specifically Figure 1 and 2, in the form of transition tables:

$x_1 \in X$	$x_2 \in X \setminus \{x_1\}$ with $x_1 \rightarrow_{\{i\}} x_2$
MFN	-
CU(i,j)	MFN
CU(j,k)	-
CU(k,i)	MFN
CUGFT	CU(j,k)
FTA(i,j)	MFN
FTA(j,k)	-
FTA(k,i)	MFN
FTAHub(i)	MFN, FTA(i,j), FTA(k,i)
FTAHub(j)	FTA(j,k)
FTAHub(k)	FTA(j,k)
FTAGFT	FTA(j,k), FTAHub(j), FTAHub(k)
MTA(i,j)	MFN
MTA(j,k)	-
MTA(k,i)	MFN
MTAGFT	MTA(j,k)

(A) The transition table for coalition  $\{i\}$ ,  $i \in N$ .

$x_1 \in X$	$x_2 \in X \setminus \{x_1\}$ with $x_1 \rightarrow_{\{i,j\}} x_2$
MFN	CU(i,j), FTA(i,j), MTA(i,j)
CU(i,j)	MFN, FTA(i,j), MTA(i,j)
CU(j,k)	MFN, CU(i,j), FTA(i,j), MTA(i,j)
CU(k,i)	MFN, CU(i,j), FTA(i,j), MTA(i,j)
CUGFT	MFN, CU(i,j), CU(j,k), CU(k,i), FTA(i,j), MTA(i,j)
FTA(i,j)	MFN, CU(i,j), MTA(i,j)
FTA(j,k)	MFN, CU(i,j), FTA(i,j), FTAHub(j), MTA(i,j)
FTA(k,i)	MFN, CU(i,j), FTA(i,j), FTAHub(i), MTA(i,j)
FTAHub(i)	MFN, CU(i,j), FTA(i,j), FTA(k,i), MTA(i,j)
FTAHub(j)	MFN, CU(i,j), FTA(i,j), FTA(j,k), MTA(i,j)
FTAHub(k)	MFN, CU(i,j), FTA(i,j), FTA(j,k), FTA(k,i), FTAGFT, MTA(i,j)
FTAGFT	MFN, CU(i,j), FTA(i,j), FTA(j,k), FTA(k,i), FTAHub(i), FTAHub(j), FTAHub(k), MTA(i,j)
MTA(i,j)	MFN, CU(i,j), FTA(i,j)
MTA(j,k)	MFN, CU(i,j), FTA(i,j), MTA(i,j)
MTA(k,i)	MFN, CU(i,j), FTA(i,j), MTA(i,j)
MTAGFT	MFN, CU(i,j), FTA(i,j), MTA(i,j), MTA(j,k), MTA(k,i)

(B) The transition table for coalition  $\{i, j\}$ ,  $i, j \in N$ ,  $i \neq j$ .

TABLE 16. The network structure as transition tables

### B.5. Additional Remarks.

B.5.1. *Pseudo-GFT.* In Appendix B.2 a special case of ‘Pseudo-GFT’ is a possibility. Namely, in the case of a hub structure with both non-hub nodes capping at zero the trade agreement amounts to the same tariff structure (and welfare) of a GFT. If it were ever part of the stable set, then it would necessarily need to be considered de facto GFT even though it is not de jure GFT. However, in our analysis this case never occurred and it is therefore a negligible oddity.

B.5.2. *A Special Case.* As can be seen in Table 15 the overall welfare is equal in case of MFN,  $CU(i^c, j^c)$ , and  $FTA(i^c, j^c)$  even though the tariff structure is different. The following explores this equivalence in order to provide an insight into the underlying mechanics. In terms of tariff structure both  $CU(i^c, j^c)$  and  $FTA(i^c, j^c)$  are the same and therefore it is sufficient to compare MFN with  $CU(i^c, j^c)$  when only interested in (effects on) welfare. Now, Table 17 shows us the differences in the welfare (components) both on the individual as well as on the joint/overall level, which are computed from the expressions in Table 18 and 19.

$\Delta(MFN, CU(i^c, j^c))$	
$TR_i$	$1/24(e_j + e_k)(2e_j - e_k)$
$CS_i$	$(32e_i^2 + 64e_i e_k - 13e_j^2 - 26e_j e_k + 19e_k^2)/1152$
$PS_i$	$1/12e_i(-e_i - e_k)$
$W_i$	$(-64e_i^2 - 32e_i e_k + 83e_j^2 + 22e_j e_k - 29e_k^2)/1152$
$TR_j$	$1/24(e_i + e_k)(2e_i - e_k)$
$CS_j$	$(-13e_i^2 - 26e_i e_k + 32e_j^2 + 64e_j e_k + 19e_k^2)/1152$
$PS_j$	$1/12e_j(-e_j - e_k)$
$W_j$	$(83e_i^2 + 22e_i e_k - 64e_j^2 - 32e_j e_k - 29e_k^2)/1152$
$TR_k$	0
$CS_k$	$19(-e_i^2 - 2e_i e_k - e_j^2 - 2e_j e_k - 2e_k^2)/1152$
$PS_k$	$1/24e_k(e_i + e_j + 2e_k)$
$W_k$	$(-19e_i^2 + 10e_i e_k - 19e_j^2 + 10e_j e_k + 58e_k^2)/1152$

(A) The difference in the individual welfare (components) depending on endowments

$\Delta(MFN, CU(i^c, j^c))$	
$TR_i + TR_j$	$1/24(2e_i^2 + e_i e_k + 2e_j^2 + e_j e_k - 2e_k^2)$
$CS_i + CS_j$	$19(e_i^2 + 2e_i e_k + e_j^2 + 2e_j e_k + 2e_k^2)/1152$
$PS_i + PS_j$	$1/12(-e_i^2 - e_i e_k - e_j^2 - e_j e_k)$
$W_i + W_j$	$(19e_i^2 - 10e_i e_k + 19e_j^2 - 10e_j e_k - 58e_k^2)/1152$
$\sum_{n \in N} TR_n$	$1/24(2e_i^2 + e_i e_k + 2e_j^2 + e_j e_k - 2e_k^2)$
$\sum_{n \in N} CS_n$	0
$\sum_{n \in N} PS_n$	$1/24(-2e_i^2 - e_i e_k - 2e_j^2 - e_j e_k + 2e_k^2)$
$\sum_{n \in N} W_n$	0

(B) The difference in the joint/overall welfare (components) depending on endowments

TABLE 17. The difference in the welfare (components) depending on endowments

Welfare Components	
MFN	
$TR_i$	$(e_j + e_k)^2/32$
$CS_i$	$(18e_i^2 + 13e_j^2 + 13e_k^2 + 8e_j e_k + 18e_i(e_j + e_k))/128$
$PS_i$	$-e_i(-16\alpha + 6e_i + 3e_j + 3e_k)/8$
$TR_j$	$(e_i + e_k)^2/32$
$CS_j$	$(13e_i^2 + 18e_j^2 + 13e_k^2 + 8e_i e_k + 18e_j(e_i + e_k))/128$
$PS_j$	$-e_j(-16\alpha + 3e_i + 6e_j + 3e_k)/8$
$TR_k$	$(e_i + e_j)^2/32$
$CS_k$	$(13e_i^2 + 13e_j^2 + 18e_k^2 + 8e_i e_j + 18e_k(e_i + e_j))/128$
$PS_k$	$-e_k(-16\alpha + 3e_i + 3e_j + 6e_k)/8$
CU( $i^c, j^c$ )	
$TR_i$	$-(5e_j - 7e_k)(e_j + e_k)/96$
$CS_i$	$(65e_i^2 + 65e_j^2 + 49e_k^2 + 49e_j e_k + e_i(81e_j + 49e_k))/576$
$PS_i$	$-e_i(-48\alpha + 16e_i + 9e_j + 7e_k)/24$
$TR_j$	$-(5e_i - 7e_k)(e_i + e_k)/96$
$CS_j$	$(65e_i^2 + 65e_j^2 + 49e_k^2 + 49e_i e_k + e_j(81e_i + 49e_k))/576$
$PS_j$	$-e_j(-48\alpha + 9e_i + 16e_j + 7e_k)/24$
$TR_k$	$(e_i + e_j)^2/32$
$CS_k$	$(17e_i^2 + 17e_j^2 + 25e_k^2 + 25e_i e_k + 25e_j e_k + 9e_i e_j)/144$
$PS_k$	$-e_k(-24\alpha + 5e_i + 5e_j + 10e_k)/12$

(A) The individual welfare (components) depending on endowments

Welfare Components	
MFN	
$TR_i + TR_j$	$1/32(e_i^2 + 2e_i e_k + e_j^2 + 2e_j e_k + 2e_k^2)$
$CS_i + CS_j$	$1/128(31e_i^2 + 36e_i e_j + 26e_i e_k + 31e_j^2 + 26e_j e_k + 26e_k^2)$
$PS_i + PS_j$	$1/8(-6e_i^2 - 6e_i e_j - 3e_i e_k - 6e_j^2 - 3e_j e_k) + 2\alpha(e_i + e_j)$
$\sum_{n \in N} TR_n$	$1/16(e_i^2 + e_i e_j + e_i e_k + e_j^2 + e_j e_k + e_k^2)$
$\sum_{n \in N} CS_n$	$11/32(e_i^2 + e_i e_j + e_i e_k + e_j^2 + e_j e_k + e_k^2)$
$\sum_{n \in N} PS_n$	$1/4(-3e_i^2 - 3e_i e_j - 3e_i e_k - 3e_j^2 - 3e_j e_k - 3e_k^2) + 2\alpha(\sum_{n \in N} e_n)$
CU( $i^c, j^c$ )	
$TR_i + TR_j$	$1/96(-5e_i^2 + 2e_i e_k - 5e_j^2 + 2e_j e_k + 14e_k^2)$
$CS_i + CS_j$	$1/288(65e_i^2 + 81e_i e_j + 49e_i e_k + 65e_j^2 + 49e_j e_k + 49e_k^2)$
$PS_i + PS_j$	$1/24(-16e_i^2 - 18e_i e_j - 7e_i e_k - 16e_j^2 - 7e_j e_k) + 2\alpha(e_i + e_j)$
$\sum_{n \in N} TR_n$	$1/48(-e_i^2 + 3e_i e_j + e_i e_k - e_j^2 + e_j e_k + 7e_k^2)$
$\sum_{n \in N} CS_n$	$11/32(e_i^2 + e_i e_j + e_i e_k + e_j^2 + e_j e_k + e_k^2)$
$\sum_{n \in N} PS_n$	$1/24(-16e_i^2 - 18e_i e_j - 17e_i e_k - 16e_j^2 - 17e_j e_k - 20e_k^2) + 2\alpha(\sum_{n \in N} e_n)$

(B) The joint/overall welfare (components) depending on endowments

TABLE 18. The welfare (components) depending on endowments

Trade Agreement	Individual/Joint/Overall Welfare
MFN	
$\hookrightarrow W_i$	$1/128(-78e_i^2 + 256\alpha e_i - 30e_i e_j - 30e_i e_k + 17e_j^2 + 16e_j e_k + 17e_k^2)$
$\hookrightarrow W_j$	$1/128(17e_i^2 - 30e_i e_j + 16e_i e_k - 78e_j^2 + 256\alpha e_j - 30e_j e_k + 17e_k^2)$
$\hookrightarrow W_k$	$1/128(17e_i^2 + 16e_i e_j - 30e_i e_k + 17e_j^2 - 30e_j e_k - 78e_k^2 + 256\alpha e_k)$
$\hookrightarrow W_i + W_j$	$1/128(-61e_i^2 - 60e_i e_j - 14e_i e_k - 61e_j^2 - 14e_j e_k + 34e_k^2) + 2\alpha(e_i + e_j)$
$\hookrightarrow \sum_{n \in N} W_n$	$1/32(-11e_i^2 - 11e_i e_j - 11e_i e_k - 11e_j^2 - 11e_j e_k - 11e_k^2) + 2\alpha(\sum_{n \in N} e_n)$
CU( $i^c, j^c$ )	
$\hookrightarrow W_i$	$1/576(-319e_i^2 + 1152\alpha e_i - 135e_i e_j - 119e_i e_k + 35e_j^2 + 61e_j e_k + 91e_k^2)$
$\hookrightarrow W_j$	$1/576(35e_i^2 - 135e_i e_j + 61e_i e_k - 319e_j^2 + 1152\alpha e_j - 119e_j e_k + 91e_k^2)$
$\hookrightarrow W_k$	$1/288(43e_i^2 + 36e_i e_j - 70e_i e_k + 43e_j^2 - 70e_j e_k - 190e_k^2 + 576\alpha e_k)$
$\hookrightarrow W_i + W_j$	$1/288(-142e_i^2 - 135e_i e_j - 29e_i e_k - 142e_j^2 - 29e_j e_k + 91e_k^2) + 2\alpha(e_i + e_j)$
$\hookrightarrow \sum_{n \in N} W_n$	$1/32(-11e_i^2 - 11e_i e_j - 11e_i e_k - 11e_j^2 - 11e_j e_k - 11e_k^2) + 2\alpha(\sum_{n \in N} e_n)$

TABLE 19. The welfare depending on endowments

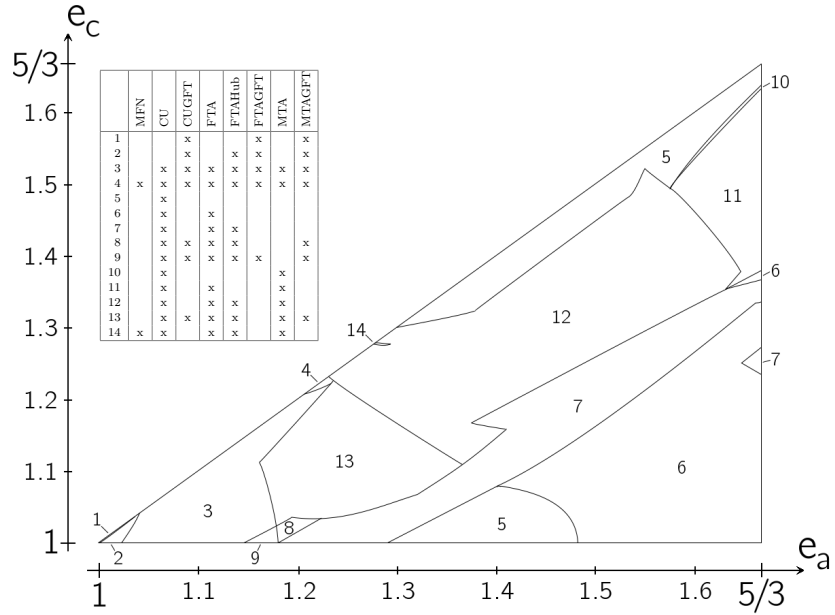
Now, recall that capping at the MFN-tariff for both members of a customs union, in this case CU( $i^c, j^c$ ), occurs when the endowment of the non-member is above a minimal value determined by the endowments of the members,  $\max\{e_i, e_j\} < \frac{11}{13}e_k$  (Appendix B.2). Using this together with the general assumptions on the relation of endowments, the following effects on welfare (components) take place, where each expression of the type ‘+ $c$ ’ for some  $c$  is positive and each ‘- $c$ ’ negative:

$\Delta(MFN, CU(i^c, j^c))$		
Country $i$	Country $j$	Country $k$
<i>TR</i>		
$+\tau_i$	$+\tau_j$	0
$+\tau_{ij}$		0
$+\tau$		
<i>CS</i>		
$+\gamma_i$	$+\gamma_j$	$-\gamma_k$
$+\gamma_{ij}$		$-\gamma_k$
0		
<i>PS</i>		
$-\rho_i$	$-\rho_j$	$+\rho_k$
$-\rho_{ij}$		$+\rho_k$
$-\rho$		
<i>W</i>		
$\pm\omega_i$	$\pm\omega_j$	$+\omega_k$
$-\omega_{ij}$		$+\omega_k$
0		

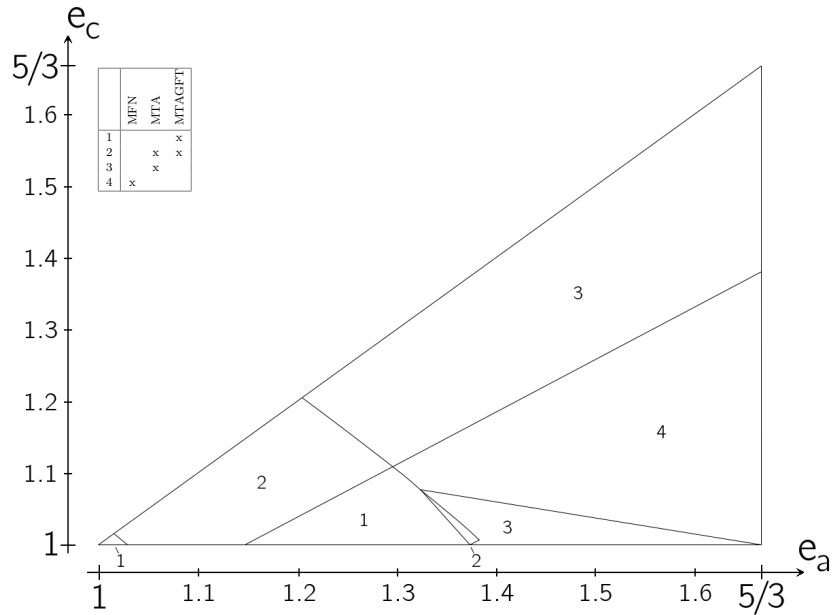
TABLE 20. The effect on the welfare (components)

APPENDIX C. ANALYSIS

C.1. **Additional Graphics.** The following provides detailed figures:



(A) Overall Stability with PTAs



(B) Overall Stability without PTAs

FIGURE 21. Overall Stability with and without PTAs

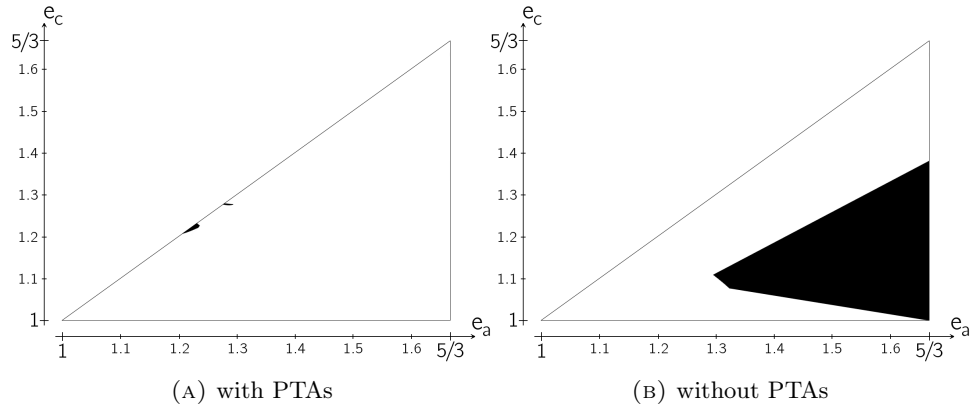


FIGURE 22. Stability of MFN

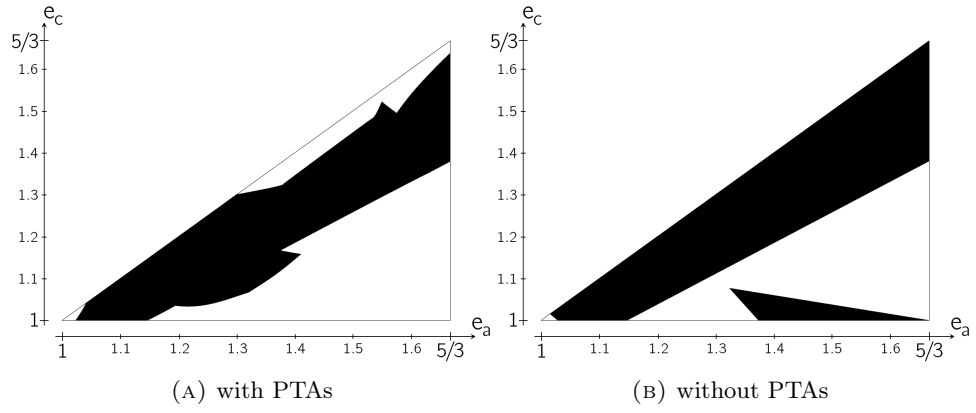


FIGURE 23. Stability of MTA

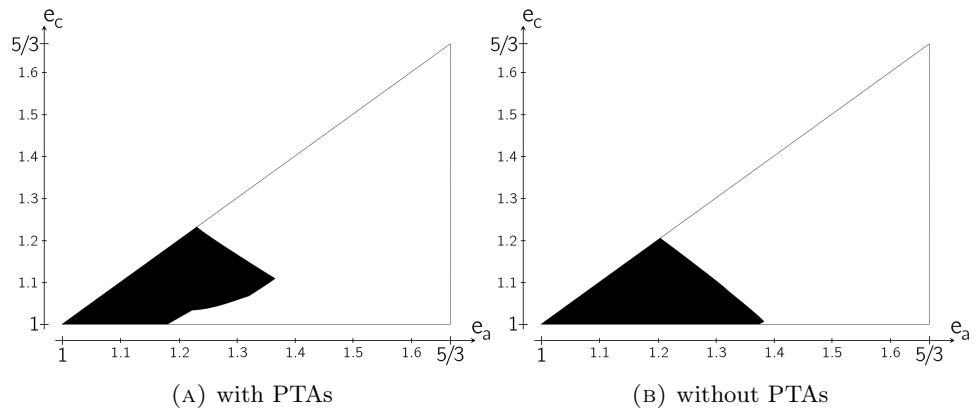


FIGURE 24. Stability of MTAGFT

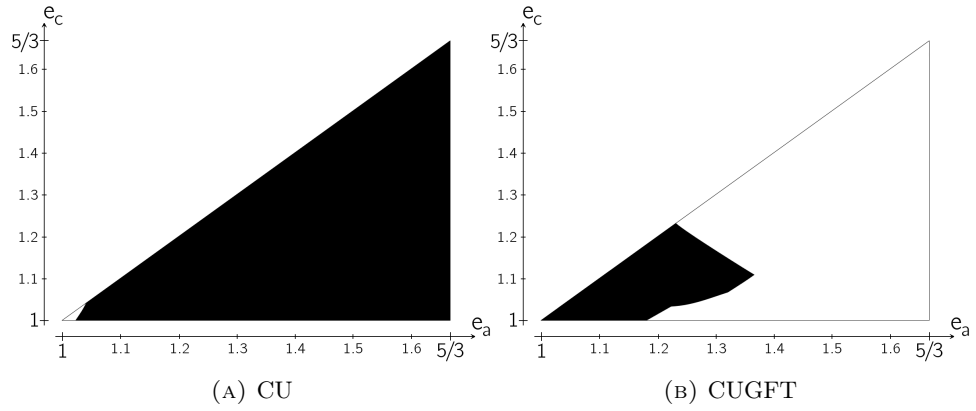


FIGURE 25. Stability of CU

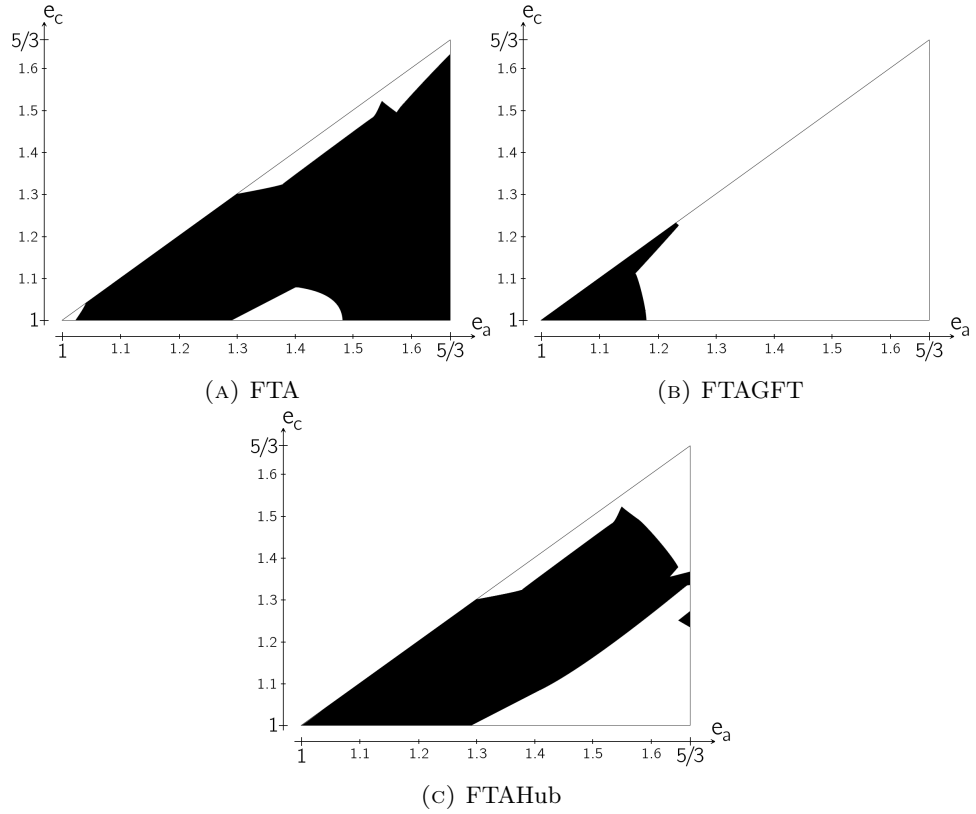


FIGURE 26. Stability of FTA

**C.2. Exact Intervals.** The table here lists the exact intervals where each specific trade agreement is part of the stable set (for the border of the parameter space):

Trade Agreement	Exact Interval(s)
$e_b = e_{\min} \leq e_c \leq e_{\max} = e_a$	
CU(b,c)	[1.0000000000000000,1.3380093520374081]
CU(c,a)	[1.3259853039412157,1.6666666666666667]
FTA(b,c)	[1.0000000000000000,1.3807615230460921]
FTA(c,a)	[1.3353373413493654,1.6305945223780896]
FTAHub(c)	[1.2364729458917836,1.2698730794923179]
	[1.3353373413493654,1.3647294589178356]
MTA(c,a)	[1.379425517702071,1.635938543754175]
$e_b = e_c = e_{\min} \leq e_a \leq e_{\max}$	
CU(a,b)	[1.0240480961923848,1.1108884435537743]
CU(b,c)	[1.0240480961923848,1.6666666666666667]
CU(c,a)	[1.0240480961923848,1.1108884435537743]
CUGFT	[1.0000000000000000,1.1803607214428857]
FTA(a,b)	[1.0240480961923848,1.2404809619238477]
	[1.0240480961923848,1.29124916499666]
FTA(b,c)	[1.483633934535738,1.6666666666666667]
FTA(c,a)	[1.0240480961923848,1.2404809619238477]
FTAHub(b)	[1.0013360053440215,1.29124916499666]
FTAHub(c)	[1.0013360053440215,1.29124916499666]
FTAGFT	[1.0000000000000000,1.1803607214428857]
MTA(a,b)	[1.0240480961923848,1.1469605878423514]
MTA(c,a)	[1.0240480961923848,1.1469605878423514]
MTAGFT	[1.0000000000000000,1.1803607214428857]
$e_b = e_{\min} \leq e_a = e_c \leq e_{\max}$	
MFN	[1.2044088176352705,1.2297929191716768]
	[1.2752171008684035,1.276553106212425]
CU(a,b)	[1.0454241816967267,1.2498329993319974]
CU(b,c)	[1.0454241816967267,1.2498329993319974]
CU(c,a)	[1.0494321977287908,1.6666666666666667]
CUGFT	[1.0000000000000000,1.2297929191716768]
FTA(a,b)	[1.0494321977287908,1.2925851703406814]
FTA(b,c)	[1.0494321977287908,1.2925851703406814]
FTA(c,a)	[1.0454241816967267,1.2925851703406814]
FTAHub(a)	[1.0454241816967267,1.276553106212425]
FTAHub(b)	[1.0454241816967267,1.2925851703406814]
FTAHub(c)	[1.0454241816967267,1.276553106212425]
FTAGFT	[1.0000000000000000,1.2297929191716768]
MTA(a,b)	[1.0494321977287908,1.2244488977955912]
MTA(b,c)	[1.0494321977287908,1.2244488977955912]
MTA(c,a)	[1.0454241816967267,1.2925851703406814]
MTAGFT	[1.0000000000000000,1.2297929191716768]

TABLE 27. The exact intervals of stability with PTAs



Trade Agreement	Exact Interval(s)
$e_b = e_{\min} \leq e_c \leq e_{\max} = e_a$	
MFN	[1.0000000000000000,1.3780895123580494]
MTA(b,c)	[1.0000000000000000,1.0000000000000000]
MTA(c,a)	[1.379425517702071,1.6666666666666667]
$e_b = e_c = e_{\min} \leq e_a \leq e_{\max}$	
MFN	[1.6666666666666667,1.6666666666666667]
MTA(a,b)	[1.0307281229124916,1.1469605878423514]
MTA(b,c)	[1.0307281229124916,1.1469605878423514]
	[1.3754175016700068,1.6666666666666667]
MTA(c,a)	[1.0307281229124916,1.1469605878423514]
MTAGFT	[1.0000000000000000,1.3740814963259853]
$e_b = e_{\min} \leq e_a = e_c \leq e_{\max}$	
MTA(a,b)	[1.0160320641282565,1.1202404809619237]
MTA(b,c)	[1.0160320641282565,1.1202404809619237]
MTA(c,a)	[1.0160320641282565,1.6666666666666667]
MTAGFT	[1.0000000000000000,1.203072812291249]

TABLE 28. The exact intervals of stability without PTAs