WHY ARE TRADE AGREEMENTS REGIONAL?

A Theory Based on Noncooperative Networks

Ben Zissimos

No 652

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

THE UNIVERSITY OF WARWICK

Why Are Trade Agreements Regional?

A Theory based on Noncooperative Networks[1]

Ben Zissimos

University of Birmingham

First draft: October 2002

ABSTRACT: This paper argues that free trade agreements (FTAs) are regional because, in their absence, optimal tariffs are higher against (close) regional partners than (distant) countries outside the region. Optimal tariffs shift rents from foreign firms to domestic citizens. Lower transport costs imply higher rents and therefore higher tariffs. So regional FTAs have a higher payoff than non-regional FTAs. Therefore, only regional FTAs may yield positive gains when sponsoring a FTA is costly. To analyze equilibrium, standard theory of non-cooperative networks is extended to allow for asymmetric players. Naive best response dynamics show that 'trade blocks can be stepping blocks' for free trade.

KEYWORDS. free trade, globalization, networks, noncooperative games, protection, regionalization, trade agreement, trade block, trade liberalization.

JEL CLASSIFICATION NUMBERS: F02, F13, F15, C73.

[1]I would like to thank Ruth Baldry, Iwan Barankay, Gilluame Haeringer, Anne van den Nouweland and Myrna Wooders for helpful comments and conversations about this paper. Financial support from the ESRC's "Understanding the Evolving Macroeconomy Programme" and the Warwick Centre for Public Economics is gratefully acknowledged.

1. Introduction

There is a sizeable literature on the economic implications of regional trade agreements. Yet almost the entire literature leaves aside the question of *why* it is that trade agreements are regional. The purpose of this paper is to present a theoretical explanation of why countries tend to form trade agreements with other countries in the same region rather than with more distant nations.

In referring to 'regional trade agreements' it is generally recognized that members are geographically close to one another. Prominent examples are the North American Free Trade Agreement (NAFTA) and European Union (EU). In both cases, members share common borders (The British Isles and Sweden are separated by sea, but are otherwise contiguous to other members). Wider evidence that trade blocks are predominantly regional is provided by WTO (2000), a report titled "Mapping of Regional Trade Agreements", in which each of the 150 agreements notified to the WTO is represented in map form. It shows that member countries tend to be geographically close in the majority of cases.

Krugman (1991) argues that it is 'natural' for trade blocks to exist between countries that are close if distance makes inter-regional trade uneconomical. But Frankel, Stein and Wei (1995) use a gravity model to show empirically that countries behave preferentially towards close neighbors; trade volumes in the Western Hemisphere and elsewhere are greater than could be explained by 'natural determinants' such as distance, size and common languages.

In the present paper, a theoretical model is set up which can be used to explain why trade blocks are regional. An equilibrium is demonstrated in which countries form trade agreements with (close) countries in the same region, but have no trade agreements with countries outside the region. To construct the argument, two developments of existing theory are made in the paper. The first is to extend Brander and Spencer's (1984) model of optimal tariff setting to allow for variation in the distance between countries. The second is to extend Bala and Goyal's (2000) theory of non-cooperative networks to allow for discrimination in network formation across different types of player, in this instance countries of different regions. These new theories of 'tariff differentiation by distance'

and 'discrimination in non-cooperative network formation' are linked in the model by an assumption that countries can be grouped into regions. Countries in a region are closer to each other than to countries in other regions.[2]

Brander and Spencer (1984) use a profit shifting argument to motivate optimal tariffs. The higher the rents made by a foreign firm in the domestic market, the more scope there is for shifting rents to domestic citizens through the use of higher tariffs. And because trading costs increase with distance, firms make higher rents in nearby markets than those that are further away. So in the absence of an agreement, optimal tariffs are higher on imports from countries in the same region than on imports from countries of other regions. It follows that a bilateral free trade agreement (FTA) between two close neighbors brings about larger production and trade gains than between distant countries because the former entails a larger mutual tariff reduction.

Whilst Brander and Spencer's model provides a basis for individual tariff setting, the structure of trade agreements in the world as a whole is formalized by adapting Bala and Goyal's (2000) model of noncooperative network formation. Bala and Goyal bring the communication networks previously modelled by others, notably Myerson (1977) and Jackson and Wolinski (1996), into a noncooperative setting.

In communications networks, players benefit from being linked to each other directly and indirectly. For example, if you know someone is the friend of a friend, you can ring up the mutual friend for their phone number. As pointed out on many occasions previously, when communications networks are formed on a cooperative basis they can suffer from coordination failures. The problem is illustrated most clearly in the present setting of

[2]Other papers in the literature have had similar concerns to the present paper, or used modes of analysis that are technically similar. Bond (1999) is closest in the question that he addresses. He compares the sustaintability of multilateral versus regional trade agreements in a repeated game setting, where both types of agreement are sustained through trigger strategies. Bond finds that optimal tariffs are higher between closer neighbours, and that this makes regional agreements easier to sustain using trigger strategies. Whilst some of Bond's results are related, his approach is quite different, not using profit shifting to motivate tariffs, nor the notion of non-cooperative networks to determine equilibrium. The approach of the present paper allows a wider range of dynamic equilibria to be characterised, as discussed below. Other papers, by Goyal and Joshi (2000) and Furusawa and Konishi (2002) are technically similar, in modelling trade agreements as networks. Both papers show that free trade will not necessarily arise. In the case of Goyal and Joshi (2000) this is due to coordination failure. Furusawa and Konishi (2002) show that free trade fails when countries form customs unions. Both papers take a cooperative rather than a non-cooperative approach to the modelling of trade agreements as networks, and neither paper has a regional dimension.

trade agreements by Goyal and Joshi (2000). They model FTAs in the manner of a communication network and network formation is cooperative. As a result, whilst free trade is the most efficient Nash equilibrium, it is by no means the unique Nash equilibrium. Other less efficient FTA structures can be an equilibrium because countries may simply fail to coordinate on membership. In the present context it is important to rule out such possibilities. Otherwise it would be possible to have equilibria with only regional trade blocks resting on nothing more than failures of coordination.

Bala and Goyal address the problem of coordination failure by making individual agents responsible for the cost of coordinating a network. Through their sponsorship, individual agents can form a network if it is in their interest without being encumbered by the need to coordinate with other agents. Then a Nash network is one where no agent can do any better by sponsoring any other network or withdrawing their support for the networks that they sponsor, taking as given networks that they do not sponsor.

By taking a noncooperative network approach, coordination failures are ruled out as a possible cause for regional FTAs. In order to model FTA formation in the setting of a noncooperative network, we will say that each FTA must have a sponsor. There are of course many different types of cost associated with setting up a FTA. These include the cost of bringing together policy-makers and officials at the outset, costs imposed by interest groups opposed to the FTA, costs of designing the administrative system required to run the agreement and the ongoing cost of maintaining it. As the framework of this present paper is essentially a dynamic network formation game, it is the ongoing period-by-period running costs of maintaining an agreement that are invoked to justify the costs of FTA formation in this stylized setting. These might include border controls, verification systems, government customs and excise departments. Such mechanisms of support and verification are a necessary part of a FTA. In any period of the game, the country sponsoring these mechanisms can deviate by withdrawing their financial support for any of the FTAs that it sponsors. If a country undertakes to sponsor a FTA then all the proposed partners accept because they anticipate (and realize) production-trade gains.

The main result of the paper concerns the characterisation of the equilibrium FTA structure that emerges over time under different levels of sponsorship cost. Not surpris-

ingly, if sponsorship costs are above a certain level then no FTAs will form at any point on the equilibrium path, and if they are below a certain level then world free trade will emerge straight away. It is when sponsorship costs are at an intermediate level that regionalism arises and can persist over time. Perhaps most interesting of all, a range of sponsorship costs is identified at which regionalism emerges first before free trade can be reached, providing an answer to Bhagwati's (1992) famous question, "Are trade blocks stepping blocks are stumbling blocks in the path to free trade?"

The model is, of course, highly stylized. In practice FTA formation is significantly more complicated. One complication is that sponsorship of an agreement is likely to be more balanced between members. However, it seems fair to argue that one country normally takes a leading role in getting an agreement off the ground, particularly in terms of its financing. The US played such a leadership role in the setting up of NAFTA, for example, and Germany has been the biggest financial supporter of the EU. As it stands, the equilibrium analysis covers the full spectrum of sponsorship costs, showing how the equilibrium agreement structure changes as sponsorship costs are varied. If it were possible to treat sponsorship costs in a more subtle way, allowing countries to share the cost of an agreement with a sponsor or proposer paying a larger share than the others, then the level of costs at which the equilibrium outcome altered from one structure to another might change, but the range of possible outcomes would probably not.

One element of network formation required for the present model is the property that payoffs vary with different types of player. In the present context, countries receive different payoffs from agreement formation depending on the distance to the FTA partner. Discrimination across different types of player does not feature in previous models of noncooperative networks. However, Slikker and van den Nouweland (2000) have examined discrimination in a cooperative model of network formation. The way that they partition the total set of players by type is used in the model of this present paper. But otherwise their approach is quite different. They have an objective hierarchy over players, with a higher payoff being derived from network formation with members of a particular group. In the present model, variation is subjective. Payoffs are differentiated from the perspective of a given individual country over the geographical distance of its FTA partners. This extension to the standard model of noncooperative network formation is not significant

in itself. But it does allow the analysis of micro-founded differences in the payoff to links with different types of player.

The fact that the gains from different types of network formation can be analyzed, and that they are derived from an underlying micro-model is a new development of the present paper worth emphasizing. The relative benefit to a regional agreement comes *not* simply because benefits to regional agreements are assumed to be higher. It is instead because they are derived to be higher. The paper develops a way of linking these different micro-founded gains to the payoff structure of a network formation game.

This new approach to the analysis of network formation with different types of player potentially makes it possible to study a range of different situations that are of interest in economics. Perhaps the most famous example is due to Coase (1960), who points out that firms exerting relatively large externalities on one another are better candidates for mergers motivated by internalization. This situation examined by Coase mirrors that analyzed in the present paper in that different types of player exert externalities of differing size on one another. The substantive difference is that the externality discussed by Coase is environmental rather than terms-of-trade.

The paper proceeds as follows. In the next section transport costs are introduced to a model of production and trade based on Cournot oligopolistic competition. This is then used to derive optimal tariffs which vary according to the distance between countries. Section 3 sets up the model of regions and trade agreements as a noncooperative network, allowing the payoffs of network formation to vary depending on the distance between members. Section 4 then establishes the main results of the paper for a simplified three region model. It is here that the possibility of regional trade agreements is demonstrated, as well as the fact that trade blocks can be stepping blocks to free trade. Section 5 concludes.

2. A Model of Optimal Tariffs where Distance Matters

The purpose of this section is to present a model of tariff setting which exhibits the property that distance between countries has an effect on the optimal level of protectionism. In particular, it will be shown that optimal tariffs are higher between close neighbors.

Brander and Spencer's (1984) profit shifting argument is used to motivate optimal tariffs. Tariffs shift profits from the foreign firm to the domestic consumer. With lower transport costs between close neighbors, more rents can be shifted through the use of tariffs. So unlike in conventional models, where each country sets a common tariff on all others, in the present model each country sets tariffs that vary, and are declining with distance.

2.1. Country Location in Regions

The set $\mathcal{N} = \{1, ..., n\}$ is the set of countries and it is finite. The number of countries is given by $|\mathcal{N}|$. The *regional structure* $\mathcal{P} = \{R_1, R_2, ..., R_m\}$ on $\mathcal{N}$ is a partition of the set of countries $\mathcal{N}$ into regions, where a *region* is a set $R_k \subseteq \mathcal{N}$: $R_i \cap R_j = \emptyset$ for $i \neq j$ and $\cup_{i=1}^{m} R_i = \mathcal{N}$. Each region is assumed to have the same number of countries in it; $|R_i| = r$, for all $R_i \in \mathcal{P}$, and $r > 1$. To avoid trivialities, there is more than one region; $|\mathcal{P}| > 1$.

To make the differences between intra-regional versus inter-regional trade concrete, suppose that each country $i \in \mathcal{N}$ can be located by the co-ordinates (x_i, y_i).[3] Therefore, the *distance* d_{ij} between any two countries i and j can then be measured by a (Euclidean) distance function.

In order to make precise the distinction between countries by region, assume $(x_i, y_i) = (x_j, y_j)$ for $i, j \in R_k$, $i \neq j$; all countries in the same region have the same location. Also assume that $(x_i, y_i) \neq (x_j, y_j)$ for all $i \in R_i$, $j \in R_j$, $i \neq j$. Assume that $d_{ij} = d_{ji} \geq d > 0$ for all $i \in R_i$, $j \in R_j$, $i \neq j$, and that d_{ij} is finite. (It is already immediate that $d_{ij} = d_{ji} = 0$ for $i, j \in R_k$.)

If the distance relationship between countries across regions has some regularity to it, being based on a regular shape for example, then it will help to be able to summarize the information on distances between countries. So, for country $i \in R_i$, let $D_i = \{(d_1, \delta_{1i}), ..., (d_z, \delta_{zi})\}$ be the set of pairs (d_k, δ_{ki}), where there exists at least one (other) country $j \in R_j$, $i \neq j$, for which $d_{ij} = d_k$, and δ_{ki} gives the number of countries at that distance from i.

[3] That is, each country can be located in Euclidean $\mathcal{R}_2$-space.

2.1.1. Example: Three Regions on an Equilateral Triangle

To keep the analysis relatively simple, the main results of the paper will be established using a three region model. Three regions are enough to capture the interactions we are interested in, whilst avoiding extensive notation. A three region model is obviously appealing because it captures the interactions between the three most important regions in economic terms, The Americas, Europe and Asia.

To fix ideas, consider the three region example ($|\mathcal{P}| = 3$), with $r = 3$, where each region is located at a distinct vertex of an equilateral triangle. Label the regions R_a, R_b and R_c. Consider country i located in region R_a. If the sides of the triangle are of length d, then countries in R_b and R_c are all at distance d from country i. Then the set D_i has a single element, $D_i = \{(d, 6)\}$, where $(|\mathcal{P}| - 1)\, r = 6$ gives the number of countries not in R_a.

2.2. Production and International Trade with Distance

Each country has a single firm. Firms compete in Cournot competition. They are able to segment international markets by country. For example, the firm in country j segments the markets of all countries $i \in \mathcal{N}$, choosing the quantity to produce (and export if $j \neq i$) in order to maximize profits π_{ij} in each. Firm j's problem is formalized in the usual way:

$$\text{Max}_{x_{ij}}\ \pi_{ij} = (p_{ij} - c_{ij})\, x_{ij} \tag{2.1}$$

where p_{ij}, c_{ij} and x_{ij} are the price, cost and quantity in country i of the good produced in country j. The cost to the firm in country j of producing a unit of output for sale in country i is given by the function

$$c_{ij} = c + t_{ij} + d_{ij}, \tag{2.2}$$

where c is the basic per-unit production cost, which is the same for all firms, t_{ij} is the tariff levied by country i on imports from country j. It will be assumed as usual that tariff revenue is transferred in lump-sum to consumers. Also, it must be assumed that transport costs are paid to an agent in the model. To keep the model simple let domestic

firms compete perfectly to bring goods to the home market, so that they deliver at cost.[4]

The basis for optimal tariff setting in a model of Cournot competition is familiar. But distance has not been introduced to a general model of tariff setting before, and this potentially makes the model intractable. However, optimal tariffs with transport costs can be solved for under the assumption that goods enter preferences independently. Let

$$u_i = e \sum_{j \in \mathcal{N}} x_{ij} - \frac{1}{2} \sum_{j \in \mathcal{N}} x_{ij}^2 + m_i, \tag{2.3}$$

where m_i is the numeraire. Countries are endowed with equal quantities of the numeraire, the role of which is to ensure that trade accounts are balanced in equilibrium.

Taking tariffs as given, firms choose their respective output levels simultaneously. They sell their output in country i at the market clearing price. The inverse demand curve of consumer i is obtained by differentiating (2.3) with respect to x_{ij}:

$$p_{ij} = \frac{du}{dx_{ij}} = e - x_{ij}, \tag{2.4}$$

Using (2.4) in (2.1), and expanding,

$$\pi_{ij} = e x_{ij} - x_{ij}^2 - c_{ij} x_{ij}.$$

The function π_{ij} is thus differentiable and strictly concave because $-x_{ij}^2$ is concave, and so firm j's problem has a unique maximum.

Firm j's first order condition in country i is thus given by

$$\frac{\partial \pi_{ij}}{\partial x_{ij}} = e - 2x_{ij} - c_{ij} = 0,$$

or equivalently,

$$p_{ij} - c_{ij} - x_{ij} = 0. \tag{2.5}$$

We can rearrange (2.5) to get

$$x_{ij} = p_{ij} - c_{ij}$$

[4] In a symmetric model, an equivalent assumption would be that the world market for transportation is competitive and that each firm from every country has an equal share of the market.

From this, note the following convenient property; in equilibrium, profits can be written as

$$\pi_{ij} = (p_{ij} - c_{ij})\, x_{ij} = x_{ij}^2. \tag{2.6}$$

The fact that profits can be represented in this way arises as a result of the linear structure of the model.

Use (2.2) and (2.4) in (2.5), then rearrange to get the solution for output by firm j for country i:

$$x_{ij} = \frac{e - c - t_{ij} - d_{ij}}{2}, \text{ all } i, j \in \mathcal{N} \tag{2.7}$$

Note from (2.7) that x_{ij} is decreasing in t_{ij} and d_{ij}. To maintain the assumption that all firms are active on all markets, $e-c$ can be made large enough to ensure that $x_{ij} > 0$. For the domestic market, $t_{ij} = d_{ij} = 0$. Therefore, the weakest possible condition necessary and sufficient to ensure strictly positive output by the domestic firm for the domestic market is $e - c > 0$. This condition will be assumed to hold throughout.

2.3. Production-Trade Payoffs

The payoffs to the FTA formation game depend directly on the structure of trading arrangements, that is tariff setting across all countries, and the reciprocal impact on production. For this reason, gains to production and trade will be referred to as production-trade payoffs. They are given this name to distinguish them from (net) payoffs to FTA formation once the cost of sponsoring agreements is taken into account.

The representative citizen in country i receives their trade-production payoff through five economic components: domestic consumer surplus (CS_i), the domestic firm's profit at home and abroad (π_{ii} and π_{ji}, $j \neq i$ respectively), tariff revenue (TR_i), and net profits from transportation (DR_i):

$$w_i = CS_i + \pi_{ii} + \sum_{j \in \mathcal{N}/\{i\}} \pi_{ji} + TR_i + DR_i. \tag{2.8}$$

The optimal tariff $\hat{t}_{ij}$ is derived by maximizing this expression with respect to t_{ij}. To do this, w must be expressed in terms of model variables; the subject of the next result.

Lemma 1. Let $CS_i = \sum_{j\in\mathcal{N}} \frac{1}{2}(p_{ij} - c_{ij})\, x_{ij}$, $TR_i = \sum_{j\in\mathcal{N}/\{i\}} t_{ij}x_{ij}$, and $DR_i = \sum_{j\in\mathcal{N}/\{i\}} d_{ij}x_{ij}$. Then

$$\begin{aligned} w_i &= CS_i + \pi_{ii} + \sum_{j\in\mathcal{N}/\{i\}} \pi_{ji} + TR_i + DR_i. \\ &= (e-c)\sum_{j\in\mathcal{N}} x_{ij} - \frac{1}{2}\sum_{j\in\mathcal{N}} x_{ij}^2 - \sum_{j\in\mathcal{N}/\{i\}} x_{ij}^2 + \sum_{j\in\mathcal{N}/\{i\}} x_{ji}^2. \end{aligned}$$

With payoffs of the representative citizen in country i as given by Lemma 1, it is straightforward to solve for the tariff that maximizes the representative citizen's payoff.

2.4. Optimal Tariffs with Distance

The solution to the optimal tariff problem is given in the following result:

Proposition 1. The unique optimal external tariff set by country i on imports from country j takes the form

$$\hat{t}_{ij} = \frac{e-c}{3} - d_{ij}.$$

The key thing to notice is that the optimal tariff is decreasing in distance. The closer a country is, the higher the optimal tariff levied on its imports. The intuition is simple. Higher rents are made in nearby markets because a smaller share of revenue is lost in transportation costs to serve those markets. Consequently, there are more rents available to shift to domestic consumers using the tariff.

As was the case in the solution for x_{ij} given by (2.7), it is always possible to set $e-c$ high enough to ensure that $\hat{t}_{ij} > 0$. If $\hat{t}_{ij} < 0$ then the optimal trade intervention is a subsidy. In that case, free trade is not necessarily welfare maximizing. In the present analysis we will be focusing on the standard case where free trade is best. For the optimal tariff to be positive, it is necessary and sufficient to make the following assumption:

A1. $0 < d_k \leq (e-c)/3$ for all $d_k \in D_i$, all $i \in \mathcal{N}$.

This will be assumed to hold throughout. Intuitively, one would generally expect trade agreements which entailed the removal of tariffs to bring about an improvement of welfare. The following result shows this intuition to hold in the present model.

Proposition 2. Assume A1. Assume that in the absence of a FTA between countries i and j then country i sets optimal tariffs on imports from country j - $t_{ij} = \hat{t}_{ij} = (e-c)/3 - d_{ij}$, and vice versa. Assume that if countries i and j have a FTA then country i adopts free trade on imports from country j - $t_{ij} = 0$ - and vice versa.

(i) The trade-production gain to two countries i and j from a bilateral FTA is given by

$$\Delta w_i = \frac{1}{72}\left(7\left(e-c\right)+3d_{ij}\right)\left(\left(e-c\right)-3d_{ij}\right)$$

and is positive.

(ii) Trade-production gains from a bilateral FTA are decreasing in the distance between members i and j: $d\left(\Delta w_i\right)/d\left(d_{ij}\right) = -\frac{1}{4}\left(e-c+d_{ij}\right)$.

This proposition tells us that trade agreements are good for welfare, and that a higher gain in welfare results from a larger tariff reduction. In the absence of a trade agreement, Proposition 1 shows that tariffs between regional members will be higher than between countries of different regions. Proposition 2 shows that FTA formation between regional members will yield higher production-trade gains than between countries of different regions.

If production-trade gains were all that mattered, then Proposition 2 suggests the world would move straight to free trade. Anecdotal discussions often reflect surprise that the process of regionalism has not led more quickly towards free trade. One explanation is that the costs of coordinating such agreements holds the process back. Sponsorship costs, formalized in the next section, play exactly this role in the present model. But by themselves such costs do not explain why FTAs are formed within a region. Proposition 2 indicates that regional agreements tend to form because they are worth more to their members than non-regional agreements. Equivalently, greater benefits to regional FTA formation can be set against the costs of sponsoring an agreement. This is the central insight that will be developed in the following sections.

3. Regions and Trade Agreements in a Non-cooperative Network

This section formalizes FTA formation in a world of regions. It shows how the formal language of network formation can be adapted to model FTA formation when there is a regional dimension to the model. The distinction that countries make between FTA members in their own region and those from other regions is formalized by an adaptation of Slikker and van den Nouweland's (2000) partitioning of players in a network formation game.

3.1. FTAs as Networks

The overall FTA structure is described by the *graph* $(\mathcal{N}, g)$, a pair of disjoint sets, where g is a set of links in a (directed) network. The FTA structure of each region is described by the *subgraph* (R_k, g_{R_k}), where g_{R_k} is a set of links called a *subnetwork* between countries of region R_k.

If country i sponsors a FTA with a set of other countries $\mathcal{A}_j = \{j_1, ..., j_l\}$ then it sets $g_{ij_k} = 1$ for all $j_k \in \mathcal{A}_j$. If $g_{ij_k} = 1$ (or if $g_{j_k i} = 1$) then there is said to be a *link* between i and j_k. A *strategy* of country $i \in \mathcal{N}$ is a row vector $g_i = (g_{i1}, ..., g_{in})$. The strategies of all countries forms $g = \{g_1, ..., g_n\}$. Since $\mathcal{N}$ can be partitioned into regions, we can also have $g = \{g_{R_1}, ..., g_{R_m}\}$, where g_{R_k} is the set of links between members of region R_k.

There is a *path* from i to j_k in g if i and j are linked, or if there exist countries $i_1, ..., i_m$ distinct from each other such that they are all linked. A path in g between i and j is denoted $i \overset{g}{\leftrightarrow} j$.[5]

A set $\mathcal{A}_k \subseteq \mathcal{N}$ is a *component* of g if for all i and j in $\mathcal{A}_k$ there is a path between them, and there does not exist a path between a country in $\mathcal{A}_k$ and one in $\mathcal{N} \backslash \mathcal{A}_k$.

A network g is called *connected* if it has a unique component $\mathcal{A}$, with all $i \in \mathcal{A} = \mathcal{N}$. A network that is not connected is referred to as *disconnected*. A network is called *empty* if $g_{ij} = g_{ji} = 0$ for all $i, j \in \mathcal{N}$.

[5] This notation emphasises that for i and j to be linked there can either be a link from i to j, or from j to i or both. This is sometimes referred to as a non-directed link.

Definition 1. (FTA Membership) A component $\mathcal{A}_k \subseteq \mathcal{N}$ of g is a FTA: for all $i \in \mathcal{A}_k$, if $j \in \mathcal{A}_k$ then $t_{ij} = 0$. If $j \notin \mathcal{A}_k$ then $t_{ij} = \hat{t}_{ij} = (e - c)/3 - d_{ij}$.

Definition 2. (World FTA) If the network g is connected, then there is a *world FTA*.

Definition 3. (Only-regional FTA, complete-regional FTA and extra-regional FTA) If there is a component for which all elements are in the same region, $i, j \in \mathcal{A}_k \subseteq R_k$, then $\mathcal{A}_k$ is an *only-regional FTA*; if the network g_k is connected, then we say there is a *complete-regional FTA*; if $\mathcal{A}_k \subsetneq R_k$ then $\mathcal{A}_k$ is an *extra-regional FTA*.

Definition 1 gives FTA membership a definition in graph notation, and a convenient graphical representation. Definition 2 then says that if all countries are in the same component then there is a world FTA. If, on the other hand, all members of a component are in the same region then the component is an only-regional FTA and if all countries in a region are in the same FTA then we say there is a complete-regional FTA. Finally, if FTA membership spans regions then there is said to be an extra-regional FTA. A country that has no links with other countries is said to be in its own *singleton component*.

3.2. The Sponsor of a FTA

For a trade agreement to come about, we will say that it must have a sponsor. A sponsor must pay a *sponsorship cost* for setting up a FTA. If *none of the countries are already in an FTA*, then the sponsorship cost is proportional to the number of proposed members. The function $\kappa_i()$ measures the sponsorship fee paid by country i for an agreement, and it is assumed to be linear in the number of FTA members. Formally, country i pays a cost $\kappa_i(1)$ for each link that it forms. So the sponsorship fee for an agreement with countries $\mathcal{A}_j = \{j_1, ..., j_l\}$ is $\kappa_i(|\mathcal{A}_j|)$. It is understood that the FTA is multilateral. So each $j_k \in \mathcal{A}_j$ adopts free trade with each other member of $\mathcal{A}_j$, not just with country i.

What happens if the set of countries $\mathcal{A}_j = \{j_1, ..., j_l\}$ with whom country i proposes to sponsor an FTA are *themselves already in an FTA?*. Then we will say that country i only has to pay the cost of a single link $\kappa_i(1)$. If country i is *itself already in a FTA*, entailing the set of countries $\mathcal{A}_k = \{i_1, .., i, .., i_l\}$, then we will say that country i sponsors

a FTA between $\mathcal{A}_j$ and $\mathcal{A}_k$ and that again the sponsorship cost is just $\kappa_i(1)$.[6]

Because a country pays a sponsorship fee for each agreement that it proposes, we will want a way of keeping track of the total amount that each country pays in sponsorship fees. To this end, define $\eta_i^{\sigma}(g) = |\{k \in \mathcal{N} \,|g_{ik} = 1\}|$ as the number of countries with which i maintains direct links. Then the total sponsorship cost paid by a country is given by $\kappa_i(\eta_i^{\sigma}(g))$ or $\eta_i^{\sigma}(g)\,\kappa_i(1)$.

As well as wishing to calculate the costs of FTA formation to each country, we will also want to calculate the benefits. Because in general these vary across regions, we will need to distinguish between the number of members in each. Define $\eta_i(g) = \left|\left\{i \in R_k, j \in R_k \,\middle|\, j \overset{g}{\leftrightarrow} i\right\} \cup \{i\}\right|$ $i \neq j$ as the number of countries in the same region as country i and on the same path. Recall that D_i contains the set of distinct distances d_k of other countries from country i. Define $\eta_i^{d_k}(g) = \left|\left\{i \in R_k, j \in R_j \,\middle|\, d_{ij} = d_k, j \overset{g}{\leftrightarrow} i\right\}\right|$,$i \neq j$, as the number of other countries $j \neq i$ at a distance $d_{ij} = d_k > 0$ on the same path as country i. Let $\left\{\eta_i(g), \eta_i^{d_1}(g), ..., \eta_i^{d_k}(g), ..., \eta_i^{d_z}(g)\right\} = H_i(g)$ be the complete set of *membership variables* $\eta_i^{d_k}(g)$.

In the three region model, where each region is assumed to be at the vertex of an equilateral triangle, we have a particularly simple representation. All countries not in the same region as country i are at the same distance away. So if $i \in R_k$ then $d_{ij} = d > 0$ for all $j \notin R_k$ and a single scalar which we can call $\eta_i^{d}(g)$ gives the total number of countries not in R_k with which country i is linked.

The process of agreement formation will be much easier to formalize if we know

[6] Care should be taken to justify these assumptions about sponsorship costs. At its simplest, when a country sponsors a bilateral agreement it pays for a set of border controls that verify the origin of goods from an FTA partner. (In the absence of an agreement no such system is needed because all goods carry duty.) Once goods are verifiable by one partner then they are assumed to be verifiable by all at no extra cost. Hence if country i sponsors an agreement with $|\mathcal{A}_j|$ other countries then the sponsorship cost is assumed to be $\kappa(|\mathcal{A}_j|)$ and not $\kappa\left(|\mathcal{A}_j|^{(|\mathcal{A}_j|+1)}\right)$. If country i sponsors an agreement with another FTA, then it only has to pay for the cost of making its own good verifiable at the border of the FTA to which it applies, so the cost is just $\kappa(1)$. If country i sponsors an agreement between its own FTA and another one, then it must pay to make its own goods verifiable to the new FTA. But since a standardised system is already operated within its own FTA, country i only has to pay to make the single standard verifiable by the new FTA, again at a cost of $\kappa(1)$. Finally, in a world of free trade, there will be no need for verification of origin because all goods will pass through borders duty free. But there will be need of a verification system to ensure that the system is not being cheated upon. Implicitly it is assumed that the costs of operating such a system are the same as the costs of verification, an admittedly bold but simplifying assumption.

that any agreement which some country i proposes to sponsor will be accepted by all the proposed partners. Then the proposal of an agreement will be synonymous with its formation. Under the assumptions of Proposition 2, the production-trade gains to a FTA are always positive. The following result uses this fact to establish that the proposes of a FTA will always accept it.

Proposition 3. Assume A1. If country i proposes to sponsor a FTA with a set of other countries $\mathcal{A}_j = \{j_1, ..., j_l\}$ then all countries in the set $\mathcal{A}_j$ would obtain positive production-trade payoffs from the proposed FTA and would therefore accept. This holds whether or not the set of countries $\mathcal{A}_j = \{j_1, ..., j_l\}$ are themselves already in an FTA and whether or not country i is itself already in an FTA.

The following result shows that the network structure gives rise to an ordinary coalition structure of the form $\mathcal{C} = \{\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_m\}$.

Lemma 2. An FTA structure is a partition of the set of countries $\mathcal{N}$: $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset$; $\cup_{i \in \mathcal{C}} \mathcal{A}_i = \mathcal{N}$.

Thus, an FTA structure is like the *cooperation structure* modelled by Myerson (1977). We have a conventional coalition structure rather than a network. But as we shall see, the network terminology of link formation is helpful because it enables us to model equilibrium very conveniently in the manner of a non-cooperative network.

As acknowledged in the introduction, this formalization of FTAs is highly stylized. Now that the model has been introduced, it should be clear that one of the simplifications made here could be dealt with by making the sponsorship of agreements more balanced between members, although the modelling framework would become significantly more complex to take this on board.

A second complication not mentioned in the introduction is that in the real world the network of agreements is much more complex, with partners of an FTA being members of other mutually exclusive FTAs, as shown diagrammatically by Bhagwati and Panagaraya's (199X) "spaghetti bowl". The simplified structure that results in the present model can essentially be traced to the assumption that sponsors negotiate on behalf of all their FTA partners to join other FTAs. Without it, countries i and j in a given FTA could each

sponsor agreements with two other separate FTAs, potentially resulting in much more complex equilibrium paths. However, this assumption reflects the widely held view that it is easier to negotiate between blocks than between individual countries. Indeed Article XXIV of the GATT charter, which has now been adopted by the WTO and sets out the rules on FTA formation, states exactly this rationale for allowing trade blocks. The analysis of this paper present will show how regional trade blocks can prevail even under the GATT/WTO's assumed modus operandi.

3.3. Payoffs to FTA Formation

The production-trade payoffs w_i, given by the function (2.8), will now be adapted for use as a payoff function in a FTA formation game. Let $i \in \mathcal{A}_i$.

Lemma 3. Let x_{ij} be given by (2.7). Then the function (2.8) can be expressed in the form

$$w_i = w\left(\eta_i(g), \eta_i^{d_1}(g), ..., \eta_i^{d_k}(g), ..., \eta_i^{d_z}(g); \boldsymbol{\gamma}\right),$$

or equivalently

$$w_i = w\left(H_i(g); \boldsymbol{\gamma}\right).$$

To gain greater insight into this result, look at the expanded form of the payoff function $w\left(H_i(g); \boldsymbol{\gamma}\right)$:

$$\begin{aligned} & w\left(\eta_i(g), \eta_i^{d_1}(g), ..., \eta_i^{d_z}(g), ...; \boldsymbol{\gamma}\right) \\ = \; & \eta_i(g)\left(\frac{3}{8}(e-c)^2\right) \\ & + \left(|\mathcal{N}| - |\mathcal{A}_i|\right)\left(\frac{5}{18}(e-c)^2\right) \\ & + \eta_i^{d_1}(g)\left(\frac{1}{8}\left(3(e-c)+d_1\right)(e-c-d_1)\right) \\ & \vdots \\ & + \eta_i^{d_z}(g)\left(\frac{1}{8}\left(3(e-c)+d_z\right)(e-c-d_z)\right). \end{aligned}$$

The expression is parametric except for the membership variables because tariffs have been substituted for, using either the optimal tariff formula (Proposition 1) where no agreement exists, or zero tariffs where an agreement exists.

The first line shows the payoff to production and trade with FTA members that are in the same region, the number of which is given by $\eta_i(g)$. No distance parameter appears on this line because countries in the same region are assumed to have the same location. The second line gives the payoff to production and trade with all countries *not* in country i's FTA. Notice that when optimal tariffs are in place, the volume of trade is exactly the same between all non-members of the FTA, regardless of their distance. The remaining lines measure the production-trade gains from FTA members at all distances $d_k \in D_i$.[7]

Having derived a convenient short-hand to write w_i in terms of regional FTA membership, it is now possible to evaluate the gains to country i from FTA formation with other regional members and countries from outside the region. The following result shows that as long as optimal tariffs are positive then an increase in membership of country i's FTA carries production-trade gains to country i. The result also shows that more production-trade gains are derived the closer are the new members.

Let $\Delta\eta_i(g)$ denote a unit increase in $\eta_i(g)$ *from any level* and let $\Delta\eta_i^{d_k}(g)$ denote a unit increase in $\eta_i^{d_k}(g)$ from any level. Let $\Delta w/\Delta\eta_i(g)$ and $\Delta w/\Delta\eta_i^{d_k}(g)$ measure the impact on $w(H_i(g);\boldsymbol{\gamma})$ of a unit increase in $\eta_i(g)$ and $\eta_i^{d_k}(g)$ respectively.

Lemma 4. Assume A1.

(i) The terms $\Delta w/\Delta\eta_i(g)$ and $\Delta w/\Delta\eta_i^{d_k}(g)$ are positive and constant, (where d_k is any element of the set D_i) and are independent of $\eta_i(g)$ and $\eta_i^{d_k}(g)$.

(ii) Let d_j be the smallest element of D_i, let $d_l \in D_i$ be the largest element, and let $d_k \in D_i$ be any other element such that $d_j < d_k < d_l$. Then

$$\Delta w/\Delta\eta_i(g) > \Delta w/\Delta\eta_i^{d_j}(g) > \Delta w/\Delta\eta_i^{d_k}(g) > \Delta w/\Delta\eta_i^{d_l}(g) \geq 0.$$

[7] To see why no distance parameters appear in the second line showing production-trade gains with non-members of the FTA, use the expression for the optimal tariff (Proposition 1) in the expression for output (2.7) and notice that the distance parameter cancels.

If optimal tariffs are removed and tariffs are set to zero then the distance parameters appear. This explains why the distance parameters do appear on the remaining lines.

Lemma 4 extends Proposition 2 from bilateral to multilateral agreements. Because there are no terms of trade externalities in the model, Lemma 4 essentially establishes that the production-trade payoff to forming multilateral FTAs is a multiple of the production-trade payoff to a bilateral agreement.

Assumption A1 ensures that optimal tariffs are positive. Part (i) of Lemma 4 establishes that $\Delta w/\Delta\eta_i(g)$ and $\Delta w/\Delta\eta_i^{d_k}(g)$ are constant and do not depend on the initial levels of $\eta_i(g)$ and $\eta_i^{d_k}(g)$. This is convenient because it means that the production-trade benefit of changes in FTA membership can be evaluated depending only on the distance between members. Consequently, the production-trade gains of *any* change in FTA membership, regional or non-regional, can be captured using the notation $\Delta w/\Delta\eta_i(g)$ and $\Delta w/\Delta\eta_i^{d_k}(g)$. For example, the effect of an increase in $\eta_i(g)$ from y^1 to y^2 is given by $(y^2-y^1)\,\Delta w/\Delta\eta_i(g)$.

Part (ii) of Lemma 4 shows that production-trade gains of trade block expansion are greater for closer countries. It is possible to understand why by looking at the payoff function $w\left(\eta_i(g),\eta_i^{d_1}(g),...,\eta_i^{d_z}(g),...;\boldsymbol{\gamma}\right)$. Notice that the production trade payoff from regional members, at $\frac{3}{8}(e-c)^2$, is greater than the production-trade payoff from countries where there is no agreement, at $\frac{5}{18}(e-c)^2$. The proof then shows that for any two countries joining an FTA, the production-trade payoff lies between these two levels. The intuition is straight forward. Because tariffs between closer countries are higher, their removal brings about a relatively large increase in production-trade gains. Formally, this follows from Proposition 1, which shows optimal tariffs to be declining in distance, and Proposition 2(ii), which shows in turn that the production-trade gains to a bilateral agreement are declining in distance. As long as signing a FTA entails removal of positive tariffs, then there must be a gain from forming an agreement. But this is ensured by assumption A1, which guarantees that optimal tariffs are positive.

3.3.1. Three Regions on an Equilateral Triangle Again

As mentioned above, when using the three region model only two parameters are needed to describe FTA membership from the point of view of country i. These parameters are

$\eta_i(g)$ and $\eta_i^d(g)$, which give regional and non-regional membership respectively. Nothing more complicated than the 3-region model, based on an equilateral triangle, is needed to motivate the tendency to form regional agreements, and the main results will be based on this simplified special case. The model will be developed at a full level of generality in the next version of this paper.

3.4. FTA Formation with Sponsorship Costs

Now that the production-trade payoffs of FTA formation have been determined, these can be used in the payoffs of a noncooperative network formation game.

The parameters in the vector $\boldsymbol{\gamma}$ are held constant throughout, so from now on the function $w(H_i(g);\boldsymbol{\gamma})$ will be written $w(H_i(g))$. Using this brief form for production trade payoffs, define each country's *overall payoff function* $\Psi_i : \mathcal{G} \to \mathcal{R}$ as follows:

$$\begin{aligned} \Psi_i(g) &= \psi(H_i(g), \eta_i^{\sigma}(g)) \\ &= w(H_i(g)) - \kappa_i(\eta_i^{\sigma}(g)) \end{aligned} \tag{3.1}$$

When optimal tariffs are positive $w()$ is increasing in $\eta_i(g)$, and weakly increasing in $\eta_i^{d_k}(g) \in H_i(g)$. So for the purposes of the analysis $(H_i(g), \eta_i^{\sigma}(g))$ is increasing in $\eta_i(g)$ and weakly increasing in $\eta_i^{d_k}(g)$. The function $\kappa_i()$ is a linear function of $\eta_i^{\sigma}(g)$, the scalar measuring the number of FTAs sponsored by country i. Sponsorship costs are invariant to the distance between members.[8] Hence the overall payoff to the network formation game is given by the balance between the production-trade payoffs to an agreement and its sponsorship costs.

3.5. FTAs in Equilibrium as Nash Networks

With payoffs to the network formation game now specified, we can define the notion of equilibrium, which will be that of a Nash network. Given a network $g \in \mathcal{G}$, let g_{-i} denote the network obtained when all of country i's links are removed. Then the network g can be written as $g = g_i \oplus g_{-i}.$, where $\oplus$ denotes that g is formed as the union of the links in

[8] Asymmetries in the value of links across players have also been considered in network formation models by Myerson (1980) and Slikker and van den Nouweland (2000), but in a cooperative network framework.

g_i and g_{-i}. The strategy g_i is a *best response* of country i to g_{-i} if there does not exist a strategy g' for which

$$\Psi_i\left(g_i' \oplus g_{-i}\right) \geq \Psi_i\left(g_i \oplus g_{-i}\right) \text{ for all } g_i' \in \mathcal{G}_i.$$

The set of all country i's best responses to g_{-i} is denoted $BR_i\left(g_{-i}\right)$. A network $g = (g_1, ..., g_n)$ is a *Nash network* if $g_i \in BR_i\left(g_{-i}\right)$ for each i.

This definition of equilibrium was introduced by Bala and Goyal (2000). It is a straightforward application of the standard notion of Nash equilibrium to a noncooperative network setting. A network is in a state of equilibrium if none of the agents, countries in the setting of this present paper, has an incentive to deviate. In the present setting, deviation would entail a country breaking a link by withdrawing its sponsorship of a FTA.

Bala and Goyal show how this notion of equilibrium can be used in a dynamic setting. Here in this present paper the focus will be exclusively on a dynamic equilibrium path where a Nash network exists in each period. The dynamic process is very simple. It is initialized with the empty network. Countries are assumed to make naive best responses, taking as given the network in the previous period. Initializing with the empty network g in the period $t = 0$, along the *equilibrium path*, a Nash network exists in every period $t = 1, 2, ..., \infty$.

4. The 3-Region FTA in Equilibrium

This section uses the simple 3-region model to present the main results of the paper. Nothing more complex than the 3-region model is needed to show why trade blocks may be regional. So let $|\mathcal{C}| = 3$. Countries are located at the vertices of an equilateral triangle. For all countries not sharing the same region, $i \in R_i, j \in R_j$, $i \neq j$, let the distance between them be given by the same parameter $d_{ij} = d > 0$. Then the variable $\eta_i^d\left(g\right)$ measures the total number of non-regional members in country i's FTA. (The variable $\eta_i\left(g\right)$ gives the number of regional members as before.)

4.1. Production-Trade Payoffs, Sponsorship Costs and Overall Payoffs in a Network Game

One of the main advantages of using a 3-region model is that it keeps the overall payoff function as simple as possible. The overall payoff function takes the form

$$\psi\left(H_i\left(g\right),\eta_i^{\sigma}\left(g\right)\right)=\psi\left(\eta_i\left(g\right),\eta_i^{d}\left(g\right),\eta_i^{\sigma}\left(g\right)\right).$$

Equilibrium analysis will centre on showing network configurations from which there is no incentive to deviate, given assumptions about the relationship between production-trade benefits to network formation and sponsorship costs. So we will want a method of examining the change in payoffs to all possible strategic alternatives open to a country.

To develop such a method, let $\Delta\eta_i^{\sigma}\left(g\right)$ denote a unit increase of $\eta_i^{\sigma}\left(g\right)$. Then the change in the overall payoff to the sponsorship of any given agreement can be evaluated. For example, suppose that country i has already sponsored agreements with z countries. To be clear, formally country i has sponsored z links with other countries. And through these agreements country i is in a FTA with y^1 other regional countries and y_d^1 countries outside the region. Then the payoff to the sponsorship of an additional agreement, which will enlarge the FTA to include $y^2>y^1$ countries from the region and $y_d^2>y_d^1$ from outside the region is given by

$$\begin{aligned}&\psi\left(y^2,y_d^2,z+\Delta\eta_i^{\sigma}\left(g\right)\right)-\psi\left(y^1,y_d^1,z\right)\\=\;&\left(y^2-y^1\right)\Delta w/\Delta\eta_i\left(g\right)+\left(y_d^2-y_d^1\right)\Delta w/\Delta\eta_i^{d}\left(g\right)-\kappa_i\left(\Delta\eta_i^{\sigma}\left(g\right)\right).\end{aligned}$$

The left hand side takes the difference between overall payoffs under the two network structures. The first term on the right hand side shows the production-trade gains to an increase in regional members of the FTA. The second term shows the production-trade gains to an increase in non-regional members. The third term shows the sponsorship costs of setting up the additional agreement. Taken together, these terms show how the production-trade gains balance against the sponsorship costs of an agreement.

It begins to become clear when payoffs are presented in this way that the sponsorship of some agreements will more than compensate for the sponsorship costs whilst others will not. Recall from Lemma 4 that the production-trade payoffs to a regional FTA are higher than to a non-regional FTA. From this, it is easy to imagine sponsorship costs at a level

where regional agreements of a given size are worthwhile but non-regional agreements are not. The following result formalizes this idea by looking at sponsorship costs across a range of levels and their implications for the incentive to sponsor FTAs, regional and non-regional. It is important to keep in mind when looking at this result, however, that it evaluates the incentive to form a FTA where none are pre-existing. As we shall see later, the incentive to form new FTAs from existing ones are greater than the incentive to get FTAs off the ground in the first place. A *complete-regional* FTA is one which includes all member so the region but no other countries; $\mathcal{A}_k = R_k$ for all $\mathcal{A}_k \in \mathcal{C}$, $R_k \in \mathcal{P}$. A *world FTA* is one for which there is a single component that contains all countries; $\mathcal{A} = \mathcal{N}$.

Lemma 5. Assume A1. Assume that in period $t = 0$ the network g is empty.

(i) Let the production-trade payoff to a bilateral agreement with a country *in the same region* be *lower* than the sponsorship cost. If there are no existing FTAs then *no FTA* is worth sponsoring. Formally, $\Delta w/\Delta\eta_i(g) < \kappa(1) \Rightarrow \psi\left(y_1, y_1^d, \left(y_1 + y_1^d - 1\right)\right) < \psi(1,0,0)$ for $1 < r \leq y^1$, $0 \leq y_d^1 \leq n - r$.

(ii) Let the production-trade payoff to a bilateral agreement with a country *in a different region* be *higher* than the sponsorship cost. Even if there are no FTAs, then the payoff to sponsorship of a *world FTA* is higher than the payoff to sponsorship of any other FTA. Formally, $\Delta w/\Delta\eta_i^d(g) > \kappa(1) \Rightarrow \psi(r, n-r, (n-1)) \geq \psi\left(y_1, y_1^d, \left(y_1 + y_1^d - 1\right)\right)$, for $r \geq y^1 \geq 1$, $n - r \geq y_d^1 \geq 0$, holding with strict inequality if and only if $y^1 < r$ and/or $y_d^1 < n - r$.

(iii) Let the production-trade payoff to a bilateral agreement with a country *in the same region* be *higher* than the sponsorship cost. But let the production-trade payoff to a bilateral agreement with a country *in a different region* be *lower* than (or equal to) than the sponsorship cost. If there are no existing FTAs then sponsorship of a *complete-regional FTA* yields a higher payoff than sponsorship of any other agreement. Formally, $\Delta w/\Delta\eta_i^d(g) < \kappa(1) < \Delta w/\Delta\eta_i(g) \Rightarrow \psi(r, 0, r-1) > \psi(r, n-r, n-1)$ and $\psi(r, 0, r-1) > \psi\left(y_1, y_1^d, \left(y_1 + y_1^d - 1\right)\right)$, for $r > y^1 \geq 1$, $n - r \geq y_d^1 \geq 1$.

In a situation where there are no FTAs already existing, Lemma 5 shows the FTA structure that will yield the highest payoff from sponsorship. If the production-trade payoffs of a bilateral agreement are lower than the sponsorship cost even for an agreement

with regional neighbors, then no country has an incentive to sponsor a FTA. There will then certainly exist no incentive to sponsor a FTA with countries outside the region, because this entails the removal of lower tariffs, and therefore smaller production-trade gains. Conversely, if the production-trade payoffs are higher than the sponsorship cost of a FTA with a country in another region then a FTA with close regional neighbors will certainly be worth sponsoring.

It is when costs are at an intermediate level that the incentives show scope for regionalism. The highest payoff is yielded when a country sponsors a FTA that consists only of its regional neighbors and not more distant nations. In part (iii) of Lemma 5 it is assumed that the costs of production-trade payoffs of sponsoring a FTA with a country in the same region are above the sponsorship costs. Therefore, it is immediately clear that it will be worth sponsoring a FTA with regional neighbors. But sponsorship costs are above the production-trade payoffs of a FTA with countries outside the region. So from a situation where a country did sponsor an extra-regional FTA, it would gain more from withdrawing its sponsorship of an agreement with those more distant nations than from the production-trade gains of maintaining it. In this situation, the sponsorship costs lie between the relatively large gains from removing higher mutual tariffs with close neighbors and the smaller gains from removing lower mutual tariffs with countries that are further away.

Lemma 5 focuses exclusively on the payoffs to a country when it is the sole sponsor of a FTA. It will become clear in the analysis of equilibrium that the analysis reaches much further than just this restrictive situation. It will be shown that in equilibrium any given FTA can only have one sponsor. But to make analysis of the equilibrium path easier, it will be helpful to look at how the incentives to sponsor a world FTA change when starting not from a situation where there are no FTAs but from one where there are complete-regional FTAs already in existence. The incentives to sponsor an extra-regional FTA, given that a complete-regional FTA already exists, are analyzed in the next result.

Lemma 6. Assume A1.

Let the production-trade payoff to a bilateral agreement with a *single* country *in a different region* be *lower* than the sponsorship cost. Assume that a complete-regional

agreement exists in every region $R_k \in \mathcal{P}$. An extra-regional FTA (formed with a single link) is worth sponsoring if the production-trade payoffs to a FTA with *more than one* country *in a different region* are *higher* than the sponsorship cost. (If no FTA already exists in another region R_j then an extra-regional agreement is not worth sponsoring.)

Formally, $y_1^d \Delta w/\Delta\eta_i^d(g) > \kappa(1) > \Delta w/\Delta\eta_i^d(g) \Rightarrow \psi\left(y_1, ay_1^d, y_1 + a - 1\right) > \psi(y_1, 0, y_1 - 1)$, for $r \geq y^1 \geq 1$, $r \geq y^1 > 1$, $a \geq 1$

If there are enough other countries from another region already in a FTA then the production-trade benefits may overcome the sponsorship costs, even though these costs are too high to make an agreement with a single other country in that region worthwhile. The last part of the Lemma, shown in brackets, is a re-statement of Lemma 5(ii), to emphasize the contrasting outcomes depending on whether or not a FTA exists in the other region.

4.2. Equilibrium Paths; Are FTAs Stepping Blocks or Stumbling Blocks?

This subsection takes its title from the famous question posed by Bhagwati (1992). In the way that it will be answered below, the question should in fact be posed as 'When are FTAs stepping stones and when are they stumbling blocks in the path to free trade?' As argued in the introduction of this present paper, trade blocks in the real world are regional. In this light, the question is whether the regional blocks presently existing will ultimately promote world free trade.

The term regionalism usually describes a situation where countries in a region form a club or agreement, but where membership does not extend beyond regional boundaries. For a corresponding analytical definition that will be useful in the present context, let *regionalism* be a situation where all regions have a complete FTA but where there are no extra-regional FTAs; in the network g there is a connected subnetwork g_k for each $R_k \in \mathcal{C}$, but $g_{ij} = g_{ji} = 0$ for all $i \in R_i$, $j \in R_j$. The next proposition presents the main result of the paper.

Proposition 4. Assume A1.

(i) If the production-trade payoff to a bilateral agreement with a country *in the same*

region is *lower* than the sponsorship cost then on the equilibrium path no FTA will exist at any point in time.

(ii) If the production-trade payoff to a bilateral agreement with a country *in a different region* is *higher* than the sponsorship cost then on the equilibrium path there is world free trade at every point in time.

(iii) On the equilibrium path there is regionalism in the first period if the following conditions hold:

(a) production-trade payoff to a bilateral agreement with a country *in the same region* is *higher* than the sponsorship cost

(b) production-trade payoff to a bilateral agreement with a country *in a different region* is *lower* than the sponsorship cost.

(iv) (Regional trade blocks are stepping blocks to free trade) On the equilibrium path there is regionalism in the first period followed by world free trade from the second period onwards if the following conditions hold:

(a) the production-trade payoff to an extra-regional FTA with *all* countries *in a different region* is *higher* than the sponsorship cost;

(b) the production-trade payoff to a bilateral agreement with *a single* country *in a different region* is *lower* than the sponsorship cost.

(c) the production-trade payoff to a bilateral agreement with *a single* country *in the same region* is *higher* than the sponsorship cost.

If the production-trade payoff to an extra-regional FTA with *all* countries *in a different region* is *lower* than the sponsorship cost then on the equilibrium path there is regionalism at every point in time.

Proposition 4 shows that the equilibrium path to free trade may indeed exhibit a period of regionalism (Proposition 4(iii)&(iv)), presenting the possibility of an encouraging answer to Bhagwati's question. This outcome depends on sponsorship costs being high enough so that extra-regional agreements are not worth sponsoring between individual nations, but are worth sponsoring between existing FTAs. With costs slightly higher,

even an agreement between regional FTAs is not worth sponsoring and the FTA structure stalls at regionalism (Proposition 4(iii)&(iv)). Obviously, if the sponsorship costs are prohibitive of even a FTA between close regional neighbors then none will be sponsored at all (Proposition 4(i)). On the other hand, with sponsorship costs that are sufficiently low there will be a move straight to free trade, bypassing regionalism altogether (Proposition 4(ii)).

5. Conclusions

The main purpose of this paper has been to show that regionalism can arise in equilibrium. That is, countries may choose to form regional trade agreements rather than move all the way to free trade. Depending on the level of costs associated with sponsoring a FTA, regionalism may be a temporary phenomenon or it may lead on to free trade. Less interestingly, when costs are high then no FTAs will form and when costs are low there will be world free trade.

The analysis presented here appears to present a fairly optimistic picture for the future of trade liberalization through regional trade block formation. One significant caveat should be noted to this conclusion. Terms-of-trade benefits to expanding the *relative* size of an agreement have been suppressed in the present analysis by the assumption that goods enter preferences independently. In the analysis of Furusawa and Konishi (2002), customs union formation can prevent free trade because it is not in the interests of a large powerful block of countries that obtain relative terms-of-trade gains through the size of their agreement.[9] It appears that this type of terms-of-trade effect could overthrow the prediction of free trade as an outcome in the present analysis.

There are a number of extensions to this work that suggest themselves immediately. One straightforward extension to appear in the next version of this paper is to present the results of this paper for any number of regions. A more substantive extension would be to loosen the admittedly stringent assumptions concerning sponsorship costs of FTAs

[9] Customs union formation entails the joint maximisation of welfare. Note that when goods enter preferences independently then customs union formation is no different from FTA formation. A country can fully internalise the gains through tariff setting because it exports a good to the world market which has no substitute.

to enable an agreement to be sponsored by more than one country. Another would be to generalize the utility function that gives the production-trade payoffs to FTA formation. Also, if some way of allowing terms-of-trade effects to enter the analysis could be found then this might enable the conclusions of the present paper, where low costs of trade block formation lead to free trade, to be overturned.

References

[1] Bhagwati, J. (1992); "Regionalism versus Multilateralism." *The World Economy*, 15: 535-555.

[2] Bhagwati, J. D. Greenaway and A. Panagaraya (1998) "Trading Preferentially: Theory and Policy." *Economic Journal* 108(449): 1128-48 .

[3] Bala, V. and S. Goyal (2000); "A Noncooperative Model of Network Formation." *Econometrica*, 68(5): 1181-1229.

[4] Bond, E. (1999) "Multilateralism, Regionalism, and the Sustatainability of 'Natural' Trading Blocs." Penn State University mimeograph.

[5] Brander, J. and B. Spencer (1984); "Tariff Protection and Imperfect competition." chapter in H. Kierkowski (ed.) *Monopolistic Competition and International Trade*, Oxford University Press, Oxford.

[6] Coase, R.H. (1960); "The Problem of Social Costs", *Journal of Law and Economics*.

[7] Frankel, J., E. Stein and S. Wei (1995); "Trading Blocs and the Americas: The Natural, the Unnatural and the Super-natural." *Journal of Development Economics*, 47: 61-95.

[8] Furusawa, T. and H. Konishi (2002) "Free Trade Networks." Yokohoma National University and Boston College mimeograph.

[9] Goyal and Joshi (2000) "Bilateralism and Free Trade." Erasmus University and George Washington University mimeograph.

[10] Jackson, J. (1989); *The World Trading System*. MIT Press, Cambridge Massachusetts, 417 pages.

[11] Krugman, P. (1991); "Is Bilateralism Bad?" in E. Helpman and A. Razin, (eds.), *International Trade and Trade Policy*, MIT Press, Cambridge.

[12] Myerson, R. (1977); "Graphs and Cooperation in Games." *Mathematics of Operations Research* 2: 225-229.

[13] Slikker, M. and A. van den Nouweland (2000); "Communication Situations with Asymmetric Players." *Mathematical Methods of Operations Research*, 52: 39-56.

A. Appendix

Proof of Lemma 1. First rearrange the expression for CS_i as follows:

$$\begin{aligned} CS_i &= \sum_{j\in\mathcal{N}} \frac{1}{2}\left(p_{ij} - c_{ij}\right) x_{ij} \\ &= \frac{1}{2}\sum_{j\in\mathcal{N}} \left(e - x_{ij} - c_{ij}\right) x_{ij} \\ &= \frac{1}{2}\sum_{j\in\mathcal{N}} \left(e - c_{ij}\right) x_{ij} - \frac{1}{2}\sum_{j\in\mathcal{N}} x_{ij}^2 \\ &= \frac{1}{2}\sum_{j\in\mathcal{N}} \left(e - c - t_{ij} - d_{ij}\right) x_{ij} - \frac{1}{2}\sum_{j\in\mathcal{N}} x_{ij}^2 \end{aligned}$$

where the second line follows by (2.4), and the fourth line follows by (2.2).

Using this, the expressions for TR_i, DR_i, and (2.6) in (2.8) yields

$$\begin{aligned} w &= \frac{1}{2}\sum_{j\in\mathcal{N}}(e-c-t_{ij}-d_{ij})\,x_{ij} - \frac{1}{2}\sum_{j\in\mathcal{N}}x_{ij}^2 \\ &\quad +x_{ii}^2 + \sum_{j\in\mathcal{N}}t_{ij}x_{ij} + \sum_{j\in\mathcal{N}}d_{ij}x_{ij} + \sum_{j\in\mathcal{N}/\{i\}}x_{ji}^2 \\ &= \frac{1}{2}\sum_{j\in\mathcal{N}}(e-c)\,x_{ij} - \frac{1}{2}\sum_{j\in\mathcal{N}}t_{ij}x_{ij} - \frac{1}{2}\sum_{j\in\mathcal{N}}d_{ij}x_{ij} - \frac{1}{2}\sum_{j\in\mathcal{N}}x_{ij}^2 \\ &\quad +x_{ii}^2 + \sum_{j\in\mathcal{N}}t_{ij}x_{ij} + \sum_{j\in\mathcal{N}}d_{ij}x_{ij} + \sum_{j\in\mathcal{N}/\{i\}}x_{ji}^2 \\ &= \frac{1}{2}\sum_{j\in\mathcal{N}}(e-c)\,x_{ij} - \frac{1}{2}\sum_{j\in\mathcal{N}}t_{ij}x_{ij} - \frac{1}{2}\sum_{j\in\mathcal{N}}d_{ij}x_{ij} - \frac{1}{2}\sum_{j\in\mathcal{N}}x_{ij}^2 \\ &\quad +x_{ii}^2 + \sum_{j\in\mathcal{N}}t_{ij}x_{ij} + \sum_{j\in\mathcal{N}}d_{ij}x_{ij} + \sum_{j\in\mathcal{N}/\{i\}}x_{ji}^2 \end{aligned}$$

Now rearranging terms,

$$\begin{aligned} w &= \sum_{j\in\mathcal{N}}(e-c)\,x_{ij} - \frac{1}{2}\sum_{j\in\mathcal{N}}(e-c-t_{ij}-d_{ij})\,x_{ij} - \frac{1}{2}\sum_{j\in\mathcal{N}}x_{ij}^2 \\ &\quad +x_{ii}^2 + \sum_{j\in\mathcal{N}/\{i\}}x_{ji}^2 \\ &= \sum_{j\in\mathcal{N}}(e-c)\,x_{ij} - \frac{1}{2}\sum_{j\in\mathcal{N}}(e-c_{ij})\,x_{ij} - \frac{1}{2}\sum_{j\in\mathcal{N}}x_{ij}^2 + x_{ii}^2 + \sum_{j\in\mathcal{N}/\{i\}}x_{ji}^2 \\ &= \sum_{j\in\mathcal{N}}(e-c)\,x_{ij} - \frac{1}{2}\sum_{j\in\mathcal{N}}(p_{ij}-c_{ij})\,x_{ij} - \sum_{j\in\mathcal{N}}x_{ij}^2 + x_{ii}^2 + \sum_{j\in\mathcal{N}/\{i\}}x_{ji}^2 \\ &= (e-c)\sum_{j\in\mathcal{N}}x_{ij} - \frac{1}{2}\sum_{j\in\mathcal{N}}x_{ij}^2 - \sum_{j\in\mathcal{N}/\{i\}}x_{ij}^2 + \sum_{j\in\mathcal{N}/\{i\}}x_{ji}^2. \end{aligned}$$

where the third line uses (2.2). □

Proof of Proposition 1. Using the expression for w obtained in Lemma 1, the government of country i solves the following problem to set the optimal tariff on imports from country j;

$$\max_{t_{ij}} w = (e-c)\sum_{j\in\mathcal{N}}x_{ij} - \frac{3}{2}\sum_{j\in\mathcal{N}}x_{ij}^2 + \sum_{j\in\mathcal{N}}x_{ji}^2.$$

Because, by (2.7), x_{ji} is not a function of t_{ij} (the tariff set by country i does not affect production in other countries), the derivatives with respect to t_{ij} of all the terms under

the last summation are equal to zero. Given that $-(x_{ij})^2$ is concave, and by (2.7) x_{ij} is a linear function of t_{ij}, the objective function w is concave in t_{ij}, all $j \in \mathcal{N}$. So there must exist a unique solution for $\hat{t}_{ij}$. Write the first order condition as

$$\frac{dw}{dt_{ij}} = (e-c)\frac{dx_{ij}}{dt_{ij}} - 3\sum_{j\in\mathcal{N}}(x_{ij})\frac{dx_{ij}}{dt_{ij}} = 0.$$

Then, using the fact that $dx_{ij}/dt_{ij} = -1/2$, simplifying and rearranging obtains $\hat{t}_{ij} = (e-c)/3 - d_{ij}$. It is immediate that if $d < (e-c)/3$ then $\hat{t}_{ij} > 0$. □

Proof of Proposition 2. (i) Country i's production-trade payoff is given by

$$w_i = (e-c)\sum_{j\in\mathcal{N}} x_{ij} - \frac{1}{2}\sum_{j\in\mathcal{N}} x_{ij}^2 - \sum_{j\in\mathcal{N}/\{i\}} x_{ij}^2 + \sum_{j\in\mathcal{N}/\{i\}} x_{ji}^2.$$

By (2.7), if $d_{ij} = d_{ji}$ and $t_{ij} = t_{ji}$ then $x_{ij} = x_{ji}$ and the last two terms cancel, leaving

$$w_i = (e-c)\sum_{j\in\mathcal{N}} x_{ij} - \frac{1}{2}\sum_{j\in\mathcal{N}} x_{ij}^2.$$

Using (2.7),

$$w_i = \sum_{j\in\mathcal{N}} \frac{1}{8}\left(3(e-c) + t_{ij} + d_{ij}\right)\left(e - c - t_{ij} - d_{ij}\right).$$

Now let countries i and j form a FTA. Using $t_{ij} = \hat{t}_{ij}$ for the pre-agreement tariff, and $t_{ij} = 0$ for the post agreement tariff in w_i, take discrete differences to work out the welfare gain:

$$\Delta w_i = \frac{1}{72}\left(7(e-c) + 3d_{ij}\right)\left((e-c) - 3d_{ij}\right).$$

Under the assumptions that $e - c > 0$, and $t_{ij} = \hat{t}_{ij} = (e-c)/3 - d_{ij} > 0$, so $\Delta w_i > 0$.

(ii) Immediate by differentiation. □

Proof of Proposition 3. By Definition 1, all countries in $\mathcal{J}$ set t_{ij} with all other countries in j and with the sponsor, country i. By Proposition 1, each country in j gains $\Delta w_i > 0$ for each other country in the agreement, and pays no sponsorship cost. Therefore, it is in the interest of each country in the set $\mathcal{J}$ to join the FTA proposed by i. □

Proof of Lemma 2. Suppose not. $\cup_{i\in\mathcal{C}}\mathcal{A}_i = \mathcal{N}$ is trivial as unilateralists are singleton components. To see that $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset$, suppose not. Suppose that $i \in \mathcal{A}_i$ and $i \in \mathcal{A}_j$. This may be the case for one of the following reasons. Either i proposed to sponsor $\mathcal{A}_i$ and $\mathcal{A}_j$. But in that case all members of $\mathcal{A}_i \setminus \{i\}$ and $\mathcal{A}_j \setminus \{i\}$ must be in the same FTA; a contradiction. Or i was already in one agreement, assume $\mathcal{A}_i$, and proposed to sponsor an agreement with the members of $\mathcal{A}_j$. But then, by assumption, if i's proposal were accepted all members of $\mathcal{A}_i$ must have joined $\mathcal{A}_j$ at the same time; a contradiction. Finally, suppose that i was already in one agreement, assume $\mathcal{A}_i$, but came into $\mathcal{A}_j$ as a result of an agreement proposed by another country. But then if i entered $\mathcal{A}_j$ in this way, then so must all $\mathcal{A}_i \setminus \{i\}$; a contradiction. □

Proof of Lemma 3. For convenience, define the following piece of notation. Let $\overline{g}_{ij} = \max\{g_{ij}, g_{ji}\}$. Note that, by (2.7), $t_{ij} = t_{ji}$ and $d_{ij} = d_{ji}$, it is the case that $x_{ij} = x_{ji}$ for all $i, j \in \mathcal{N}$ (independent of whether $g_{ij} = 0$ or $g_{ij} = 1$). Consequently, the function w can be written in the form

$$w_i = (e - c)\sum_{j\in\mathcal{N}} x_{ij} - \frac{1}{2}\sum_{j\in\mathcal{N}} x_{ij}^2.$$

Let $x_{ij}^{d_k}\left(\overline{g}_{ij}\right)$ represent (2.7) where the superscript d_k denotes that country j is at distance $d_k > 0$ from country i; $i \in R_i$, $j \in R_j$, $i \neq j$. We substitute for (2.7) explicitly in the step after this. But to see how the structure of the new function arises, it is helpful to note the following intermediate step. Recall that $g_{ij} \in \{0, 1\}$, where t_{ij} is set optimally according to Proposition 1 if $\overline{g}_{ij} = 0$ and free trade is adopted if and only if $\overline{g}_{ij} = 1$. As

$x_{ij}^{d_k}\left(\overline{g}_{ij}\right)$ depends only on $\overline{g}_{ij}$ and d_k, the function w_i can be partitioned accordingly:

$$\begin{aligned}
& w_i = w\left(\eta_i(g), \eta_i^{d_1}(g), ..., \eta_i^{d_k}(g), ..., \eta_i^{d_z}(g); \boldsymbol{\gamma}\right) \\
= \; & \eta_i(g)\left((e-c)\, x_{ij}(1) - \frac{1}{2} x_{ij}(1)^2\right) \\
& + (r - \eta_i(g))\left((e-c)\, x_{ij}(0) - \frac{1}{2} x_{ij}(0)^2\right) \\
& + \eta_i^{d_1}(g)\left((e-c)\, x_{ij}^{d_1}(1) - \frac{1}{2} x_{ij}^{d_1}(1)^2\right) \\
& + \left(\delta_{1i} - \eta_i^{d_1}(g)\right)\left((e-c)\, x_{ij}^{d_1}(0) - \frac{1}{2} x_{ij}^{d_1}(0)^2\right) \\
& \vdots \\
& + \eta_i^{d_z}(g)\left((e-c)\, x_{ij}^{d_z}(1) - \frac{1}{2} x_{ij}^{d_z}(1)^2\right) \\
& + \left(\delta_{zi} - \eta_i^{d_z}(g)\right)\left((e-c)\, x_{ij}^{d_z}(0) - \frac{1}{2} x_{ij}^{d_z}(0)^2\right)
\end{aligned}$$

where the absence of a superscript in the terms $x_{ij}(1)$ and $x_{ij}(0)$ denotes that $i, j \in R_k$.

Now substitute explicitly for (2.7). First note that if $i, j \in R_k$ then $d_{ij} = 0$. Also, if $g_{ij} = 1$ then $t_{ij} = 0$ and, by (2.7), $x_{ij}(1) = (e-c)/2$. If $\bar{g}_{ij} = 0$ then $t_{ij} = (e-c)/3$ and so $x_{ij}(0) = (e-c)/3$. Analogously, by (2.7), $x_{ij}^{d_k}(1) = (e-c-d_k)/2$. And, by Proposition 1, use $\hat{t}_{ij} = \frac{e-c}{3} - d_k$ and (2.7) to obtain $x_{ij}^{d_k}(0) = (e-c)/3$. Making these substitutions, we can rewrite the function $w\left(\eta_i(g), \eta_i^{d_1}(g), ..., \eta_i^{d_z}(g), ...; \boldsymbol{\gamma}\right)$ as

$$\begin{aligned}
& w\left(\eta_i(g), \eta_i^{d_1}(g), ..., \eta_i^{d_z}(g), ...; \boldsymbol{\gamma}\right) \\
= \; & \eta_i(g)\left(\frac{3}{8}(e-c)^2\right) \\
& + (r - \eta_i(g))\left(\frac{5}{18}(e-c)^2\right) \\
& + \eta_i^{d_1}(g)\left(\frac{1}{8}(3(e-c) + d_1)(e-c-d_1)\right) \\
& + \left(\delta_{k1} - \eta_i^{d_1}(g)\right)\left(\frac{5}{18}(e-c)^2\right) \\
& \vdots \\
& + \eta_i^{d_z}(g)\left(\frac{1}{8}(3(e-c) + d_z)(e-c-d_z)\right) \\
& + \left(\delta_{kz} - \eta_i^{d_z}(g)\right)\left(\frac{5}{18}(e-c)^2\right)
\end{aligned}$$

Now, using the facts that $r + \sum_{d_k \in D_i} \delta_{ki} = |\mathcal{N}|$ and $\eta_i(g) + \sum_{d_k \in D_i} \eta_i^{d_k}(g) = |\mathcal{A}_i|$ we can simplify further by writing

$$\begin{aligned} & w\left(\eta_i(\bar{g}), \eta_i^{d_1}(\bar{g}), ..., \eta_i^{d_z}(\bar{g}), ...; \boldsymbol{\gamma}\right) \\ = \; & \eta_i(\bar{g})\left(\frac{3}{8}(e-c)^2\right) \\ & + (|\mathcal{N}| - |\mathcal{A}_i|)\left(\frac{5}{18}(e-c)^2\right) \\ & + \eta_i^{d_1}(\bar{g})\left(\frac{1}{8}(3(e-c)+d_1)(e-c-d_1)\right) \\ & \vdots \\ & + \eta_i^{d_z}(\bar{g})\left(\frac{1}{8}(3(e-c)+d_z)(e-c-d_z)\right). \end{aligned}$$

□.

Proof of Lemma 4: (i) To show that $\Delta w / \Delta \eta_i(g)$ is constant, begin by noting that although $\eta_i(g)$ is a discrete variable, the function $w\left(\eta_i(g), \eta_i^{d_1}(g), ...\eta_i^{d_z}(g); \boldsymbol{\gamma}\right)$ is continuous in $\eta_i(g)$. Treating $\eta_i(g)$ as a continuous variable in a compact set, it is possible to calculate the derivative of $w\left(\eta_i(g), \eta_i^{d_1}(g), ...\eta_i^{d_z}(g); \boldsymbol{\gamma}\right)$ with respect to $\eta_i(g)$;

$$\frac{\partial w}{\partial \eta_i(g)} = \frac{7}{72}(e-c)^2$$

As the expression for $\partial w / \partial \eta_i(g)$ is parametric, the effect on w of a discrete change in η_i is given by

$$\Delta w = \left(\frac{7}{72}(e-c)^2\right)\Delta \eta_i(g)$$

This holds at *any* $\eta_i(g)$, as required. As $(e-c) > 0$ by assumption, $\Delta w / \Delta \eta_i(g) > 0$.

To show that $\Delta w / \Delta \eta_i^{d_k}(g)$ is constant, follow the same procedure. Calculate the derivative of $w\left(\eta_i(g), \eta_i^{d_1}(g), ...\eta_i^{d_z}(g); \boldsymbol{\gamma}\right)$ with respect to $\eta_i^{d_k}(g)$:

$$\frac{\partial w}{\partial \eta_i^{d_k}(g)} = \frac{1}{72}(7(e-c)+3d_k)(e-c-3d_k)$$

As the expression for $\partial w / \partial \eta_i^{d_k}(g)$ is parametric, the effect on w of a discrete change in $\eta_i^{d_k}$ is given by

$$\Delta w = \frac{1}{72}(7(e-c)+3d_k)(e-c-3d_k)\Delta \eta_i^{d_k}(g)$$

Again, this holds at *any* $\eta_i(g)$ as required. By A1, $(e-c) > 3d_k$ and therefore $\Delta w/\Delta\eta_i^{d_k}(g) > 0$.

(ii) Show that $\Delta w/\Delta\eta_i(g) > \Delta w/\Delta\eta_i^{d_j}(g) > \Delta w/\Delta\eta_i^{d_k}(g) > \Delta w/\Delta\eta_i^{d_l}(g) \geq 0$. First establish that $\Delta w/\Delta\eta_i(g) > \Delta w/\Delta\eta_i^{d_j}(g)$. From (i) we know that in general

$$\frac{\Delta w}{\Delta\eta_i^{d_k}(g)} = \frac{1}{72}\left(7(e-c)+3d_k\right)(e-c-3d_k)$$

where d_k is any element of D_i. Expanding the brackets,

$$\frac{\Delta w}{\Delta\eta_i^{d_k}(g)} = \frac{1}{72}\left(7(e-c)^2 - 18(e-c)d_k - 9d_k^2\right)$$

Notice that $\Delta w/\Delta\eta_i^{d_k}(g)$ is declining in d_k, attaining its maximum for $d_k = 0$. So $\Delta w/\Delta\eta_i(g) > \Delta w/\Delta\eta_i^{d_k}(g)$ for all $d_k \in D_i$ and, in particular, $\Delta w/\Delta\eta_i(g) > \Delta w/\Delta\eta_i^{d_j}(g)$.

Next, establish that $\Delta w/\Delta\eta_i^{d_j}(g) > \Delta w/\Delta\eta_i^{d_k}(g) > \Delta w/\Delta\eta_i^{d_l}(g)$. But this follows immediately by the fact that $\Delta w/\Delta\eta_i^{d_k}(g)$ is declining in d_k, and that by assumption $d_j < d_k < d_l$.

Finally, it must be established that $\Delta w/\Delta\eta_i^{d_l}(g) \geq 0$. The root for $\Delta w/\Delta\eta_i^{d_k}(g) = 0$ is $d_k = (e-c)/3$. To see this, use $d_k = (e-c)/3$ in $\Delta w/\Delta\eta_i^{d_k}(g)$ to obtain

$$\frac{\Delta w}{\Delta\eta_i^{d_k}(g)} = \frac{1}{72}\left(7(e-c)^2 - \frac{18}{3}(e-c)^2 - 9\left(\frac{e-c}{3}\right)^2\right) = 0.$$

But by A1, $0 < d_k \leq (e-c)/3$. The result follows. □

Proof of Lemma 5. By A1, $\Delta w/\Delta\eta_i(g) > \Delta w/\Delta\eta_i^d(g) \geq 0$ (Lemma 4).

(i) Assume that initially g is the empty network and suppose to the contrary that there does exist a FTA that is worth sponsoring. For this to be the case the overall payoff to sponsoring such an agreement must be higher than autarchy. Then there exist values of y^1 and y_d^1, where $r \geq y^1 > 1$, $n-r \geq y_d^1 \geq 0$, for which $\psi\left(y_1, y_1^d, \left(y_1+y_1^d-1\right)\right) > \psi(1,0,0)$. This implies

$$\begin{aligned} & \psi\left(y_1, y_1^d, \left(y_1+y_1^d-1\right)\right) - \psi(1,0,0) \\ = \; & (y_1-1)\,\Delta w/\Delta\eta_i(g) + y_1^d \Delta w/\Delta\eta_i^d(g) - \kappa\left(y_1+y_1^d-1\right) > 0. \end{aligned}$$

But

$$\begin{aligned}
& (y_1-1)\,\Delta w/\Delta\eta_i(g)+y_1^d\Delta w/\Delta\eta_i^d(g)-\kappa\left(y_1+y_1^d-1\right)\\
= \;& (y_1-1)\,\Delta w/\Delta\eta_i(g)+y_1^d\Delta w/\Delta\eta_i^d(g)-\left(y_1+y_1^d-1\right)\kappa(1)\\
< \;& (y_1-1)\,\Delta w/\Delta\eta_i(g)+y_1^d\Delta w/\Delta\eta_i(g)-\left(y_1+y_1^d-1\right)\kappa(1)\\
= \;& \left(y_1+y_1^d-1\right)\left(\Delta w/\Delta\eta_i(g)-\kappa(1)\right),
\end{aligned}$$

and by assumption $\Delta w/\Delta\eta_i(g)<\kappa(1)$ so $\left(y_1+y_1^d-1\right)\left(\Delta w/\Delta\eta_i(g)-\kappa(1)\right)<0$; contradiction.

(ii) Assume that initially g is the empty network and suppose to the contrary that sponsorship of some FTA other than the world FTA yields a higher payoff. Then $\psi(r,n-r,(n-1))<\psi\left(y_1,y_1^d,\left(y_1+y_1^d-1\right)\right)$ for all values of y^1 and y_d^1, where $r\geq y^1\geq 1$, $n-r\geq y_d^1\geq 0$. This implies

$$\begin{aligned}
& \psi(r,n-r,(n-1))-\psi\left(y_1,y_1^d,\left(y_1+y_1^d-1\right)\right)\\
= \;& (r-y_1)\,\Delta w/\Delta\eta_i(g)+\left(n-r-y_1^d\right)\Delta w/\Delta\eta_i^d(g)-\kappa\left(n-y_1-y_1^d\right)<0.
\end{aligned}$$

But

$$\begin{aligned}
& (r-y_1)\,\Delta w/\Delta\eta_i(g)+\left(n-r-y_1^d\right)\Delta w/\Delta\eta_i^d(g)-\kappa\left(n-y_1-y_1^d\right)\\
= \;& (r-y_1)\,\Delta w/\Delta\eta_i(g)+\left(n-r-y_1^d\right)\Delta w/\Delta\eta_i^d(g)-\left(n-y_1-y_1^d\right)\kappa(1)\\
> \;& (r-y_1)\,\Delta w/\Delta\eta_i^d(g)+\left(n-r-y_1^d\right)\Delta w/\Delta\eta_i^d(g)-\left(n-y_1-y_1^d\right)\kappa(1)\\
= \;& \left(n-y_1-y_1^d\right)\left(\Delta w/\Delta\eta_i^d(g)-\kappa(1)\right),
\end{aligned}$$

and by assumption $\Delta w/\Delta\eta_i^d(g)>\kappa(1)$ so $\left(n-y_1-y_1^d\right)\left(\Delta w/\Delta\eta_i^d(g)-\kappa(1)\right)\geq 0$; contradiction. Clearly, if $y^1=r$ and $y_d^1=n-r$ then $\left(n-y_1-y_1^d\right)\left(\Delta w/\Delta\eta_i^d(g)-\kappa(1)\right)=0$ and $\psi(r,n-r,(n-1))=\psi\left(y_1,y_1^d,\left(y_1+y_1^d-1\right)\right)$. But if $y^1<r$ and/or $y_d^1<n-r$ then $\left(n-y_1-y_1^d\right)\left(\Delta w/\Delta\eta_i^d(g)-\kappa(1)\right)>0$. The result follows.

(ii) Assume that initially g is the empty network and suppose to the contrary that sponsorship of some FTA other than the complete-regional FTA yields a higher payoff. Then $\psi(r,0,(r-1))<\psi\left(y_1,y_1^d,\left(y_1+y_1^d-1\right)\right)$ for some values of y^1 and y_d^1, where $r>y^1\geq 1$, and $n-r\geq y_d^1\geq 0$. This implies

$$\begin{aligned}
& \psi(r,0,(r-1))-\psi\left(y_1,y_1^d,\left(y_1+y_1^d-1\right)\right)\\
= \;& (r-y_1)\,\Delta w/\Delta\eta_i(g)+\left(n-r-y_1^d\right)\Delta w/\Delta\eta_i^d(g)-\kappa\left(n-y_1-y_1^d\right)<0.
\end{aligned}$$

But

$$\begin{aligned}
& (r-y_1)\,\Delta w/\Delta\eta_i(g) + \left(n-r-y_1^d\right)\Delta w/\Delta\eta_i^d(g) - \kappa\left(n-y_1-y_1^d\right) \\
= \;& (r-y_1)\,\Delta w/\Delta\eta_i(g) + \left(n-r-y_1^d\right)\Delta w/\Delta\eta_i^d(g) - \left(n-y_1-y_1^d\right)\kappa(1) \\
> \;& (r-y_1)\,\Delta w/\Delta\eta_i^d(g) + \left(n-r-y_1^d\right)\Delta w/\Delta\eta_i^d(g) - \left(n-y_1-y_1^d\right)\kappa(1) \\
= \;& \left(n-y_1-y_1^d\right)\left(\Delta w/\Delta\eta_i^d(g) - \kappa(1)\right),
\end{aligned}$$

and by assumption $\Delta w/\Delta\eta_i^d(g) > \kappa(1)$ so $\left(n-y_1-y_1^d\right)\left(\Delta w/\Delta\eta_i^d(g) - \kappa(1)\right) \geq 0$; contradiction. Clearly, if $y^1 = r$ and $y_d^1 = n-r$ then $\left(n-y_1-y_1^d\right)\left(\Delta w/\Delta\eta_i^d(g) - \kappa(1)\right) = 0$ and $\psi(r, n-r, (n-1)) = \psi\left(y_1, y_1^d, \left(y_1+y_1^d-1\right)\right)$. But if $y^1 < r$ and/or $y_d^1 < n-r$ then $\left(n-y_1-y_1^d\right)\left(\Delta w/\Delta\eta_i^d(g) - \kappa(1)\right) > 0$. The result follows.

(iii) Assume that initially g is the empty network and suppose to the contrary that sponsorship of some FTA other than the complete-regional FTA yields a higher payoff. Then either $\psi(r, 0, (r-1)) < \psi(r, n-r, (n-1))$ or $\psi(r, 0, (r-1)) < \psi\left(y_1, y_1^d, \left(y_1+y_1^d-1\right)\right)$ for all values of y^1 and y_d^1, where $r > y^1 > 1$, and $n-r \geq y_d^1 \geq 0$. But the first inequality implies

$$\begin{aligned}
& \psi(r, 0, (r-1)) - \psi(r, n-r, (n-1)) \\
= \;& -(n-r)\,\Delta w/\Delta\eta_i^d(g) + \kappa((n-r)) \\
= \;& -(n-r)\left(\Delta w/\Delta\eta_i^d(g) - \kappa(1)\right) < 0.
\end{aligned}$$

and by assumption $\Delta w/\Delta\eta_i^d(g) < \kappa(1)$ and so $-(n-r)\left(\Delta w/\Delta\eta_i^d(g) - \kappa(1)\right) > 0$; contradiction.

The second inequality implies

$$\begin{aligned}
& \psi(r, 0, (r-1)) - \psi\left(y_1, y_1^d, \left(y_1+y_1^d-1\right)\right) \\
= \;& (r-y_1)\,\Delta w/\Delta\eta_i(g) - y_1^d\Delta w/\Delta\eta_i^d(g) - \kappa\left(r-1-y_1-y_1^d+1\right) < 0.
\end{aligned}$$

But

$$\begin{aligned}
& (r-y_1)\,\Delta w/\Delta\eta_i(g) - y_1^d\Delta w/\Delta\eta_i^d(g) - \kappa\left(r-1-y_1-y_1^d+1\right) \\
= \;& (r-y_1)\,\Delta w/\Delta\eta_i(g) - y_1^d\Delta w/\Delta\eta_i^d(g) - \left(r-y_1-y_1^d\right)\kappa(1) \\
> \;& (r-y_1)\,\Delta w/\Delta\eta_i^d(g) - y_1^d\Delta w/\Delta\eta_i^d(g) - \left(r-y_1-y_1^d\right)\kappa(1) \\
= \;& \left(r-y_1-y_1^d\right)\left(\Delta w/\Delta\eta_i^d(g) - \kappa(1)\right),
\end{aligned}$$

and by assumption $\Delta w/\Delta\eta_i^d(g) < \kappa(1)$ so $\left(r - y_1 - y_1^d\right)\left(\Delta w/\Delta\eta_i^d(g) - \kappa(1)\right) > 0$ for $r < y_1 + y_1^d$. Now $\left(r - y_1 - y_1^d\right)\left(\Delta w/\Delta\eta_i^d(g) - \kappa(1)\right) < 0$ for $r > y_1 + y_1^d$. But in addition

$$\begin{aligned} & (r - y_1)\,\Delta w/\Delta\eta_i(g) - y_1^d \Delta w/\Delta\eta_i^d(g) - \left(r - y_1 - y_1^d\right)\kappa(1) \\ < \;& (r - y_1)\,\Delta w/\Delta\eta_i(g) - y_1^d \Delta w/\Delta\eta_i(g) - \left(r - y_1 - y_1^d\right)\kappa(1) \\ = \;& \left(r - y_1 - y_1^d\right)\left(\Delta w/\Delta\eta_i(g) - \kappa(1)\right), \end{aligned}$$

and by assumption $\kappa(1) < \Delta w/\Delta\eta_i(g)$, so $\left(r - y_1 - y_1^d\right)\left(\Delta w/\Delta\eta_i(g) - \kappa(1)\right) > 0$ for $r > y_1 + y_1^d$. □

Proof of Lemma 6. Let there be a regional agreement of size $r \geq y_1^d > 1$ in region R_j ($i \in R_i$, $i \neq j$). The proof is in two parts. (i) Show that if $y_1^d \Delta w/\Delta\eta_i^d(g) > \kappa(1) > \Delta w/\Delta\eta_i^d(g)$ then country i does find it worth sponsoring an extra-regional agreement with the FTA in R_j. (ii) Show that this does not hold if $\kappa(1) > y_1^d \Delta w/\Delta\eta_i^d(g)$.

(i) Assume $y_1^d \Delta w/\Delta\eta_i^d(g) > \kappa(1) > \Delta w/\Delta\eta_i^d(g)$. Suppose to the contrary that an extra-regional agreement with the FTA in R_j is not worth sponsoring. This implies

$$\begin{aligned} & \psi\left(y_1, y_1^d, y_1\right) - \psi(y_1, 0, y_1 - 1) \\ = \;& y_1^d \Delta w/\Delta\eta_i^d(g) - \kappa(1) < 0 \end{aligned}$$

But by assumption $y_1^d \Delta w/\Delta\eta_i^d(g) > \kappa(1)$; contradiction.

(ii) Now assume $\kappa(1) > y_1^d \Delta w/\Delta\eta_i^d(g)$ in order to see that country i does not find it worth sponsoring an extra-regional agreement with the FTA in R_j. Suppose to the contrary that such an agreement is worth sponsoring. This implies $\psi\left(y_1, y_1^d, y_1\right) - \psi(y_1, 0, y_1 - 1) = y_1^d \Delta w/\Delta\eta_i^d(g) - \kappa(1) > 0$; contradiction. □

Proof of Proposition 4.

(i) Suppose to the contrary that there exists a period in which at least one FTA is sponsored. Let $t = s$ be the first period in which at least one FTA is sponsored. Then by the assumption that the network is empty at $t = 0$, the network must be empty at $t = s - 1$. Taking as given the empty network g at $t = s - 1$, the payoff to sponsoring an agreement with y_1 countries in the same region and y_1^d countries outside the region is $\psi\left(y_1, y_1^d, \left(y_1 + y_1^d - 1\right)\right)$. By assumption, production-trade payoffs and sponsorship costs are in the same relation as in Lemma 5(i). But by Lemma 5(i),

$\psi(1,0,0) > \psi\left(y_1, y_1^d, \left(y_1 + y_1^d - 1\right)\right)$ and therefore any country sponsoring an agreement could gain by deleting all its links. So the network must be empty at $t = s$ as well. As $t = s$ is any period $t \geq 1$, the network g must be empty at every period $t \geq 1$.

(ii) Suppose to the contrary that there exists a period in which there is not a world FTA. Let $t = s$ be a period in which the Nash network g is either empty or not connected. By definition of equilibrium, there must exist a world FTA at $t = 1$. By assumption, production-trade payoffs and sponsorship costs are in the same relation as in Lemma 5(ii). Then by Lemma 5(ii), $\psi(r, n-r, (n-1)) \geq \psi\left(y_1, y_1^d, \left(y_1 + y_1^d - 1\right)\right)$, for $r \geq y^1 \geq 1$, $n - r \geq y_d^1 \geq 0$ holding with strict inequality if and only if $y^1 < r$ and/or $y_d^1 < n - r$. So the empty network cannot be Nash; country i would receive a payoff $\Psi_i = \psi(1,0,0) < \psi(r, n-r, (n-1))$ and has an incentive to deviate by forming links with all other countries. Similarly, if country i sponsors a FTA that is not a world FTA, then it can increase its payoff by forming links, again contradicting Nash. By definition of equilibrium, only one country sponsors the world FTA. If not then a second sponsor could withdraw from sponsorship, gaining the sponsorship cost and not losing any production-trade payoffs, contradicting Nash.

Given that the Nash network g is connected at $t = 1$, then it must be connected at $t = 2$. If not, then the sponsor of the agreement, country i, must have deleted some or all of its links. But deviating in this way would yield a lower payoff than maintaining all links; $\psi\left(y_1, y_1^d, \left(y_1 + y_1^d - 1\right)\right) < \psi(r, n-r, (n-1))$, $r > y^1 \geq 1$, $n - r > y_d^1 \geq 0$, contradicting equilibrium. By induction, taking as given a world FTA in period $t = s-1$, it is a best response for the sponsor, country i, to maintain its sponsorship of the world FTA at $t = s$. As $t = s$ is any period $t > 1$, and as the network g is connected at $t = 1$, it must be connected at every period $t \geq 1$.

(iii) By definition, there is regionalism in the network g if there is a connected subnetwork g_k for each $R_k \in \mathcal{C}$, but $g_{ij} = g_{ji} = 0$ for all $i \in R_i$, $j \in R_j$, $i \neq j$. Suppose to the contrary that the Nash network g does not exhibit regionalism. There are two (mutually inclusive) possibilities. One is that the Nash network g contains links $g_{ij} = 1$ or $g_{ji} = 1$ for some $i \in R_i$, $j \in R_j$, $i \neq j$. The other is that the subnetwork g_k is not connected for some $R_k \in \mathcal{C}$. Contradictions for these two possibilities are found in turn.

Suppose to the contrary that at $t = 1$ the Nash network g contains links $g_{ij} = 1$ or $g_{ji} = 1$ for some $i \in R_i$, $j \in R_j$, $i \neq j$. By assumption, production-trade payoffs and sponsorship costs are in the same relation as in Lemma 5(iii). Then by Lemma 5(iii), $\psi(r, 0, r-1) > \psi\left(r, y_1^d, \left(r + y_1^d - 1\right)\right)$, for $r > y^1 \geq 1$, $n - r \geq y_d^1 \geq 1$. Therefore, if country $i \in R_i$ sponsors any links of the form $g_{ij} = 1$ with $j \in R_j$, $i \neq j$, then it can gain by deleting them, so the network g cannot be Nash. By the same argument, country i has an incentive to break links if it sponsors a world FTA.

Now suppose that in the Nash network g of period $t = 1$, country i sponsors a FTA that is not a complete-regional FTA; that is where $y^1 < r$. But then again by Lemma 5(iii) $\psi(r, 0, r-1) > \psi\left(y_1, y_1^d, \left(r + y_1^d - 1\right)\right)$, for $r > y^1 \geq 1$, $n - r > y_d^1 \geq 0$. Therefore, country i could gain by linking to the other countries j for which $i, j \in R_i$, so the network g cannot be Nash. It follows that for each $R_k \in \mathcal{C}$ the subnetwork g_k must be connected. By definition of equilibrium, only one country sponsors the complete-regional FTA. If not then a second sponsor could withdraw from sponsorship, gaining the sponsorship cost and not losing any production-trade payoffs, contradicting Nash.

(iv) Conditions (b) and (c) are exactly as in (iii) so from (iii) we know that there must be regionalism at $t = 1$. Take the network from period 1 as given, where the subnetworks g_k are connected for the elements of all $R_k \in \mathcal{C}$. Suppose to the contrary that at $t = 2$ the Nash network g is not connected. By assumption, production-trade payoffs and sponsorship costs are in the same relation as in Lemma 5(iii). If at $t = 2$ any subnetwork g_k is not connected then by Lemma 5(iii) there is an incentive to deviate by forming links to other countries within R_k so g cannot be Nash. Moreover, condition (a) implies that production-trade payoffs and sponsorship costs are in the same relation as in Lemma 6. So by Lemma 6, if there does not exist a link between country $i \in R_i$ and $j \in R_j$, $i \neq j$, then country i could gain by forming a link with a country in another region, $\psi\left(r, ay_1^d, r + a - 1\right) > \psi(r, 0, r-1)$, $0 < ay_1^d \leq (n - r)$, $a \geq 1$, contradicting Nash.

Finally, suppose that condition (a) does not hold, so that the payoff from linking to a FTA in another region is not greater than the sponsorship cost. Then the Nash network g cannot be complete, because if country $i \in R_i$ sponsors any links to countries $j \in R_j$, then it could gain by breaking those links. However, given that conditions (b) and (c)

continue to hold, and given regionalism at $t = 1$, there is no incentive for any country to deviate at $t = 2$. By (iii), any sponsor of a regional agreement could not gain by deleting links. So there is regionalism in at $t = 1$. Under these same conditions, given regionalism at $t = s - 1$, there must be regionalism at $t = s$. So when (a) fails to hold there must be regionalism at all points on the equilibrium path. □