

Preferential Trade Liberalization:  
The Traditional Theory and New Developments

Arvind Panagariya\*

February 1999

Revised: September 1999

\*Department of Economics, University of Maryland, College Park MD 20742. Email: [panagari@econ.umd.edu](mailto:panagari@econ.umd.edu). I am indebted to Jagdish Bhagwati, John McMillan and three anonymous referees for numerous helpful comments. Thanks are also due to Kyle Bagwell, Rupa Duttagupta, Pravin Krishna and Robert Staiger for useful suggestions.

## Table of Contents

1. Introduction.....	1
2. The Multilateral Trade Policy Framework.....	4
3. The Traditional Welfare Analysis .....	6
3.1 Trade Creation and Trade Diversion.....	7
3.2 The Revenue-Transfer Effect in a Customs Union .....	12
3.3 Extension to Free Trade Areas .....	14
Case 1: The Total Supply by A and B is Less than the Demand in A.....	15
Case 2: The Total Supply by A and B Equals the Demand in A.....	16
Case 3: The Total Supply by A and B Exceeds the Demand in A.....	18
3.4 The Meade-Lipsey General-Equilibrium Model .....	18
3.5 A More General Analysis of the Small-Union Case .....	22
3.6 The Large-union Case.....	25
3.7 A Differentiated Products Model .....	27
3.8 Transport Costs.....	28
3.9 Economies of Scale and Imperfect Competition .....	30
4. Welfare-Increasing CUs and FTAs .....	33
4.1 Customs Unions .....	33
4.2 Free Trade Areas .....	36
4.3 Customs Unions with Non-economic Objectives.....	37
5. Exogenous Division of the World into Blocs .....	37
5.1 Symmetric Blocs .....	38
5.2 Asymmetric Blocs.....	40

5.3	"Natural" and "Unnatural" Blocs.....	40
6.	Endogeneity of Policy .....	43
6.1	The Decision to Form an FTA.....	43
	A Perfect Competition, Small-Union Model .....	43
	A Cournot Oligopoly Model.....	48
6.2	The Extra-Union Tariff.....	49
	Lobbying and the External tariff in Small-Union Models.....	50
	Turning Trade Diversion into Tariff Revenue .....	51
	Increased Monopoly Power of Trade Blocs.....	52
6.3	Evidence .....	53
7.	Regionalism and Multilateralism.....	54
7.1	Strangers: Bloc Expansion.....	54
7.2	Impact of Regionalism on Multilateralism: Friends or Foes?.....	57
	Stumbling Blocks: A Median Voter Model .....	57
	Stumbling Blocks: A Cournot Oligopoly Model.....	61
	Insidious Regionalism .....	61
	Stumbling Blocks in Transition but Building Blocks in the Long Run: FTAs .....	62
	Building Blocks in Transition but Stumbling Blocks in the Long Run: CUs .....	63
7.3	The Impact of Multilateralism on Regionalism.....	64
	Multilateral Liberalization Making PTAs More Sustainable .....	64
	Liberalization in North Leading to North-South PTAs .....	65
8	Theoretical Considerations in Empirical Assessments of PTAs.....	67
9.	Concluding Remarks .....	73

## 1. Introduction

The current wave of preferential trade arrangements, like the first one in the 1950s and 1960s, has given rise to a lively debate between the free trade economists who view the arrangements as harmful and others who see them as beneficial. To the old concerns relating to welfare effects, captured in Jacob Viner's (1950) influential "static" concepts of trade creation and trade diversion, the current debate has added what Jagdish Bhagwati (1993) calls the "dynamic" time-path issue.

The effective regional arrangements during the first wave did not spread beyond Western Europe.<sup>1</sup> Consistent with this reality, the debate at the time, and the literature it spawned, remained largely confined to the question whether regional arrangements resulted in higher or lower welfare for their members. Today, with "trade blocs" being vigorously sought by virtually all countries in the world, economists and policy analysts are also focusing on the implications of such blocs for the global trading system. In terms of Bhagwati's (1991) memorable phrase, they are asking whether trade blocs serve as "building blocks" or "stumbling blocks" for worldwide freeing of trade.

The purpose of this essay is to bring together the key theoretical contributions addressing both the old and new themes. Two features distinguish this essay from others that have appeared in recent years.<sup>2</sup> First, its emphasis is almost exclusively on theory. In

---

<sup>1</sup> Throughout this essay, the term "regional arrangements" refers to preferential trade arrangements, defined more precisely below.

<sup>2</sup> Survey articles include Richard Baldwin and Anthony Venables (1995), Bhagwati and Arvind Panagariya (1996a, 1996b), Alan Winters (1996), Bhagwati, David Greenaway and Panagariya

spite of a number of recent books and survey articles, we lack a single source synthesizing the large body of the theoretical literature on the subject. Second, rather than simply report the results derived in various contributions, the essay offers a deeper treatment of them. A special effort is made to provide an intuitive but rigorous explanation of the mechanism underlying many of the contributions reviewed.

My concern in the essay is solely with the literature on *preferential* liberalization of tariffs on goods. Thus, I do not review many of the recent contributions on "trade agreements" in which trade preferences play no explicit role. Also excluded from consideration are issues such as preferential trade in services, the role of investment in regional arrangements, and harmonization of domestic policies. Though these issues figure in the current policy debate, they have not been seriously addressed in the theoretical literature.<sup>3</sup>

Three terms are used frequently in the essay: preferential trade area (PTA), free trade area (FTA) and customs union (CU). Throughout, a PTA refers to a union between two or

---

(1998), Raquel Fernandez and Jonathan Portes (1998) and Panagariya (1999a). Recent book-length treatments are Kym Anderson and Richard Blackhurst (1993), Jaime de Melo and Panagariya (1993), Jeffrey Frankel (1997, 1998) and Panagariya (1999b). Many of the contributions reviewed in this paper can be readily found in Bhagwati, Pravin Krishna and Panagariya (1999).

<sup>3</sup> Even within this narrow definition of regional arrangements, space constraints preclude full coverage. For instance the analysis of the effects of a policy change (e.g., a reduction in the common external tariff) on individual union members within an existing common market with internal factor mobility, pioneered by Richard Brecher and Bhagwati (1981), is not considered.

more countries in which lower tariffs are imposed on goods produced in the member countries than on goods produced outside. An FTA is a PTA with tariffs eliminated entirely on goods produced in member countries. A customs union (CU) is an FTA with all members imposing a common external tariff on a given good. The term PTA being wider, it is used to include the arrangements with limited tariff preferences, FTAs and CUs.

In Section 2, I discuss briefly the multilateral trade policy framework within which PTAs are formed. In Section 3, I develop carefully the traditional welfare analysis, adding several new twists to it. In Section 4, I discuss the Kemp-Wan-Ohyama-Vanek theorem on CUs and its recent extension to FTAs by Panagariya and Pravin Krishna (1997) that cuts through the ambiguities of trade creation and trade diversion and offers a clear approach to identifying welfare-enhancing FTAs. In Section 5, I turn to the more recent literature that focuses on the welfare effects of a simultaneous, exogenous division of the world into several blocs. The emphasis of this literature is on the relationship between the number of blocs and welfare. In Section 6, I turn to the literature on endogenous policy. Two sets of questions are discussed. First, if the decision to form an FTA is endogenous, under what circumstances are they likely to be accepted and under what conditions are they likely to be rejected? Second, what is the impact of FTAs and CUs on the outside tariff? In section 7, I turn to the models addressing directly the relationship between regionalism and multilateralism. In section 8, I briefly examine some theoretical issues underlying the empirical approaches to resolving the ambiguity in welfare outcomes. In section 9, I conclude the paper.

## 2. The Multilateral Trade Policy Framework

International trade in goods is governed by the General Agreement on Tariffs and Trade (GATT). Signed in 1947, this agreement was incorporated into the 1994 Marrakesh Agreement establishing the World Trade Organization (WTO). The centerpiece of GATT is the Most Favored Nation (MFN) principle enunciated in its Article I. Accordingly, in matters of trade policy, each WTO member is to grant to all members the same advantage, privilege, favor or immunity that it grants to any other country. A key implication of this provision is that member countries are not to discriminate in their tariff policy across other members.

PTAs are, thus, in conflict with Article I of GATT and had to be accommodated through a variety of additional provisions. As it stands currently, there exist three alternative provisions for trade preferences within the GATT/WTO system. First, developed countries can give developing countries *one-way* trade preferences. This provision is the basis of the Generalized System of Preferences, designed to promote the exports from developing to developed countries. Second, under the Enabling Clause, developing countries can exchange virtually any trade preferences to which they agree. This provision is intended to promote trade among developing countries themselves. Under the Enabling Clause, preferences need not lead to a full free trade area; partial preferences across a subset of goods are permitted. All PTAs among developing countries, including the prominent ones such as the Southern Cone Common Market, popularly known by its Spanish acronym MERCOSUR, and ASEAN Free Trade Area (AFTA), have been formed under this

provision.<sup>4</sup> Finally, under Article XXIV of GATT, any two or more members of WTO can form an FTA or CU. European Economic Community (EEC) and its various association agreements and the North American Free Trade Agreement (NAFTA) have been concluded under Article XXIV.

Article XXIV offers the only avenue to PTAs in which developed countries are *recipients* of trade preferences. Given the large share of developed countries in both world trade and world GDP, this fact gives Article XXIV a central role in any discussion of PTAs. A key requirement of Article XXIV is that the exchange of preferences not be partial. Instead, it should result in the formation of an FTA or CU with duties and other restrictive regulations of commerce eliminated with respect to ‘substantially all the trade’ in products originating in union members. In the case of free trade areas, external tariffs of member countries must not be raised. In the case of CUs, the incidence of the common external tariff on outside countries’ trade is not to exceed that of individual tariffs of union members prior to the formation of the union.

To-date, GATT and WTO have not actively enforced the provisions of GATT Article XXIV. For one thing, the definition of ‘substantially all trade’ has never been clarified satisfactorily. Thus, even though trade in agriculture was not freed within EEC, GATT never censured that arrangement. EEC’s association agreements with its neighbors have also carried frequently exceptions with respect to product coverage but these have not been questioned either. In effect, matters have been left to the member countries themselves.

---

<sup>4</sup> ASEAN stands for the Association of South-East Asian Nations.



This is not to suggest, however, that Article XXIV has not influenced the design of PTAs in which developed countries are participants. Despite aberrations and exceptions, EEC and NAFTA are closer to free trade among union members than any of the PTAs concluded under the Enabling Clause. Article XXIV has also kept in check the temptation to exchange partial trade preferences when member countries have no intention to form an FTA or CU. Thus, in the absence of Article XXIV, it is conceivable that some developed-country members of Asia Pacific Economic Cooperation (APEC) forum would have opted for an exchange of partial trade preferences but were restrained by this provision.

### 3. The Traditional Welfare Analysis

Though the current wave of regionalism has given rise to new concerns, the old concerns have remained alive. Therefore, it is appropriate to begin our review with the traditional, static welfare analysis pioneered by Viner (1950) in his classic work, *The Customs Union Issue*. I first spell out the broad structure of the model to be used throughout this section. Within this framework, I then employ a set of partial-equilibrium models to identify the key factors, which determine the welfare outcome. The modifications implied by general equilibrium considerations are brought into the picture only at the last stage of the analysis.

Assume three countries, A, B and C, which trade a product, steel, with each other. Countries A and B are potential union partners and C represents the rest of the world. Between the union members, A is the importer of steel and B the exporter.<sup>5</sup> If an FTA is

---

<sup>5</sup> Based on the analysis to be presented, the reader can analyze the remaining, less interesting, cases in which A and B are both importers or exporters of steel.

formed, each member sets the external tariff at its pre-union level. If a CU is formed, the common external tariff is set equal to the pre-union tariff of A, the importing member of the union.<sup>6</sup> Other details, relating to demand, supply, trade and tariffs are spelt out in the context of specific models.

### 3.1 *Trade Creation and Trade Diversion*

Any discussion of the welfare effects of PTAs must inevitably begin with the influential concepts of trade creation and trade diversion, introduced by Viner (1950). As James Meade (1955, Ch. 2) noted in his seminal contribution, *The Theory of Customs Unions*, these concepts are best introduced within a model exhibiting infinite supply elasticities and zero demand elasticities. This model avoids some of the ambiguities that arise in more general models.

Let us then begin by representing country A's demand for steel by the vertical line  $D_A D_A$  in Figure 1a. Firms in A, B and C supply at constant prices shown by  $P_A$ ,  $P_B$  and  $P_C$ , respectively. Under perfect competition, these prices also represent the constant average and marginal costs of production in the three countries. By assumption, A is the least efficient supplier of steel and C the most efficient one. Thus, we have  $P_A > P_B > P_C$ .

---

<sup>6</sup> In the case of a CU, it will be assumed that B keeps a non-negative tariff on the books even though it is non-binding in the initial equilibrium. As long as this tariff is no lower than A's, equating the common external tariff to the latter is consistent with the GATT Article XXIV. But in other cases, this traditional assumption made implicitly or explicitly by all analysts employing the partial-equilibrium framework is inconsistent with GATT Article XXIV.

[Figure 1 here]

We assume that countries B and C do not trade with each other. This will be true, for example, if B applies a per-unit tariff higher than  $P_B - P_C$  on imports. Initially, country A imposes a non-discriminatory tariff at rate  $t$  *per-unit* on steel.<sup>7</sup> The tariff rate is chosen such that  $P_A > P_C + t > P_B$ . The entire quantity demanded,  $OQ_0$ , is imported from C. The price paid by consumers is  $P_C + t$ , with area  $e+f$  collected in tariff revenue by A's government.

Suppose now that country A eliminates the tariff on B but retains it on C. Given  $P_C + t > P_B$ , A now purchases its imports from B rather than C at price  $P_B$ . Because the change creates no new trade and merely substitutes the less efficient B for the more efficient C, in Viner's terminology, the union is "trade diverting". Country A loses the tariff revenue  $e+f$ , with  $e$  used up to pay for the higher production cost in B and  $f$  becoming a part of A's consumers' surplus. The net loss to A and the world from the union is area  $e$ .

Next, suppose the initial non-discriminatory tariff in A is  $t'$ , where  $t'$  is sufficiently high to result in  $P_A < P_C + t' < P_B + t'$ . Thus, the high tariff prices out both B and C from A's market. The entire demand for steel,  $OQ_0$ , is satisfied by A's own firms at price  $P_A$ . Once again, let A remove the tariff on B but not C. This change leads to a switch in the source of supply from A to B. The price of steel paid by A's buyers drops from  $P_A$  to  $P_B$ , yielding a gain in consumers' surplus equal to  $f+g$ . Because the union creates new trade between A and B and is associated with a switch from higher-cost suppliers in A to lower-cost suppliers

---

<sup>7</sup> A *per-unit* tariff rate is employed mainly to simplify the figures. Unless otherwise noted, replacing this rate by an *ad valorem* rate will not change any of the conclusions.

in B, in Viner's terminology, the union is "trade creating."<sup>8</sup> Welfare of A and the world rises by  $f+g$  while that of B and C is unchanged.

Within the confines of the model under consideration, trade diversion is associated with a welfare loss and trade creation with a welfare gain. Viner argued that since a union is trade creating in some products and trade diverting in others, in general, we could not say whether it increases or decreases welfare. The answer depends on the relative magnitudes of trade creation and trade diversion.

But as Meade (1955, ch. 2) has rightly pointed out, even the relative magnitudes of trade creation and trade diversion alone are insufficient to determine the welfare effect of the union for two reasons. First, benefits of preferential liberalization depend on not only the extent of trade creation, but also the magnitude by which costs are reduced on each unit of newly created trade. Similarly, losses are determined not just by the amount of trade diversion but also the magnitude of the increase in costs due to trade diversion. In terms of Figure 1a, the benefit from trade creation, area  $f+g$ , equals  $OQ_0.P_AP_B$  while the loss due to trade diversion, area  $e$ , equals  $OQ_0.P_BP_C$ . We cannot infer the gain or loss from  $OQ_0$  alone.

The second problem, formalized subsequently by Franz Gehrels (1956-57) and Richard Lipsey (1957) within a one-factor, general-equilibrium model, is that once we drop

---

<sup>8</sup> Observe that even though the lowest-cost source of supply is C, this union is trade creating since the switch is to a lower-cost source. Viner (1950, p. 43) was quite explicit about this possibility. To quote him, "This shift in the locus of production as between the two countries is a shift from a high-cost to a lower-cost point, a shift which the free-trader can properly approve, as at least a step in the right direction, even if universal free trade would divert production to a source with still lower costs."

the unrealistic assumption of zero elasticity of demand in A, even a wholly trade-diverting union may lead to a net increase in welfare. This can be demonstrated by replacing the vertical demand curve in Figure 1a by a downward-sloped demand curve. Thus, in Figure 1b, let the demand curve in A,  $D_A D_A$ , be negatively sloped. The initial non-discriminatory tariff is set at  $t$  with country A importing  $OQ_0$  from C. A removal of the tariff on B but not C prices out the latter, the least-cost producer of steel, but allows an expansion of imports from  $OQ_0$  to  $OQ_1$ . The result is a loss of area e on the original imports but a gain of area h on new imports. Area f is a redistribution of tariff revenue to consumers in A (ignore area k for now). In principle, area h can be larger than area e, establishing the possibility that a wholly trade diverting union can lead to an improvement in welfare.<sup>9</sup>

---

<sup>9</sup> Because quantity  $Q_1 Q_2$  is new trade rather than a replacement of old trade by the partner, it is not entirely clear whether Viner would have called it trade diversion. There is at least one statement in Viner (1950, p. 44) that contradicts the Meade-Gehrels-Lipsev interpretation: "It will be noted that for the free-trader the benefit from a customs union to the customs union area as a whole derives from that portion of the new trade between the member countries which is wholly new trade, whereas each particular portion of the new trade between the member countries which is a substitute for trade with third countries he must regard as a consequence of the custom union which is injurious for the importing country, for the external world, and for the world as a whole, and is beneficial only to the supplying member." Given the references to the gains and losses to the partner and the outside country, this statement also undermines the interpretation that Viner thought purely in terms of a constant-costs model. The difficulty, however, is that non-constant costs do not sit well with the bulk of the analysis in the book.

Bhagwati (1971) makes the further point that even with a zero demand elasticity, a Trade-diverting union can lead to an improvement in welfare provided the supply elasticity of steel in country A is positive but finite. To see this, revert to Figure 1a and imagine that country A's supply curve is upward sloped, starting below  $P_B$  and meeting  $D_A D_A$  above  $P_C+t$ . Under a nondiscriminatory tariff, the price in A is  $P_C+t$  with steel supplied partially by C and partially by A. A free trade area with B leads to a replacement of C by B as the foreign supplier, which is trade diverting. Nevertheless, the internal price falls so that inefficient domestic production is partially replaced by imports. The net effect on welfare depends on the magnitude of this gain in efficiency relative to the loss from replacing higher-cost source B for the original quantity of imports from C. As Bhagwati (1971) correctly concludes, to eliminate the possibility of a trade-diverting union leading to welfare gains, we must assume the elasticity of demand for *imports* in A to be zero and the elasticity of supply from B and C to be infinity.

To Meade's caveats, we may add one further limitation. When union members are themselves large, even leaving aside the two considerations noted by Meade, we can rely on trade creation and trade diversion to infer the welfare effects of preferential trade liberalization only if we are interested in the world welfare. If the focus of the analysis is the welfare of the union instead, trade diversion is likely to be beneficial due to the improvement in the terms of trade it brings. Likewise, trade creation, which enhances the union's income at constant world prices, could generate a harmful second-round effect through deterioration in the terms of trade. Meade himself did not have to confront this problem since he analyzed PTAs exclusively from the viewpoint of global welfare.

Despite these limitations, trade creation and trade diversion have remained central to policy debates on PTAs.<sup>10</sup> This is presumably because economists have found these terms to be highly effective tools of focusing policy makers' attention on the ambiguous welfare effects of PTAs.

### 3.2 *The Revenue-Transfer Effect in a Customs Union*

Even after we allow for downward-sloped demand and upward-sloped supply in A, the model just considered remains unrealistic in one key respect: it necessarily implies that all of A's imports come from either B or C but not both. To capture the realistic case, in which imports come from the union partner as well as the outside country, we must introduce a finite elasticity of supply in at least one of B and C.

As shown originally in Panagariya (1996) and elaborated further in Bhagwati and Panagariya (1996a), the introduction of a finite supply elasticity on the part of one or both of B and C leads to a fundamental change in the effects of preferential trade liberalization on welfare. Thus, continuing to assume that A is the potential importer of steel, subtract its supply from demand and obtain its import-demand curve. Similarly, assuming that B is an exporter of steel, subtract its demand from supply and obtain its export-supply curve. In Figure 2, represent these curves by  $M_A M_A$  and  $E_B E_B$ , respectively, and C's infinitely elastic supply by  $P_C P_C$ .

[Figure 2 here]

---

<sup>10</sup> Carsten Kowalczyk (1990) has argued in favor of replacing these terms by volume-of-trade and terms-of-trade effects but without impact.

Initially, A imposes a per-unit tariff at rate  $t$  on both B and C. As viewed by buyers and sellers in A, this tariff shifts export-supply curves of B and C to  $E_B^t E_B^t$  and  $P_C^t P_C^t$ , respectively. The internal price in country A comes to a rest at  $P_C^t$  with imports from B and C equaling  $OM_1$  and  $M_1 M_3$ , respectively. A's gains from trade (relative to autarky, of course) are represented by triangle KGS plus rectangle GSNH. The triangle is the net change in the consumers' and producers' surplus while the rectangle represents tariff revenue. Since B exports steel, the internal price there equals  $PC$  even if the country happens to have a positive tariff on the books.

Precisely how a preferential freeing of trade by A with respect to B changes the equilibrium depends on the level of the external tariff on steel imports in B in the post-union equilibrium. Initially, consider the simpler, CU case in which B's external tariff on steel coincides with A's. Freeing up of trade between the two countries leads to a single union-wide price,  $P_C+t$ . Country B's export-supply curve, as perceived by agents in A, drops down to  $E_B E_B$ . Since, by construction, imports continue to come from C in the post-CU equilibrium, the domestic price in A remains unchanged at  $P_C^t$ . We have a case of pure trade diversion with imports  $M_1 M_2$  diverted from the more efficient C to less efficient B. No new trade is created.

Though *extra-union* terms of trade are fixed by assumption, *intra-union* terms of trade shift in favor of B by the full amount of the tariff. For A, this shift is manifested in the *transfer of tariff revenue*, GFLH, to exporters in B.<sup>11</sup> Of the total revenue transferred, GFUH

---

<sup>11</sup> This revenue-transfer effect is also present in the general-equilibrium analyses of Eitan Berglas (1979) and Raymond Riezman (1979).



becomes an addition to the gains from trade for B while FLU pays for the higher cost of production of  $M_1M_2$  in B over C. The latter constitutes a deadweight loss. On a net basis, A loses GFLH, B gains GFUH and the union as a whole and world lose FLU.

### 3.3 *Extension to Free Trade Areas*

Traditionally, the analysts focusing on FTAs have assumed that the price facing consumers and producers in each member country is the world price plus its own external tariff.<sup>12</sup> If we were to make this assumption, the extension of the above analysis to FTAs will be straightforward. But, as Martin Richardson (1994) has pointed out recently, this is problematic since producers are free to sell their output anywhere within the union. If the price is higher in country A, producers in B will sell all their output in that country and let the demand in B be satisfied entirely by imports. Gene Grossman and Elhanan Helpman (1995) and Bhagwati and Panagariya (1996a) have further elaborated on this point, showing that it makes the analysis of FTAs more cumbersome than traditionally recognized.

To explain, begin as before with a nondiscriminatory tariff in A at a per-unit rate of  $t_A$ . Make the tariff in B explicit now and denote it by  $t_B$ , where  $t_B < t_A$ . Since we continue to assume that B is a net exporter of steel, the pre-FTA price of steel in that country continues to coincide with the world price. In Figure 3,  $E_B^t E_B^t$  is B's export supply curve, inclusive of  $t_A$  along the vertical axis. Imports come partially from B and partially from C, with each paid the net price of PC. Suppose now that A and B form an FTA with A setting its external

---

<sup>12</sup> For example, see Peter Lloyd (1982) and John McMillan and Ewen McCann (1981).

tariff at  $t_A$  and B at  $t_B$ . Three cases may be distinguished, based on the total supply of steel by A and B in relation to the demand in A in the post-FTA equilibrium.<sup>13</sup>

[Figure 3 here]

Case 1: The Total Supply by A and B is Less than the Demand in A.

Let us first consider the simpler case when  $t_B = 0$ . Since country B is an exporter of steel, this is a plausible assumption. The combined supply of steel in A and B being insufficient to satisfy the demand in A, the latter must necessarily import the product from C. This, in turn, implies that the price in A must settle at  $P_C + t_A$ . With imports into B being free of duty, the price there is  $P_C < P_C + t_A$ . As a result, producers in B divert their entire

---

<sup>13</sup> The following analysis assumes that FTAs are supported by rules of origin that ensure that a lower-tariff member does not import goods from outside to re-export them to a higher-tariff member. If such trade deflection was permitted, ignoring internal transport costs, all imports into the union will be routed through the member with the lowest tariff and the FTA will be turned into a CU. In practice, when goods cross the common border between two FTA members, they qualify for a tariff-free entry only upon presentation of documents proving a within-union origin. Because a product is rarely produced in its entirety in a single country, the rules defining within-union origin can be manipulated to effectively deny a union partner's good the tariff preference. Though the rules of origin have been criticized for their protectionist effects, especially by Anne Krueger (1999), in principle, they can lead to an improvement in efficiency by reversing the trade-diverting effect of a tariff preference on the final good. Rod Falvey and Goeff Reed (1997), Jiandong Ju and Kala Krishna (1998), Panagariya (1999a), and Rupa Duttagupta (1999) offer further discussion of the analytic aspects of the rules of origin.

supply of steel to A. In terms of Figure 3, the FTA leads to a replacement of  $E_B^t E_B^t$  by B's *total* supply curve,  $S_B^1 S_B^1$ .

As in the CU case, we have a revenue-transfer effect from A to B that now equals rectangle EFGH, with no new trade created. Country A necessarily loses. Country B gains rectangle EFGH minus the triangle marked f. It is readily verified that the net gain to B is strictly positive. The loss to the union as a whole is strictly positive and is represented by triangle f.

The conclusion that FTA hurts A which gives the tariff preference, benefits B which receives the preference and hurts the union as a whole remains valid even if we allow  $t_B$  to be strictly positive. Recall that since B exports steel, this tariff is initially redundant. The pre-FTA price in B continues to be  $P_C$ . The formation of the FTA has effects identical to those just discussed for the case when  $t_B = 0$  plus a triangular welfare loss to B and the union as a whole. The additional loss arises because the redundant tariff in B becomes effective in the post-FTA equilibrium. With all of B's supplies diverted to A, its own demand is satisfied by imports from C. But given  $t_B > 0$ , the domestic price rises above  $P_C$  to  $P_C + t_B$ , leading to the triangular welfare loss just mentioned. The reader may verify that even after we take this loss into account, on a net basis, B necessarily gains from the tariff preferences received from A.

#### Case 2: The Total Supply by A and B Equals the Demand in A.

Let us revert back to the assumption  $t_B = 0$ . If B's supply curve lies sufficiently far to the right to cross  $M_A M_A$  below point S, as shown by  $S_B^2 S_B^2$  in Figure 3, the FTA eliminates C as a source of imports to A. This de-links the steel price in A from  $P_C^t$ .

Instead, it is determined by the intersection of A's import-demand curve and B's supply curve.

The net effect of the tariff preference by A is now ambiguous on itself and the union as a whole and, as before, non-negative on B. With the decline in the internal price, new trade in the amount KL is created, which is associated with a rise in the union's welfare equal to triangle SLU. At the same time, since A's protection is extended to B's firms, there is harmful trade diversion: the cost of production of units ZV (previously imported from C) in B exceeds that in C by area UVZ. The union as a whole gains or loses as area SLU is larger or smaller than area UVZ. The farther to the right does  $S_B^2 S_B^2$  lie, the larger is SLU and smaller UVZ. In the limit, if  $S_B S_B$  crosses  $P_C P_C$  at or to the right of V, area UVZ disappears altogether and the union as a whole necessarily benefits.

Turning to A, it gains area SKL from trade creation but loses area WKVH due to tariff-revenue transfer to B's exporters. The remaining part of tariff revenue, GSKW, becomes a part of A's own consumers' and producers' surplus. The farther to the right B's supply curve lies, the closer is A's internal price to PC and more likely that it will be a net gainer. In the limit, if the internal price drops to PC, no revenue transfer to B takes place and there is benefit from trade creation, implying a net gain.

Finally, the effect on B's welfare is non-negative. As drawn in Figure 3, the price received by its exporters as well as the quantity of exports rises. It benefits on both counts, receiving a net gain of WLZH. In the limiting case when B's supply is sufficiently large that the price in A drops to  $P_C$ , it makes no gain but it also does not lose.

This analysis is modified if  $t_B$  is positive. As long as  $P_C + t_B$  is less than the height of point L in Figure 3, the modification is minor. We continue to obtain the effects just

discussed but, in addition, have a triangular welfare loss in B. As in Case 1, this results from the initially redundant tariff becoming effective. If  $P_{C+t_B}$  exceeds the height of point L, the modification is more substantive. The reason is that the price in B's market in this case exceeds the height of point L. Therefore, no producer within the union will sell at the price indicated by point L. In effect, the price in A cannot fall below  $P_{B+t_B}$ . With the union-wide price settling at  $P_{B+t_B}$ , the combined supply of A and B exceeds the demand in A, bringing us to Case 3.

Case 3: The Total Supply by A and B Exceeds the Demand in A.

If B's supply curve intersects MAMA at a price below  $P_{C+t_B}$ , the union-wide price settles at  $P_{C+t_B}$ . The key difference with the previous case is that producers in B are now indifferent between markets in A and B. But welfare effects are unchanged qualitatively: B benefits while A and the union as a whole may or may not benefit. The lower is  $t_B$ , the more likely that the union as a whole and A benefit. In the limit, as  $t_B$  approaches zero (and B's supply curve, therefore, crosses  $M_A M_A$  below point R), the FTA degenerates into the free-trade equilibrium, with the price in both A and B dropping to  $P_C$ . In this limiting case, A and the union benefit while B neither gains nor loses.

### 3.4 *The Meade-Lipsey General-Equilibrium Model*

The above analysis suggests that if we are seeking unambiguous gains from a CU or FTA, we must look for sectors in which the partner country is the sole source of imports even at the initial equilibrium. In such sectors, there is no outside trade to be diverted in the first place. Maintaining the small-union assumption, the point is illustrated in Figure 4. Given  $P_C$  as the price in the rest of the world, the export-supply curve of B facing country A

is HUEB. Under a nondiscriminatory tariff, the supply curve, as perceived by consumers and producers in A, is  $GFE_B^t$ . Country A imports GS from B and collects rectangle f in tariff revenue. Country B exports GS to A and SF to C. There is no trade between A and C. As A lowers the tariff on B,  $GFE_B^t$  shifts down with the internal price in A declining by the full amount of tariff reduction. Country A's trade with B expands and welfare in A rises every step of the way. With country C essentially out of the picture—though not entirely since it is needed to fix the external price at  $P_C$ —the possibility of trade diversion as well as revenue-transfer effect is ruled out.

By itself, this case is uninteresting since it avoids trade diversion by assumption and leads to an FTA that is effectively equivalent to free trade. The case can be made to yield something more interesting, however, if it is embedded in a general-equilibrium model. This was demonstrated by Lipsey (1958), using Meade's three-good framework.<sup>14</sup> Thus, suppose there are three goods, 1, 2 and 3. Assume that A specializes completely in good 1 and exports it to B and C while B specializes completely in good 2 and exports it to A and C. Country C produces all three goods and exports good 3 to A and B. Country C is sufficiently large that A and B act as price takers in its market. By appropriate choice of units, we can set the prices of all three goods in C at unity.

---

<sup>14</sup> The three-good model to be outlined below originated in Meade (1955). But whereas Meade focused on the effects of preferential trading in this model on the world welfare, Lipsey (1958) analyzed the effects on the member countries assuming the small-union context. The small-union model has been explored further by McMillan and McCann (1981) and Lloyd (1982).

Consider now country A. Suppose it initially imposes tariffs at rates  $t_2$  and  $t_3$  on goods 2 and 3, respectively, where  $t_2 = t_3 \equiv \bar{t}$ . Given all prices in C equal unity, prices in A for goods 1, 2 and 3 are 1,  $1+t_2$  and  $1+t_3$ , respectively. Preferential trade liberalization involves lowering  $t_2$  without lowering  $t_3$ .

The effect of a small reduction in  $t_2$  in sector 2 can be gleaned from Figure 4. Preferential liberalization lowers the price of good 2 in A and leads to trade creation in this sector. Denoting the rise in the imports of good 2 by  $dM_2$ , the associated welfare gain is represented by  $t_2 dM_2 > 0$  as shown by the shaded strip in Figure 4.

[Figure 4 here]

But this is not the end of the story. Assuming the demand for good 2 exhibits substitutability with goods 1 and 3, the reduction in the price of the former leads to a reduction in the demands for the latter. Imports of good 3 fall and exports of good 1 rise. Since good 3 is imported from C, the decline in its imports can be characterized as trade diversion. Moreover, since good 3 is subject to a tariff, the diversion is associated with a welfare loss. For a small change in  $t_2$ , this welfare effect can be written  $t_3 dM_3 < 0$ . The net welfare effect depends on whether  $t_2 dM_2 + t_3 dM_3$  is positive or negative.

It can be shown that, starting with  $t_2 = t_3 \equiv \bar{t}$  and assuming substitutability between goods 1 and 2 (the good being subject to liberalization and the exportable, respectively), for a small reduction in  $t_2$ , the benefit from trade creation dominates the loss from trade diversion. Recall that the tariff reduction lowers the demand for good 1 thereby releasing goods for exports. Since exports rise, the trade-balance condition implies that total imports, valued at world prices, must rise as well. Given that all world prices have been normalized

to unity, the rise in imports of good 2 is larger than the decline in imports of good 3. That is to say,  $dM_2 > -dM_3$ , which, given  $t_2 = t_3 = \bar{t}$ , implies  $t_2 dM_2 + t_3 dM_3 > 0$ .

Though a small preferential reduction in the tariff is, thus, beneficial, pushing preferential liberalization all the way to free trade may be harmful. After the initial reduction in  $t_2$ , we have  $t_2 < t_3 = \bar{t}$  so that  $dM_2 > -dM_3$  no longer necessarily implies  $t_2 dM_2 + t_3 dM_3 > 0$ . Indeed, as  $t_2$  approaches 0, the weight of the positive term in this expression also approaches zero. The likely pattern of welfare as  $t_2$  moves from  $t_2 = t_3 = \bar{t}$  towards  $t_2 = 0$  is shown in Figure 5. There is no guarantee that welfare at  $t_2 = 0$  will be higher than at  $t_2 = t_3$ . That is to say, the FTA may lower or raise welfare.<sup>15</sup>

[Figure 5 here]

Though the Meade-Lipsey model has been influential in the literature on preferential trading, it suffers from four key limitations. First, in this model, the member countries have no incentive whatsoever to coordinate their liberalization through an FTA. The liberalization by one member has no effect whatsoever on the economy of the other member and vice versa. The formation of the FTA is identical to a unilateral trade reform.<sup>16</sup> Second,

---

<sup>15</sup> Indeed, in general, we cannot even be sure that the initial tariff preference is welfare improving. If good 2 exhibits complementarity with good 1, exported by A, the tariff preference increases the demand for good 1 in A and, thus, lowers its exports. Via trade balance condition, this change leads to a greater decline in the imports of good 3 than the increase in the imports of good 1 and makes  $t_2 dM_2 + t_3 dM_3 < 0$  even at  $t_2 = t_3 = \bar{t}$ .

<sup>16</sup> Not surprisingly, the result just discussed is identical to the concertina theorem of piecemeal trade reform according to which a reduction in the highest tariff to the next highest one by a small country



the model assumes a very specific structure of trade flows. Recall that our partial-equilibrium analysis in Sections 3.2 and 3.3 emphasized the importance of the case in which the good imported from the partner is also imported from the rest of the world. The Meade-Lipsey model rules out this case entirely by assumption. Third, as traditionally analyzed, it rules out the arbitrage in producer prices within the union by imposing the assumption that the price in each union member equals the world price plus its own tariff. Finally, the analysis is conducted in terms of infinitesimally small changes in tariffs. As such, it does not offer clear conditions for an improvement in welfare following a move to complete FTA.

### 3.5 *A More General Analysis of the Small-Union Case*

In a recent paper, Panagariya (1999c) offers a more general analysis of the small-union case, which is able to overcome all these limitations. He allows for all possible patterns of trade by incorporating goods that are exported by both partners, those that are imported by both partners and those that are exported by one partner and imported by the other. He also considers large changes in tariffs and compares pre- and post-FTA equilibria directly. All goods may be produced in all countries with producer prices fully arbitrated across the union in the post-FTA equilibrium. Finally, on the demand side, the only restriction imposed is that all goods are normal in consumption. Even this restriction is a

---

is beneficial provided the good with the highest tariff exhibits substitutability with respect to the exportable.

sufficient rather than necessary condition for the results to hold. In production, as usual, constant returns to scale are assumed.

This level of generality is achieved at the cost of two key restrictive assumptions, however. First, there are no redundant tariffs in the initial equilibrium. For example, if a good is exported in the pre-FTA equilibrium, no tariff is imposed on it. This assumption is commonly made in general-equilibrium analyses of PTAs but is not entirely innocent as we saw in Section 3.3.<sup>17</sup> Second, the union member with the higher tariff on a good continues to import that good after the formation of the FTA. This ensures that post-FTA prices remain linked to world prices via the relevant tariff rates. Once again, this assumption is commonly made in general-equilibrium analyses but rules out Cases 2 and 3 discussed above.

Panagariya (1999c) demonstrates that an FTA between two countries increases or reduces the union's joint welfare as it increases or reduces the value of the union-wide output at world prices. If the production of the numeraire good (exported by both member countries) requires only labor and the production of all other goods requires a specific factor and labor, the FTA necessarily lowers the value of the union's output at world prices and hence its joint welfare.

Next, revert back to the general production structure and suppose a small country removes partially or wholly its tariffs on another small country, retaining them at their original levels against the rest of the world. Assume as before that, after the preferential

---

<sup>17</sup> In defense of the assumption, it may be noted that if a political-economy process such as that in Grossman-Helpman (1994) generates initial tariffs, no redundant tariffs are obtained.

liberalization, the country continues to import each good from the rest of the world. Then the country's welfare necessarily declines, with the loss in real income equaling the lost tariff revenue on the imports from the union partner. The more the country imports from the partner and the greater the magnitude of tariff preference, the more it loses. If we further assume that the numeraire good uses only labor and other goods use a specific factor and labor, the partner country necessarily benefits from the tariff preference. And the more it exports to the union member giving the tariff preference and the higher the margin of preference, the more it gains. These results imply that if two small countries with approximately balance bilateral trade form an FTA, the member with higher tariff is likely to lose.

These results contradict the so-called "natural-trading-partners" hypothesis, enunciated by Paul Wonnacott and Mark Lutz (1989) and espoused by Lawrence Summers (1991) and Paul Krugman (1993). According to this hypothesis, the more two countries trade with each other relative to the outside world, the less likely that a union between them will be harmful. The results relating to the joint welfare of the union are independent of the volume of trade. Instead, they depend on the value of output at world prices. In the specific-factors case, the decline in the union's welfare is independent of the volume of trade as well. From the viewpoint of an individual union member, the volume of trade can work against it. *Ceteris paribus*, the more it imports from the partner, the larger its losses or smaller its gains.<sup>18</sup>

---

<sup>18</sup> A systematic and compelling critique of the natural trading partners' hypothesis can be found in Bhagwati and Panagariya (1996a).

These results are modified if we abandon the assumption that imports continue to come from the outside country into the member with higher tariff after the FTA is formed. Allowing the partner to become the sole supplier of one or more products takes us into the realm of Cases 2 and 3 discussed above. The internal price in the country, which gives the tariff preference, is no longer tied to the world price through its tariff. Instead, it declines, permitting new trade to be created, which contributes positively to welfare. Trade diversion is also smaller due to a smaller difference between the union's and C's price. The larger the price decline, the larger the positive effect or the smaller the negative effect on the union.

### 3.6 *The Large-union Case*

Some of the strong conclusions in the previous section are also modified if the outside country's export supplies are less than perfectly elastic. As noted in Section 3.1, in this large-union case, trade diversion is beneficial to the union: the diversion of demand away from C as a consequence of preferential liberalization is likely to improve the union's terms of trade relative to the outside country.

Interestingly, however, even in this case, the effects on the welfare of the preference-granting and preference-receiving country remain asymmetric along the lines indicated in Section 3.5. To see this, let us return briefly to the Meade-Lipsey model, which, despite the limitations noted earlier, remains suitable for analyzing the implications of preferential trade liberalization for the terms of trade. Suppose country A gives a one-way tariff preference to country B. At constant border prices, the change increases A's demand for good 2 and reduces it for goods 1 and 3. The price of good 2 increases relative to goods 1 and 3, which implies that B's terms of trade improve with respect to both trading partners. The effect on the relative price between goods 1 and 3 is ambiguous. Therefore, A's intra-union terms of

trade deteriorate while its extra-union terms of trade may improve or worsen. Thus, the conflict between the interests of the country offering tariff preference and the one receiving it, central to the discussion in Sections 3.3 and 3.5, is resurrected even in the model in which the good imported from the partner is not imported from the rest of the world.

In a neglected but important paper, Robert Mundell (1964) has formally studied the Lipsey-Meade model with flexible terms of trade. Assuming import demands for all goods exhibit gross substitutability and that initial tariffs are low, he reached the following conclusions:

"(1) A discriminatory tariff reduction by a member country improves the terms of trade of the partner country with respect to both the tariff reducing country and the rest of the world, but the terms of trade of the tariff-reducing country might rise or fall with respect to third countries.

(2) The degree of improvement in the terms of trade of the partner country is likely to be larger the greater is the member's tariff reduction; this establishes the presumption that a member's gain from a free-trade area will be larger the higher are initial tariffs of partner countries." [Mundell (1964), 8].

Interestingly, the revenue-transfer effect emphasized in Sections 3.2 and 3.3 comes back to dominate the outcome. Intra-union terms of trade move against a country and in favor of the partner when the country offers a tariff preference.<sup>19</sup>

---

<sup>19</sup> Panagariya (1997a) has recently extended Mundell's (1964) analysis by decomposing the total welfare effect into a pure efficiency effect, an intra-union terms-of-trade effect and an extra-union terms-of-trade effect.

### 3.7 *A Differentiated Products Model*

So far, we have assumed that goods are homogeneous. It is sometimes asserted that the results derived from homogeneous-goods models are dramatically altered once we allow for differentiated goods.<sup>20</sup> At least for the problem at hand, this is an incorrect assertion. The main complication in the presence of differentiated goods is that we can no longer use the simplifying, small-union assumption. Each country has monopoly power over its products and can influence its terms of trade. For example, as Daniel Gros (1987) has demonstrated, in this setting, the optimum tariff for a country, no matter how small, is strictly positive and finite.

It is easy to see that Mundell's analysis, quoted above, can be brought to bear on the differentiated-products case. Assume, as in Krugman (1980), that there is a single good in the economy with a large number of potential varieties. The consumer preferences are symmetric and CES over these varieties. Furthermore, there is a single factor of production, labor, and the cost function of a representative variety is characterized by a fixed cost and a constant marginal cost. Free entry drives all profits to zero. We know from Krugman (1980) that, in this model, the equilibrium output of each variety is fixed and, for a given labor force, the equilibrium number of varieties is also fixed.

---

<sup>20</sup> The implications of differentiated goods should be distinguished from those of economies of scale. Though both are present in the Krugman (1980) monopolistic competition model on which this subsection is based, economies of scale do not play a substantive role in it. The implications of economies of scale are discussed below separately. They also figure in some of the contributions discussed in Section 7 (e.g. Philip Levy 1997 and Wilfred Ethier 1998).

Remembering that the CES form of the utility function implies substitutability in demand, this model reduces to a special case (in terms of the generality of the utility function) of Mundell's model discussed in Section 3.4. The only cosmetic difference is that each country produces several varieties. But since each country's varieties are symmetric in all respects, they can be aggregated into a single product and Mundell's analysis invoked.<sup>21</sup>

### 3.8 *Transport Costs*

Perhaps guided by the observation that, *in practice*, PTAs often form among countries that are geographically proximate, some analysts have gone on to argue that low transport costs make them more likely to be beneficial.<sup>22</sup> This is new development. For example, in his comprehensive work, Viner (1950) had noted the presence of departures from the Most Favored Nation (MFN) principle in commercial pacts between countries within Europe going as far back as the nineteenth century. But rather than link these departures to low transport costs within Europe, he attributed them to 'close ties of sentiment and interest arising out of ethnological, or cultural, or historical political affiliations'.

---

<sup>21</sup> Product differentiation and its implications for trade creation and trade diversion will appear more explicitly later in Section 5 when we consider Krugman's analysis of a simultaneous division of the world into two or more blocs. Also see Section 3.9 below.

<sup>22</sup> In particular, see Wonnacott and Lutz (1989), Krugman (1991b, 1993), Summers (1991) and Frankel, Ernesto Stein and Shang-Jin Wei (1995).

There is little basis for giving transport costs a special treatment in evaluating PTAs.<sup>23</sup> Leaving aside the possibility that sufficiently high transport costs can eliminate the scope for mutually beneficial trade between countries, the principle of comparative advantage and the proposition on the optimality of nondiscriminatory free trade (from the global standpoint) are valid with and without these costs. Lower transport costs may give a proximate trading partner a cost advantage over a distant one, but this may be outweighed by lower production costs in the latter. This is certainly demonstrated by the ability of distant countries in East Asia to compete effectively with Latin American countries in the U.S. market.

Indeed, even a *ceteris paribus* proposition that PTAs among proximate partners are superior to distant ones is not valid in general. Thus, Bhagwati and Panagariya (1996a) provide an example in which, between two otherwise identical potential partners, a country achieves a superior outcome by giving the trade preference to the distant one. The reason is that, with an initial nondiscriminatory tariff, the country imports less from the distant partner. A preference to that partner leads to a smaller transfer of tariff revenue than to the proximate one.

---

<sup>23</sup> For detailed critiques, see Bhagwati and Panagariya (1996a) and Panagariya (1998). Paul Wonnacott and Ronald Wonnacott (1981) give a special role to transport costs but, as shown by Berglas (1983) and further discussed in Panagariya (1998), their examples require transport costs to be sufficiently high to rule out the distant partner as a trading partner either entirely or in the pre-union equilibrium.



### 3.9 *Economies of Scale and Imperfect Competition*

Let us now turn to a discussion of the implications of economies of scale and imperfect competition for the theory of preferential trading. The first point to note is that even though many models incorporate economies of scale and imperfect competition simultaneously into the analysis, they can be treated separately. Indeed, the discussion in this section begins with a simple model of economies of scale in which the assumption of perfect competition is maintained. Likewise, several contributions discussed later in the essay incorporate imperfectly competitive market structure without recourse to economies of scale.

Consider then a partial-equilibrium model with economies of scale where the good in question, steel, is homogeneous.<sup>24</sup> To maintain the assumption of perfect competition, assume that scale economies are external to the firm. Make the further simplification that scale economies derive from the industry-wide output of steel. These assumptions allow us to represent the industry-wide average cost of production by a downward sloping curve such

---

<sup>24</sup> Corden (1972) first analyzed the implications of economies of scale for preferential trading in a homogeneous goods model. The following discussion is heavily influenced by his contribution but differs from it in details. Corden assumes a single producer of the product in each country with price determined by the world price plus external tariff. He also allows for transport costs, which introduce a wedge between the prices at which steel can be imported from and exported to the outside country. By contrast, I assume external economies and average-cost pricing and do not allow for transport costs.

as AA' in Figure 6. The precise form of the scale economy is such that the cost curve first declines, reaches a minimum and becomes horizontal thereafter.

[Figure 6 here]

To bring out the implications of scale economies most sharply, it is best to focus on the case in which countries A and B are identical in all respects. Thus, let AA' represent the average cost of steel production in both countries. Similarly, let DD' be the demand in each country. UU', the horizontal sum of the two demand curves, represents the combined demand for steel by union members in the absence of trade barriers between them. Initially, the two countries levy a non-discriminatory tariff at the same rate.

The first point to note is that the outcome in this setting depends critically on the price prevailing in C,  $P_C$ . If this price is above the minimum average cost along AA', trade barriers in A and B are redundant. Producers in each country produce the maximum quantity of steel they can, satisfy the domestic demand at  $P_C$  and export the remainder of steel. The formation of an FTA, which happens to be equivalent to a CU given the same external tariff in the two countries, leaves the equilibrium entirely undisturbed.

If  $P_C$  is below the minimum cost of production in A and B as shown by  $P_C P_C$  in Figure 6, however, the outcome is different. Consider first the case in which C's tariff inclusive price,  $P_C^t$ , is between points E and F. In the initial equilibrium, both countries import steel from C. In each country, the area under DD' up to price  $P_C^t$  represents the consumers' surplus while area a+c represents tariff revenue. The formation of an FTA permits one of the two union members to enter production by exploiting the larger union-wide market. Though trade is diverted from C, declining costs lead to a decline in the union's internal price. The former effect is harmful and is accompanied by the usual

revenue loss but the latter is beneficial. Each country loses area  $c$  due to trade diversion while gaining area  $b$ . The net effect on welfare is ambiguous.

Outcomes more favorable to PTAs can arise under higher initial tariffs, however. Thus, suppose  $P_C^t$  lies above point  $E$ . Then, in the initial equilibrium, domestic production is viable and both countries produce and consume their own steel. If they now form an FTA, setting the external tariff at its original level, one of the two countries ceases to produce steel. The union-wide price declines to  $P_U$  and each member gains a trapezium-shaped area. With no trade with  $C$  initially, there is no scope for trade diversion and the CU is unambiguously welfare improving. Due to the decline in the average cost, made possible by economies of scale, the gain is larger than the traditional welfare triangles.

A key difference between this model and those considered before is that we now have a stronger tendency towards the elimination of country  $C$  as a source of imports in the post-FTA equilibrium. As is common with scale-economy models, we get an all or nothing outcomes with respect to imports: either all steel comes from the outside country or none does. Such outcomes are also accompanied by a de-linking of the internal price with the outside country's price (precisely as in Cases 2 and 3 above) and generate gains that counteract revenue losses from trade diversion.

An alternative approach to analyzing the implications of economies of scale is to combine them with imperfect competition and, often, product differentiation. Some of the key applications of in this category of models have been developed in the context of the more recent issues such as the welfare implications of a simultaneous division of the world into several trade blocs and the expansion of a bloc through addition of new members. These applications will be discussed later in the context of the relevant questions. The

survey by Baldwin and Venables (1995), on the other hand, discusses the implications of imperfect competition and scale economies in the three-country framework employed in the present section. While these authors carefully discuss the scale and cost effects of FTAs, identifying the relevant channels, they do not explicitly derive the welfare effects. It is presumably this complexity which has led some researchers in this strand of the literature to resort to numerical simulations.<sup>25</sup>

#### 4. Welfare-Increasing CUs and FTAs

The preceding analysis fixes the pre-union external *tariffs* and allows external *trade* flows to adjust endogenously as intra-union trade barriers are removed. The welfare effects on the union in this setting turn out to be either negative or ambiguous but never unambiguously positive. Remarkably, if we take the opposite approach, fixing the initial extra-union *trade* flows and letting the external *tariffs* adjust endogenously, the outcome is essentially the opposite. Regardless of whether potential members are small or large, neither the union as a whole nor the rest of the world can lose from a CU or FTA and the union is likely to benefit.

##### 4.1 *Customs Unions*

This result was first stated for a CU by Murray Kemp (1964) and Vanek (1965) independently and proved by Ohyama (1972) and Kemp and Henry Wan (1976) and will be

---

<sup>25</sup> Thus, see Alasdair Smith and Venables (1988) and Diego Puga and Venables (1995).

called the Kemp-Wan-Vanek-Ohyama theorem in this paper.<sup>26</sup> The logic behind the theorem is simple. Freezing the net trade vector of A and B with the rest of the world ensures that the rest of the world can be made neither better off nor worse off by the union. Then, taking the external trade vector as a constraint, the joint welfare of A and B is maximized by equating the marginal rate of transformation (MRT) and marginal rate of substitution (MRS) for each pair of goods across all agents in the union. This is, of course, accomplished by eliminating all intra-union trade barriers and setting the common external tariff (CET) vector at a level just right to hold the extra-union trade vector at the pre-union level.

To get an idea of the CET and welfare effects on member countries, let us consider a diagrammatic illustration of the Kemp-Wan-Vanek-Ohyama theorem.<sup>27</sup> In Figure 7, various curves have the same interpretation as in Figure 2 except that we do not show the export supply of C, which may or may not be horizontal. Let  $P_A$  be the pre-CU domestic price in country A, with quantity GV imported from B and VS from C. Since UV is per-unit tariff initially, the price in B and C is given by the height of point U.

[Figure 7 here]

As a part of the CU, A and B eliminate all trade barriers between themselves and set the CET at a level freezes their joint quantity of imports from C at VS. To derive the

---

<sup>26</sup> In their survey, Baldwin and Venables (1995) attribute the result to Meade (1955). In Panagariya (1997b), I argue conclusively that this is an error.

<sup>27</sup> Srinivasan (1997) derives the external tariff in the Kemp-Ohyama-Kemp-Wan CU within a two-sector general-equilibrium model. Also see the comment on this paper by Davis (1997).

resulting equilibrium, at every price, subtract  $VS$  from  $M_A M_A$  and obtain  $M'_A M'_A$  as the import demand to be satisfied by B. Since no tariffs apply to B now, its export supplies are given along  $E_B E_B$ . The market-clearing price, in turn, is  $P'_A$ . Country A imports  $LN (> GV)$  from B and  $NT (= VS)$  from C. Since the imports from C are unchanged, the price in that country is still given by the height of point U, yielding  $UF (< UV)$  as the CET per-unit. The expansion of intra-union trade leads to a net gain for the union equal to area  $f+g$ . The reader can verify that, holding imports from C fixed at  $VS$ , this is the best the union can do.

Observe that the external tariff falls due to the fact that, at constant tariff rate, trade would be diverted from C, causing imports from the latter to decline. To maintain the imports from C at their original level, the external tariff must fall.

It can be seen that, in the spirit of Sections 3.2 and 3.3, country B necessarily gains while country A may or may not gain despite total absence of trade diversion. With outside imports held fixed, the internal price declines by the full amount of the decline in the external tariff. The revenue lost on the imports from C (due to the reduction in the external tariff) is redistributed to A's consumers. But, to the extent of the tariff preference, the revenue lost on imports from B is redistributed to the latter's exporting firms. The loss to A on this account is area  $h$  which must be compared against the gain  $f$  on the new trade created with B. The gain to B is  $h+g$  with  $h$  being the redistribution from A and  $g$  the gain on new intra-union trade. This analysis suggests that in a multi-commodity setting, if trade is approximately balanced between the partners, the member with high initial tariffs will lose and the member with low initial tariffs will benefit from the Kemp-Ohyama-Kemp-Wan CU.

## 4.2 Free Trade Areas

Proving an analogous result for FTAs is tricky. If we freeze member countries' *individual* trade vectors with the rest of the world, the resulting external tariff vectors will, in general, be different for different member countries. This means the condition  $MRT = MRS$  for each pair of countries across all agents in the union cannot be satisfied in general.

Recently, Panagariya and Krishna (1997) have overcome this difficulty, however. Fully in spirit of the Kemp-Ohyama-Kemp-Wan theorem, they show that if two or more countries form an FTA, freezing their initial, *individual* trade vectors via country-specific tariff vectors, welfare of neither the union nor the rest of the world falls and that of the former is likely to rise.

The key to explaining this result lies in the analysis in Section 1.3. For the products for which within-union supply is sufficiently large that the union-wide price coincides with the price in the lower-tariff country (i.e., when B's supply curve in Figure 3 crosses  $M_A M_A$  below point W), we effectively obtain a CU with the  $MRT = MRS$  condition satisfied union-wide. For products for which the union-wide supply is smaller, the domestic price is lower in the lower-tariff country. But in this case, the entire union-wide output is sold in the high-tariff country so that the marginal rates of transformation are equalized across the union members. Furthermore, the marginal rate of substitution in the high-tariff country is also equalized to this marginal rate of transformation. Only the marginal rate of transformation in the lower-tariff country is lower. But given the requirement that *individual* import vectors be frozen, this is also the best that can be done. Any move from FTA necessarily lowers the union's joint welfare.

#### 4.3 *Customs Unions with Non-economic Objectives*

Recently, Krishna and Bhagwati (1997) have shown that if two or more countries are pursuing certain non-economic objectives, they can still form a CU between themselves and be jointly better off. The result relates to an old issue discussed in the development literature: given any level of import substitution vis-a-vis the developed countries, can the developing countries open up trade preferentially among themselves and reduce the cost of their individual import substitution? At the time, an affirmative answer had been given by authors but relying on the presence of economies of scale. Krishna and Bhagwati (1997), by contrast, show that scale economies are not essential to the argument. The solution involves a Kemp-Ohyama-Kemp-Wan CU complemented by tax-cum-subsidies to achieve the non-economic objectives of member states as indicated by the theory of optimal intervention in the presence of non-economic objectives.

### 5. Exogenous Division of the World into Blocs

The massive wave of regional arrangements that started in the 1980s raises the question how the welfare of the world, individual blocs and individual countries will change if the world is divided simultaneously into several blocs. We may also ask how the welfare of these entities changes with the number and size of blocs.

We have already seen that the analysis of preferential liberalization turns out to be complex even when we consider the formation of a single bloc in isolation and impose the small-union assumption. Therefore, it should come as no surprise that when several blocs are allowed to form simultaneously, with inter-bloc terms of trade allowed to change, strong assumptions must be made.



### 5.1 *Symmetric Blocs*

The simplest approach to the problem is to imagine that all countries are identical and then consider their division into two or more identical blocs. This simplification makes the problem tractable since it implies that the welfare of each country and bloc must move in the same direction. It also allows us to relate the number of blocs to welfare in a straightforward manner. Thus, following Krugman (1991a), let us postulate a world consisting of a large number of small, identical units, called 'provinces.' Each province specializes in the production of a distinct good. Products of all provinces enter symmetrically into the utility function with an identical, constant elasticity of substitution between each pair of products.

Assume that the world is divided into  $B$  identical blocs where  $B$  is exogenous. There are no barriers on within-bloc trade and a common external tariff on extra-bloc trade. Given complete symmetry, the external tariff of each bloc is the same. Though each bloc acts as a Nash player and chooses the external tariff optimally, since this endogeneity is not crucial to the results [Krugman (1993)], it is best not to introduce it at this stage. A key point to bear in mind is that, given complete symmetry, a change in the number of blocs and, hence, the size of each bloc, generates no terms of trade effects. Welfare outcomes are driven entirely by the effect of bloc expansion on efficiency. This is a very special feature of a model with endogenous terms of trade.

Begin with a single bloc initially so that we have worldwide free trade. Given no distortions, this naturally maximizes the welfare of each province and the world. Suppose that we divide this bloc into two equal-sized blocs. This division leads to trade diversion: each province trades more with the provinces in the same bloc at the expense of the

provinces in the other bloc. With no trade creation to offset this effect, welfare necessarily declines.

Suppose next that we take one third of the provinces of each existing bloc and create a third bloc. We now have trade creation as well as trade diversion. Seen from the viewpoint of the provinces within an existing bloc, trade diversion results from the decline in trade with the provinces that have just been moved outside to create the new bloc. Trade creation results from an expansion of trade with provinces that were already outside and subject to the external tariff. The net effect on welfare is ambiguous. Given the symmetry of blocs, what applies to provinces within an existing bloc also applies to provinces in the newly created bloc. We need not analyze the latter separately.

It can be shown that as the number of blocs grows, the trade creation effect must come to dominate the trade diversion effect. Given a large number of blocs, the representative bloc is small and most of its trade is with outside provinces. Therefore, when another bloc is created, the expansion of trade with these outside provinces dominates the contraction of trade with the provinces that are moved out to create the new bloc. Welfare must rise.

This basic story is reinforced when external tariffs are chosen endogenously, with each bloc acting non-cooperatively. As the number of blocs rises, each bloc becomes smaller and its optimum tariff declines. The trade creation effect is reinforced.

Remembering that the initial division of the world into two blocs necessarily lowers welfare, this discussion implies that welfare exhibits a U-shaped pattern as a function of the number of blocs. Krugman (1993) simulates the model for an elasticity of substitution of four and finds the number of blocs at minimum-welfare point to be 2 for a tariff rate of 10

percent and 3 for tariff rates of 20 and 30 percent. In view of the fact that the world may be dividing into precisely two or three blocs, this is a provocative result.

### 5.2 *Asymmetric Blocs*

Given the special, highly symmetric structure of the model just considered, its results turn out to be fragile. Thus, Alan Deardorff and Robert Stern (1994) provide a simple example in which a small number of blocs lead to the maximization of world welfare. They assume that all goods are homogeneous so that trade is of inter-industry type. Economies of scale are ruled out. Assuming further that there are  $n/2$  types of different countries where  $n$  is the total number of countries, we can divide the world into two identical blocs such that each of them consists of exactly one country of each type. This allows each bloc to exploit all the gains from trade without any trade whatsoever with the other bloc. Alternative examples in which welfare bears no relationship to the number of blocs can also be constructed. T.N Srinivasan (1993) does this within a Ricardian, constant costs model.

### 5.3 *"Natural" and "Unnatural" Blocs*

Following the suggestion in Krugman (1991b) that proximity between member countries minimizes trade diversion, Frankel, Stein and Wei (1995, 1998) have gone on to extend the Krugman (1993) model to incorporate transport costs. Like Krugman (1993), these authors assume a highly symmetric world: identical continents with identical and equal number of countries that follow identical trade policies before as well as after the formation of CUs. The only difference with Krugman is that there are (identical) positive transport costs of moving goods between countries on different continents but not those on

the same continent. By analogy with the melting of iceberg, transport costs take the form of a fraction of each unit of a good being lost in transit.

Frankel et al. consider two types of blocs: (i) Continental blocs such that each bloc consists of all countries on the same continent but no others; and (ii) Across-continent blocs such that each bloc consists of exactly one country from each continent. The authors call the former "natural" and the latter "unnatural" blocs. In either case, each bloc has the same common external tariff that equals the initial tariff.

This model works entirely like the Meade-Lipsey model discussed in Section 3.4. Because of the symmetry, bloc formation has no effect whatsoever on the terms of trade. Therefore, bloc formation is like the formation of a CU between two small countries. Moreover, products imported from the partner are not imported from the rest of the world and vice versa just as in the Meade-Lipsey model. Finally, the substitutability between own and partner-country products also obtains since preferences are symmetric. Applying the logic outlined in Section 3.4, welfare follows an inverted-U path as in Figure 5. In general, welfare of a member country when all within-bloc tariffs have been eliminated may be higher or lower than at the initial equilibrium.

This result applies to both "natural" and "unnatural" blocs and is confirmed by the simulations done by Frankel et al. Assuming three continents, two countries per continent, and an external tariff of 30 percent, they find that if transport costs (between continents) result in the "melting away" of 15% or less of the product in transit, bloc formation reduces welfare whether blocs are natural or unnatural. Unnatural blocs do consistently worse than natural ones in these simulations, however. Moreover, as transports costs rise above 15% of the product, natural blocs become welfare superior to the initial equilibrium. As expected,

when transport costs rise to 100 percent of the product, thereby precluding trade between continents, natural blocs lead to the same outcome as global free trade. Unnatural blocs always remain inferior to the status quo though the harm done by them is less and less as transport costs rise.

From Figure 5, we can deduce that a partial tariff preference up to the point where welfare is maximized ( $t_2 = t_2^{opt}$ ) will yield a superior outcome in the Frankel et al. model than either the initial nondiscriminatory tariff or a full CU. This is confirmed by various simulations that the authors undertake. Based on this result, Frankel et al. conclude, "some degree of preferences along natural continental lines...would be a good thing,..."

This is a questionable conclusion. What the authors have done is to provide an *example* (with highly unrealistic assumptions, I might add) in which sufficiently high transport costs make blocs among neighbors welfare superior to status quo. But it is equally possible to construct examples showing the opposite result, as done in Panagariya (1998). Thus, consider a two-good Ricardian world. Assume there are two continents, each consisting of two countries. Transport costs are zero within each continent but positive, *albeit* non-prohibitive, across continents. The opportunity costs of production are identical between countries on the same continent but different between countries on different continents. If blocs are now formed between countries on the same continent, there is no change in welfare. But if they are formed between countries located on different continents, welfare improves and is the same as under global free trade. The point here is that benefits from trade, in general, and preferential trade, in particular, depend on differences in costs between trading partners irrespective of the sources of these differences. As argued in Section 3.8, transport costs are not special.

## 6. Endogeneity of Policy

So far, we have assumed that the decision to form trade blocs as well as the choice of the external tariff is exogenous.<sup>28</sup> In view of the developments on endogenous choice of policy, it is natural to ask under what circumstances countries are likely to exercise the option to form FTAs and, alternatively, how is the decision to form an FTA likely to affect the choice of the external tariff. These questions take us into the realm of political-economy-theoretic analyses.

### 6.1 *The Decision to Form an FTA*

Let us revert back to the three-country set up of Section 3 but make the decision to form FTA endogenous. The central question we now ask is whether an FTA is more likely to form when it is largely trade creating and hence welfare improving or when it is trade diverting and therefore welfare reducing. We also ask whether the exclusion of certain sectors from preferential liberalization and the rules of origin make an otherwise infeasible FTA feasible.

#### A Perfect Competition, Small-Union Model

The natural starting point for the analysis of these questions is the small-union model we discussed in Sections 3.3 and 3.5. What we need to do is to embed that essential model

---

<sup>28</sup> Two exceptions were Kemp-Wan-Vanek-Ohyama CU and Krugman (1991a). But even in these cases, the endogeneity is based on welfare maximization rather than an explicit political-economy model.

into a general equilibrium model and construct a political-economy model around it that trade policy to be determined endogenously. This is the approach taken by Grossman and Helpman (1995). To outline their model, consider two potential union members, which are both small relative to the outside country. Each of them produces  $n$  non-numeraire goods that use a sector-specific factor and labor and a numeraire good that uses only labor. This structure of factor use makes all non-numeraire goods independent of each other in production (i.e., neither substitutes nor complements). Assume further that preferences are additively separable and the numeraire good enters into them linearly. This makes the demands for non-numeraire goods independent of each other as well. The net result is that each non-numeraire good behave exactly like steel in the partial-equilibrium model discussed in Section 3.3.

By assumption, neither country imposes a tariff on the numeraire good. The political-economy model developed in Grossman and Helpman (1994) determines the initial tariffs on non-numeraire goods. The objective function each government pursues in this model is a weighted sum of campaign contributions from the lobbies and overall welfare of voters. Each lobby represents the owners of a sector-specific factor and maximizes their welfare. The campaign contribution is made in return for the lobby's desired action by the government on tariffs. In the game, the lobbies move first and the government second. In equilibrium, the contribution is made only if the government takes this desired action. This model generates zero tariffs on exportables (export subsidies are ruled out by assumption) and non-negative tariffs on import-competing goods. The precise structure of tariffs depends on the parameters of the model including the relative weight placed by the government on campaign contributions.

Taking an initial set of tariffs determined by this process as given, let us introduce the possibility of an FTA. The FTA involves the removal of tariffs by A and B on goods produced within the union but leave the external tariffs at their initial levels. The only decision facing the (incumbent) governments in the two countries is to either accept the FTA agreement or reject it.

The solution to the problem is complex but it is possible to give some flavor of the results with the help of the analysis associated with Figures 3 and 4 in Section 3. Thus, letting  $n$  be an even number, suppose  $n/2$  of the non-numeraire products are imported by A and exported by B while the reverse holds for the remaining  $n/2$  products. Assume complete symmetry so that we can speak in terms of representative imports of A and B. Let the import demands for the representative imports of A be  $M_A$  in Figure 3 or 4 and imagine an identical curve for the representative import of B. The import good of A is subject to a tariff  $t_A$  by itself and zero by B. Analogously, the import good of B is subject to a tariff  $t_B = t_A$  by itself and zero by A. Consider two extreme cases shown in Figures 3 and 4, respectively, based on the exporting country's supply curve.

First, suppose the supply curve of the exporting country is as shown by  $S_B^1$  in Figure 3. We know that, in this case, exporting firms benefit and import-competing firms are unaffected by the FTA. Given that each country exports  $n/2$  products, exporting firms in each member benefit while import-competing firms are unaffected. Welfare of the union as a whole declines which, given the symmetry, implies that welfare of each union member also declines. It follows that the governments of the two countries will accept the FTA agreement provided the lobbying contributions from exporters who stand to gain outweigh the political cost of the decline in welfare. In initial game of free trade versus protection, if



firms prevailed and got high tariffs, it is likely that they will also prevail in the FTA versus status quo game. Thus, high initial tariffs in this case, coincide with the acceptance of the FTA. Ex post acceptance of the FTA is associated with large increases in exporters' profits and large trade diversion.

Second, suppose the export supply curve of the exporting country crosses  $M_A M_A$  below point R as in Figure 4. In this case, the FTA hurts the import-competing firms and does not benefit the exporting firms. But welfare rises in each member country. The FTA can be rejected in this case provided lobbying contributions by import-competing firms, who stand to lose from the FTA, more than outweigh the political benefit to the government from the gain in welfare. Once again, if, in the initial game for free trade versus protection, firms prevailed and got high tariffs, they are also likely to prevail in the FTA versus status quo game so that the FTA will be rejected. Ex post, a trade creating FTA is rejected.

These are, of course, highly simplified examples. But they are sufficient to suggest some of the problems that will be faced in asymmetric cases. For example, taking the basic setting in the first example, if one of the countries happens to be the exporter of many more than half the products, the government of the other country will likely refuse to accept the FTA proposal. For in that case, exporting firms are unlikely to make campaign contributions in equilibrium and the FTA lowers welfare. Reflecting this underlying logic, Grossman and Helpman reach the following important conclusion: "A free-trade agreement requires the assent of both governments. We have found that this outcome is most likely when there is relative balance in the potential trade between the partner countries and when the agreement affords enhanced protection rather than reduced protection to most sectors."

Enhanced protection here refers to trade diversion (as in Figure 3) whereas reduced protection refers to trade creation (as in Figure 4).<sup>29</sup>

Grossman and Helpman (1994) further show that the allowance for the exclusion of certain sectors from FTA agreement can make a previously infeasible FTA feasible. More recently, Duttagupta (1999) has introduced an intermediate input into the Grossman-Helpman model and addressed frontally the role of rules of origin in a general equilibrium model. She assumes that one partner exports the input to the other and imports from the latter the final good using the input. Thus, the union members are necessarily asymmetric. She shows that the introduction of a rule of origin in this setting can make acceptable an FTA that is otherwise rejected though, under some circumstances, the reverse may also happen. The former possibility arises because the country that exports the input and votes against the FTA in the absence of a rule of origin switches its vote in the presence of such a rule. As regards the welfare of the union, there are two notable possibilities. First, an FTA that lowered the joint welfare of the union and was rejected in the absence of the rules of origin becomes feasible upon the inclusion of such rules. Second, an FTA that improved joint welfare of the union but was nevertheless rejected in the absence of the rules of origin becomes feasible but the rules of origin are so distortionary that FTA becomes welfare inferior relative to the status quo.

---

<sup>29</sup> In Figure 3, the FTA extends A's higher tariff to producers in B thereby "enhancing" overall protection. In Figure 4, the FTA extends B's free trade price to A's market, "reducing" overall protection.

## A Cournot Oligopoly Model

So far, our analysis of the decision to form an FTA has been conducted in the context of a perfect competition model. In view of the growing popularity of imperfectly competitive models, we may ask whether the key results just derived remain valid in the presence of imperfect competition. Pravin Krishna (1998) performs this task using a Cournot oligopoly model in which firms belonging to three countries compete in one another's market. Asymmetries across countries are admitted both in terms of the market size and number of firms in a given country. Producers are given the decisive role in determining the policy outcome via the assumption that governments base their policy decisions on the home firms' profits. Initially, each country imposes a nondiscriminatory tariff on imports from all sources. The tariff is the same across all countries. Two countries, A and B, must decide whether or not to form an FTA which, given equal initial tariffs, is equivalent to a customs union. For the FTA to be accepted by both governments, profits of home firms must rise in each potential member.

Like Grossman and Helpman (1995), Krishna (1998) finds that the greater the degree of trade diversion, the more likely that the FTA will be accepted. The intuition behind the result is straightforward. When an FTA is formed, each member benefits (in terms of profits of its firms) from obtaining preferential access to the partner's market but loses from giving a similar access to the partner in its own market. In the absence of trade diversion, this is more or less a zero-sum game. But if the members can capture a part of the outside country's share in the union's market (trade diversion) without a corresponding loss of their share in the outside market, they can generate positive net benefits. The FTA is more likely to be accepted.

Krishna's analysis is based on what is essentially a one sector, partial equilibrium model. Therefore, he is unable to address the role of sectoral exclusions in making FTAs more or less acceptable. Likewise, intermediate inputs being absent as well, the role of the rules of origin is not considered.

## 6.2 *The Extra-Union Tariff*

Let us now turn to the second question raised at the beginning of this section: how does the decision to form a PTA by a country impact its choice of external tariff in a choice-theoretic model? In raising the question initially, Bhagwati (1993) expressed the concern that such a decision may result in a rise in the extra-union trade barriers either via an increase in tariff or more vigorous implementation of anti-dumping measures against outside countries.<sup>30</sup> This may turn even an initially trade-creating union into a trade-diverting union. He argued that within the traditional three-country framework, increased imports from the PTA partner that threaten a member country's firms will lead the latter to seek higher tariffs on imports from the outside country.<sup>31</sup> The issue can be analyzed from a variety of viewpoints, however.

---

<sup>30</sup> For developed countries, GATT bindings do not permit an increase in tariff so that increased barriers must take the form of safeguard measures such as anti-dumping or voluntary export restraints. For virtually all developing countries, tariff bindings have water so that they can raise tariffs.

<sup>31</sup> A formal presentation of this point can be found in Bhagwati and Panagariya (1996a).

## Lobbying and the External tariff in Small-Union Models

One way to address this issue is to begin with a model in which industry-specific lobbies play a decisive role in the determination of tariffs and FTA is introduced as an exogenous institutional change. Panagariya and Ronald Findlay (1996) take this approach. Following them, consider a three-good, Meade-Lipsey model in which the country under consideration imports two goods and exports the remaining one. Each good is produced using a sector-specific factor and labor. One of the imports comes from the partner and the other from the outside country. The tariff in each sector is determined by the amount of labor employed by that sector's lobby. The lobby, in turn, represents the interests of the factor specific to the sector.

The grant of a tariff preference is modeled as an institutional change that lowers the effectiveness of lobbying in gaining protection against imports from the partner country. The change reduces the level of lobbying in the sector competing with the partner country's good and releases labor into the economy. This, in turn, puts downward pressure on the wage rate and makes lobbying in the sector competing against the outside country's good less costly. The magnitude of lobbying in the latter sector rises, as does the external tariff. The impact of the tariff preference on welfare, which may have been positive at constant external tariff, is now ambiguous.<sup>32</sup>

---

<sup>32</sup> A referee has argued that labor employed in lobbying constitutes such a small fraction of total labor force that any changes in it are unlikely to have significant labor-market effects. The analysis can be resurrected, however, by introducing a lobbying-specific factor, lawyers. A shift in lobbying function that alters the return to this specific factor—a reasonable outcome—will yield results similar to those discussed in the text.

Cadot, de Melo and Olarreaga (1996) also use the Meade-Lipsey, three-good model in which tariffs are determined endogenously via the Grossman-Helpman (1995) political-economy process. They find that in FTA arrangements without rules of origin whereby goods destined to a high-tariff member can be imported through the low tariff member, competition for tariff revenue may lead to competitive reductions in external tariffs until they are removed completely. In a CU setting, by contrast, lobbies may cooperate on a union-wide basis and win increased external protection.

#### Turning Trade Diversion into Tariff Revenue

As hinted by the tariff-revenue-competition result just mentioned, under some circumstances, FTAs may actually lead to a reduction in the external tariff. Richardson (1993) shows this in an elegant paper. His argument can be illustrated with the help of Figure 1b where country A's import demand is downward-sloped and export-supplies of B and C are horizontal. In the initial equilibrium, A has an FTA with B, which is the less efficient supplier of the product in comparison with C.

In this setting, if A reduces the tariff on C to  $P_B P_C - \varepsilon$  where  $\varepsilon$  is infinitesimally small, it can switch all its imports from the less efficient B to more efficient C and collect areas e and k in tariff revenue. If the government maximizes a political support function to which tariff revenues contribute positively, it will, in fact, take that course. Richardson shows that this basic argument is valid in a general equilibrium setting.

A key limitation of this argument, however, is its reliance on a model in which imports come from either B or C but not both. As already discussed in Section 3, this specialization in the source of imports results from the assumption that supply curves of both B and C are infinitely elastic. A moment's reflection shows that the result break down

as soon as we allow for imports from both B and C by assuming B's supply curve to be upward sloped. As is easily verified with the help of Figure 3, lowering the tariff on C just enough to allow it to sell in A at a price slightly below  $P_C^t$  yields no more than the conventional gains from unilateral liberalization. At the margin, B is able to compete with C at a slightly lower level of output and cannot be eliminated as a source of supply.

### Increased Monopoly Power of Trade Blocs

The simplest model in which the formation of a CU can lead to a rise in the external tariff is the one in which the external tariff is chosen to maximize the union's welfare. This is, indeed, the outcome in Krugman's (1991a) model: as the world is consolidated into fewer and fewer symmetric blocs, the external tariff rises monotonically. As each bloc gets larger, the proportion of income spent by the rest of the world on its exports rises, giving it a greater market power.

But this outcome is not inevitable. Eric Bond and Constantinos Syropoulos (1996) modify Krugman's model slightly and are able to generate an ambiguous effect of larger blocs on the outside tariff. In the Krugman model, each province is endowed with a fixed amount of one product that it exports and none of any other product. Bond and Syropoulos (1996) modify this assumption and postulate that each province is endowed with some of each product plus a little more of the one it exports. Formally, province  $i$  is endowed with  $x+z$  of good  $i$  and  $x$  of all other products ( $x, z > 0$ ). The Krugman model is, thus, a special case in which  $x = 0$ . The authors show that, in this model, the relationship between bloc size and the external tariff is ambiguous due to two opposing effects at work. As the representative bloc grows larger, as in Krugman, the share of outside blocs' income spent on the goods exported by the bloc increases and gives it more market power. But the increase

in bloc size also increases the representative bloc's share in the total world endowments of the goods it exports which reduces its market power.<sup>33</sup>

### 6.3 *Evidence*

Are FTAs and CUs largely trade diverting as political-economy-driven models suggest? And do these arrangements make member countries more or less protectionist with respect to the rest of the world? Since I devote the entire Section 6 to the first question, let me confine myself to the second one here.

Though evidence is subject to alternative interpretations, the fact is that unilateral trade liberalization has come to a virtual standstill in Latin America where the forces of regionalism are most strongly at work. The advocates of regional arrangements argue that this suspension of liberalization has little to do with regionalism. When tariffs are high, trade liberalization is simply easier politically. But once they reach the 10-20% range, political costs of further liberalization become prohibitive.

Multilateralists, on the other hand, argue that the slowdown is the result of expectations of securing preferential access to other countries' markets in exchange for giving similar access to one's own market. If countries eliminated their barriers unilaterally, there will be no preferences left to be given.

Quite apart from the suspension of unilateral trade liberalization, there is evidence that countries raise their external trade barriers following the conclusion of regional

---

<sup>33</sup> In a Cournot oligopoly model, Sang-Seung Yi (1996) also finds that the external tariff may rise or fall as an only existing bloc expands.



arrangements. As documented in Panagariya (1999a), Mexico, Israel, and MERCOSUR have raised their external tariffs since entering into regional arrangements. There is even evidence that the European Union has reacted to internal liberalization by implementing anti-dumping more vigorously against outside countries (Brian Hindley and Patrik Messerlin 1993).

## 7. Regionalism and Multilateralism

Indirectly, we have already begun to explore the relationship between regionalism and multilateralism by asking in the previous section whether PTAs lead to a rise or decline in the external tariff. But this approach is at best incomplete since it does not consider explicitly the role of multilateral process itself in the determination of the outcome.

Following Bhagwati (1993) and Bhagwati and Panagariya (1996a), the implications of regionalism for multilateralism can be addressed along two separate lines. First, assuming regional and multilateral processes do not interact, i.e., they are strangers, will one or more trade blocs continue to expand until they encompass the entire world? Second, if these processes interact, will the option to form regional blocs make the success of multilateral process more or less likely, i.e., will the two processes act as friends or foes. To these, the recent literature has added a third question: what is the impact of multilateralism on regionalism? We take these questions in turn.

### 7.1 *Strangers: Bloc Expansion*

Bloc expansion depends on the willingness of the existing members to offer entry and the incentives facing outsiders to seek entry. Baldwin (1995) analyzes formally the incentive of outsiders to seek entry. He assumes that potential entrants face "non-economic"

costs of acceding to a bloc. The entrants can be indexed along the real line such that a rising value of the index is associated with a country with higher non-economic cost of entry. This means that successive countries require larger and larger economic incentive to seek entry.

Baldwin takes a variant of the Helpman-Krugman (1985, chapter 10) model of economic geography and combines it with the Grossman-Helpman (1994) political-economy model. Trade barriers in this model take the form of transport costs and entry into an existing bloc is modeled as a reduction in the transport cost. At the initial equilibrium, the economic benefit of membership to the last member in the bloc equals its non-economic cost. Baldwin disturbs this equilibrium by introducing an exogenous shock, which he calls an idiosyncratic event and likens it to the European Single Market initiative. The shock increases relative profitability within the bloc thereby encouraging the firms in the outside country at the margin to lobby their government harder for entry.<sup>34</sup> As this country accedes to the bloc, the potential economic benefits of entry for the next country on the outside margin rise and may offset the higher non-economic costs of entry it faces. Thus, bloc expansion generates a "domino" effect. Unless non-economic costs rise faster than the benefits of entry, given Baldwin's assumption of open entry, the bloc can come to encompass the entire world and, hence, global free trade.<sup>35</sup>

---

<sup>34</sup> From the viewpoint of the firms in an outside country, the protected market within the bloc becomes larger while the outside market becomes smaller.

<sup>35</sup> Baldwin does assume that non-economic costs of entry rise faster so that the process comes to an end before global free trade is achieved.

There are two key limitations of Baldwin's otherwise elegant analysis. First, as already noted, working in the tradition of economic-geography models, he formalizes trade barriers as transport costs. As such, accession to the PTA becomes equivalent to a reduction in transport costs. The revenue aspect of trade barriers, central to traditional models, is completely absent in his analysis. It is not clear whether his result will remain valid once transport costs are replaced by tariffs and, hence, the revenue-transfer effect of entry into the bloc is taken into account. Second, even if we ignore this problem, Baldwin (1995) assumes that "insiders" have no incentive to block entry. It may be conjectured that even within his own model, once the bloc reaches a certain size, insiders will have an incentive to block further entry.

This is indeed the message of a recent paper by Soamiley Andriamananjara (1999) that explicitly models the incentives facing outsiders to seek entry and willingness of insiders to give entry. He uses a Cournot oligopoly model of identical countries in which the outside tariff is fixed by assumption and decisions to seek and offer entry are driven by profits. He shows that in this model as the CU expands, profits of insiders first rise, reach a maximum and then decline. Moreover, the maximum-profit point reaches before the CU comes to encompass all countries. Profits of outsiders, on the other hand, decline monotonically as the CU expands. Thus, while outsiders have an increasing incentive to seek entry, insiders stop short of taking all of them into the club. The CU fails to expand into a global bloc.

Bond and Syropoulos (1996) ask this same question, albeit in slightly circuitous manner, using the model discussed in Section 6. They hypothesize a world that is initially divided into several identical blocs. They then allow one of these blocs to expand by

drawing one country at a time from each of the remaining blocs, with Nash-optimum tariffs applied at all times by all blocs. With the help of simulations, they show that as this bloc expands, the welfare of its members peaks before it absorbs all members of other blocs.<sup>36</sup>

## 7.2 *Impact of Regionalism on Multilateralism: Friends or Foes?*

Next, let us turn to the issue whether regionalism serves as a building block or stumbling block for multilateralism. A number of different approaches can be distinguished.

### Stumbling Blocks: A Median Voter Model

There are two key questions we may ask: (i) Can the option to form a trade bloc make a previously infeasible multilateral liberalization feasible; and (ii) Can this additional option render a previously feasible multilateral liberalization infeasible? Levy (1997) addresses this question in a median voter model. The answer to the first question is a straightforward “no.” The initial unfeasibility of multilateral liberalization implies that the median voter enjoys higher utility under autarky than under free trade. The option to form a bloc is exercised only if it increases the voter’s utility further. But this raises his reservation utility and must make him even less willing to accept multilateral liberalization.

The second question requires deeper analysis and necessitates spelling out the model explicitly. Levy (1997) addresses it within two popular models: a two-sector, two-factor, multi-country, Heckscher-Ohlin model and a variant of it in which one of the sectors produces a differentiated, monopolistically competitive good. He shows that in the first

---

<sup>36</sup> Yi (1996) also considers the issue of bloc expansion in a Cournot oligopoly model.

model, the option of a trade bloc cannot block a previously feasible multilateral accord but, in the second one, it can.

The analysis requires some strong assumptions. It is assumed that when blocs are formed, they adopt total free trade with each other but maintain complete autarky vis-a-vis the rest of the world. Furthermore, endowments of the countries in the world are sufficiently similar to permit factor price equalization when trade is free between two or more countries. These assumptions permit factor prices to be determined by the overall endowment ratio of the region within which trade is free.

Focusing on the Heckscher-Ohlin setting first, consider three countries, A, B and C. Letting  $k_0^A$  be the capital-labor-endowment ratio of the median voter in country A, his utility exhibits the pattern shown by curve  $U^A U^A$  in Figure 8a (ignore  $U^B U^B$  for now) with respect to the capital-labor ratio of the economy in which he operates. Under autarky, the relevant endowment ratio coincides with the endowment ratio of country A, under FTA, with that of A and B combined, and under global free trade, with that of the world. The key point is that when the median voter's endowment ratio coincides with that of the economy in which he operates, his utility is minimized. When it differs from the latter, utility is higher because he can benefit from "trading" with the rest of the economy.

[Figure 8 here]

Introduce country B now. For ease of exposition, consider the highly special case in which the endowment ratio of the median voter in each of A and B coincides with the country's endowment ratio. Let  $k^A$  and  $k^B$  denote the endowment ratios of A and B so that  $k_0^A = k^A$  and  $k_0^B = k^B$ . Since the case  $k^A = k^B$  is uninteresting, without loss of generality, assume  $k_0^A = k^A < k^B = k_0^B$ . The utility curves of median voters in A and B are then as

shown by  $U^A U^A$  and  $U^B U^B$ , respectively, in Figure 8a. The trade bloc's endowment ratio must lie somewhere between  $k^A$  and  $k^B$  and is shown by  $k^{AB}$ . By assumption, autarky minimizes each median voter's utility. Therefore, the bloc necessarily increases their utility. From this, it would seem that an agreement to form the bloc will succeed. But the story is more complicated, requiring the introduction of the precise voting sequence between regionalism and multilateralism.

It is assumed that, in the first period, voters in both A and B decide whether they want to form a bloc. In the second period, they vote on multilateral free trade. Voters are fully informed and the two periods are sufficiently close to each other that the utility level of the second period guides the voters' decisions. What this means is that even if a bloc increases utility of a median voter, he will vote against it if multilateralism increases utility even more and he realizes that, after the bloc is formed, the other median voter will block the multilateral accord.

Recall that we assume that the multilateral accord is feasible in the absence of the option to form a bloc. In the specific case we have chosen for simplicity, since autarky minimizes the utility of the median voter, multilateral accord cannot reduce their utility and, hence, is necessarily feasible in the absence of the option to form a trade bloc. To analyze the outcome when the option of a bloc is offered, we need to specify explicitly the multilateral capital-labor endowment ratio,  $k^M$ . If  $k^M$  is no more than  $\bar{k}^M$  or no less than  $\tilde{k}^M$  in Figure 8a, the multilateral accord is at least as good as or better than trade bloc for both A and B. In this case, both countries approve the bloc in the first period and the multilateral accord in the second period. The trade bloc forms but it neither helps nor hinders multilateral accord.

Interestingly, even if  $k^M$  lies anywhere between  $\bar{k}^M$  and  $\tilde{k}^M$ , the multilateral accord survives due to the fact that one of the countries will defeat the trade bloc in period one. To see this, consider  $k^M = k^{M*}$  in Figure 8a. In this case, the median voter in A prefers multilateral accord to trade bloc while the opposite is true for the median voter in B. Knowing that B will block the multilateral accord in period two if the bloc is already in place, A blocks the trade bloc in period one in the first place.

Thus, in the standard Heckscher-Ohlin framework, regionalism neither helps nor hinders multilateralism. If one of the goods is differentiated, however, the trade bloc can become a stumbling block for multilateralism. The main difference now is that benefits from trade also arise from an increase in the variety of the differentiated product. The utility curves depend on not just the relative factor endowment of the economy in which the individual operates but also product variety.

To make the point most simply, consider Figure 8b where A and B are identical in all respects, including absolute size and relative endowments. Let  $k^A = k^B$  be the country-wide endowment ratio and  $k_0$  the median voter's endowment ratio (which is different now from the country's endowment ratio). Each median voter's utility curve is given by  $UU$  under autarky. The initial level of the voter's utility is given by the height of  $UU$  at  $k^A = k^B$ . A trade bloc does not change the economy's endowment ratio (since  $k^A = k^B$ ) but increases the available variety of the differentiated product. In the presence of the bloc, the utility curve is given by the dotted curve  $U^{AB}U^{AB}$  and the level of utility by  $\bar{U}^{AB}$ .

Since multilateral free trade offers an even larger variety than the bloc, it shifts the utility curve further up to, say,  $U^M U^M$ . But if the multilateral accord also alters the economy's endowment ratio to anywhere between  $\bar{k}^M$  and  $\tilde{k}^M$ , it yields a lower utility to

both median voters than the trade bloc. Thus, even though both median voters would have accepted the multilateral accord in the absence of the trade bloc, they will reject it in its presence.

#### Stumbling Blocks: A Cournot Oligopoly Model

An alternative approach to the "friends versus foes" issue is in terms of an oligopoly model in which the decisions are driven by producer profits. This is the setting of the Pravin Krishna (1994) model discussed earlier in the context of the decision to form an FTA. Using that model, we can ask the same question asked by Levy: Does an initially feasible multilateral liberalization remain necessarily feasible after two of the three countries have formed an FTA? Krishna addresses this question and answers it in the negative. He finds, in particular, that the more the FTA benefits (in terms of the firms' profits) from trade diversion, the more likely it will turn into a stumbling bloc. Through a multilateral liberalization, union members obtain tariff free access to the third country's market in return for offering it access to their own market on equal terms. But if the FTA was heavily trade diverting to begin with, the benefit in terms of the government's objective function from the former change is less than the loss due to the latter change.

#### Insidious Regionalism

An entirely different approach to the question at hand is taken by McLaren (1998) who models regionalism as a coordination failure in a world with sector-specific sunk costs and 'friction' in trade negotiations. Based on the expectation that a regional bloc is likely to form, private agents make investments that make potential bloc members more specialized toward each other but, together, less specialized relative to nonmembers. These



investments, assumed to be irreversible, reduce the demand for multilateral free trade *ex post*. Thus, the expected supply of regionalism generates its own demand, creating a Pareto-inferior equilibrium.

#### Stumbling Blocks in Transition but Building Blocks in the Long Run: FTAs

Kyle Bagwell and Robert Staiger (1997a, 1997b) investigate how multilateral tariff cooperation is impacted by the formation of FTAs and CUs during the *transition* period. A distinguishing feature of their approach is the assumption that countries cannot make binding commitments to enforce the international bargaining outcomes.<sup>37</sup> They are, therefore, limited to self-enforcing multilateral arrangements that balance short-term gains from deviation against the cost of an ensuing trade war.

The set up chosen by Bagwell and Staiger (1997a) is different from the traditional three-country set up. They assume two countries, called Home and Foreign, which cooperate on reciprocal tariffs subject to the above-mentioned incentive constraint. The objective is to maximize welfare as represented by the sum of consumers' and producers' surplus and tariff revenue.

Trade relations between the two countries have three phases. In phase 1, they trade with each other with tariffs set cooperatively via a dynamic tariff game. Phase 2 corresponds to a transition phase, in which trade between Home and Foreign continues but

---

<sup>37</sup> For earlier contributions in this tradition, see Jensen and Thursby (1984), Avinash Dixit (1987), Bagwell and Staiger (1990) and Rodney Ludema (1992).

each country has begun discussions about future free trade agreements with other (unmodeled) countries that are assumed to exist in the background. In phase 3, the free-trade agreements are fully implemented. Home and Foreign countries now trade less with one another since they divert some trade to their respective FTA partners and reset the cooperative tariffs. The new trade patterns and tariffs are stationary into the infinite future.

The authors focus on the impact of the negotiations for the FTA on tariff cooperation during phase 2. Their key result is that the emergence of FTAs is associated with temporarily heightened multilateral trade tensions between Home and Foreign. The tension arises because the current trade flows between the two countries have not changed (since FTAs are implemented in phase 3) but expected future flows have declined due to trade diversion. The former fact implies that the short-term gains from deviation have not changed but the latter one implies that the cost of a future trade war between them has declined. This leads to a temporary rise in the multilateral tariff. In phase 3, as the agreement is implemented fully, cooperation resumes and the tariff declines below the phase 1 tariff partially because of the reduced volume of trade between Home and Foreign.

#### Building Blocks in Transition but Stumbling Blocks in the Long Run: CUs

In Bagwell and Staiger (1997b), the authors consider a variation of this model and focus on the impact of customs unions on tariff cooperation during transition. Home and Foreign are now interpreted as regions with each of them consisting of several customs unions. There are two goods with one exported by Home CUs and the other by Foreign CUs. Acting as independent units, Home CUs negotiate tariffs with Foreign CUs. Starting with phase 1 cooperative tariffs, the possibility of consolidating each of Home CUs and

Foreign CUs into larger CUs is then introduced in phase 2. Once again, the agreement is actually implemented in phase 3.

In addition to the trade-diversion effect (which the authors choose not to highlight), there is now a market-power effect. The agreement to consolidate each of Foreign and Home into larger CUs implies that the market power of participants in phase 3 has gone up. In phase 2, this means that the cost of a future trade war has gone up. This leads to a reduction in the multilateral tariff in phase 2. In phase 3, reflecting increased market power, the multilateral tariff rises above the phase 1 tariff.

### *7.3 The Impact of Multilateralism on Regionalism*

So far, the focus of the analysis in this section has been on how the option to form FTAs impact decisions regarding multilateral liberalization. Let us now turn to the opposite question: How does multilateral liberalization impact the decision of countries to exercise the regionalism option. There are two principal contributions addressing this reverse relationship.

#### Multilateral Liberalization Making PTAs More Sustainable

The current wave of regionalism has been launched in the wake of a more liberal trading environment than the first wave during 1950s and 1960s. This wave also promises to be more sustainable as was predicted by Bhagwati (1993). We may, therefore, ask whether greater openness may imply greater sustainability of PTAs. Caroline Freund (1998) uses symmetric, three-country, repeated games Cournot oligopoly model to analyze this issue. Initially, each country levies the same multilateral tariff on the other two countries. She shows that, in this setting, the welfare gain from joining a PTA is greater than the gain

from a move to free trade when the multilateral tariff is low while the reverse is true when it is high. She goes on to show that this feature makes PTAs more sustainable when multilateral tariffs are low. Hence, PTAs may proliferate and be sustained as a result of multilateral freeing of trade.

The logic behind Freund's result can be best understood by considering the case when the initial multilateral tariff is near autarky. In this case, when two countries form an FTA, there is no room for exploiting the third country via better terms of trade: at near zero trade with the latter, the gain from improved terms of trade is also near zero. Thus, under the PTA, the benefits are limited to those arising from mutual liberalization by partners. But under multilateral liberalization, benefits also accrue from the liberalization of the third country.

When the multilateral tariff is initially low, however, the partner countries can benefit from mutual liberalization as well as the improvement in the terms of trade with respect to the third country that accompanies preferential liberalization. Under multilateral liberalization, by contrast, no terms-of-trade benefits accrue: the benefits are limited to the conventional efficiency triangles. These factors increase the attractiveness of preferential liberalization over multilateral liberalization at low tariffs.

#### Liberalization in North Leading to North-South PTAs

Some of the recent PTAs including NAFTA and the association agreements of the European Union with some of the North African countries have formed along North-South lines. As already stated, these agreements have also been concluded in the wake of considerable liberalization among the countries in North. Inspired by these observations,

Ethier (1998) constructs a model in which preferential liberalization by South is the result of multilateral liberalization by North and has a happy coexistence with it.

A highly simplified account of Ethier's basic story can be given as follows. The world is divided into two regions to be called here North and South. Each region consists of several countries. Northern countries are all symmetric. Each Northern country produces one non-traded good, which uses skilled and unskilled labor, and one variety of a traded, differentiated good, which uses human capital and an intermediate good. Only North has human capital so that the differentiated good can be produced only in that region. The intermediate good uses skilled labor and can be produced anywhere. A key feature, which drives many of the results, is the presence of an (international) external economy in the production of the intermediate input.<sup>38</sup>

Southern countries have skilled labor and could produce the intermediate input, trading it for the differentiated good with North. But they face resistance to openness. Initially, this resistance is sufficiently strong even the country with least resistance is in autarky. Each Southern country produces and consumes a rudimentary good, which is a (poor) substitute for North's differentiated good. Northern countries trade initially but impose the Nash optimum tariff on the imports of the differentiated good from other countries. Because the countries are symmetric, the tariff is the same for all of them.

---

<sup>38</sup> Ethier defines the production of the input in another country as direct foreign investment, though no investment, technology flows or repatriation of earnings are associated with the shift in location. For simplicity I will call the location of production of intermediate input in South as "production" rather than foreign investment.

Suppose now that multilateral cooperation leads to a reduction in Northern tariffs. This leads to an expansion of the intermediate input and differentiated goods sectors in each Northern country. The international externality lowers the production cost of intermediate input and allows some Southern countries to overcome resistance to openness. The production of the intermediate input moves partially to the reforming Southern countries. This opening up itself creates opportunities for North-South regional arrangements. Some Northern countries give a tariff preference to the intermediate input produced in the Southern partner in return for exclusive access to the Southern partner's market for the differentiated good.<sup>39</sup>

## 8 Theoretical Considerations in Empirical Assessments of PTAs

Let us now return to the welfare issue, focusing this time on whether, *in practice*, FTAs and CUs lead to increased or reduced welfare. Broadly speaking, empiricists have taken two approaches to sort out this issue. First, they have conducted counter-factual analyses, based on partial- or general-equilibrium models. The idea here is to assume a certain model structure, with specific functional forms and parameter values, to represent the economies in a base year prior to the formation of the union. The model is then shocked by a preferential removal of tariffs and the welfare (and other) effects calculated. Second, empiricists have carried out *ex post* studies of the arrangements to measure the extent of trade creation and trade diversion. The typical approach here has been to estimate

---

<sup>39</sup> By assumption, Southern partners are unable to distinguish between different varieties of the differentiated good.

econometrically the so-called "gravity" equation which represents bilateral trade flows as a function of incomes and populations of trading partners, distance between them and membership in a common regional arrangement. Summaries of these studies can be found in Adela de la Torre and Morgan Kelly (1982), Srinivasan, John Whalley and Ian Wooton (1993) and Frankel (1997).

Unfortunately, paralleling the theoretical predictions, these studies generate ambiguous answers. After reviewing a large number of studies, Srinivasan, Whalley and Wooton (1993) conclude, "We, therefore, see these studies as shedding somewhat incomplete and at times conflicting light on the effects of post-war RIAs [Regional Integration Agreements] on trade and welfare, to say nothing of what might be the likely effects of prospective RIAs. There seems to be near unanimity that trade creation occurred in Europe, but its size and the precise contribution of the RIAs relative to other factors is unclear. Nor is it clear that significant trade creation from RIAs has occurred elsewhere."

There are sufficiently serious problems with both empirical approaches that the results based on them are unlikely to change the minds on either side of the regionalism debate. Consider first the simulation approach. It is relatively easy to manipulate the structure of the model, functional forms and parameter values in these models to obtain one's desired results.<sup>40</sup> Let me note just two factors.

First, most modelers rely on the so-called Armington assumption according to which goods are assumed differentiated by the country of origin. They then proceed to combine

---

<sup>40</sup> The critique of CGEs in this section is based on Panagariya and Duttagupta (1999), which also provides numerical examples to illustrate the points discussed here.

this assumption with the small-union assumption. But there is an inherent contradiction between these two assumptions: being the sole producer of its product, each country has some monopoly power in the world market. Again, the assumption plays a key role in determining the outcome. With the Armington assumption ruling out the imports of the product from the outside country and the small-union assumption ruling out the terms-of-trade effects, as in Figure 4, each country benefits solely from its own liberalization. Not surprisingly, so many studies of NAFTA predict high-tariff Mexico gaining much more relative to its GDP than the United States. If, instead, the theoretically correct, large-union assumption is employed, we find the low-tariff member (United States) benefiting from preferential liberalization by the high-tariff member (Mexico), as predicted by Mundell's (1964) analysis (see Section 3.4 above).

Second, even accepting the co-existence of the Armington structure and the small-union model for the sake of argument, the functional forms and parameter values can be exploited to obtain particular results and rule out others. For example, it is not uncommon to use Stone-Geary utility function or the linear expenditure system to represent demand.<sup>41</sup> This greatly limits the possibilities of substitution. For instance, it can be shown that if the partner's product shows a high degree of substitutability with that of the outside country but low substitutability with the product of the home country (as is likely, for example, for Mexico in the NAFTA context), an FTA is likely to be harmful (Panagariya 1997a, pp. 482-

---

<sup>41</sup> The wide use, rather than econometric evidence, has often been also cited as evidence that the assumption is "reasonable".



3). Even the widely used, standard CES utility function rules out this possibility by assumption.

Turning next to the *ex post* approach, a key problem here is that investigators have tried to calculate simply *total* quantities of trade creation and trade diversion. As Meade demonstrated as far back as 1955 (see Section 3.1 above), however, aggregate trade creation and trade diversion are insufficient to infer the welfare effects of PTAs. We need to know trade creation and trade diversion by sector and, in each case, use the information on the decline in the prices of imports to evaluate the benefit from trade creation and the height of trade barriers to measure the damage from trade diversion.<sup>42</sup> The information requirements of such calculations are far too demanding for them to be carried out in practice.

McMillan (1993) has tried to cut through this Gordian knot by suggesting that at least from the viewpoint of the GATT rules, the criterion for evaluating the regional arrangements should be the welfare of nonmember countries. To quote him, (McMillan 1993, p. 295), "Trade theorists have usually evaluated RIAs either from the point of view of the world as a whole (asking whether the trade creation outweighs the trade diversion) or from the point of view of the members (asking how to maximize the gains from trade creation). I suggest that, for the rules of international trade, the size of any trade creation among member countries is irrelevant. In practice, it is possible that some member countries will not benefit from a RIA. But it seems reasonable to have a hierarchy of concerns: to put preventing harm to third countries ahead of preventing members from hurting themselves."

---

<sup>42</sup> Even this is valid only if the changes in question are small. Otherwise, the knowledge of the entire structure of the model will be necessary.

Taking the welfare of nonmembers as the sole criterion, McMillan goes on to argue, by appeal to the Kemp-Wan-Vanek-Ohyama theorem, that outside countries will be protected from being harmed provided the union's total imports from them do not decline after the formation of the union. If correct, this criterion can serve as a simple basis for distinguishing desirable unions from undesirable ones, at least *ex post*.

It can be shown, however, that the McMillan test is insufficient to guarantee nonmembers their pre-union welfare. Thus, for instance, imagine a substantial deterioration in the terms of trade of nonmembers following the formation of an FTA or CU. Assuming no distortions in nonmembers, this change will lower their welfare. Yet, it is entirely possible that they now export more to the newly formed union in exchange for the same or smaller basket of imports than before. Though the McMillan test is met, the formation of the union hurts nonmembers.<sup>43</sup>

Assuming that trade imbalance is exogenous, utility depends on the current consumption and there are no domestic distortions or tariffs in nonmembers, a sufficiency condition for them not to suffer a welfare loss is

$$(1) \quad \mathbf{p}^1 \mathbf{e}^0 \geq \mathbf{p}^1 \mathbf{e}^1$$

Here  $\mathbf{p}$  denotes the price vector,  $\mathbf{e}$  the net exports vector, and superscripts 0 and 1 identify pre- and post-union equilibria. The elements in  $\mathbf{e}^0$  and  $\mathbf{e}^1$  are positive in the case of

---

<sup>43</sup> Winters (1997) criticizes McMillan for his focus on the union's imports from nonmembers as the criterion for ensuring no harm to the latter. But he ends up arguing (incorrectly) that McMillan should have used the expansion of union's exports as the criterion. As I argue immediately below, neither aggregate exports nor aggregate imports are appropriate to answer the question at hand.

exportables and negative in the case of importables. According to (1), welfare of nonmembers improves provided their pre-union net exports vector generates a larger trade surplus than their post-union net exports vector at post-union prices. The Kemp-Wan-Vanek-Ohyama theorem freezes the trade vector of nonmembers at its pre-union level. This, in turn, freezes the price vector and condition (1) is automatically satisfied as equality.

Inequality (1) admits the possibility of trade deficit or surplus in both pre- and post-union equilibria. If we impose the trade-balance condition in the post-union equilibrium, however, the inequality requires that nonmembers be able to buy their pre-union import bundle with their pre-union export bundle at post-union prices. This condition can be viewed as saying that the union should not lead to a deterioration of the terms of trade of outside countries.<sup>44</sup>

The relationship of (1) to the terms of trade becomes more explicit if we impose the trade-balance condition in both pre and post-union equilibria. Trade balance in the pre-union equilibrium yields  $\mathbf{p}^0 \mathbf{e}^0 = 0$ . Subtracting this equality from (1) and recognizing that trade balance in the post-union equilibrium implies  $\mathbf{p}^1 \mathbf{e}^1 = 0$ , the condition for no loss of welfare reduces to

---

<sup>44</sup> In his critique of McMillan (1997), Winters does mention the role of the terms of trade in ensuring that nonmembers do not suffer a loss in welfare. But, relying on the two-good model, he winds up arguing in favor of increased exports by the union as the criterion for welfare improvement. But making the plausible assumption that the import-demand elasticity in outside countries is no less than unity, in the two-goods model, increased exports are necessarily accompanied by increased imports. Thus, the test favored by Winters coincides with that suggested by McMillan.

$$(2) \quad (\mathbf{p}^1 - \mathbf{p}^0)\mathbf{e}^0 \geq 0.$$

This inequality is the traditional definition of an improvement in the terms of trade in the multi-good model.

In principle, (1) or (2) can serve as the simple test sought by McMillan to sort out desirable FTAs and CUs from undesirable ones. But, in practice, both of these conditions suffer from two key limitations. First, their ability to guarantee no harm to nonmembers is based on the assumption of no distortions in nonmembers. Second, and more importantly, they assume that all changes in post-union prices are due to the formation of the union. In practice, observed prices will reflect the impact of many other changes that are likely to take place independently of the union.

But the prospects for the McMillan criterion (that nonmembers not be harmed) need not be so bleak. Under one set of empirically relevant conditions, theory gives us a strong indication of the impact of the formation of a union on outside countries' terms of trade. If import demands exhibit gross substitutability and initial tariffs are low, the terms of trade of outside countries are highly likely to deteriorate if within-union barriers are lowered, holding the extra-union barriers at their original levels. Since gross substitutability is not an especially strong assumption in the present context and most FTAs (as opposed to CUs) leave their external barriers at pre-union level, the McMillan criterion will reject all FTAs involving large countries. It will also accept all small unions since these neither harm nor help nonmembers.

## 9. Concluding Remarks

As this review demonstrates, trade theorists have responded quickly to the challenges thrown by the current wave of regionalism. Within less than a decade, a solid

body of scientific work, shedding light on the political economy of regional arrangements and their impact on external tariffs and multilateral freeing of trade has been created.

While there remain sharp divisions among economists and policy makers on the merits of PTAs, a consensus appears to be emerging on one issue of great policy relevance. Proliferation of FTAs is leading to the creation of what Bhagwati (1995) has called a “spaghetti bowl” of tariffs whereby a country subjects the same product to different tariff rates depending on its ostensible origin. There are two sources of this discrimination in tariff rates. First, with each country participating in multiple FTA agreements, the tariff during transition to full internal free trade depends on the FTA member from which the product is imported. Second, in the long run, even after FTAs have been fully implemented, varying degrees of discrimination across products and countries will remain due to differences in the rules of origin across FTA agreements.<sup>45</sup> Thus, ironically, free-trade intentions threaten to reproduce the chaos in the tariff regime that was created in the 1930s by protectionism and the absence of the MFN principle in trade policy. There is now a general agreement among free trade economists that the best solution to this problem is to

---

<sup>45</sup> Sapir (1998) notes that the European Union currently applies its MFN tariff to barely six countries (Australia, Canada, Japan, New Zealand, Taiwan and the United States) which account for approximately one third of its total imports. On other trading partners, it imposes a variety of different rates depending on its relationship with them.

speed up MFN liberalization. Once external tariffs drop to zero, tariff preferences and the spaghetti bowl created by them will automatically disappear.

In the meantime, on the theoretical front, at least two major gaps remain with respect to the theory of preferential trading. Theory remains almost non-existent on the relationship among regional, multilateral and unilateral liberalization in trade in services. Formal models of PTAs deal almost exclusively with border barriers, which do not capture the reality of much of the trade in services. Yet, regional arrangements have now begun to focus on trade in services.

Equally, in the policy debate, direct foreign investment is frequently cited as a key reason for signing FTAs and CUs. Yet, there is little theoretical work drawing the link between these two phenomena. Issues such as why regional arrangement might be a superior instrument of bringing foreign investment than multilateral liberalization have yet to be addressed.

## References

- Anderson, Kym and Richard Blackhurst, eds. 1993. *Regional Integration and the Global Trading System*, New York: St. Martin's Press.
- Andriamananjara Soamiley. 1999. "On the Size and Number of Regional Integration Arrangements: A Political Economy Model," University of Maryland, mimeo.
- Bagwell, Kyle and Robert W. Staiger. 1990. "A Theory of Managed Trade," *Amer. Econ. Rev.*, 80:4, pp. 779-785.
- Bagwell, Kyle and Robert W. Staiger. 1997a. "Multilateral Tariff Cooperation During the Formation of Free Trade Areas," *Int. Econ. Rev.*, 38:2, May, pp. 291-319.
- Bagwell, Kyle and Robert Staiger. 1997b. "Multilateral Tariff Cooperation During the Formation of Customs Unions," *J. of Int. Econ.* 42:1-2 , pp. 91-123.
- Baldwin, Richard. 1995. "A Domino Theory of Regionalism." in *Expanding Membership of the European Union*. Richard Baldwin, P. Haaparnata, and J. Kiander, eds. Cambridge, U.K: Cambridge University Press, pp. 25-53.
- Baldwin, Richard and Anthony Venables. 1995. "Regional Economic Integration," in *Handbook of International Economics Volume III*. Grossman, Gene and Ken Rogoff, eds. Amsterdam: North Holland, pp. 1597-1644.
- Berglas, Eitan. 1979. "Preferential Trading: The n Commodity Case," *J. of Polit. Econ.*, 87: 21, pp. 315-331.
- Berglas, Eitan. 1983. "The Case for Unilateral Tariff Reductions: Foreign Tariffs Rediscovered," *Amer. Econ. Rev.* 73:5, pp. 1142-43.
- Bhagwati, Jagdish. 1971. "Trade-Diverting Customs Unions and Welfare Improvement: A Clarification," *Econ. J.*, 81:323, pp. 580-587.

- Bhagwati, Jagdish. 1991. *The World Trading System at Risk*. Princeton, N.J.: Princeton University Press.
- Bhagwati, Jagdish. 1993. "Regionalism and Multilateralism: An Overview," in *New Dimensions in Regional Integration*. Jaime de Melo and Arvind Panagariya, eds. pp. 22-51.
- Bhagwati, Jagdish. 1995. "U.S. Trade Policy: The Infatuation with Free Trade Areas," in *The Dangerous Drift to Preferential Trade Agreements*. Bhagwati, Jagdish and Anne O. Krueger, eds. Washington, D.C.: American Enterprise Institute for Public Policy Research.
- Bhagwati, Jagdish, David Greenaway and Arvind Panagariya. 1998. "Trading Preferentially: Theory and Policy," *Econ. J.*, 108:449, pp. 1128-1148.
- Bhagwati, Jagdish, Pravin Krishna and Arvind Panagariya, eds. 1999. *Trading Blocs: Alternative Approaches to Analyzing Preferential Trade Agreements*, Cambridge, MA: MIT Press.
- Bhagwati, Jagdish and Arvind Panagariya. 1996a. "Preferential Trading Areas and Multilateralism: Strangers, Friends or Foes?" in *The Economics of Preferential Trade Agreements*. Jagdish Bhagwati and Arvind Panagariya, eds. Washington, D.C: AEI Press, pp. 1-78.
- Bhagwati, Jagdish and Arvind Panagariya. 1996b. "The Theory of Preferential Trade Agreements: Historical Evolution and Current Trends." *Amer. Econ. Rev.* 86:2, pp. 82-87.
- Bond, Eric W. and Constantinos Syropoulos. 1996. "The Size of Trading Blocs, Market Power and World Welfare Effects," *J. of Int. Econ.* 40:3-4, pp. 411-437.



- Brecher, Richard and Jagdish Bhagwati. 1981 "Foreign Ownership and the Theory of Trade and Welfare." *J. of Polit. Econ.*, 89:3, pp. 497-511.
- Cadot, Olivier, Jaime de Melo and Marcelo Olarreaga, 1996. "Regional Integration and Lobbying for Tariffs Against Non-Members," mimeo, University of Geneva. Forthcoming in *Int. Econ. Rev.*
- Davis, Donald. 1997. "Common External Tariffs of a Customs Unions: Alternative Approaches. Comment," *Japan and the World Economy* 9:4, pp. 467-70.
- Deardorff, Alan and Robert Stern. 1994. "Multilateral Trade Negotiations and Preferential Trading Arrangements," in *Analytical and Negotiating Issues in the Global Trading System*. Alan Deardorff and Robert Stern, eds. Ann Arbor, MI: University of Michigan Press, pp. 53-85.
- De la Torre, Adela and Morgan Kelly. 1992. "The Regional Trading Arrangements," Occasional Paper No. 93. Washington: International Monetary Fund.
- Dixit, Avinash. 1987. "Strategic Aspects of Trade Policy," in *Advances in Economic Theory: Fifth World Congress*. T. F. Bewley, eds., New York: Cambridge University Press, 329-62
- Duttagupta, Rupa. 1999. "Rules of Origin in Preferential Trade Agreements: Implications for the Viability of the Agreements and Welfare," University of Maryland at College Park, mimeo.
- Ethier, Wilfred, J. 1998. "Regionalism in a Multilateral World, *J. of Polit. Econ.* 106:6, pp. 1214-45.

- Ethier, Wilfred, J. and Henrik Horn. 1984. "A New Look at Economic Integration," in *Monopolistic Competition and International Trade*. Kierzkowski, Henryk, ed. Oxford: Clarendon Press, pp. 207-229.
- Falvey, Rod and Geoff V. Reed. 1997. "Rules of Origin as Commercial Policy Instruments," University of Nottingham, mimeo.
- Frankel, Jeffrey A. 1997. *Regional Trading Blocs in the World Trading System*. Washington, D.C.: Institute for International Economics.
- Frankel, Jeffrey A., ed. 1998. *The Regionalization of the World Economy*. Chicago: University of Chicago Press.
- Frankel, Jeffrey A., Ernesto Stein and Shang-Jin Wei. 1995. "Trading Blocs and the Americas: The Natural, the Unnatural and the Supernatural," *J. of Devel. Econ.* 47:1, pp. 61-96.
- Fernandez, Raquel and Jonathan Portes. 1998. "Returns to Regionalism: An Analysis of Nontraditional Gains from Regional Trade Agreements," *World Bank Econ. Rev.*, 12:2, pp. 197-220.
- Freund, Caroline. 1998. "Multilateralism and the Endogenous Formation of PTAs" Board of Governors of the Federal Reserve System, International Finance Discussion Paper #614, Washington, D.C. Forthcoming in *J. of Int. Econ.*
- Gehrels, Franz. 1956-57. "Customs Union from a Single-Country Viewpoint," *Rev. of Econ. Studies*, 24:6, pp. 61-64.
- Gros, Daniel. 1987. "A Note on the Optimal Tariff, Retaliation and the welfare Loss from Tariff Wars in a Framework with Intra-Industry trade," *J. of Int. Econ.* 23:3-4, pp. 357-367.

- Grossman, Gene and Elhanan Helpman. 1994. "Protection for Sale," *Amer. Econ. Rev.*, 84:4, pp. 835-50.
- Grossman, Gene and Elhanan Helpman. 1995. "The Politics of Free Trade Agreements," *Amer. Econ. Rev.* 85:4, pp. 667-690.
- Helpman, Elhanan. and Paul R. Krugman. 1985. *Market Structure and Foreign Trade. Increasing Returns, Imperfect Competition, and the International Economy.* Cambridge, Mass.: MIT Press.
- Hindley, Brian. and Patrik Messerlin. 1993. "Guarantees of market Access and Regionalism," in *Regional Integration and the Global Trading System.* K. Anderson and R. Blackhurst, eds. London: Harvester Wheatsheaf.
- Jensen, R. and M. Thursby. 1984. "Free Trade: Two Noncooperative Approaches," Ohio State University, Working Paper.
- Ju, Jiandong and Kala Krishna. 1998. "Firm Behavior and Market Access in a Free Trade Area with Rules of Origin," NBER Working Paper, No. 6857, December.
- Kemp, Murray, C. 1964. *The Pure Theory of International Trade.* Englewood Cliffs, N.J. Prentice-Hall, pp. 176-177.
- Kemp, Murray and Henry Jr. Wan. 1976. "An Elementary Proposition Concerning the Formation of Customs Unions", *J. of Int. Econ.* 6:1, pp. 95-98.
- Kowalczyk, Carsten., 1990, "Welfare and Customs Unions," NBER Working Paper No. 3476.
- Krishna, Pravin. 1998. "Regionalism and Multilateralism: A Political Economy Approach," *Quart. J. of Econ.*, 113:1, pp. 227-251.

- Krishna, Pravin and Jagdish Bhagwati. 1997. "Necessarily Welfare-enhancing Customs Unions with Industrialization Constraints," *Japan and the World Economy* 9:4, pp. 441-446.
- Krueger, Anne O. 1999. "Free Trade Agreements as Protectionist devices: Rules of Origin," in *Trade Theory and Econometrics : Essays in Honour of John S. Chipman*. James R. Melvin, James C. Moore and Raymond Riezman, eds. New York: Routledge.
- Krugman, Paul. 1980. "Scale Economies, Product Differentiation, and the Pattern of Trade," *Amer. Econ. Rev.*, 70:5, pp. 950-959.
- Krugman, Paul. 1991a. "Is Bilateralism Bad?" in E. Helpman and A. Razin (eds.), *International Trade and Trade Policy*, Cambridge, Mass.: MIT Press, pp. 9-23.
- Krugman, Paul. 1991b. "The Move to Free Trade Zones," in *Policy Implications of Trade and Currency Zones*. Symposium Sponsored by the Federal Reserve Bank of Kansas City, pp. 7-41.
- Krugman, Paul. 1993. "Regionalism versus Multilateralism: Analytical Notes," in *New Dimensions in Regional Integration*. Jaime de Melo and Arvind Panagariya eds. Cambridge, Great Britain: Cambridge University Press, pp. 58-84.
- Levy, Philip. 1997. "A Political-Economic Analysis of Free-Trade Agreements," *Amer. Econ. Rev.*, 87:4, pp. 506-519.
- Lipsey, Richard. 1957. "The Theory of Customs Unions: Trade Diversion and Welfare," *Economica* 24:93, pp. 40-46.
- Lloyd, Peter J. 1982. "3x3 Theory of Customs Unions," *J. of Int. Econ.* 12:1-2, pp. 41-63.
- Ludema, Rodney. 1992. "On the Value of Preferential Trade Agreements in Multilateral Negotiations," unpublished manuscript.

- McLaren, John. 1998, "A Theory of Insidious Regionalism," Department of Economics, Columbia University, mimeo.
- McMillan, John. 1993. "Does Regional Integration Foster Open Trade? Economic Theory and GATT's Article XXIV," in *Regional Integration and the Global Trading System*, Kym Anderson and Richard Blackhurst, eds. New York: St. Martin's Press, pp. 292-310.
- McMillan, John and Ewen McCann. 1981. "Welfare Effects in Customs Union," *Econ. J.* 91:363 , pp. 697-703.
- Meade, James E. 1955. *The Theory of Customs Unions*. Amsterdam: North-Holland.
- Melo, Jaime de and Arvind Panagariya eds. 1993. *New Dimensions in Regional Integration*, Cambridge, Great Britain: Cambridge University Press.
- Mundell, Robert A. 1964. "Tariff Preferences and the Terms of Trade," *Manchester School of Economic and Social Studies*, pp. 1-13.
- Ohyama, Michihiro. 1972. "Trade and Welfare in General Equilibrium," *Keio Econ. Studies* 9, pp. 37-73.
- Panagariya, Arvind. 1996. 'The Free Trade Area of the Americas: Good for Latin America?' *World Economy* 19:5, pp. 485-515.
- Panagariya, Arvind. 1997a. "Preferential Trading and the Myth of Natural Trading Partners," *Japan and the World Economy* 9:4, pp. 471-489.
- Panagariya, Arvind. 1997b. "The Meade Model of Preferential Trading: History, Analytics and Policy Implications," in *International Trade and Finance: New Frontiers for Research. Essays in Honor of Peter B. Kenen*. Cohen, B.J., ed., (Cambridge University Press, New York) pp. 57-88.

- Panagariya, Arvind. 1998. "Do Transport Costs Justify *Regional* Preferential Trade Arrangements? No," *Weltwirtschaftliches Archiv* 134:2, pp. 230-301.
- Panagariya, Arvind. 1999a. "The Regionalism Debate: An Overview," *World Economy* 22:4, pp. 477-511.
- Panagariya, Arvind. 1999b. *Regionalism in Trade Policy: Essays on Preferential Trading*. Singapore: World Scientific Press Co..
- Panagariya, Arvind. 1999c. "Preferential Trading and Welfare: The Small-Union Case Revisited," University of Maryland, mimeo.
- Panagariya, Arvind and Rupa Duttagupta. 1999. "The "Gains" from Preferential Trade Liberalization in the CGEs: Where From do They Come?" University of Maryland, mimeo. Forthcoming in *Regionalism and Globalization: Theory and Practice*. Sajal Lahiri, ed. London: Routledge.
- Panagariya, Arvind and Ronald Findlay. 1996. "A Political Economy Analysis of Free Trade Areas and Customs Unions," in *The Political Economy of Trade Reform: Essays in Honor of Jagdish Bhagwati*. Robert Feenstra, Douglas Irwin and Gene Grossman, eds. Cambridge, Mass.: MIT Press, pp. 265-287.
- Panagariya, Arvind and Pravin Krishna. 1997. "On the Existence of Necessarily Welfare-Enhancing Free Trade Areas," Working Paper No. 32, Center for International Economics, Department of Economics, University of Maryland.
- Puga, Diego and Anthony J. Venables. 1995. Preferential Trading Arrangements and Industrial Location, CEPR Discussion Paper No.1309, December.
- Richardson, Martin. 1993. "Endogenous Protection and Trade Diversion," *J. of Int. Econ.* 34:3-4, pp. 309-24.

- Richardson, Martin. 1994. "Why a Free Trade Area? The Tariff Also Rises," *Economics & Politics* 6:1, pp. 79-95.
- Riezman, Raymond. 1979. "A 3x3 Model of Customs Unions," *J. of Int. Econ.* 9:4, pp. 341-354.
- Sapir, Andre. 1998. "The Political Economy of the EC regionalism," *Europ. Econ. Rev.* 42:3-5, pp. 717-732.
- Smith, Alasdair and Anthony J., Venables. 1988. "Completing the Internal Market in the European Community: Some Industry Simulations," *Europ. Econ. Rev.* 32:, pp. 1501-1525.
- Srinivasan, T.N. 1993. "Regionalism versus Multilateralism: Analytical Notes. Comment," in *New Dimensions in Regional Integration*, Jaime de Melo and Arvind Panagariya eds. Cambridge, Great Britain: Cambridge University Press, pp. 84-89.
- Srinivasan, T. N. 1997. "Common External Tariffs of a Customs Unions: Alternative Approaches," *Japan and the World Economy* 9:4, pp. 447-70.
- Srinivasan, T.N., John Whalley and Ian Wooton. 1993. "Measuring the Effects of Regionalism on Trade and Welfare," in *Regional Integration and the Global Trading System*. Kym Anderson and Richard Blackhurst, eds. New York: St. Martin's Press, pp. 52-79.
- Summers, Lawrence. 1991. "Regionalism and the World Trading System," in *Policy Implications of Trade and Currency Zones*. Symposium Sponsored by the Federal Reserve Bank of Kansas City, pp. 295-301.
- Vanek, Jaroslav. 1965. *General Equilibrium of International Discrimination. The Case of Customs Unions*. Cambridge, Mass.: Harvard University Press.

- Viner, Jacob. 1950. *The Customs Union Issue*. New York: Carnegie Endowment for International Peace.
- Winters, L. Alan. 1996. "Regionalism versus Multilateralism," Policy Research Working Paper 1687, Washington, D.C: World Bank.
- Winters, Alan. 1997. "Regionalism and the Rest of the World: The Irrelevance of the Kemp-Wan Theorem," *Oxford Economic Papers* 49:2, pp. 228-234.
- Wonnacott, Paul and Mark Lutz. 1989. "Is There a Case for Free Trade Areas?" in *Free Trade Areas and U.S. Trade Policy*. Schott, Jeffrey, Washington, D.C.: Institute for International Economics, pp. 59-84.
- Wonnacott, Paul and Ronald J. Wonnacott. 1981. "Is Unilateral Tariff Reduction Preferable to a Customs Union? The Curious Case of the Missing Foreign Tariff," *Amer. Econ. Rev.* 71:4, pp. 704-14.
- Yi, Sang-Seung. 1996. "Endogenous Formation of Customs Unions under Imperfect Competition: Open Regionalism is Good," *J. of Inter. Econ.* 41:1-2, pp. 153-77.





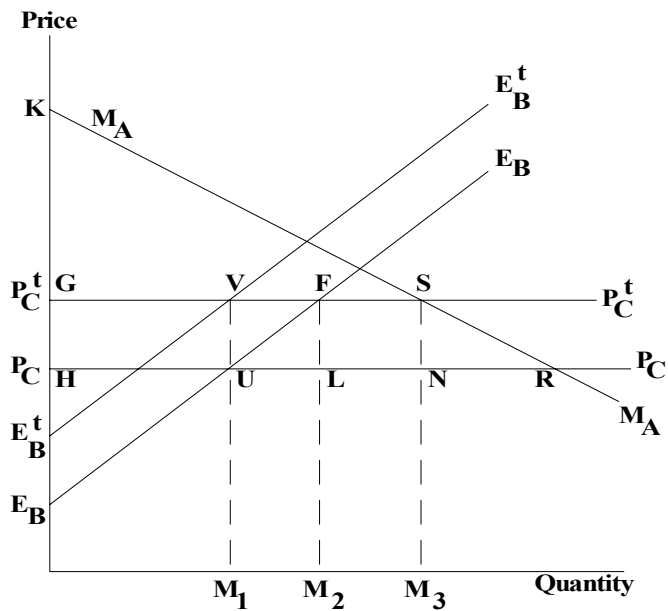


Figure 2. CU: A Preferential removal of a Tariff by A Leads to a Loss of  $GFLH$  to Itself, a Gain of  $GFUH$  to Partner B, and a Net Loss of  $F$  to the Union as a Whole.

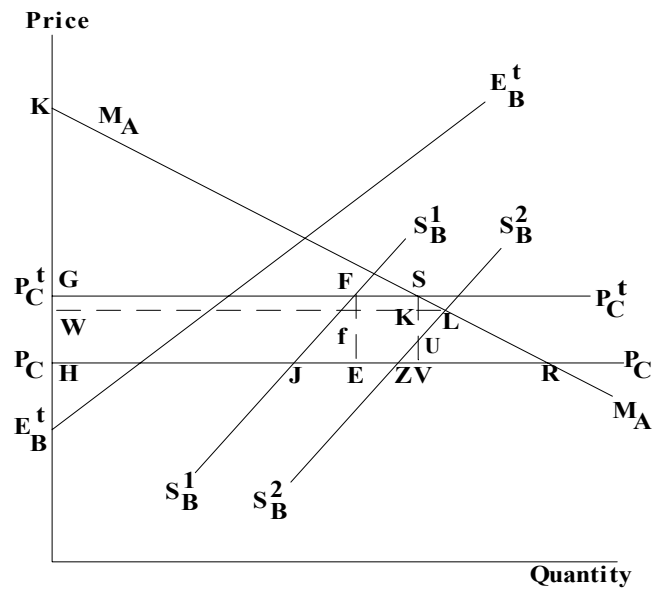


Figure 3. The Effect of Preferential Removal of Tariff by A under FTA

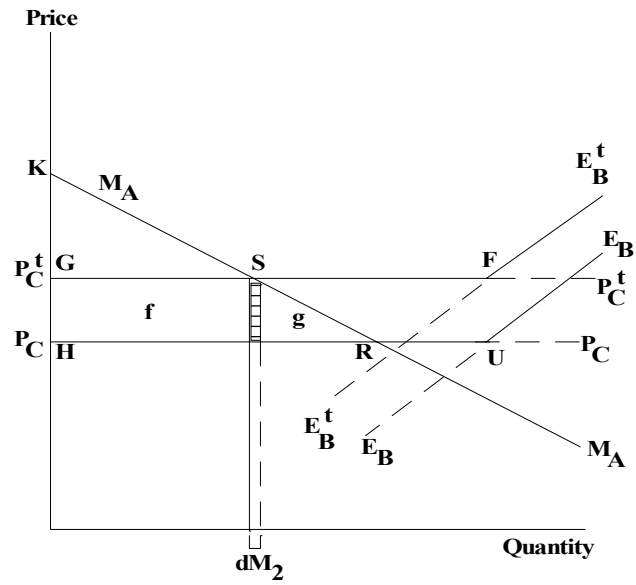


Figure 4. FTA: No Imports Come from Country C. Preferential Liberalization Coincides with Non-discriminatory Liberalization

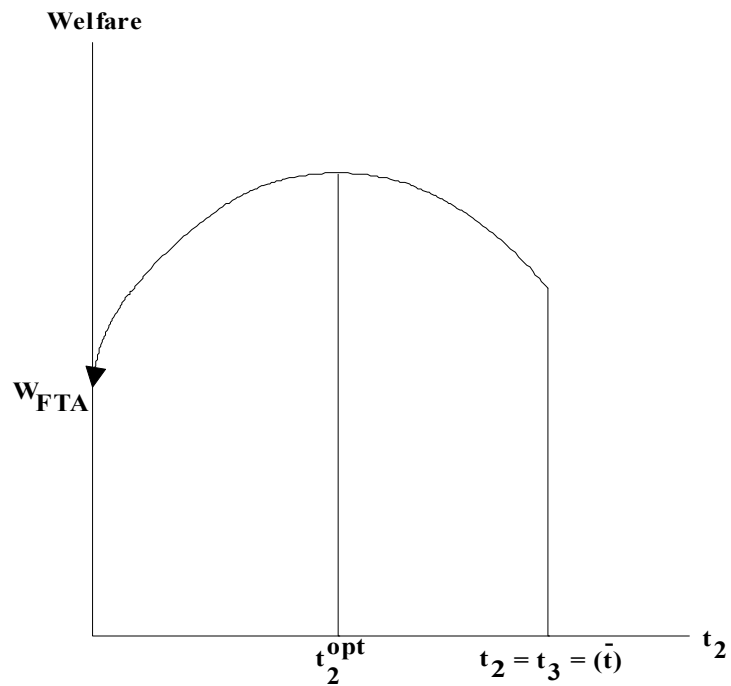


Figure 5. Meade-Lipsey Model: The Change in Welfare as the Tariff on the Partner ( $t_2$ ) is reduced from  $t_2$  to 0

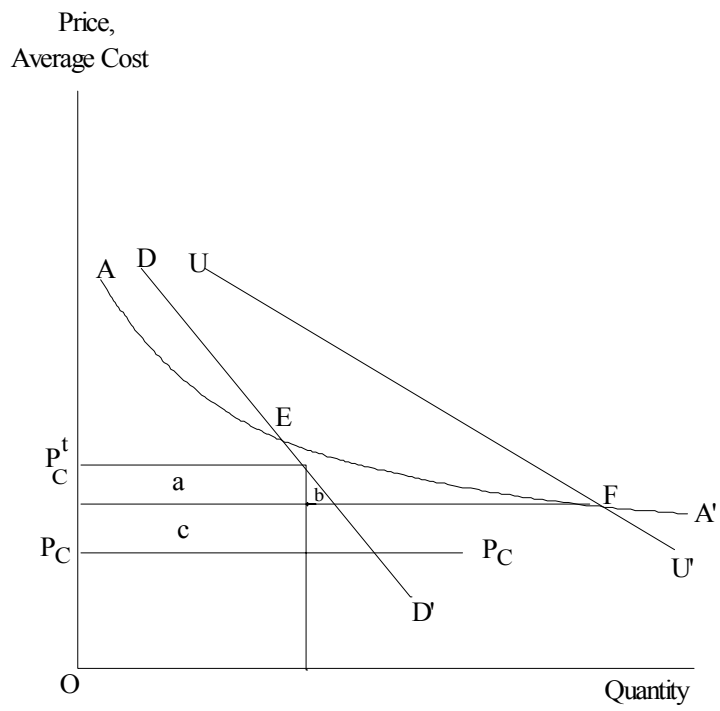


Figure 6: Economies of Scale and Preferential Trading

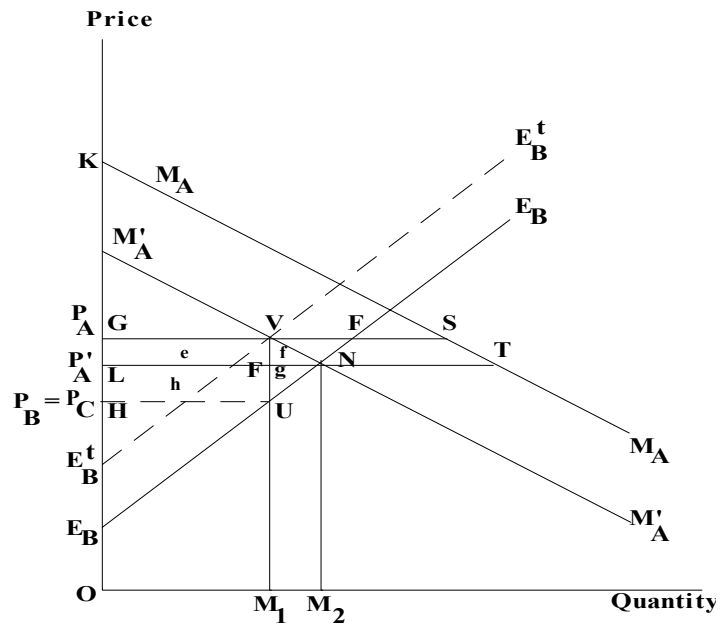


Figure 7::The Kemp-Ohyama-Kemp-Wan Customs Union

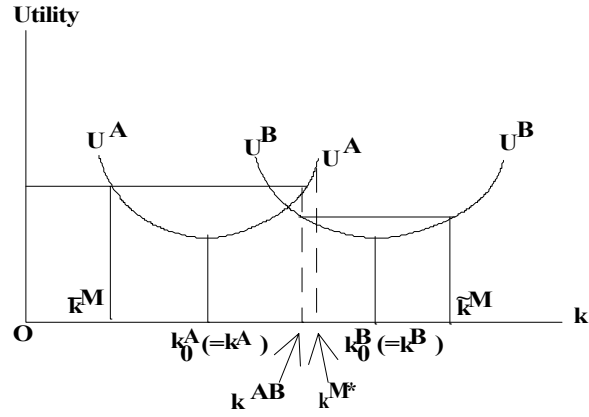


Figure 8a. In the Heckscher-Ohlin Model, Trade Bloc Neither Helps Nor Hurts the Multilateral Accord

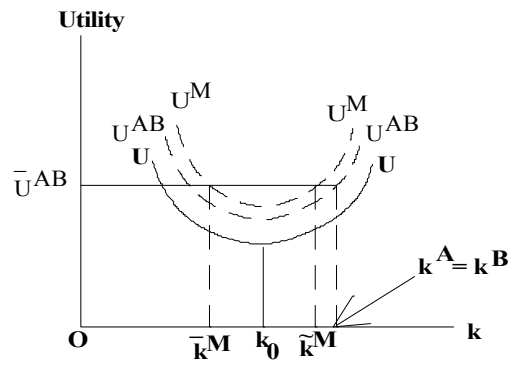


Figure 8b. With Product Differentiation Added to the Model, a Trade Bloc can Block the Multilateral Accord