

Berens, Stefan; Chochua, Lasha; Willmann, Gerald

**Working Paper**

## The farsighted stability of global trade policy arrangements

Kiel Working Paper, No. 2193

**Provided in Cooperation with:**

Kiel Institute for the World Economy – Leibniz Center for Research on Global Economic Challenges

*Suggested Citation:* Berens, Stefan; Chochua, Lasha; Willmann, Gerald (2021) : The farsighted stability of global trade policy arrangements, Kiel Working Paper, No. 2193, Kiel Institute for the World Economy (IfW), Kiel

This Version is available at:

<https://hdl.handle.net/10419/246589>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# KIEL WORKING PAPER

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



No. 2193 July 2021

*Stefan Berens, Lasha Chochua, and Gerald Willmann*

# ABSTRACT

## THE FARSIGHTED STABILITY OF GLOBAL TRADE POLICY ARRANGEMENTS\*

*Stefan Berens, Lasha Chochua, and Gerald Willmann*

We study the stability of trade policy arrangements under two regulatory scenarios, with/without PTAs. Unlike previous papers, we consider an extensive set of trade policy constellations and allow for unlimited farsightedness of negotiating parties. We find global free trade (GFT) to be uniquely stable under symmetry, with/without PTAs. When two countries are smaller(larger) than the third, availability of PTAs decreases (increases) the stability of GFT. Away from symmetry, GFT is not attainable, and without PTAs the non-cooperative MFN regime is the only stable outcome. The effect of (dis-)allowing PTAs thus depends on the size asymmetry of countries.

**Keywords:** Trade Policy Arrangements, Stability, Unlimited Farsightedness

**JEL classification:** F13; F55

### **Stefan Berens<sup>†</sup>**

Bielefeld University  
Universitätsstraße 25  
33615 Bielefeld, Germany

**Email:**

[sbehrens@math.uni-bielefeld.de](mailto:sbehrens@math.uni-bielefeld.de)  
[www.uni-bielefeld.de](http://www.uni-bielefeld.de)

### **Lasha Chochua<sup>†</sup>**

Bielefeld University  
Universitätsstraße 25  
33615 Bielefeld, Germany

**Email:**

[lasha.chochua@uni-bielefeld.de](mailto:lasha.chochua@uni-bielefeld.de)  
[www.uni-heidelberg.de](http://www.uni-heidelberg.de)

### **Gerald Willmann<sup>‡</sup>**

Bielefeld University and  
Kiel Institute for the World Economy  
Kiellinie 66  
D-24105 Kiel, Germany

**Email:**

[gerald.willmann@ifw-kiel.de](mailto:gerald.willmann@ifw-kiel.de)  
[www.ifw-kiel.de](http://www.ifw-kiel.de)

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

## 1. INTRODUCTION

Following the General Agreement on Tariffs and Trade (GATT) of 1947, an ever increasing number of signatory countries liberalised 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. Starting in force in the 1980ies, the world has seen an ever-increasing number of Preferential Trade Agreements (PTAs) concluded by bilateral negotiations. Currently, about forty percent of all countries/territories are a member of more than five PTAs, and about a quarter participates in more than ten.<sup>1</sup>

The World Trade Organization (WTO), successor of the GATT since 1995, provides the set of rules for the trade liberalisation process of a significant number of countries.<sup>2</sup> Its Article I acts as the foundation for any multilateral trade liberalisation 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) in the spirit of said principle.<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 has been the focus of many contributions.<sup>6</sup> In this tradition, the primary purpose of this paper is the analysis of the stability of different trade policy arrangements in two scenarios: one with PTAs in line with the current WTO rules, and an alternative scenario without PTAs (modified WTO rules). In addition, we 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 set of trade agreements or assumes limited farsightedness of the negotiating countries. While this allows for a cleaner description of the model and interpretation of its results, it raises the question whether these restrictions significantly influence the analysis, and to what

---

<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>).

<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>We interchangeably use the terms trade policy arrangements, trade agreements, trade policy constellations.

<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 the paper provides further information on the related literature.

extent such frameworks capture reality. In our view, certain empirical observations favour an extensive choice set of trade policy regimes and full farsightedness. In fact, over past rounds of multilateral trade negotiations, many countries were simultaneously involved in several trade liberalisation processes.<sup>7</sup> Moreover, such trade negotiations are usually complicated processes with significant impact on the countries' economies, and thus decisionmakers usually commission studies to quantify the effects, and try to anticipate strategic actions by the trade partners involved.<sup>8</sup> Extending the analysis to a broad set of trade policy arrangements, and taking into account the farsightedness of decision-makers is one of the contributions of this paper.

In our analysis, we thus consider an extensive set of trade agreements, containing PTAs — both 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' as stable outcomes — a concept developed by Chwe (1994). As a result, our work expands the set of trade agreements under consideration and also extends the farsightedness of the negotiating parties in comparison to the existing literature. In fact, to the best of our knowledge, no other paper considers a choice set as extensive as ours.

Our analysis shows that the effect of PTAs on trade liberalisation depends on the size distribution of the countries in terms of endowments.<sup>10</sup> As long as the countries are almost equal in size, Global Free Trade (GFT) emerges as the unique stable outcome under both the existing and the alternative institutional arrangement (with vs. without PTAs). However, when two countries are considerably smaller than the third, a modified WTO arrangement that does not allow PTAs would facilitate the stability of GFT. This results from the suppression of the exclusion incentive (when PTAs are available) of the small countries versus the large. By contrast, if two countries are relatively larger than the third, a modified WTO without PTAs would actually reduce the extent to which GFT is stable. The argument here is the free-riding incentive of the small country when PTAs are not available, and we show that this is due to the non-availability of CUs. When the world is further away from symmetry, full trade liberalisation is not attainable at all, and abolishing the exception for PTAs can result in the worst possible outcome from the perspective of overall world welfare, the non-cooperative MFN regime.

<sup>7</sup>Maggi (2014) showcases the importance of an extensive set of trade policy 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 Maggi and Yildiz (2010), which itself is a modification of the one in Bagwell and Staiger (1997). The modified one is also used in Maggi, Woodland and Yildiz (2013).

<sup>10</sup>We use the endowments as a measure of size, as these are exogeneously given.

Our work is related to the so-called ‘rules-to-make-rules’ literature, as Maggi (2014) has called it, that tries to determine the role of PTAs in the global trade liberalisation process. A number of relevant studies are the work of Kamal Saggi and Halis Yildiz with various co-authors. Saggi and Yildiz (2010) consider a three-country trade model where the degree and nature of trade liberalisation, 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) analyses the effect of both forms of PTAs, i.e. CUs and FTAs, on attaining global free trade, but does not consider MTAs. In a sense, this research has studied ‘2 out of 3’ permutations of trade agreements.

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. In his approach, he uses a three-country dynamic model where a 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 analyses the effect of country asymmetries on global trade liberalisation.

Furthermore, several papers focus primarily on analysing the effect of FTAs. Goyal and Joshi (2006) consider several countries, each producing one good, and study different degrees of asymmetry across countries. They employ the notion of ‘Pairwise Stability’ proposed by Jackson and Wolinsky (1996) as the solution method. Furusawa and Konishi (2007) use similar methods but introduce a continuum of 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) who replace the concept of Pairwise Stability with Pairwise Farsighted Stability by Herings, Mauleon and Vannetelbosch (2009), thereby comparing myopia with farsightedness in an otherwise fixed framework. Connected to this is the paper of Zhang et al. (2014), which uses the work of Goyal and Joshi (2006) as a benchmark and analyses the evolutionary effect of the number of countries in a dynamic framework featuring random perturbations. While all 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 by them, 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.

We use the stability concept put forward by 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

---

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

include unlimited consideration of the future by the participants while simultaneously avoiding emptiness of the stable set that plagues other (more) restrictive solution concepts.<sup>12</sup> It is closely related to the stability concept proposed in Herings, Mauleon and Vannetelbosch (2009) and its extension (HMY (2014)). As noted by the authors, their criterion constitutes a stricter version, which in specific cases (that include our model) coincides.

The remainder of the paper is organised as follows. Section 2 specifies the model. Section 3 analyses the findings, and Section 4 offers concluding remarks.

## 2. MODEL

In this section, we start by presenting the trade policy arrangements we consider, and introduce the transition graphs that show how countries can move from one policy regime to another, individually or in coalitions. Subsequently, we provide the necessary theoretical and intuitive details of the stability notion we use, including the solution concept. Finally, we present the underlying trade model that determines countries' preferences over trade policy regimes and parametrise the model.

**2.1. Trade Policy Arrangements.** As is standard in the literature, we consider a three-country world. Let  $N = \{a, b, c\}$  denote the set of three countries. These countries trade with one another, and each country can (possibly) impose tariffs on imported goods following the rules of the GATT. Let  $t_{ij}$  denote the tariff (vector) imposed by country  $i$  on imports from country  $j$ , where  $i, j \in N$  are distinct. The decision-makers in each country maximize national welfare. We denote by  $W_i$  the welfare of country  $i \in N$ . This basic setup allows us to study all possible trade policy constellations under articles I and XXIV of the GATT. In subsection 2.4, we provide further details on the underlying trade model which then allows us to determine the tariffs chosen and the resulting welfare levels, conditional on trade policy regime, which in turn determine countries' preferences over these regimes.

All trade policy arrangements we consider are of the following four types: most favored nation (MFN), customs union (CU), free trade area (FTA), and multilateral trade agreement (MTA). Each type (except for MFN) induces combinations of insiders and outsiders. For a CU, (single) FTA, and MTA, we have three combinations each of two members, plus one comprising the full set of countries. Note that we take Global Free Trade to come in three different forms: a CU, FTA, or MTA of three members.<sup>13</sup> In addition, there is the case of two over-lapping FTAs in a hub-and-spoke structure — adding another three regimes. In total,

<sup>12</sup>It is also resistant to the criticism of Ray and Vohra (2015) about the sovereignty of coalitions as their main issues concerned with feasibility and distribution do not apply to our framework. Furthermore, their specific criticism about the explanatory power of Chwe's approach only applies to transferable utility games.

<sup>13</sup>While their welfare level is equal, their position in the network (cf. Section 2.2) need not be, and for our concept of stability it is important which group of countries can create or destroy specific trade agreements (see Appendix C.2). Where applicable, we group all three together and refer to them as 'GFT'.

we thus have 16 different global trade policy regimes and define the set of global trade policy constellations as  $X \equiv \{MFN, CU(i, j), CU(i, k), CU(j, k), CUGFT, FTA(i, j), FTA(i, k), FTA(j, k), FTAHub(i), FTAHub(j), FTAHub(k), FTAGFT, MTA(i, j), MTA(i, k), MTA(j, k), MTAGFT\}$ .

Under each of these trade agreements, tariffs are bounded by zero from below, and from above by the MFN-tariff, which we discuss in more detail in Appendix B.2. The national policymaker in country  $i$  thus chooses  $t_{ij} \in T_i \equiv [0, t_i^{MFN}] \forall j$ .<sup>14</sup> Conditional on trade agreement, the choice of tariffs is restricted further, as we discuss now.

In the baseline case, which we denote by MFN, countries do not liberalise their trade relations, but satisfy the non-discrimination principle. Each country  $i$  unilaterally chooses its (optimal) tariffs, solving  $\max_{(t_{ij}, t_{ik}) \in T_i^{MFN}} W_i$  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  $T_i$ .

When countries  $i$  and  $j$  form a customs union  $CU(i, j)$  together, each of them removes any trade restriction on the partner country, and they jointly impose an optimal tariff on country  $k$ . Their optimization problem amounts to  $\max_{(t_{ij}, t_{ik}) \in T_i^{CU}, (t_{ji}, t_{jk}) \in T_j^{CU}} W_i + W_j$  with  $T_i^{CU} = \{(t_{ij}, t_{ik}) \in T_i^2 \mid t_{ij} = 0\}$  and analogously for  $T_j^{CU}$ . Country  $k$ , finally, simply follows and applies the MFN principle (as before). Once all three countries enter a single CU together, the (common) optimisation problem is trivial because internal tariffs are restricted to zero. We denote this scenario by CUGFT.

In case countries  $i$  and  $j$  form an  $FTA(i, j)$ , each of them again removes any trade restriction on the other country, and unilaterally imposes an optimal tariff on the outsider country  $k$ . The (representative) optimisation problem of country  $i$  (and  $j$ ) 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 optimisation problem of country  $k$  is identical to that of the third country in case of a CU. Furthermore, if country  $i$  forms an FTA each with countries  $j$  and  $k$ , we call this  $FTAHub(i)$ , then both tariffs of country  $i$  are set to zero by nature of its trade relation with both partner countries. Each of the other two countries operates as before: Country  $j$  ( $k$  analogous) solves  $\max_{(t_{ji}, t_{jk}) \in T_j^{FTA}} W_j$  where  $T_j^{FTA} = \{(t_{ji}, t_{jk}) \in T_j^2 \mid t_{ji} = 0\}$ . Note that in terms of decision problem, it does not matter for a country whether its partner also forms a trade agreement with the other country. Finally, if all three countries form an FTA, denoted by FTAGFT, the optimization problem is identical to the case of CUGFT, but differs in terms of structure and network position (cf. Section 2.2).

We turn finally to the case where countries  $i$  and  $j$  form a multilateral trade agreement  $MTA(i, j)$ . In this case, they jointly change their tariffs vis-à-vis each other, and also for the third country. The corresponding optimization problem 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 analogously for  $T_j$ . As before, the optimization problem of country  $k$  is identical to that of the outsider

<sup>14</sup>Note that the upper bound is uniform across trading partners by the very nature of MFN.



country in the case of a CU. When all three countries enter a single MTA together, which we denote by MTAGFT, the optimization problem is identical to the case of CUGFT and FTAGFT, but the regime differs in terms of network position (cf. Section 2.2).

**2.2. Transition Graphs.** We now consider transition possibilities from one global trade policy constellation to another. Countries can implement trade policy arrangements, or leave an existing arrangement, depending on whether they act individually or in a coalition with other countries. We represent these possibilities by means of directed graphs, with trade agreements as vertices and the transition between them as directed edges.

To start with, consider the case of a single country  $i \in N$  (this being a coalition of one), with  $j, k \in N \setminus \{i\}$ ,  $j \neq k$  denoting the other two countries. The digraph for this case is provided in Figure 1, leaving aside loops at every vertex for illustrative clarity. As we can see from the diagram, the baseline MFN regime is reachable from a number of different global trade policy regimes, which country  $i$  can decide to quit, but not from the trinity of GFT, i.e. CUGFT, FTAGFT and MTAGFT, because quitting those would leave a two-country arrangement between  $j$  and  $k$  in place (or possibly a FTA hub-structure, if  $i$  were to quit only one FTA). The three variants of GFT thus form separate groups of connected trade agreements, and hence the overall transition graph (in Figure 1, abstracting from loops), consists of four sub-graphs.

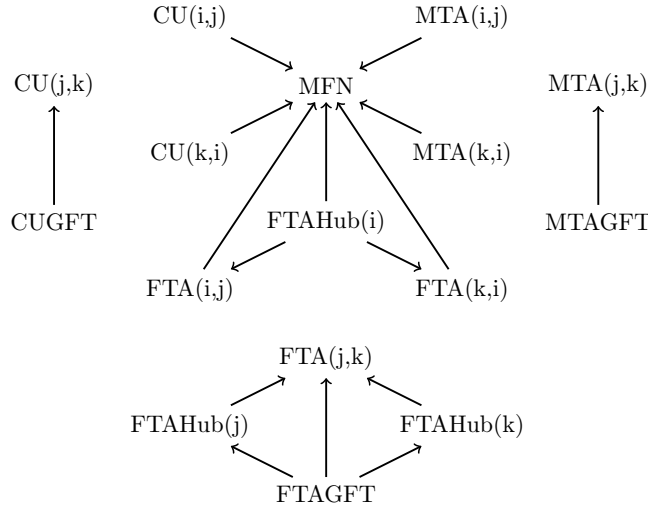


FIGURE 1. The transition graph for coalition  $\{i\}$ ,  $i \in N$ .

Next, consider the transition graph for a coalition of two countries  $i, j \in N$ , where  $i \neq j$  and  $k \in N \setminus \{i, j\}$  denotes the other country. The transition graph for this case is provided in Figure 2, again leaving aside loops. As we see here, the four regimes MFN, CU(i,j), FTA(i,j), and MTA(i,j) are all interconnected. This is because the coalition of  $i$  and  $j$  can move from and to any two-country agreement between themselves and MFN at will. In addition, any regime connected to one of the four is automatically connected to all four of them. This

group of four represents a complete directed sub-graph, which we illustrate as a dotted (super) node in the figure. In contrast to the previous case (compare Figure 1), the coalition of  $i$  and  $j$  can move from GFT to MFN or any two-country arrangement between themselves (the super node). They can also transition from any form of GFT to a corresponding two-country arrangement between one of them and the outside country, if only one of them decides to quit, plus from FTAGFT to any FTA hub-and-spoke structure.

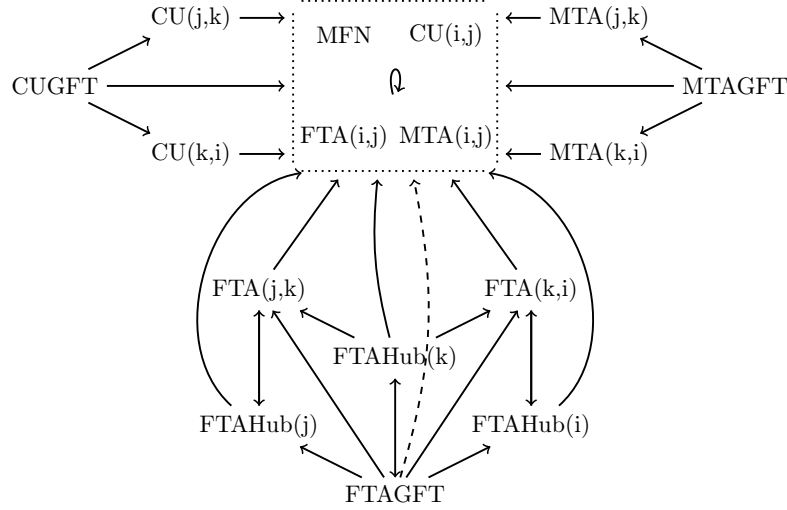


FIGURE 2. The transition graph for coalition  $\{i, j\}$ ,  $i, j \in N$ ,  $i \neq j$ .

Finally, consider the coalition of (all) three countries. Together, these three can implement any trade policy constellation. Therefore, the corresponding directed graph for the case of the grand coalition is a complete directed graph with loops (for short, a complete loop-digraph).

**2.3. Stability Concept.** In terms of stability concept, we use the approach of Chwe (1994), as he combines far-sightedness with deviations by a wider, more realistic set of coalitions.<sup>15</sup>

Consider the tuple  $\Gamma = (N, X, \{\prec_i\}_{i \in N}, \{\rightarrow_S\}_{S \subseteq N, S \neq \emptyset})$  that lists the set of countries or players, the set of global trade policy constellations, the preferences of players over these regimes, and the effectiveness relation of possible transition paths. To understand the meaning of the latter, let  $x \in X$  be the status quo trade agreement at the start. Next, each coalition  $S \subseteq N$ ,  $S \neq \emptyset$  (including one-country coalitions) is able to make  $y \in X$  the new status quo as long as  $x \rightarrow_S y$ . Continue with  $y$  as the new status quo. If a status quo  $z \in X$  is reached without any coalition moving away, then the state is actually realised and each country receives their corresponding welfare.<sup>16</sup> By consequence, any coalition only favors following through on their ability to move from  $x \rightarrow_S y$ , if it prefers 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:

<sup>15</sup>Consult the paper of Chwe (1994) for proofs of the propositions that are presented here.

<sup>16</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.

**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 direct dominance implies indirect dominance, i.e. if  $x_1 < x_2$  for some  $x_1, x_2 \in X$ , then automatically  $x_1 \ll x_2$ .

Using this definition, we introduce the concept of the ‘consistent set’, a (sub-)set that exhibits internal stability in the sense of a lack of incentives 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 need not be unique, but the following proposition allows us to focus on the unique ‘largest consistent set’, i.e. the (consistent) set that contains all consistent sets:

**Proposition 1.** *There exists a unique  $Y \subseteq X$  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.*

Similar to the internal stability captured by the definition of consistent sets, external stability is captured by 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 characterises one setting of relevance 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.*

Importantly, this result applies to our setting, so the (unique) LCS is non-empty and satisfies external (as well as internal) stability.<sup>17</sup>

<sup>17</sup>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.

The largest consistent set is going to be the focus of our analysis. Any trade agreement is considered to be ‘(potentially) stable’ if it is in the LCS, and ‘unstable’ otherwise. Note that 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)).

**2.4. Underlying Trade Model.** To study the stability of different global trade policy constellations, we use a three-country, general equilibrium trade model. We follow the literature in using the model employed by Saggi and Yildiz (2010), which features competition in exports. We would like to emphasize that our approach allows for the use any type of underlying trade model. Our choice is guided solely by the widespread use of the competing exporters model in the literature, so that the use of this model facilitates the comparison of results.

Recall that  $N = \{a, b, c\}$  denotes the set of countries. We denote by  $G = \{A, B, C\}$  three non-numéraire goods, so that each country  $i$  is endowed with zero units of good  $I$  (the corresponding capital letter), and  $e_i$  units of the other two goods. Country  $i$  will end up importing good  $I$  and exporting goods  $J$  and  $K$ , with  $J, K \neq I$ . Each good will thus be exported by two countries, e.g. good  $I$  will be exported by countries  $j$  and  $k$ , which has given the model its name, ‘competing exporters model’. In order to guarantee the ‘competing exporters’-structure, we need to impose a condition on the degree of asymmetry in terms of the respective endowments:

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

This condition is sufficient to ensure that exports remain non-negative.

On the demand side, let preferences of individuals in each country be identical. The demand for any non-numéraire good  $L \in G$  in country  $i \in N$  is given by  $d(p_i^L) = \alpha - p_i^L$  with  $p_i^L$  the price of good  $L$  in country  $i$  and  $\alpha$  the (universal) reservation price.<sup>18</sup> As pointed out before, each country can impose tariffs on imports. Let  $t_{ij}$  denote the tariff imposed by country  $i$  on the import of good  $I$  from country  $j$ . The prices and tariffs of good  $I \in G$  across countries are linked 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 the model at hand, prices together with the corresponding endowments are the only determinants of imports and exports. In particular, the level of imports  $m_i^I$  of good  $I$  into country  $i$  is completely determined

---

<sup>18</sup>The demand function can be derived from a utility function that is additively separable and quadratic in each non-numéraire good.

by the above demand function (depending on the price):  $m_i^I = d(p_i^I) = \alpha - p_i^I$ . In turn, the exports  $x_j^I$  of good  $I$  from country  $j$  are the combination of the demand function (or prices) and the corresponding endowment,  $x_j^I = e_j - d(p_j^I) = e_j + p_j^I - \alpha$ . Market-clearing for any good  $I$  requires that country  $i$ 's imports equal the exports of countries  $j$  and  $k$  (again  $i, j, k \in N$  pairwise distinct) combined:

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

Ultimately, the objective function of the benevolent policymaker in country  $i$  is its national welfare<sup>19</sup>, denoted  $W_i$ , which includes consumer surplus (CS), producer surplus (PS), and tariff revenue (TR):

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

Using the no-arbitrage (1) and market-clearing (2) conditions, we can compute the equilibrium prices:

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

Given equilibrium prices, we can deduce imports, exports, and the welfare of each country up to the value of the tariffs (Appendix B.1). The maximisation of welfare with respect to tariffs will then be constrained by the trade agreement under consideration (see Section 2.1). The full equilibrium of our model is then computed as follows: Fix a trade agreement and thereby the restrictions on tariffs. Compute the best-response functions for each country (with respect to the tariffs) and determine the optimal choices. While Section 2.1 contains all information on the trade agreements that is necessary to compute the equilibria, the actual results are presented in Appendix B.2. An overview of the (resulting) overall welfare is provided in Appendix B.3.

**2.5. Algorithm and Parameters.** The contribution in this paper of considering an extensive set of trade agreements and unlimited farsightedness comes at the cost of a relatively complex computational problem. We solve this problem numerically by means of an algorithm, the pseudocode of which we provide in Appendix A.<sup>20</sup> To solve the model numerically by means of the algorithm, we need to specify and discretize the parameter space.

Recall that the endowments have to 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 (cf. Section 2.4). Without loss of generality, let us normalize one endowment to one, namely  $e_b = 1$ . Consequently, for  $i, j \in N \setminus \{b\}$ :  $e_{\min} \equiv \frac{3}{5} \leq \frac{3}{5} \max\{1, e_j\} \leq e_i$  and  $e_i \leq \frac{5}{3} \min\{1, e_j\} \leq \frac{5}{3} \equiv e_{\max}$ . The resulting parameter space is depicted by the hexagon in Figure 3. Given that the three countries are interchangeable, we can split the hexagon into six right-angled triangles, which

<sup>19</sup>Depending on the type of trade agreement entered, the (joint) objective function may include the welfare of other countries as well. See Section 2.1 for the details.

<sup>20</sup>The authors are grateful to Michael Chwe for the provision of an exemplary algorithm.

are mirror images of one another (in terms of relative endowments). Without loss of generality, we focus on one of them, namely the triangle depicted as shaded in the figure. We cover this triangle of the parameter space with a fine grid for the actual computation.<sup>21</sup>

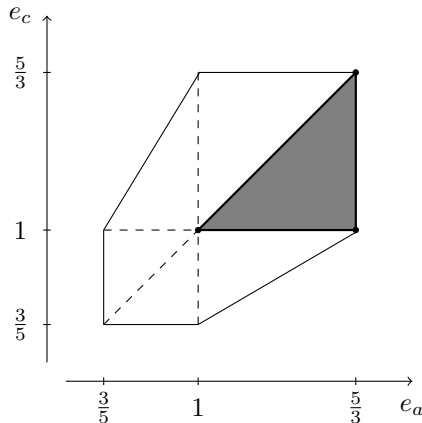


FIGURE 3. The parameter space of the endowments with  $e_b = 1$

In addition to endowments, we also need to specify the choke price *alpha* in the demand function. To obtain plausible results — that is, positive prices — the parameter  $\alpha$  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, results remain unchanged.<sup>22</sup> 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 combinations ensures that the results are plausible and comparable.<sup>23</sup>

### 3. ANALYSIS

We now turn to the analysis of stable outcomes among the global trade policy regimes that we consider. That is, for every endowment vector, we determine which regimes make up the largest consistent set. Figure 4 depicts the parameter space of endowments under consideration — it is the (marked) triangle from before. Note that different endowment vectors represent varying size combinations of countries.

For expositional purposes, we start by studying the three corners of the triangle, then turn to the connecting intervals or edges, and finally consider the entire triangle by studying the full interior. Clearly, the corners are the extrema of the edges, and the edges in turn the bounds of the interior. So the second and third step contain the preceding results as special cases, and we can think of each subsequent step as a linear combination of the endowment combinations considered previously.

<sup>21</sup>The distance is set to 0.0013360053440215, resulting in 500 points per dimension.

<sup>22</sup>The parameter  $\alpha$  enters the welfare of country  $i$  as  $2\alpha e_i$  (see Appendix B.1). Any changes above the minimal value leave the relative welfare levels and therefore the rankings unaffected.

<sup>23</sup>In our computation  $\epsilon$  is fixed at 0.01, which yields  $\alpha = 1.3988888888888888$ .

For all allowable endowment combinations, we will analyse two scenarios: The first scenario corresponds to the current institutional setup of the WTO. The constitutional status-quo includes PTA regimes, as article XXIV provides for. The second scenario that we study is a hypothetical alternative constitutional arrangement without article XXIV, i.e. the choice set of global trade policy regimes does not include any form of PTA (specifically neither CUs nor FTAs).

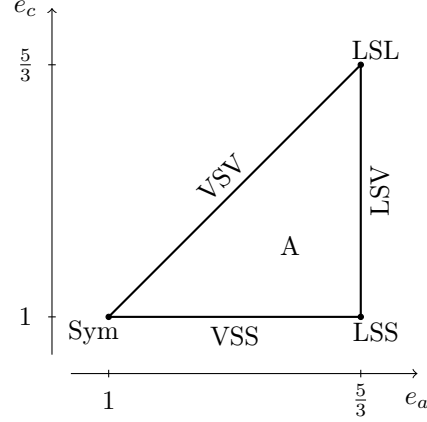


FIGURE 4. Overview of the different points, intervals and areas of interest depending on the (partially normalised) endowment tuple

**3.1. Corners of the triangle.** We start our analysis by considering the three corners of the triangle. First, thorough attention will be given to the symmetric case in order to demonstrate the working of each component of the model. Subsequently, for the other two corners, we focus on results and provide a general overview.<sup>24</sup>

In the symmetric corner, denoted by *Sym* in Figure 4, all three countries have identical endowments:  $e_a = e_b = e_c = 1 = e_{\min}$ . As the countries do not differ from one another in this case, 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 the preferences of the other countries as well). For fixed  $i, j \in N \setminus \{a\}$  with  $i \neq j$ , we obtain the following preference ranking:

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

The case where the two other countries *i* and *j* form a customs union is the least favourable trade constellation for country *a*. As an outsider to the  $CU(i, j)$ , country *a* faces the second-highest tariffs (with MFN-tariffs the highest), while the CU members *i* and *j* abolish the tariffs between 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.

<sup>24</sup>We provide additional proofs for two, non-symmetric corners in Appendix C.1.

The same applies to total imports. In other words, the ‘trade diversion’ effect is strongest for country  $a$  when we have  $CU(i,j)$ .

In general, the MFN regime favours country  $a$  as compared to  $CU(i,j)$ . The tariff revenue remains the same, while the consumer surplus is lower and the producer surplus higher — the increase in the latter offsetting the decrease of the former. The MFN regime mitigates the ‘trade diversion’ effect present in the case of  $CU(i,j)$  due to increased export values of country  $a$ .

Within the set of bilateral trade agreements where country  $a$  is an insider, the MTAs result in the lowest welfare (for country  $a$ ).  $MTA(a,i)$  generates a higher welfare for country  $a$  in comparison to the MFN regime due to increased consumer and producer surplus. The  $FTAHub(i)$  constellations result in further gains in welfare for country  $a$  through higher export values and producer surplus, even though tariff revenue and consumer surplus are lower under  $FTAHub(i)$  compared to  $MTA(a,i)$ .

However, country  $a$  does not have an incentive to remain in this constellation. The unilateral deviation from  $FTAHub(i)$  to  $FTA(i,j)$  — that is, if  $a$  quits its FTA with the hub country — comes with a decrease in consumer and producer surplus, but enough of an increase in tariff revenue to ultimately ensure higher welfare under  $FTA(i,j)$ . Yet among (single) FTAs, being an outsider is less desirable than being an insider for any of the three countries, as the drop in tariff revenue is offset by an expansion of consumer and producer surplus, resulting in higher welfare for country  $a$  in the case of  $FTA(a,i)$  compared to  $FTA(i,j)$ . As an insider, country  $a$  prefers  $CU(a,i)$  over  $FTA(a,i)$ , though. Despite the decline in the consumer surplus, welfare goes up due to 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 their fellow member and to the outsider. Country  $a$  free-rides on this characteristic of the MTA regime, and in addition attains the highest possible tariff revenue in this case.

Each country obtains the second-highest welfare level when the world reaches global free trade (that is, the trinity of MTA-, FTA-, and CU-GFT). Under full trade liberalisation, the producer surplus is at its second-highest among all trade agreements (effectively driving the ranking). It is only surpassed by that of  $FTAHub(a)$ . This constellation brings about the highest possible welfare for country  $a$ . Notice, however, that this hub-n-spoke trade agreement disproportionately favours the hub country over the other two.

The countries’ preference rankings, which we have just derived for one particular endowment combination, provide the basis for computing the largest consistent set (LCS). In the case at hand, the trinity of global free trade regimes are ranked second-best by each country, while the respective first-best options, hub-n-spoke FTA structures, are ranked considerably lower by spoke countries. The trinity of global free trade thus seems the likely equilibrium outcome. The following proposition and its proof confirm 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 (cf. Sections 2.3 and 2.2), the preference rankings from above 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 (respectively is not) 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>25</sup>

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

<sup>25</sup>Appendix A contains the pseudocode for this procedure.

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.  $Y = LCS$ .  $\square$

The proposition informs us that in the symmetric case, under the current institutional arrangement of the WTO, (the trinity of) global free trade is the only stable trade policy regime according to our framework.

Yet what would happen in the absence of article XXIV? We now consider an alternative constitutional setup. In that case, countries would not have the option to liberalise trade through the formation of CUs or FTAs, leaving MTAs as the only possibility, apart from global free trade and MFN. The preference ranking of the representative country  $a$  over the reduced set of policy regimes then looks as follows:

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

Each country achieves its maximal welfare under MTAGFT, i.e. global free trade. Note that FTAGFT and CUGFT are no longer available, nor is FTAHub, the regime that dominated the welfare ranking for the hub country in the presence of article XXIV. It is thus reasonable for us to conjecture the stability of MTAGFT, and the following proposition proves this conjecture:

**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:

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

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.  $\square$

We see that if countries' endowments are symmetric, corresponding to the lower left corner of the triangle, then the presence of article XXIV does not change the stable outcome: global free trade (or the trinity thereof, if FTAs and CUs are permissible).

We now proceed to the analysis of the other two extreme cases, which correspond to points LSS and LSL of the triangle, cf. Figure 4. The acronyms stand for country *a* being large, country *b* small (always), and country *c* either small or large.

We consider first the case of two small and one large country, i.e. point LSS. In this scenario,  $e_a = e_{max}$  and  $e_b = e_c = e_{min}$ . We obtain the following ranking of preferences for country *a* and for the small countries  $i \in \{b, c\}$  and  $j \in N \setminus \{a, i\}$ :

$$\begin{aligned}
 \text{country a:} \quad & CU(i, j), FTA(i, j) \prec_a FTAHub(i) \prec_a MFN, MTA(i, j) \prec_a FTA(a, i) \\
 & \prec_a MTA(a, i) \prec_a CU(a, i) \prec_a GFT, FTAHub(a) \\
 \text{countries b/c:} \quad & MTA(a, i) \prec_i GFT, FTAHub(a) \prec_i CU(a, i) \prec_i FTA(a, i) \prec_i CU(a, j) \\
 & \prec_i MFN, MTA(i, j) \prec_i FTAHub(i) \prec_i FTA(a, j) \prec_i MTA(a, j) \\
 & \prec_i FTAHub(j) \prec_i CU(i, j), FTA(i, j)
 \end{aligned}$$

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>26</sup> For example,  $CU(i, j)$  and  $FTA(i, j)$  result in the 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 paragraph 5 rule this out, and therefore we cap the tariffs at their  $MFN$ -level. A similar argument applies to the case of  $FTAHub(a)$ . Here, the optimal tariffs of the small countries would be negative. Restricting tariffs to be non-negative implies that  $FTAHub(a)$  corresponds to  $GFT$ , or rather a pseudo- $GFT$  (as external tariffs between spoke countries are zero). Finally, the  $MTA$  between the small countries actually coincides with the  $MFN$  regime because of identical optimal tariffs for both cases.

Using the above preference rankings to derive the largest consistent set yields the following proposition:

<sup>26</sup>Additional details on this can be found in Appendix B.2.

**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  $CU(b,c)$  and  $FTA(b,c)$  between the two small countries.*

Even though global free trade is the most desirable regime for the large country, the two small countries do not have an incentive to agree to such a constellation, and the large country cannot enforce it. By consequence, country  $a$  ends up with the worst arrangement (from its perspective). So in this scenario, the size advantage of the large country does not translate into a favourable outcome. Moreover, the case at hand demonstrates the relevance of restrictions on PTAs (remember that insiders are not allowed to raise tariffs on outsiders). The constraint renders the small countries indifferent between the two forms of PTAs.

We now turn to the hypothetical scenario without article XXIV paragraph 5. The preference rankings for the large country  $a$  and the small countries  $b$  and  $c$  (indexed by  $i$  and  $j$ ) are as follows:

$$\begin{array}{ll} \text{country a:} & MTA(i,j), MFN \prec_a MTA(a,i) \prec_a GFT \\ \text{countries b/c:} & MTA(a,i) \prec_i MTA(a,j) \prec_i MFN, MTA(i,j) \prec_i MTA(a,j) \end{array}$$

As a result, the best outcome for a small country  $i$  is the  $MTA(a,j)$  regime, i.e. an MTA between the large country and the other small country, as the PTAs are not available anymore. The next proposition presents the LCS for the reduced set of regimes:

**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 without article XXIV, the stable constellations are MFN and  $MTA(b,c)$ .*

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

We now turn to the third corner of the triangle, denoted by LSL. In this point,  $e_b = e_{min}$  and  $e_a = e_c = e_{max}$ , i.e. we have one small and two large countries. Let us start with the preference rankings for the small country  $b$  and the large countries  $i \in \{a, c\}$  and  $j \in N \setminus \{b, i\}$ :

$$\begin{array}{ll} \text{country b:} & GFT \prec_b CU(i,b) \prec_b FTAHub(i) \prec_b FTA(i,b) \prec_b FTAHub(b) \\ & \prec_b MTA(i,b) \prec_b MFN \prec_b CU(i,j) \prec_b MTA(i,j) \prec_b FTA(i,j) \\ \text{countries a/c:} & CU(j,b) \prec_i FTA(j,b) \prec_i MFN \prec_i MTA(i,b), MTA(j,b) \prec_i FTA(i,j) \\ & \prec_i MTA(i,j) \prec_i CU(i,j) \prec_i FTAHub(b) \prec_i FTAHub(j) \prec_i GFT \\ & \prec_i FTAHub(i) \prec_i FTA(i,b) \prec_i CU(i,b) \end{array}$$

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

We determine the LCS under these preference rankings in the following 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  $CU(c,a)$ .*

The small country  $b$  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 two large countries, despite being a majority, cannot impose their will on the single small country. All the large countries can achieve is the best trade agreement that they can reach without the participation of the small country, which is  $CU(a,c)$ , a customs union among themselves.

A similar story unfolds in the scenario without article XXIV paragraph 5. Under the reduced choice set, the countries' preference rankings are as follows:

$$\begin{array}{ll} \text{country } b: & MTAGFT \prec_b MTA(i,b) \prec_b MFN \prec_b MTA(i,j) \\ \text{countries } a/c: & MFN \prec_i MTA(i,b), MTA(j,b) \prec_i MTA(i,j) \prec_i MTAGFT \end{array}$$

As the logic of the respective preference rankings for 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  $MTA(c,a)$ .*

Similar to the other asymmetric corner considered before, having one small and two large countries allows for partial but not full liberalisation of world trade, irrespective of the constitutional 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 the case of  $MTA(i,j)$  relative to  $CU(i,j)$ , as it exploits the MFN obligation of the large countries. By contrast, the large countries are better off under the PTA. Therefore, while neither of the two institutional arrangements facilitate global free trade, they influence the welfare of the stable set, both globally and its distribution across countries.

**3.2. Edges of the Triangle.** Let us now turn to the edges of the triangle where the endowments of the three countries vary along a single dimension. Referring back to Figure 4, we denote the southern edge by VSS, where country  $a$  varies from small to large, and countries  $b$  and  $c$  are small, connecting the Sym and LSS corners. The eastern edge is denoted by LSV, where country  $a$  is large,  $b$  is small, and  $c$  varies from small to large, connecting the corners LSS and LSL. The third, northwestern edge, finally, we denote by VSV, where countries  $a$  and  $c$  vary in unison from small to large, and  $b$  is small, connecting the corners Sym and LSL.

Our analyses of each edge will be summarized by graphs that depict the composition of the stable set as we vary the endowment of one country (or of two in unison in the case of VSV), and we provide accompanying descriptions that explore the underlying forces at play. The exact numerical values for the (sub-)intervals of each edge are provided in Section C.3. In order to gain an intuitive understanding of the results, we identify specific trade agreements that switch from stable to unstable (or vice versa) at particular endowment tuples, and explore the underlying mechanics to understand the reason for these switches.

**3.2.1. VSS: 1 varying, 2 small countries.** We consider first the scenario where  $e_b = e_c = e_{min}$  and  $e_a \in [e_{min}, e_{max}]$  (the edge connecting corners Sym and LSS). Alongside this edge, certain trade policy arrangements never appear in the LCS, in particular the MFN, FTAHub(a), and MTA(b,c) regimes. The stable sets along this edge are presented in Figure 5.<sup>27</sup>

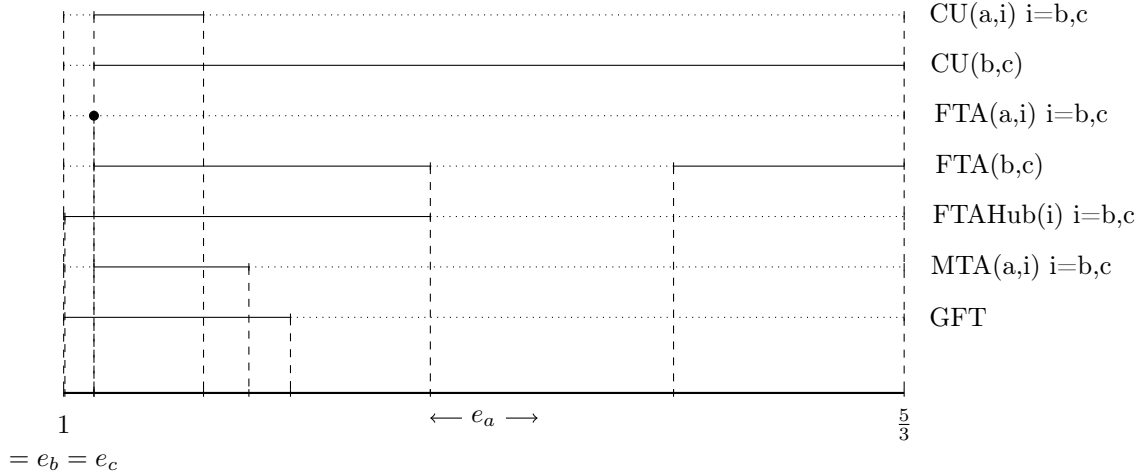


FIGURE 5. Characterisation of varying, small, and small country

In the left corner (Sym), where  $e_a$  is close to one and the size distribution close to symmetry, the GFT regime (that is, the trinity of FTAGFT, MTAGFT, and CUGFT) is the only element of the LCS. As  $e_a$  increases slightly, both FTAHub(b) and FTAHub(c) emerge as stable outcomes. Along the whole interval, a number of different PTA and MTA constellations, either between the growing country  $a$  and one of the smaller

<sup>27</sup>Where the dot marks a single point, and the height of different lines in the diagram has no meaning, except for keeping the elements of the LCS distinguishable, as we move along the edge horizontally.

countries (b/c) or between the two smaller countries b and c appear. Towards the upper end, where  $e_a$  tends to one and we reach the LSS corner, only the PTAs between the small countries b and c are still stable.

First, the spike in the number of stable constellations close to symmetry actually follows from a change in the preferences of the varying country with respect to  $\text{CU}(b,c)$  vs. the trinity of GFT regimes — it starts preferring the former over the latter as it grows larger. Furthermore,  $\text{FTA}(a,b)$  and  $\text{FTA}(c,a)$  become unstable because the small countries start to like the MFN regime more than the GFT trinity (or rather these are only stable for as long as this is not the case). When country  $a$  becomes sufficiently large, country  $b$  prefers  $\text{FTAHub}(c)$  over  $\text{CU}(a,b)$ , and  $c$  prefers  $\text{FTAHub}(b)$  over  $\text{CU}(c,a)$ . As a consequence, both of these CUs drop out of the LCS. Similarly, when countries  $b$  and  $c$  start preferring  $\text{FTAHub}(c)$  and  $\text{FTAHub}(b)$  over GFT, the latter stops being stable. A similar argument also applies to the MTA regimes. When the size of country  $a$  increases even further, both country  $b$  and  $c$  favour  $\text{CU}(b,c)$  over the respective hub constellations, which results in  $\text{FTA}(b,c)$ , as  $\text{FTAHub}(b)$  and  $\text{FTAHub}(c)$  become unstable. When the endowment of country  $a$  tends towards its maximum, the small countries are constrained by the MFN-tariffs and do not differentiate between  $\text{CU}(b,c)$  and  $\text{FTA}(b,c)$  anymore, which renders  $\text{FTA}(b,c)$  stable again.

We now turn to the hypothetical scenario without Article XXIV Paragraph 5. The interval over which the GFT (MTAGFT) regime is stable increases significantly. Moreover, over two thirds of the 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 6 demonstrates these results.<sup>28</sup>

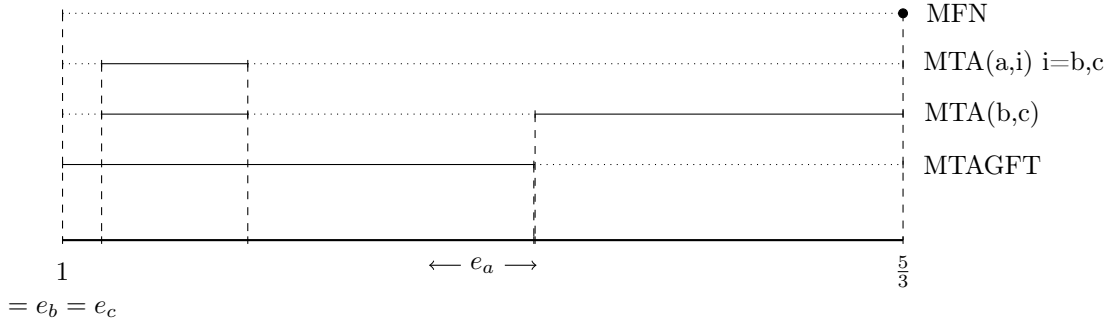


FIGURE 6. Characterisation of varying, small, and small country

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

<sup>28</sup>As before, in addition to the above mentioned trade agreements, the graphic also has MFN as a single point, see the dot, at  $e_a = 5/3$  (point LSS).

stability of the MTA among themselves. Furthermore, the GFT regime becomes unstable when the small countries start to prefer their joint MTA over MTAGFT.

To conclude, the effect of the PTAs on the stability of the GFT regime is significant: 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.

**3.2.2. VSV: 2 varying, 1 small country.** Now, let us turn to the case where  $e_b = e_{min}$  and  $e_a = e_c \in [e_{min}, e_{max}]$  (edge VSV). Along this edge, depending on the size of the larger countries ( $a/c$ ), any trade agreement can be part of the stable set. The exact composition of the LCS is indicated by Figure 7.

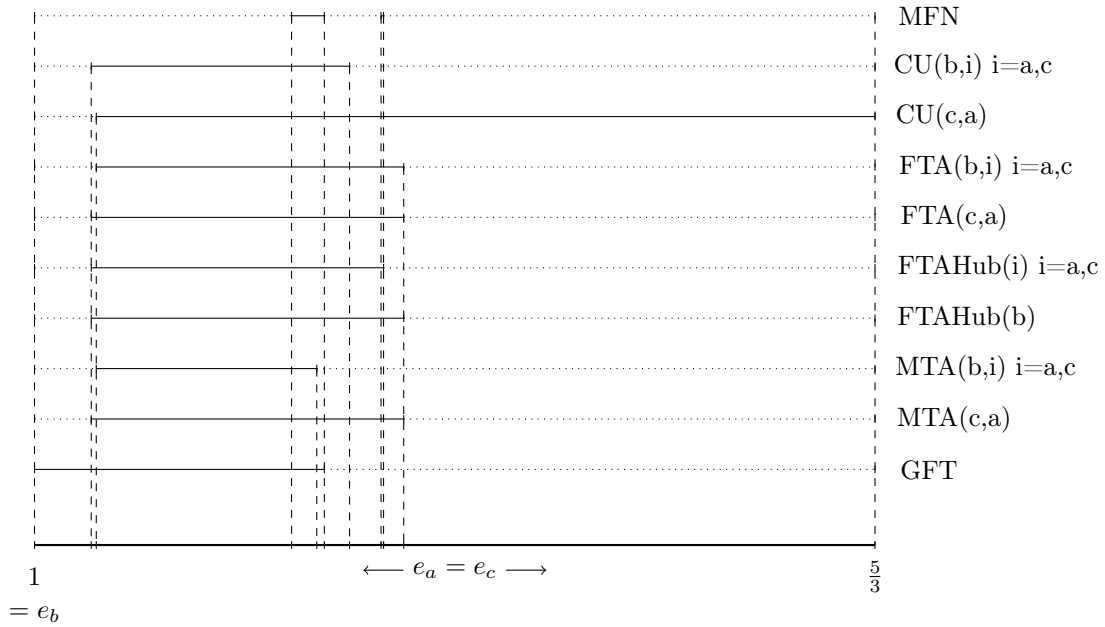


FIGURE 7. Characterisation of varying, small, and varying country

Close to the symmetric corner (Sym), the trinity of GFT regimes is stable, and remains the unique element(s) of the stable set for longer (compared to the previous case, VSS). In addition, a collection of different trade agreements is stable relatively close to symmetry. Towards the other side, as  $e_a = e_c$  tends to  $5/3$  (corner LSL), the CU between the two (varying) countries  $a$  and  $c$  is the unique stable outcome. Furthermore, the MFN regime is also an element of the LCS in two small (one tiny), separated regions.

Again, the peak in stability near symmetry stems from a shift in the preferences of the varying countries ( $a$  and  $c$ ) with respect to the CUs. At the point of the shift, both these countries start to prefer a respective CU with the small country over the trinity of GFT. The occurrence of the MFN regime follows from a preference of the small country ( $b$ ) of MFN over the trinity of GFT (the first region), and then over FTAHub( $a$ ) and FTAHub( $c$ ) respectively (the second, tiny region). As countries  $a$  and  $c$  grow larger, first MTA( $a,b$ ) and



MTA(b,c) drop out from the LCS as they rank lowest according to the preferences of country  $b$ , and then MTA(a,c) follows. »> CHECK «< The trinity of GFT becomes unstable once the small country  $b$  starts to prefer CU(c,a). Next, CU(a,b) and CU(b,c) follow in dropping out from the LCS as the small country starts to prefer the MFN regime over a CU with any of the larger two countries. As soon as CU(c,a) becomes preferred by country  $b$  over any FTA where  $b$  participates, or any hub structure with a large country as the hub, all aforementioned constellations drop out from the LCS.

We now switch off Article XXIV Paragraph 5 and consider the alternative scenario without PTAs. In contrast to the previous case (edge VSS), the interval where the GFT regime (MTAGFT) is part of the stable set decreases. However, the decrease in range is small. A similar observation holds for the range close to symmetry, where the GFT regime is the unique stable outcome. The exact composition of the LCS is shown in Figure 8.

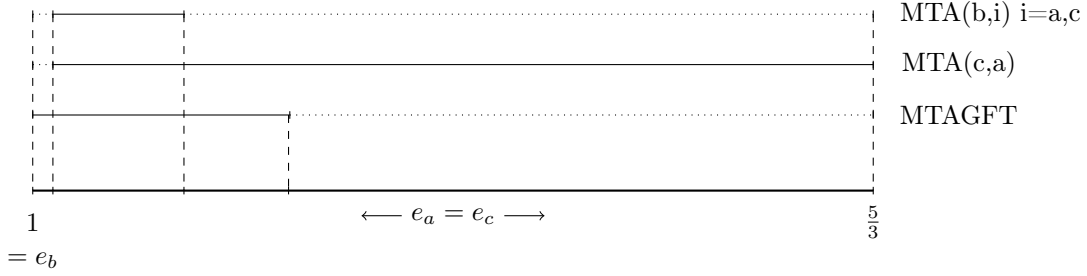


FIGURE 8. Characterisation of varying, small, and varying country

The main driving force behind the changes in the set of stable outcomes 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 compared to the other trade agreements. As soon as country  $b$  prefers MFN over another constellation, the latter drops out from the LCS. Beyond a certain point, the MTA between the large countries  $a$  and  $c$  remains the only stable outcome.

Similar to the previous edge (VSS), the case at hand renders a clear prediction of stable outcomes difficult, as it is hard to rule out regimes, especially around symmetry where, as before, almost all constellations are stable. The effect of the abolition of Article XXIV Paragraph 5 affects the stability of the GFT regime in the opposite direction when compared to the previous case (VSS).

Before moving to the analysis of the third, final edge of the triangle, it is worth to pause and try to understand the role that farsightedness plays in determining the stable set of the trade policy arrangements. The related literature does not allow a full comparison of the results due to differences in the set of trade policy regimes that are considered. In specific cases, however, we can directly compare the composition of the stable sets of our paper with those found by Saggi, Woodland, and Yildiz (2013), who apply the concept of

Coalition Proof Nash Equilibrium (CPNE) — and hence limited farsightedness — to determine the stability of trade policy constellations. Their ‘multilateralism game’ is exactly the same as our scenario of a modified institutional setup without PTAs. Compare Figures 2 and 5 of Saggi, Woodland, and Yildiz (2013) to our Figures 8 and 6 in this paper. In the case with one small country and the other two varying, both approaches predict the same stable sets near the interval’s endpoints (corners Sym and LSL). Under the CPNE approach, as soon as the small country has the incentive to deviate from GFT (MTAGFT) unilaterally, this regime stops being stable. By contrast, MTAGFT remains a part of the stable set in our work even though the small country has a "shortsighted" incentive to deviate from GFT. As long as MTAGFT indirectly dominates all other trade policy regimes, we cannot rule it out as part of the stable set. As soon as MTAGFT stops indirectly dominating the MFN constellation, it drops out from the LCS. A similar observation applies when we consider the edge VSS, the case of two small and one varying country. Again, near the endpoints of the interval, i.e. close to the corners Sym and LSS, the results coincide. Here, the appearance of MTAs in the LCS does not prevent MTAGFT from staying in the stable set (unlike the results under CPNE). Furthermore, there is an interval away from symmetry (one-third of the total interval) where MTAGFT emerges as the unique stable regime once again. We obtain this distinctive result under farsightedness because the MFN regime starts to dominate MTAs between small and large countries indirectly. Simultaneously, all above-mentioned trade policy constellations are indirectly dominated by an MTA between the two small countries and by MTAGFT, while these two regimes indirectly dominate each other. In conclusion, in the cases we have discussed, the main difference between limited (CPNE) and unlimited farsightedness boils down to the length of the intervals where MTAGFT is found to be stable. One could speculate that for certain ranges of the endowment distribution, the farsighted decision-makers are more inclined towards global free trade than those who are less farsighted.

**3.2.3. LSV: 1 large, 1 small, 1 varying.** Finally, we turn attention to the third edge of the triangle where we find a unique composition of LCS under the hypothetical scenario without PTAs. As before, though, we start by considering the status-quo institutional setup with PTAs. Along LSV,  $e_a = e_{max}$ ,  $e_b = e_{min}$ , and  $e_c \in (e_{min}, e_{max})$ . For these endowment constellations, a number of trade agreements can be completely ruled out (with respect to the LCS). The MFN and (trinity of) 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 FTA-hub structures where either the small or the large country is the hub node. As for the actual composition of the LCS along this edge, see Figure 9 for a graphical representation.

The general observation is that when the varying country  $c$  is close in size to the small country  $b$ , then the PTAs between these smaller countries appear as elements in the stable set. As the varying country  $c$  becomes

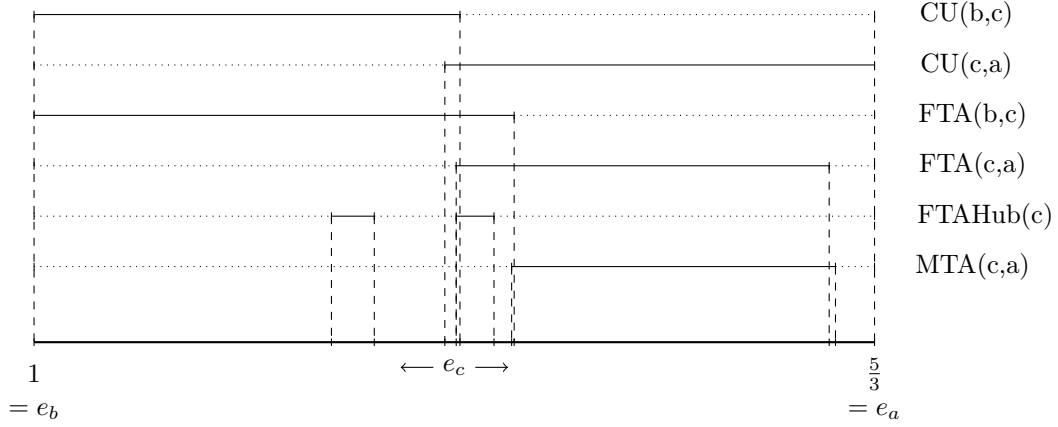


FIGURE 9. Characterisation of large, small, and varying country

larger, the trade policy constellations between the larger countries  $a$  and  $c$  replaces these. Additionally, there are two small, separate, regions in the middle of the interval where  $\text{FTAHub}(c)$  is stable.

Start with the PTAs between countries  $b$  and  $c$ , the small and the varying one. Interestingly, the only factor driving the stability of these regimes are the preferences of country  $b$  (with fixed minimal endowments). Once the MFN regime becomes more desirable than  $\text{CU}(b,c)$  for country  $b$ , the constellation  $\text{CU}(b,c)$  drops from the stable set. Now, an identical line of reasoning holds for the case of  $\text{FTA}(b,c)$ . Thus, for both constellations it only takes a single change in the preference ranking of country  $b$  to alter the stable set.

The PTAs and MTAs between countries  $a$  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 the small country  $b$ , that is, when country  $c$  is relatively small (and  $\text{CU}(b,c)$  actually is an element in the stable set). However, once it becomes preferable for country  $b$  to be the outsider instead of the insider in a CU,  $\text{CU}(c,a)$  emerges as a stable outcome (even though  $\text{CU}(b,c)$  still remains stable). Moreover, as soon as country  $c$  prefers  $\text{FTA}(c,a)$  respectively  $\text{MTA}(c,a)$  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:

$$\begin{aligned} \text{country } a: & \quad \text{FTA}(c,a) \prec_a \text{CU}(c,a) \prec_a \text{MTA}(c,a) \\ \text{country } c: & \quad \text{MTA}(c,a) \prec_c \text{FTA}(c,a) \prec_c \text{CU}(c,a) \end{aligned}$$

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

FTAHub(c) is stable in the two small, separate 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 hence agrees with country  $a$ 's preferences in this respect. Once the preferences of country  $b$  over these outcomes are 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, once 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 by consequence the corresponding hub structure is stable as well. 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 alternative institutional scenario 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 MTA(c,a) as the stable elements. Figure 10 presents these findings.<sup>29</sup>

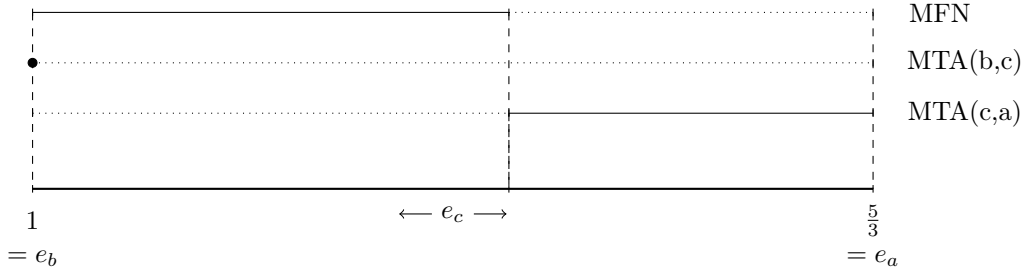


FIGURE 10. Characterisation of large, small, and varying country

Over the whole interval, country  $b$  does not have any incentive to form an MTA with either of the other countries. This is one reason why the MFN regime is stable for part of the interval. The other reason is that country  $c$  prefers not to have a trade agreement with country  $a$  as long as its own size is not large enough. Once country  $c$  is sufficiently large then MTA(c,a) presents a better option than the MFN regime. By consequence, MTA(c,a) replaces the MFN regime as the single element of the stable set.<sup>30</sup>

In conclusion, for the endowment constellations along this edge, the GFT regime never appears as part of the stable set, independent of the scenario (with/without PTAs). However, the constitutional setup does determine whether partial trade liberalisation may take place or not. The possibility of forming PTAs reduces the incentive of the small(est) country to free ride. Otherwise, without PTAs, the MFN regime is the unique stable outcome when there is one small, one large, and one rather small country, and it is the worst outcome

<sup>29</sup>In addition, to the aforementioned elements, it also depicts MTA(b,c) as a single point, see the dot, but this appears only for the sake of completeness because that point corresponds to the corner LSS discussed earlier.

<sup>30</sup>Note that MTA(b,c) is stable in the LSS corner, that is, at the lower bound of the interval.

in terms of over-all welfare. Prohibiting PTAs has this important and negative implication only along this edge.

**3.3. The interior of the triangle.** We now focus on the interior of the triangle in Figure 4. That is, we now consider the full parameter space of endowment allocations. For expositional clarity, we group both CUs and FTAs (incl. FTA-Hubs) together under the label of PTA in figures 11-13.<sup>31</sup> For the same reason, we also suppress the exact member countries of specific trade agreements, i.e. who is an insider and who stays out.

As before, we start by considering the existing institutional set-up, where PTAs are options countries can choose. Figure 11 depicts regions of the parameter space and their stable sets (or rather a simplified view of these). In a small region close to symmetry, labelled as region 1, the (trinity of) GFT regime(s) is the unique stable element. In an adjacent and a slightly more distant area, both labelled as region 2, PTAs become stable as well. The region connecting those two, which we denote as region 3, adds MTAs as another element of the stable set. In a tiny area along the diagonal, indicated as region 4, no form of trade agreement can be excluded from the stable set.<sup>32</sup> Towards the asymmetric corners LSS and LSL, we find two regions, both labelled as 5, where PTAs are the only stable trade policy constellation. In between those two lies region 6, where MTAs are also found to be stable. Finally, in another tiny area along the diagonal, indicated as region 7, MFN enters the stable set as well. In general, with a certain degree of asymmetry between countries' endowments, at most partial trade liberalisation can be expected, as (the trinity of) GFT is an element of the stable set only in regions 1 through 4 which are all characterized by relatively symmetric endowment allocations.

Next, we consider the alternative institutional set-up, where PTAs are not allowed. Figure 12 depicts regions of the parameter space and their corresponding stable sets. In a small area near symmetry as well as in a sizeable area away from it, both denoted as region 1, GFT (that is, MTAGFT here) is again the unique stable element. Adjacent to these are two areas, both denoted as region 2, where MTAs become stable as well. Moving towards the asymmetric corners, two regions denoted by 3, feature MTAs as the only stable element. In between those, in region 4, only the MFN is the unique element of the stable set.

The comparison of the two graphs allows us to deduce two compelling statements. The first noteworthy result is the extent of MFN under each scenario. In the alternative institutional setup without PTAs, the area where MFN is an element of the stable set (in fact the unique element in this scenario) is substantially larger than under the status-quo institutional rules. Note that this effect is at work sufficiently away from symmetry towards the corner LSS. Under (significant) asymmetry then, it seems that PTAs allow countries to move towards their international efficiency frontier by forming such agreements (cf. Bagwell et al. (2016)).

<sup>31</sup>However, graphs of this analysis that distinguish between the two types of PTAs can be found in Appendix C.2.

<sup>32</sup>Note that this does not imply that every CU, every FTA or FTA-Hub combination of countries is stable.

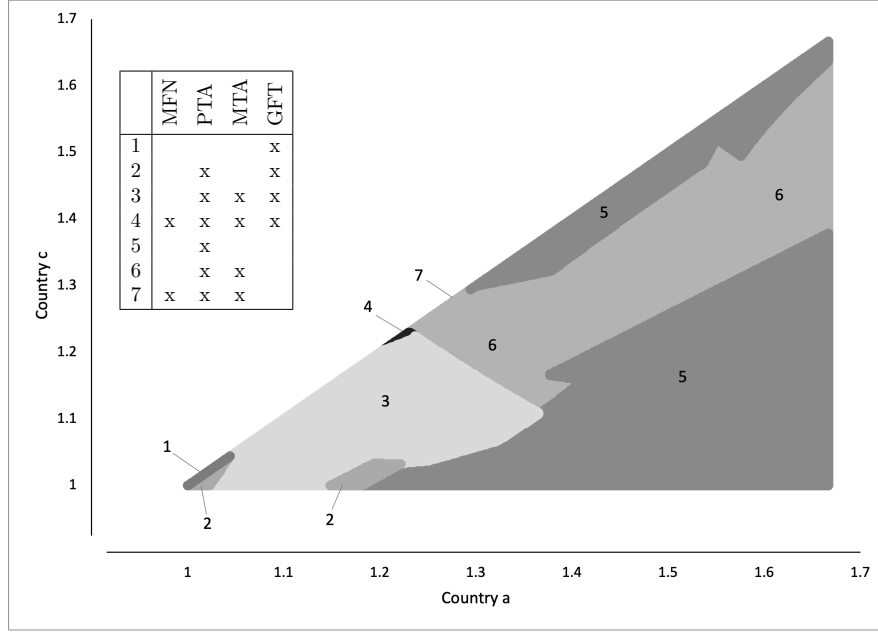


FIGURE 11. Aggregate View of Overall Stability with PTAs

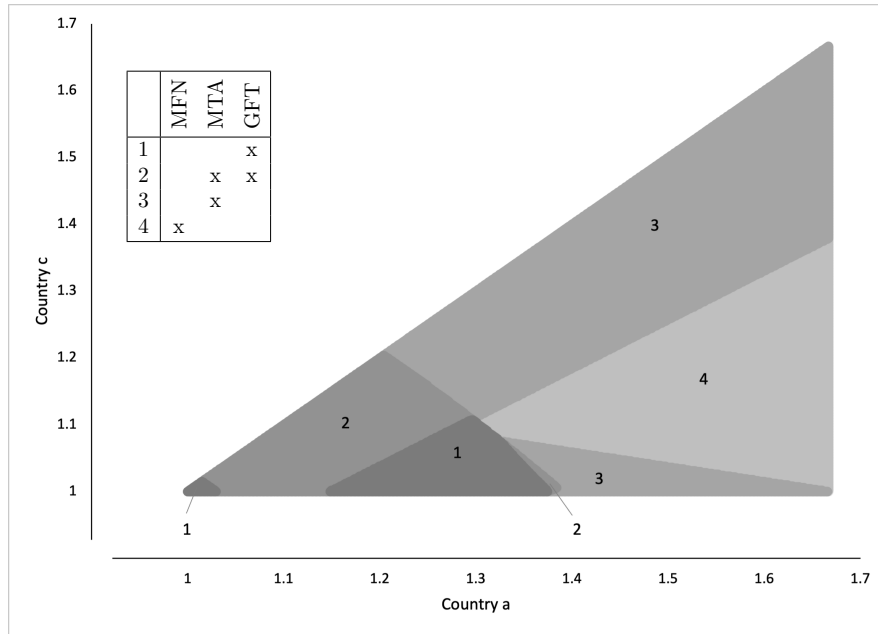


FIGURE 12. Simplified Overall Stability without PTAs

The reason for this is that PTAs allow countries to discriminate against the outsider, whereas an MTA does not allow for this.

The second interesting result is the difference in the extent of stability of GFT (trinity thereof and MTAGFT respectively) across the two regulatory scenarios. First, recall that once the degree of asymmetry exceeds a certain threshold, none of the GFT regimes remain in the stable set, irrespective of the institutional set-up.

Around symmetry by contrast, the opposite holds in that the GFT regimes are always stable (under both scenarios). In between, the effect of PTAs on the stability of GFT depends on the structure of asymmetry. Figure 13 depicts the different areas of stability of the GFT regimes, depending on the regulatory scenario — with or without PTAs. Note that region 1, close to symmetry, indicates the stability of GFT under both regulatory scenarios, as described above. Away from symmetry, in case two countries are relatively larger (but not too large), the abolishment of PTAs results in reducing the area where GFT is stable, corresponding to region 2 in the diagram. In this region, PTAs act as ‘building blocks’ on the road to GFT, as they generate sufficient costs for the small country in the case of leaving the a GFT arrangement (either CUGFT or FTAGFT here). So punishment in the form of a discriminatory external tariff of a PTA between the relatively large countries overcomes the free-riding incentive of the small country. By contrast, if two countries are relatively smaller (but not too small), the same regulatory difference yields the exact opposite effect, see region 3. Here, PTAs act as ‘stumbling blocks’. When two countries are relatively smaller, a PTA arrangement between them increases their incentive to exclude the single large country. If they form an MTA (when PTAs are not allowed), they would not have such an exclusion incentive, as the non-discriminatory principle applies. The comparison of the two different forms of asymmetry thus yield an important insight: whether PTAs turn out to be ‘building blocks’ or ‘stumbling blocks’ for intermediate levels of asymmetry depends on the relative size of two countries versus country three.

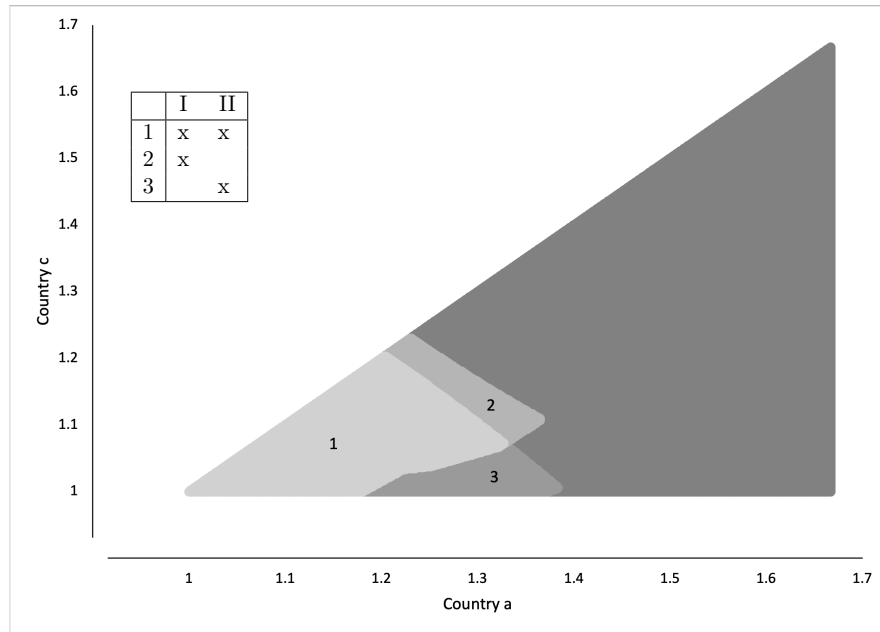


FIGURE 13. Stability of GFT regimes with (I) and without (II) PTAs

We now go one step further and ask: whenever PTAs are conducive to global free trade, which form of PTA is responsible for the result? In other words, we want to find out which form of PTA, a CU or an FTA,

contributes to the stability of GFT when PTAs are available (regions 1 and 2 of Figure 13). The subsequent figure is meant to shed light on this question. From Figure 14 one can see that the stability for GFT (due to the option of forming PTAs) in region 2 stems from the possibility to form a CU among the relatively larger countries. For endowment combinations with somewhat larger countries as in region 2 of Figure 14, an FTA would not generate enough incentive for a small country to be a member of a GFT arrangement. Under the FTA, member countries impose externalities on each other, resulting in lower tariffs vis-à-vis an outsider (the small country). While under a CU, the larger countries coordinate their decisions, and the resulting tariffs against the small country are higher. The higher tariffs incentivise the small country to be a member of CUGFT. Contrary to region 2, in region 1 of Figure 14, both forms of PTAs guarantee that a GFT regime is stable. Note that here we have always analysed an alternative scenario where both forms of PTAs are available. In the next (and final) section (of analysis), we consider the CU vs. FTA question in scenarios where we switch off one of them only, one at a time.

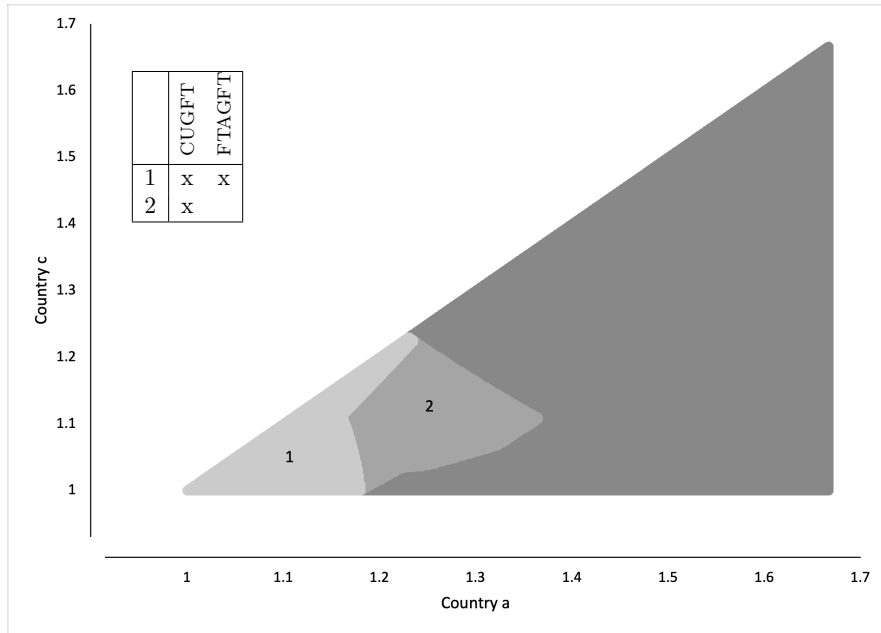


FIGURE 14. Stable area for GFT reached via CUs/FTAs

Summing up our results so far, if the world is characterized by an intermediate level of asymmetry, with two out of three countries close to identical in size while relatively smaller than the third country, the area where GFT is stable increases when prohibiting PTAs. However, when the two similar countries are larger (not too large), the availability of PTAs (in particular, a CU between the two larger countries), is conducive to the GFT regime's stability. If the world is even further away from symmetry, full trade liberalisation is not attainable at all, and an area where the MFN regime is stable appears under the scenario without PTAs.



**3.4. Hypothetical scenarios with only FTAs or CUs.** Figure 13 showed that the availability of PTAs can increase the stability of GFT (region 2). Figure 14 then indicated that this finding is primarily due to customs unions. We obtained this insight in a scenario where both forms of PTAs are available to decision-makers. Taking this one step further to gain a better understanding, we now extend the space of constitutional setups by considering two additional hypothetical scenarios: In the first scenario, FTAs are ruled out, so countries can form MTAs and CUs only, while in the second scenario, CUs are ruled out, so countries can create MTAs and FTAs only. So we now have four scenarios: status-quo, no-FTA, no-CU, and no-PTA

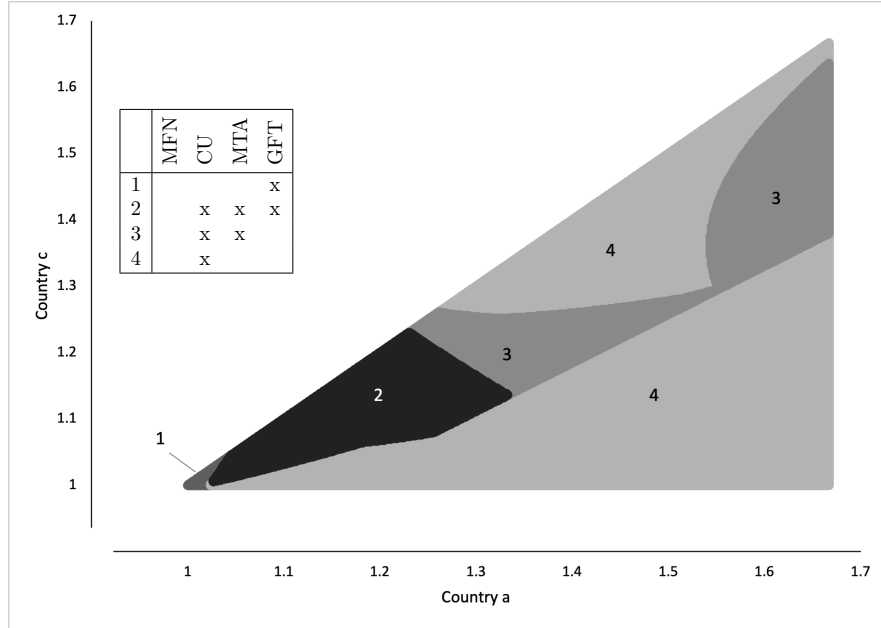


FIGURE 15. Aggregate View of Overall Stability without FTAs

Focusing on no-FTA and no-CU, we demonstrate the respective results in Figures 15 and 16. Figure 15 shows the results for the hypothetical scenario where FTAs are not available. In Figure 16 we represent the regions of the parameter space and their corresponding stable sets for the hypothetical scenario when we rule out CUs.

A few points are worth noting. First, across both scenarios, there is no endowment configuration for which MFN is a stable outcome. Second, even though the area where GFT is uniquely stable is more extensive in a scenario without FTAs compared to the one without CUs (region 1 in Figure 15 vs. region 1 in Figure 16, the latter consisting only of the point of symmetry), the total area where stability of GFT cannot be ruled out is larger without CUs (regions 1 and 2 in Figure 15 vs. regions 1, 2 and 3 in Figure 16). Once we move further away from the point of symmetry, the scenarios considered here do not generate qualitatively different results compared to the scenario with both types of PTA available (cf. Figure 11): towards the asymmetric right-hand-side of the triangle, only partial liberalisation is attainable.

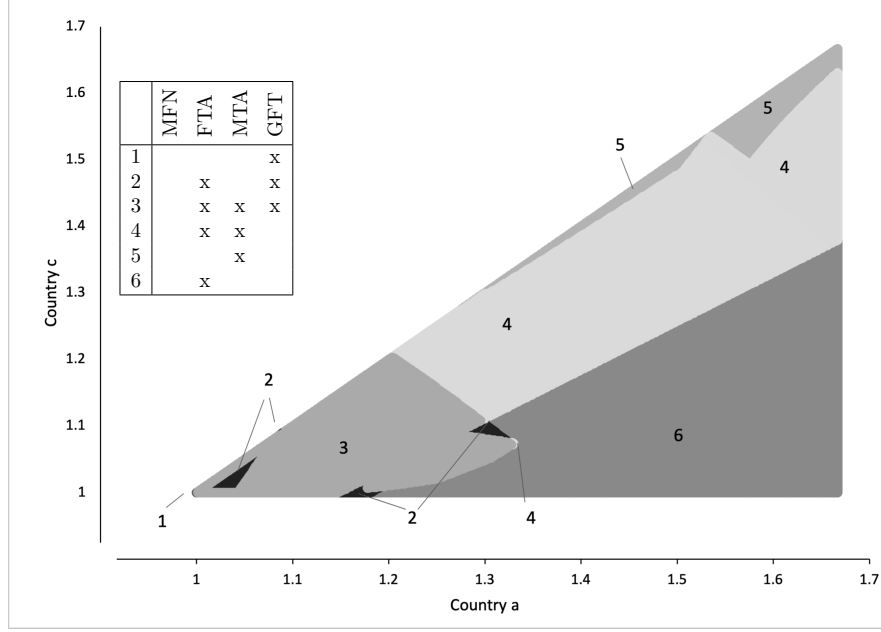


FIGURE 16. Aggregate View of Overall Stability without CUs

Overall we have considered four scenarios: a scenario with PTAs (the existing status-quo institutional set-up), the alternative scenario without PTAs (only MTAs available), and in addition a scenario without FTAs (only CUs and MTAs available) as well as a fourth scenario without CUs (only FTAs and MTAs available). Figure 17 demonstrates the regions where GFT is a stable outcome under each of the four scenarios.

Our main conclusion is that the fundamental trade-off between the stability of GFT regimes with and without PTAs (regions 2 and 3 in Figure 13) is also present in the scenarios without CUs. When we switch off the possibility of forming CUs, even though there is a slight gain in the area where GFT is stable (region 3 in Figure 13), we lose regions 5 and 6 where the GFT regime is no longer stable. When we switch off only FTAs, by contrast, there is no gain in stability for GFT, we simply lose stability of GFT in areas 2 and 6.

Next, when discussing region 2 of Figure 13 in the previous subsection (corresponding to regions 5 and 6 in Figure 17), we concluded that CUs were essential in extending stability of the GFT regime to region 2. We can now provide a more nuanced view: The stability of GFT is ensured solely by the availability of CUs (region 5 in Figure 17). In region 6 of Figure 17, however, the availability of FTAs is required for CUs to play their role. This final result demonstrates the importance of considering a comprehensive set of possible trade policy arrangements when studying their stability.

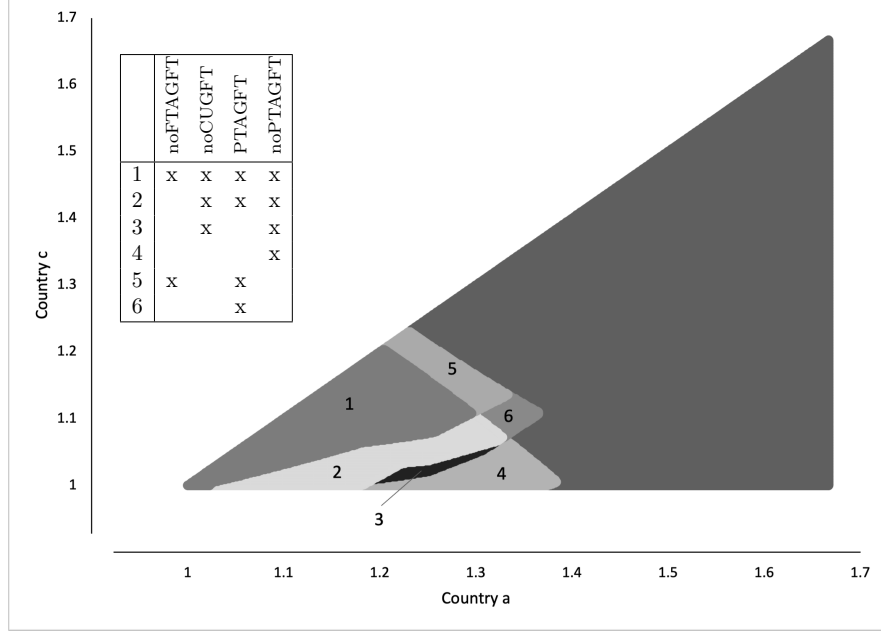


FIGURE 17. Different Scenarios and Stability of Global Free Trade Arrangements

#### 4. CONCLUSION

Under the rules of the WTO (previously GATT), a group of countries can engage in both multilateral and different forms of preferential trade liberalisation. The formation of global trade agreements is a complex game and the rules of the game affect the exact nature of the outcome. Article I of the WTO aims at creating a multilateral free trading system, while Article XXIV Paragraph 5 allows countries to seemingly bypass the multilateral liberalisation process. In this paper, our focus lies on the stability of trade policy arrangements under two different regulatory scenarios — with versus without PTAs. Based on unlimited farsightedness of the participants in trade negotiations and considering an extensive set of trade agreements, we thus link our model more closely to reality.

Whether PTAs are ‘building blocks’ or ‘stumbling blocks’ on the path towards global free trade is not straightforward to answer. In the end, our results presented here provide a mixed answer, and we find that it depends on the size distribution of countries. Close to symmetry, global free trade is the unique stable trade constellation under both regulatory scenarios (PTA vs. no-PTA). Far away from symmetry, by contrast, GFT might not be reached at all. In between, the effect of switching off Article XXIV Paragraph 5 depends on the exact form of asymmetry. In case two countries are relatively smaller than the third, prohibiting PTAs increases the area of stability of the GFT regimes. When two countries are relatively larger than the third, it reduces the area. Finally, when the world is far 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. For such size distributions, PTAs act as a mechanism that helps to avoid the MFN regime.

Our research raises several, interesting questions requiring further investigation. First, it will be interesting to study the robustness of our findings with respect to the underlying trade model. While the competing-exporters model has been the standard choice in the related literature, economists have also used both oligopoly and competing-importers models. The framework proposed here is general enough to accommodate different underlying trade models such as the ones just mentioned. Another potential area of inquiry might be an extension of the framework to allow a larger number of countries. In fact, in addition to bilateral negotiations, so-called plurilateral negotiations play an important role in the development of preferential trade liberalisation. Recent examples include the Trans-Pacific Partnership (which has evolved into the CPTPP) and the Regional Comprehensive Economic Partnership (RCEP) championed by China. Including more than three countries in our model would allow us to investigate the strategic interactions among countries whilst taking such plurilateral negotiations into account. The incorporation of political economy considerations into the underlying trade model is another area of interest,<sup>33</sup> as it might allow us to understand the nature of tariff peaks occurring after PTAs come into effect. It is our view that modifications and extensions of our framework (in line with these ideas) are directions worthy of further research.

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

## 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.
- [9] B. Douglas Bernheim, Bezalel Peleg, and Michael Whinston. Coalition-proof nash equilibria i. concepts. *Journal of Economic Theory*, 42(1):1–12, 1987.

---

<sup>33</sup>See for example Facchini et al. (2013).

- [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] Debraj Ray and Rajiv Vohra. The farsighted stable set. *Econometrica*, 83(3):977–1011, 2015.
- [26] 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.
- [27] Kamal Saggi and Halis Murat Yildiz. Bilateralism, multilateralism, and the quest for global free trade. *Journal of International Economics*, 81(1):26–37, 2010.
- [28] John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.
- [29] Jin Zhang, Zhiwei Cui, and Lei Zu. The evolution of free trade networks. *Journal of Economic Dynamics and Control*, 38:72–86, 2014.
- [30] 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  $(i,j) \in X \times X$ .

---

**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 2.5
3:    $\alpha = \alpha MinValue(E)$  ▷ See Section 2.5
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 18. 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 2.1, 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  $[0, t_i^{MFN}]$ :

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

**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.

**B.3. Overall Welfare.** The following table lists the overall welfare for each (representative) trade agreement, depending purely on endowments, computed modulo  $2\alpha \left( \sum_{n \in N} e_n \right)$ , 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.

Trade Agreement	Overall Welfare
MFN	
$\hookrightarrow$ 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)	
$\hookrightarrow$ no cap	$\frac{1}{1600}(-563e_i^2 - 550e_i e_j - 448e_i e_k - 563e_j^2 - 448e_j e_k - 704e_k^2)$
$\hookrightarrow i^c$	$\frac{1}{1600}(-563e_i^2 - 550e_i e_j - 448e_i e_k - 550e_j^2 - 550e_j e_k - 627e_k^2)$
$\hookrightarrow i^c, j^c$	$\frac{11}{32}(-e_i^2 - e_i e_j - e_i e_k - e_j^2 - e_j e_k - e_k^2)$
CUGFT	
$\hookrightarrow$ 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)	
$\hookrightarrow$ no cap	$\frac{1}{7744}(-2963e_i^2 - 2662e_i e_j - 1728e_i e_k - 2963e_j^2 - 1728e_j e_k - 3648e_k^2)$
$\hookrightarrow i_c$	$\frac{1}{23232}(-8889e_i^2 - 7986e_i e_j - 5184e_i e_k - 7865e_j^2 - 7744e_j e_k - 9344e_k^2)$
$\hookrightarrow i_c, i_c$	$\frac{1}{192}(-65e_i^2 - 66e_i e_j - 64e_i e_k - 65e_j^2 - 64e_j e_k - 64e_k^2)$
$\hookrightarrow i^c$	$\frac{1}{7744}(-2963e_i^2 - 2662e_i e_j - 1728e_i e_k - 2662e_j^2 - 2662e_j e_k - 3155e_k^2)$
$\hookrightarrow i^c, j^c$	$\frac{11}{32}(-e_i^2 - e_i e_j - e_i e_k - e_j^2 - e_j e_k - e_k^2)$
FTAHub(i)	
$\hookrightarrow$ no cap	$\frac{1}{363}(-153e_i^2 - 81e_i e_j - 81e_i e_k - 146e_j^2 - 121e_j e_k - 146e_k^2)$
$\hookrightarrow j_c$	$\frac{1}{363}(-137e_i^2 - 81e_i e_j - 121e_i e_k - 146e_j^2 - 121e_j e_k - 121e_k^2)$
$\hookrightarrow j_c, k_c$	$\frac{1}{3}(-e_i^2 - e_i e_j - e_i e_k - e_j^2 - e_j e_k - e_k^2)$
$\hookrightarrow j^c$	$\frac{1}{23232}(-8889e_i^2 - 5184e_i e_j - 7986e_i e_k - 9344e_j^2 - 7744e_j e_k - 7865e_k^2)$
$\hookrightarrow j^c, k^c$	$\frac{1}{192}(-66e_i^2 - 66e_i e_j - 66e_i e_k - 65e_j^2 - 64e_j e_k - 65e_k^2)$
FTAGFT	
$\hookrightarrow$ 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)	
$\hookrightarrow$ 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	
$\hookrightarrow$ 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 19. The overall welfare for each trade agreement depending on endowments

## APPENDIX C. ANALYSIS

## C.1. Proofs of propositions.

## C.1.1. Proof of the proposition 5.

*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 <i>CU(a, b)</i>	1	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0
3 <i>CU(b, c)</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4 <i>CU(c, a)</i>	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 <i>FTA(a, b)</i>	1	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0
7 <i>FTA(b, c)</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8 <i>FTA(c, a)</i>	1	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0
9 <i>FTAHub(a)</i>	1	0	1	0	0	1	1	1	0	1	1	0	0	1	0	0
10 <i>FTAHub(b)</i>	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
11 <i>FTAHub(c)</i>	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 <i>MTA(a, b)</i>	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0
14 <i>MTA(b, c)</i>	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
15 <i>MTA(c, a)</i>	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>34</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.  $Y = LCS$ .  $\square$

## C.1.2. Proof of the proposition 6.

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

<sup>34</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).

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

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.  $\square$

### C.1.3. Proof of the proposition 7.

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

$$\begin{array}{c}
1 \text{ } MFN \\
2 \text{ } CU(a, b) \\
3 \text{ } CU(b, c) \\
4 \text{ } CU(c, a) \\
5 \text{ } CUGFT \\
6 \text{ } FTA(a, b) \\
7 \text{ } FTA(b, c) \\
8 \text{ } FTA(c, a) \\
9 \text{ } FTAHub(a) \\
10 \text{ } FTAHub(b) \\
11 \text{ } FTAHub(c) \\
12 \text{ } FTAGFT \\
13 \text{ } MTA(a, b) \\
14 \text{ } MTA(b, c) \\
15 \text{ } MTA(c, a) \\
16 \text{ } MTAGFT
\end{array}
\begin{pmatrix}
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0
\end{pmatrix}$$

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.  $\square$

#### C.1.4. Proof of the proposition 8.

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

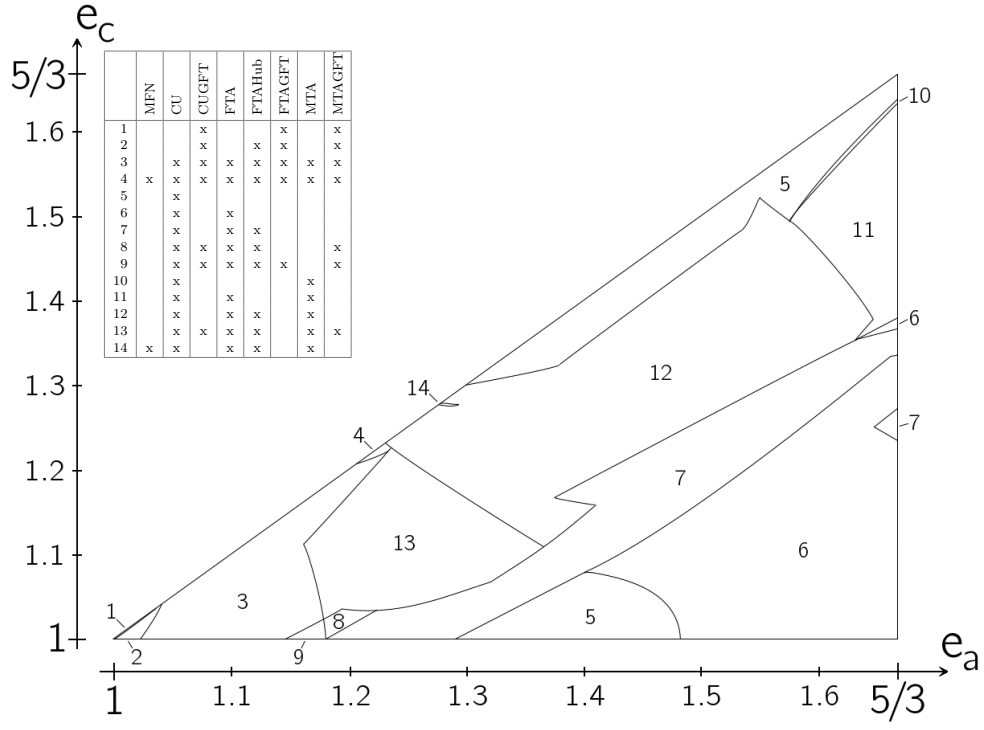
$$\begin{array}{c}
 \begin{array}{l}
 1 \text{ } MFN \\
 2 \text{ } MTA(a, b) \\
 3 \text{ } MTA(b, c) \\
 4 \text{ } MTA(c, a) \\
 5 \text{ } MTAGFT
 \end{array}
 \begin{pmatrix}
 1 & 2 & 3 & 4 & 5 \\
 0 & 0 & 0 & 1 & 0 \\
 1 & 0 & 0 & 1 & 0 \\
 1 & 0 & 0 & 1 & 0 \\
 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0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

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.

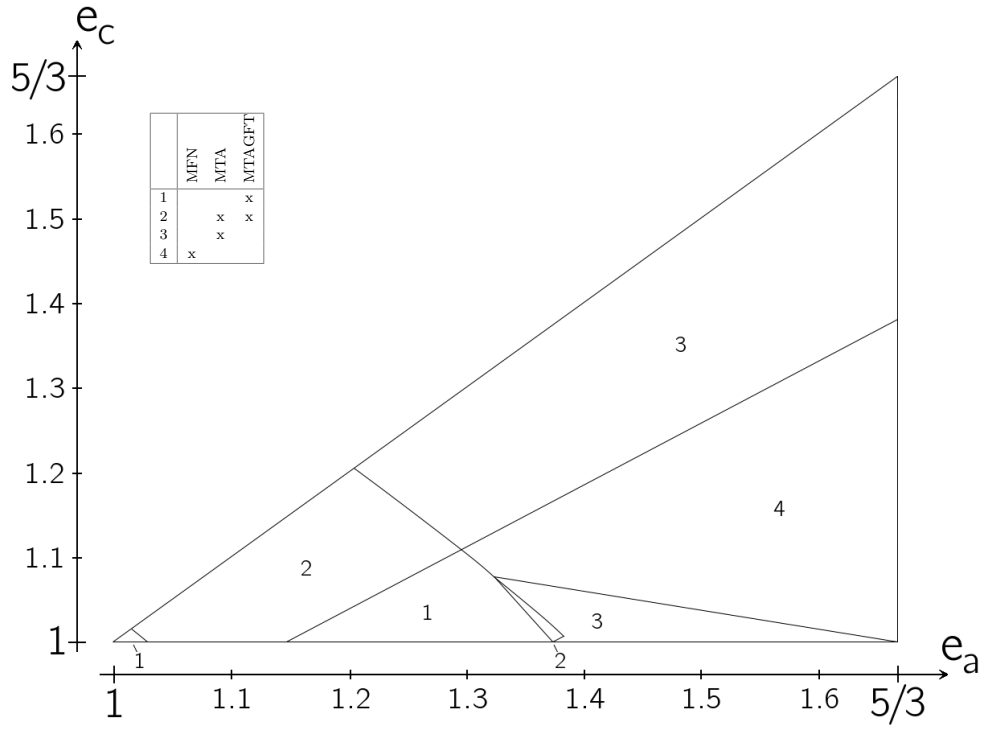
Now, analyse the case of the  $MFN$  regime. Take the following deviation:  $MFN \rightarrow_{\{a, c\}} MTA(c, a)$ . As  $MTA(c, a)$  is not indirectly dominated by any other trade agreement and  $MFN \prec_{\{a, c\}} MTA(c, a)$ , the  $MFN$  regime can not be stable as well.

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.  $\square$

C.2. **Additional Graphics.** The following provides detailed figures:



(A) Overall Stability with PTAs



(B) Overall Stability without PTAs

FIGURE 20. Overall Stability with and without PTAs

**C.3. 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)	[1.000000000000000,1.3380093520374081]
CU(c,a)	[1.3259853039412157,1.6666666666666667]
FTA(b,c)	[1.000000000000000,1.3807615230460921]
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	[1.000000000000000,1.1803607214428857]
FTA(a,b)	[1.0240480961923848,1.2404809619238477]
FTA(b,c)	[1.0240480961923848,1.29124916499666]
	[1.483633934535738,1.6666666666666667]
FTA(c,a)	[1.0240480961923848,1.2404809619238477]
FTAHub(b)	[1.0013360053440215,1.29124916499666]
FTAHub(c)	[1.0013360053440215,1.29124916499666]
FTAGFT	[1.000000000000000,1.1803607214428857]
MTA(a,b)	[1.0240480961923848,1.1469605878423514]
MTA(c,a)	[1.0240480961923848,1.1469605878423514]
MTAGFT	[1.000000000000000,1.1803607214428857]
$e_b = e_{\min} \leq e_a = e_c \leq e_{\max}$	
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	[1.000000000000000,1.2297929191716768]
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	[1.000000000000000,1.2297929191716768]
MTA(a,b)	[1.0494321977287908,1.2244488977955912]
MTA(b,c)	[1.0494321977287908,1.2244488977955912]
MTA(c,a)	[1.0454241816967267,1.2925851703406814]
MTAGFT	[1.000000000000000,1.2297929191716768]

TABLE 21. 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]
MTA(c,a)	[1.3754175016700068, 1.6666666666666667]
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 22. The exact intervals of stability without PTAs