Preferential Trade Liberalization:

The Traditional Theory and New Developments

Arvind Panagariya*

*Department of Economics, University of Maryland, College Park MD 20742. Email: panagari@econ.umd.edu

Table of Contents

Introduction

Like the first wave of regional arrangements, launched in the 1950s with the founding of the European Economic Community, the current wave has given rise to a lively debate between free traders who view preferential trade liberalization as harmful and those who see them as beneficial.[1] To the old concerns relating to welfare effects, captured in Viner's (1950) influential "static" concepts of trade creation and trade diversion, the current debate has added what Bhagwati (1993) calls the "dynamic" time-path issue.

The effective arrangements during the first wave did not spread beyond Western Europe. 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. First, its emphasis is almost exclusively on theory. In 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.[2] 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 contribution 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

---

[1]Throughout this essay, the term "regional arrangements" refers to preferential trade arrangements, defined more precisely below.

[2]Survey articles include Baldwin and Venables (1995), Bhagwati and Panagariya (1996a, 1996b), Bhagwati, Greenaway and Panagariya (1998), Panagariya (1998a) and Winters (1996). Recent book-length treatments are Anderson and Blackhurst (1993), Melo and Panagariya (1993) and Frankel (1997, 1998). Bhagwati, Krishna and Panagariya (1999) offers a collection of most of the contributions reviewed in this paper.

explicit role.   As regards matters such as preferential trade in services, role of investment in regional arrangements, and harmonization of domestic policies, though they figure in the current policy debate, they have not been seriously addressed in the theoretical literature.[3]

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 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 limited tariff preferences, FTAs and CUs.

In Section 1, I develop carefully the traditional welfare analysis, adding some new twists to it. In Section 2, I discuss the Kemp-Ohyama-Kemp-Wan theorem on CUs and its recent extension to FTAs by Panagariya and Krishna (1997) which cut through the ambiguities of trade creation and trade diversion and offer a clear approach to forming welfare-enhancing regional arrangements. In Section 3, 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 4, 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 5, I turn to the models addressing directly the relationship between regionalism and multilateralism. In section 6, I briefly examine the empirical approaches to resolving the theoretical ambiguity in welfare outcomes. I conclude the paper in section 7.

Before I proceed to the main body of the essay, a cautionary note is in order. In the policy debate, I have

---

[3]Even within this narrow definition of regional arrangements, space constraints preclude full coverage. For instance, the analysis of the distributional effect of a policy change (e.g., a reduction in the common external tariff) within an existing common market with internal factor mobility, pioneered by Brecher and Bhagwati (1981), is not considered.

sided with multilateralists who argue that regional arrangements fragment the global trading system and should, therefore, be discouraged. Though I have tried to be neutral in this largely theoretical--as opposed to policy--survey, occasional failures are inevitable. It is my hope, however, that having read the survey, the reader will be in a position to form his or her own opinion.

1. The Traditional Welfare Analysis

Though the current wave of regionalism has given rise to new concerns, the old concerns have remained alive as well. Therefore, it is appropriate to begin with the traditional welfare analysis. I begin by spelling out the broad structure of the model to be used throughout this section.

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.[4] If an FTA is 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.[5] Other details, relating to demand, supply, trade and tariffs are spelt out in the context of specific models.

1.1 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) in his classic work, The Customs Union Issue. As Meade (1955, Ch. 2) noted in his own 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.

---

[4]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.

[5]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 the other cases, this traditional assumption, made by all analysts implicitly or explicitly, is GATT inconsistent.

Constant prices at which steel is supplied by firms in A, B and C are given 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, $P_A > P_B > P_C$, implying that A is the least efficient supplier of steel and C the most efficient one.

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.[6] 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 A removes 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 quantity of steel, OQ_0, is supplied by A's own firms at price P_A. Suppose once again that A removes 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 in B, in Viner's terminology, the union is "trade creating."[7] Welfare of A and the world rises by f+g while that of B and C is

---

[6]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.

[7]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) 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

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 cannot 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, the relative magnitudes of trade creation and trade diversion alone are insufficient to determine the welfare 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 Gehrels (1956-57) and Lipsey (1957) within a one-factor, general-equilibrium model, is that once we drop 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_AD_A, be negatively sloped. The initial non-discriminatory tariff is set at t with 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.[8]

divert production to a source with still lower costs." (43)

[8]Because quantity Q_1Q_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-Lipsey interpretation: "It will be noted that for the free-trader the benefit

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 of steel in country A is positive but finite. In general, to eliminate the possibility of a trade-diverting union leading to welfare gains, one 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.

Despite these limitations, trade creation and trade diversion have remained central to policy debates. This is presumably because economists have found them an effective tool for focusing policy makers' attention on the ambiguous welfare effects of PTAs.

1.2   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 A import all its steel 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, letting B be 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 $P_C P_C$.

Initially, A imposes a per-unit tariff at rate t on both B an 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 A settles at

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.

P_C^t with imports from B and C equalling OM_1 and M_1M_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 P_C 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_BE_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_1M_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.[9] Of the total revenue transferred, GFUH 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 world loses FLU.

Assuming (i) the union's external tariff is set equal to the pre-union tariff of the importing member, (ii) all products continue to be imported from the rest of the world after the formation of the union, (iii) the union is small relative to the rest of the world, and (iv) production in union members is characterized by increasing opportunity costs, three striking results can be gleaned from the construction in Figure 2.[10]

First, the formation of the union necessarily lowers the welfare of one of the partners, union as a whole, and the world. It is possible for the welfare of every partner to decline. Only if either the external tariff is fixed

---

[9]This revenue-transfer effect is also present in the general-equilibrium analyses of Berglas (1979) and Riezman (1979).

[10]Panagariya (1999) provides formal proofs of these results.

7

below the initial tariff or the formation of the union eliminates entirely the imports from the outside country can the loss from trade diversion be partially or wholly offset by a gain. In either of these cases, the price in the importing country falls and new trade created.

Second, preferential liberalization by a country hurts itself and benefits the exporting union partners. And the more the imports from the partners in the post-union equilibrium, the larger the former's losses and the latter's gains. This result contradicts the so-called "natural-trading-partners" hypothesis, enunciated by Wonnacott and Lutz (1989) and espoused by Summers (1991) and Krugman (1993), according to which the more two countries trade with each other relative to the outside world, the less likely that a union between them will be harmful.

Finally, in a union between two countries with approximately balanced bilateral trade, the country importing products with higher external tariffs loses while the country importing products with lower external tariffs benefits. This is because the country importing goods with higher tariffs makes a larger tariff-revenue transfer to the country importing lower-tariff goods than it receives from the latter. Analogously, if the partners have equal tariffs, the country with bilateral trade deficit stands to lose from the union.

These results are modified under two circumstances. First, if $E_B E_B$ in Figure 2 lies sufficiently far to the right to intersect $M_A M_A$ below point S, thus, eliminating country C as a supplier of steel to the union, the internal price in A decline. New trade is created now, 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 and the smaller the negative effect on the union. In principle, it is possible for the positive effect to dominate. Also observe that with a price decline, a part of the lost tariff revenue now goes to augment the consumers' plus producers' surplus in A.

Second, continuing to assume that imports from C are not eliminated, benefits to the union may arise if the outside country's export supply is less than perfectly elastic. In this large-union case, the diversion of demand away from C as a consequence of preferential liberalization can improve the union members' *extra-union* terms of trade. This improvement counteracts the losses suffered by the higher-tariff union member through

the tariff-revenue-transfer effect or what is, equivalently, a deterioration in its *intra-union* terms of trade. If the former effect dominates, it may experience a net gain in welfare.

1.3 Extension to Free Trade Areas

Traditionally, the analyses of FTAs have assumed that the price facing consumers and producers in each member country is the world price plus its own external tariff.[11] But, as Richardson (1994) has pointed out recently, this is problematic since producers are free to sell their output freely anywhere within the union. If the price is higher in country A, producers in B will sell all their output in that country, letting the demand in B be satisfied entirely by imports. As Grossman and Helpman (1995) have shown and Bhagwati and Panagariya (1996a) have elaborated, this feature 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 P_C. Suppose now that A and B form an FTA with A setting its external tariff at t_A and B at t_B. Three cases must be distinguished.[12]

Case 1: Combined Supplies of A and B are Insufficient to Satisfy the Demand in A. In this case, A must still import steel from C which ensures that the price in the country must settle at $P_C + t_A$. With imports in B being subject to a lower tariff, the price there, $P_C + t_B$, is lower than $P_C + t_A$. As a result, producers in B divert their entire supply 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, $S_B{}^1 S_B{}^1$.

---

[11]For example, see Lloyd (1982) and McMillan and McCann (1981).

[12]The following analysis assumes that FTAs are supported by rules of origin which ensure that goods are not imported by a lower-tariff member for re-export 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, specially by Krueger (1993), in principle, they can lead to beneficial effects by reversing the effect of a tariff preference that was trade diverting in the first place. Krishna and Krueger (1995) and Panagariya (1998a) further discuss the analytic aspects of the rules of origin.

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 triangle marked f and a similar triangle in consumption due to the rise in the domestic price from P_C to P_C+t_B.[13] It is easily verified that the net gain to B is strictly positive. The loss to the union as a whole is represented by triangle f plus the triangular loss in consumption in B just mentioned.

It can be deduced from this analysis that, in a multi-good model, the formation of an FTA between two countries with approximately balanced trade will make the member with high external tariffs worse off and the member with low external tariffs better off. The high-tariff country makes a larger revenue transfer to the low-tariff country than it receives from the latter. Thus, qualitatively, the results of Section 1.2 are reproduced.

Case 2: Combined Supply of A and B Satisfies the Demand in A with B Selling All its Output in A. If the B's supply curve lies sufficiently far to the right to cross M_AM_A below point S, as shown by $S_B^2S_B^2$ in Figure 3, the FTA eliminates C as a source of imports and the price in A falls below P_C+t_A. Letting the height of point W equal P_C+t_B, as long as $S_B^2S_B^2$ crosses M_AM_A strictly between S and W, the price in A remains above P_C+t_B. As in Case 1, all within-union supply is sold in A and all of B's demand is satisfied by imports from C.

Though the tariff revenue collected initially by A is lost entirely now, not all of it is transferred to B. Given the price decline, a part of the lost revenue goes to augment the sum of the consumers' and producers' surplus in A. As before, there is a deadweight loss in production (triangle g in Figure 3) and consumption (under country B's demand curve, not shown in Figure 3). But this time, in addition, we have a gain in A (triangle h) generated by the new trade created by the price decline. In principle, this gain can outweigh the loss due to the revenue-transfer effect as also the deadweight loss represented by g and the triangle under the demand curve in B. But in general, the effect on A and the union is ambiguous while that on B is positive.

Case 3: Combined Supply in A and B Satisfies the Demand in A with B Selling in Both Countries. If B's supply curve intersects M_AM_A below point W, the price in A drops to the tariff-inclusive price in B, given by

---

[13]Recall that in the pre-FTA equilibrium, the domestic price in B is P_C even though it has a tariff $t_B \geq 0$ on the books.

the height of W. 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_AM_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.

1.4   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 a case, there is no outside trade to be diverted in the first place. Maintaining the small-union assumption, the point is illustrated in Figure 4 where, as before, M_AM_A is A's import-demand curve and E_BE_B B's export-supply curve. Given P_C as the price in the rest of the world, firms in B do not offer any steel for sale to A below this price. As such, under free trade, the effective export-supply curve of B, facing country A, is HUE_B. Under a nondiscriminatory tariff at rate t per-unit, the supply curve is $GFE_B{}^t$, where F lies vertically above U. The tariff-inclusive price of the rest of the world is $P_C{}^t$ ($=P_C+t$) which is also the internal price in A. 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.

What happens if country A lowers the tariff on B in this case? Curve $GFE_B{}^t$ 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. In the limit, as the tariff on B is eliminated, A gains triangle g (which includes the shaded strip, though it has no special significance for the present analysis) while B and C are unaffected by the change. 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 are ruled out.

By itself, this case is uninteresting for two reasons. First, it avoids trade diversion by assumption: A has no trade with C initially that could be diverted. Second, the FTA in this case is no different from free trade.

The tariff on C plays no role whatsoever and its elimination (along with t_B) has no impact on the outcome.

The case can be made to yield something more interesting, however, when it is embedded in a general-equilibrium model. This was demonstrated by Lipsey (1958), using Meade's three-good framework.[14] 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.

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.

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 3 (the liberalized good and the exportable, respectively), for a small reduction in t_2, the benefit from trade creation

---

[14]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).

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 $t_2 = t_3 \equiv \bar{t}$, 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.[15]

Though the Meade-Lipsey model has been influential in the literature on preferential trading, it suffers from a serious flaw: A and B have no incentive whatsoever to coordinate their liberalization through an FTA. None of the agents in B are affected by A's liberalization and vice versa. The formation of the FTA is identical to a unilateral trade reform.[16]

One way this flaw can be somewhat remedied is by relaxing the small-union assumption. The FTA can then serve as an instrument of improving the union's terms of trade vis-a-vis the rest of the world. Assuming that import demands exhibit substitutability, preferences by A and B to each other divert demands towards their goods and away from C's goods. There is a strong presumption that this change improve their terms of trade relative to the latter and confers welfare gains on them.

But even in this case, results can be asymmetric between high- and low-tariff members. To see this,

---

[15]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 $t_2 = t_3 = \bar{t}$.

[16]Not surprisingly, the result just discussed is a close cousin of the concertina theorem of piecemeal trade reform.

consider a one-way preference by A to 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 relative to goods 1 and 3 increases, implying 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 in general. 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 1.2 and 1.3, 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, 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 reaches the following conclusions:[17]

"(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 1.2 and 1.3, 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.[18]

1.5   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. At least for

the problem at hand, this is an incorrect assertion.[19] The main complication in the presence of differentiated

---

[17]As far as I can tell, by initial tariffs being low, Mundell means that they are below their optimum levels.

[18]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.

[19]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 sub-section is based,

goods is that we can no longer use the simplifying, small-union assumption. Each country has a monopoly power over its products and can influence its terms of trade. For example, as 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.

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 1.4. 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.

1.6   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.[20] This is new development. For example, Viner (1950), who was aware of the departures from the Most Favored Nation (MFN) principle in commercial pacts in Europe going as far back as the nineteenth century, attributed them to 'close ties of sentiment and interest arising out of ethnological, or cultural, or historical political affiliations' rather than any underlying economic factors.

---

economies of scale do not play a substantive role in it. If economies of scale are given a serious play, we must deal with multiple equilibria which make the analysis complex even in the standard two-country models.

   [20]In particular, see Wonnacott and Lutz (1989), Krugman (1991b, 1993), Summers (1991) and Frankel, Stein and Wei (1995).

There is, indeed, no basis in theory for giving transport costs special treatment in evaluating PTAs.[21] 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. The analysis of PTAs presented above is also valid with or without transport costs. The discussion associated with Figures 1-4 is quite consistent with transport costs; all we need to do is to interpret the supply curves in these figures as including transport costs.

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.

1.7   Summary

The analysis in this section suggests only two circumstances under which preferential trade liberalization, holding external tariffs at their initial levels, gives rise to beneficial effects for the union that may or may not outweigh the accompanying harmful effects: when the union is small and the product being liberalized is not imported from the rest of the world in the final equilibrium; and when the union is large so that preferential liberalization can improve the union's terms of trade.

2.   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

---

[21]For detailed critiques, see Bhagwati and Panagariya (1996a) and Panagariya (1998b). Wonnacott and Wonnacott (1981) give a special role to transport costs but, as shown by Berglas (1983) and further discussed in Panagariya (1998b), 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.

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.

2.1   Customs Unions

This result was first stated for a CU by Kemp (1964) and proved by Ohyama (1972) and Kemp and Wan (1976) and is referred to as the Kemp-Ohyama-Kemp-Wan theorem in this paper.[22] 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 removing all intra-union trade barriers and fixing 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-Ohyama-Kemp-Wan theorem.[23]   In Figure 6, 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.

As a part of the CU, A and B remove trade barriers between themselves and adopt a CET so as to freeze the quantity of imports from C at VS. To derive the 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

[22]As documented in Panagariya (1997b), Vanek (1965, chapter 5) also mentions this result and proves it for a two-good model. Vanek (1965) and Kemp (1964) do not refer to each other, however.

[23]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).

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 1.2 and 1.3, 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 f 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.

2.2   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 is 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.

2.3   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.

3.   Exogenous Division of the World into Blocs

The massive wave of regional arrangements that started in the 1980s has led some analysts to look at the welfare effects of a simultaneous division of the world into several blocs. A key question these models ask is how does welfare change with a change in the number of blocs and bloc size.

3.1   Symmetric Blocs

The contribution which set off this literature is Krugman (1991a), which envisages 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. The world is assumed to be divided into B identical blocs with 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 result [Krugman (1993)], the model is best understood by holding the tariff constant. The modification introduced by the endogeneity of the external tariff can be added later.

To understand the model, a key point to bear in mind is that, given complete symmetry, a change in the number of blocs and, hence, their size, generates no terms of trade effects. The 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 world-wide 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. This he finds very interesting since the current trend is precisely towards two or three large blocs.

3.2   Asymmetric Blocs

Given the special, highly symmetric structure of Krugman's model, it should come as no surprise that alternative formulations of the problem will alter the results. This has now been demonstrated by several authors. Deardorff and Stern (1994) provide a simple example in which a small number of blocs lead to the maximization of world welfare. In contrast to Krugman, 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.

Srinivasan (1993) offers an example in which the number of blocs has no relationship to welfare. Since the example is interesting in its own right, I offer a diagrammatic interpretation of it here. suppose there are two goods, 1 and 2, one factor of production, labor, and a continuum of countries. The labor-output ratio in good 1 is one everywhere but that in good 2 varies across countries. Denoting by b a constant such that $0 < b < 1$, let the labor-output ratio in good 2 be 'a,' where a is uniformly distributed over the interval (1-b, 1+b). Divide the countries into two sets: set I with countries having a à (1, 1+b) and labor endowment $\bar{L}(a) = 1$ and set II with countries having a à (1-b, 1) and labor endowment $L(a) = a$.

Globally, the countries in set I have a comparative advantage in good 1. They can produce either one unit of good 1 or $1/a$ (< 1) units of good 2. In Figure 7, the production possibilities frontiers of these countries originate from point A_I $(OA_I = 1)$ and are shown by the dotted lines. Countries in set II have a comparative advantage in good 2 and can produce either one unit of that good or a (≤ 1) units of good 1. Production possibilities frontiers of these countries originate at A_{II} $(OA_{II} = 1)$ and are shown by solid lines.

Assume identical Cobb-Douglas preferences with equal weights on the two goods. It is then easy to infer from Figure 7 that, under global free trade which maximizes world welfare, the relative price of good 1 is one. The countries in set I specialize in good 1 and those in set II in good 2 except the border line country with $a = 1$, which can produce anywhere along $A_I A_{II}$.

23

Consider alternative blocs now. Suppose we divide the world into two blocs such that one bloc consists of countries with extreme endowments 1-b and 1+b (with the innermost production possibilities curves) and the other all other countries. The relative price of good 1 in each bloc settles at one and welfare is maximized with no inter-bloc trade. We can also choose four, six, eight, etc. innermost countries to form one bloc and the rest to form the other bloc. The outcome with respect to welfare and the relative price will be unchanged. Alternatively, we can place all countries in set I into one bloc and those in set II into another. This entails a welfare loss.

It is evident that we cannot relate the number of blocs and welfare in this model. The result should not surprise us. To illustrate the special nature of Krugman's entirely symmetric model, Srinivasan has chosen here a model at the opposite extreme. The countries are as asymmetric as they can be.

3.3   "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.

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 which equals the initial tariff.

This model works entirely like the Meade-Lipsey model discussed in Section 1.4. 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

24

own and partner-country products also obtains since preferences are symmetric. Applying the logic outlined in Section 1.4, 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 (1998b). Thus, consider a two-good Ricardian world. Let there be two continents, each consisting of two countries. Assume zero transport costs within each continent but positive, *albeit* non-prohibitive, transport costs between them. 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

25

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. Transport costs are not special.

4. 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.[24] We may ask when will countries exercise the option to form FTAs and, alternatively, how will the decision to form an FTA affect the choice of the external tariff. These are political-economy-theoretic questions, requiring an explicit modelling of the decision-making process.

4.1 The Decision to Form an FTA

Two papers focus centrally on the decision to form an FTA, Grossman and Helpman (1995) and Krishna (1998). The former choose a two-member, small-union model and focus on the modelling of the political process underlying the decision to form an FTA. The latter relies on a Cournot oligopoly model.

The Small-Union Model

Grossman and Helpman (1995) take a specific-factors model with n+1 goods. One of these goods serves as the numeraire which is not subject to any trade barriers in either country. Non-numeraire goods use labor and a sector-specific factor while The numeraire good uses only labor. This assumption makes all non-numeraire goods independent of each other (i.e., neither substitutes nor complements) in production. It is further assumed that preferences are additively separable and the numeraire good enters into them linearly. Consequently, the demands for non-numeraire goods also become independent of each other. All non-numeraire goods behave exactly as in a partial-equilibrium model, allowing a direct application of the analysis in Sections 1.3 and 1.4.

Initially, the domestic price of an importable equals the world price plus the tariff on it. For an exportable, the tariff is set equal to zero. By assumption, under the FTA agreement, the external tariff on each good is set equal to the pre-FTA tariff on it. The only decision facing the (incumbent) governments in the two

---

[24]Two exceptions were Kemp-Ohyama-Kemp-Wan CU and Krugman (1991a).

countries is to either accept the FTA agreement or reject it.

The objective function each government pursues 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 the proposed trade agreement. In equilibrium, the contribution is made only if the government takes this desired action. In the game, the lobbies move first and the government second.

The solution to the problem is complex but it is possible to give some flavor of the results with the help of the analysis in Sections 1.3 and 1.4. 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 other 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 $M_A M_A$ 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 $S_1^B S_1^B$ 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 outweigh the political cost of the decline in welfare.

Second, suppose the 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. Since the FTA equilibrium in this example is equivalent to that under free trade, such

27

an outcome is all the more likely if the initial tariffs were themselves the result of a political game between the government and lobbies similar to the one played for the FTA.

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, a key conclusion of Grossman and Helpman is as follows: "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."

A Cournot Oligopoly Model

Krishna (1998) addresses the issue of endogenous FTA formation in 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.

Krishna (1998) shows 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

28

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 in this case.

4.2. The Extra-Union Tariff

Bhagwati (1993) expressed the concern that the decision by a country to enter into a PTA may result in a rise in the extra-union tariff, turning even an initially trade-creating union into a trade-diverting union. Thus, in our three-country set up, increased imports from partner B in the post-FTA equilibrium that threaten A's firms may lead the latter to seek higher tariffs on imports from C.[25]

Lobbying and the External tariff in Small-Union Models

Panagariya and Findlay (1996) offer a model parallelling more closely Bhagwati's (1993) concern. They consider a small union in which tariffs are determined through lobbying and the outcome is an increase in the external tariff following the grant of a tariff preference. As in the Meade-Lipsey model, they assume two importable goods of which one comes from the partner and the other from the outside country. The tariff in each sector is determined by the amount of labor employed in lobbying by it.

The grant of a tariff preference is modeled as an institutional change which 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. 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.

Cadot, de Melo and Olarreaga (1996) also use the Meade-Lipsey, three-good model in which tariffs are determined endogenously via the Grossman-Helpman (1994) model. They find that in FTA arrangements without rules of origin, competition for tariff revenue may lead to competitive reductions in tariffs resulting in their complete removal. In CUs, by contrast, lobbies may cooperate on a union-wide basis and win increased

---

[25]A formal presentation of this point can be found in Bhagwati and Panagariya (1996b).

external protection.

<u>Turning Trade Diversion into Tariff Revenue</u>

Richardson (1993) offers a case in which, following an FTA, a member government may actually be induced to *lower* the external tariff to capture potential tariff revenue for itself rather than pay for the inefficiency of partner countries' firms . The 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.

Richardson's key point is that, in this setting, if A reduces the tariff on C to $P_B P_C$-à where à 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 is its reliance on a model in which imports come from either B or C but not both. As already noted, this outcome follows 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.

<u>Increased Monopoly Power of Trade Blocs</u>

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.

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.[26]

4.3   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 arrangements. As documented in Panagariya

---

[26]In a Cournot oligopoly model, Yi (1996) also finds that the external tariff may rise or fall as an only existing bloc expands.

(1998a), 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.

5.   Regionalism and Multilateralism

Indirectly, we have already begun to explore the relationship between regionalism and multilateralism by asking in the pervious 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, some authors have gone on to add a third question: what is the impact of multilateralism on regionalism? We take these questions in turn.

5.1   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 profitability within the bloc thereby attracting an outside firm on the margin. 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.[27]

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 elegant paper of Andriamananjara (1999) who 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 assumed to be fixed 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 seek entry, insiders stop short of taking all of them into the club. The CU fails to expand into a global bloc.

---

[27]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.

Bond and Syropoulos (1996) ask this same question, albeit in slightly circuitous manner, using the model discussed in Section 3. 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.[28]

5.2   Impact of Regionalism on Multilateralism: Friends or Foes?

We now turn to models that address directly the impact of regionalism on multilateralism. A number of different approaches can be distinguished.

Stumbling Blocks: A Median Voter Model

Using a political-economy model in which decisions are made by a majority vote, Levy (1997) addresses two key questions: (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? The answer to the first of these questions is a straightforward no since the option to form a bloc is exercised only if it makes the median voter better off and, in that case, the median voter's reservation utility rises. If multilateralism was already infeasible, it cannot now become feasible.

The second question requires deeper analysis. Levy (1997) addresses it within two alternative 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 model, the option of a trade bloc cannot block a previously feasible multilateral accord but, in the second one, it can.

Some restrictive assumptions must be made to make the story tractable. Thus, 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

---

[28]Yi (1996) also considers the issue of bloc expansion in a Cournot oligopoly model.

the overall endowment ratio of the region within which trade is free.

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 as a function of the capital-labor ratio of the economy in which he operates (e.g., country A under autarky, A plus B under a trade bloc, and the world under global free trade) exhibits the pattern shown by curve $U^A U^A$ in Figure 8a (ignore $U^B U^B$ for now). 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.

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 $\overline{k}^M$ and no less than $\check{k}^M$ 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 $\overline{k}^M$ and $\check{k}^M$, 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 the 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 not hinders multilateralism. If one of the goods is assumed to be differentiated, however, the trade bloc can become a stumbling block to 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 $\overline{U}^{AB}$.

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 $\overline{k}^M$ and $\check{k}^M$, 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

Krishna (1998) also analyzes the "friends versus foes" issue within the oligopoly model mentioned in Section 4.1. The question is the same as Levy's: Does an initially feasible multilateral liberalization remain necessarily feasible after two of the three countries have formed an FTA? Krishna's answer is 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 from the former change is less than the loss due to the latter change.

Insidious Regionalism

A recent interesting paper by McLaren (1998) 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

Bagwell and 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.[29] They are, therefore, limited to self-enforcing multilateral arrangements that balance short-term gains from

---

[29]For earlier contributions in this tradition, see Kowalczyk and Sjostrom (1994) and Ludema (1992).

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 stationary dynamic tariff game. Phase 2 corresponds to a transition phase, in which trade between Home and Foreign continues but 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 benefits from cooperation have not changed but the latter one implies that the costs of deviation have 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 each of which consists 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

38

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.

5.3   The Impact of Multilateralism on Regionalism

Some authors have also argued that multilateral liberalization may itself be the cause of rising trend towards regionalism. We look at two contributions focusing on this theme.

Multilateral Liberalization Making PTAs More Sustainable

Freund (1998) uses a symmetric, three-country, Cournot oligopoly model in which, initially, each country levies the same 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 at low multilateral tariffs. Hence, PTAs may proliferate 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.

<u>Liberalization in North Leading to North-South PTAs</u>

Ethier (1998) constructs a model in which regionalism is an outcome of multilateral liberalization and has a happy coexistence with it. The basic story is as follows. The world is divided into two regions which I will call North and South. Each region consists of several countries. Northern countries are all symmetric with each producing one non-traded and one variety of a traded, differentiated good. The former uses unskilled and skilled labor while the latter uses human capital and skilled labor. The production of the differentiated good involves two stages: in the first stage, using only skilled labor, an intermediate input is produced and, in the second stage, the input is combined with human capital to produce the final good. A key feature, which drives many of the results, is the presence of an (international) external economy in the production of the intermediate input. The production cost of the input declines with the world-wide employment of skilled labor in it. The intermediate input can be produced anywhere but it must be shipped to the source country for the second stage of production.[30]

Initially, each Northern country imposes the Nash optimum tariff on the imports of the differentiated good from other countries.[31] Because the countries are symmetric, the tariff is the same for all of them. Southern countries can benefit by producing the intermediate input (or, in Ethier's terminology, attracting Northern firms to locate the production of the input inside their borders) and exchanging it for imports of the final good. But they face resistance to liberalization. This resistance varies across countries and is, initially, sufficiently strong even in the least resistant country to rule out a liberal regime. Therefore, all Southern countries are in autarky, producing and consuming a "rudimentary" good that is a (poor) substitute for the Northern traded good.

---

[30]Ethier defines the production of the input in another country as direct foreign investment, though neither investment nor technology flow are associated with the shift in location. Since the externality is international, there is no reason why any country wishing to produce the intermediate input could not produce it and ship it to the producer of the final good for the second-stage production. What Ethier calls "foreign investment" is more accurately interpreted as conventional international trade. At least in the data, the underlying transactions will appear as trade rather than foreign investment.

[31]The objective function of each government is a weighted sum of utility and unskilled wage.

The equilibrium is disturbed by a multilateral negotiation among Northern countries, which leads to a reduction in the tariff they impose on each other. Employment of skilled labor in the intermediate input expands everywhere and, given the externality, confers gains on all Northern countries. With the value of the externality in the intermediate-input production having gone up, some Southern countries may now be able to overcome the resistance to trade liberalization. If such a reform actually takes place, the production of the intermediate input moves partially to the reforming Southern countries. There is a further expansion of the externality effect.

Now introduce a regional arrangement. Assume that, under the arrangement, a Northern country gives a small tariff preference to the intermediate input produced in the Southern partner. The preference gives the latter an edge over all other Southern countries in the production of the input; it becomes the sole foreign source of the partner's input. In return, the Southern partner gives the Northern partner's traded variety free access while denying it entirely to nonmember Northern countries (by assumption).

The regional arrangement has the following effects. First, the Southern partner becomes the sole foreign supplier of the Northern partner's input. Second, this trade diversion (or investment diversion in Ethier's terminology) may make reforms by other countries more difficult since they lose market access. Third, the externality effect rises due to a net expansion of the input sector. Finally, the arrangement may induce other Northern countries to seek their own arrangements. Under some very strong assumptions, Ethier derives a final equilibrium in which each Northern country forms regional arrangements with all Southern countries able to liberalize successfully.

From a multilateralist perspective, one can take issue with some of Ethier's basic premises. For example, the view that the current wave of regionalism is a *friendly* response by developing countries to past multilateral liberalization by developed countries is at odds with historical evidence. It was frustration with rather than success of the multilateral process that led the United States to open negotiations with Mexico.

The view that regionalism is the central instrument of liberalization in developing countries is also at odds with reality. A considerable liberalization in developing countries, including Mexico, had already taken place before the current wave of regionalism was launched. Indeed, it was this unilateral liberalization, rather

than multilateral liberalization among developed countries, that created a favorable environment within Mexico to enter NAFTA. Furthermore, it is the liberalization in developing that has induced developed countries to consider entering into regional arrangements with them. Even today, the country that is on top of the U.S. list for entry into NAFTA is Chile, the most liberal of the Latin American countries. The developing countries that are relatively closed, mainly in South Asia and Africa, are on no developed country's list as potential partners in a PTA.

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 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 de la Torre and Kelly (1982), Srinivasan, Whalley and 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. Let me note just two factors.

First, most modelers rely on the so-called Armington assumption which says that goods are differentiated by the country of origin. They then proceed to combine 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 1.4 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, the widely used CES functional form rules out complementarities.[32] As we saw in the context of the Meade-Lipsey model in Section 1.4, if the liberalized good and exportable show complementarity, even a small tariff preference lowers welfare. It can be further shown that restricting ourselves to substitutability, 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. 482-3). 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 1.1), 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

---

[32]The wide use, rather than econometric evidence, has often been also cited as evidence that the assumption is "reasonable".

to measure the damage from trade diversion.[33] 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."

Taking the welfare of nonmembers as the sole criterion, McMillan goes on to argue, by appeal to the Kemp-Ohyama-Kemp-Wan 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.[34] 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 from them than before. Though the McMillan test is met, nonmembers are hurt by the union.

---

[33]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.

[34]Winters (1997) criticizes McMillan for his focus on the union's imports from nonmembers as the criterion for ensuring no harm to the latter and argues that he (McMillan) should have used union's exports, instead. But, as I will argue below, neither aggregate exports nor aggregate imports are appropriate to answer the question at hand.

Assuming unchanging trade balance and no distortions in nonmembers, the right test for them not to be harmed is that they do not experience a deterioration in the terms of trade at the hands of the union. Formally stated, we need to satisfy the condition

(1) $p^1 e^0 \geq p^0 e^0$

where p^0 and p^1 are world-price vectors in pre- and post-union equilibrium and e^0 is the vector of net exports by nonmembers to the union. An element in e^0 is positive if the good is exported and negative if it is imported by nonmembers. According to (1), the value of nonmembers' pre-union net exports should be no less at post-union prices than at pre-union prices. Because the Kemp-Ohyama-Kemp-Wan theorem freezes the initial trade vector of the union with nonmembers, we have $p^1 = p^0$ and condition (1) is automatically satisfied.

Imposing the trade-balance condition, $p^0 e^0 = 0$, (1) can be rewritten as

(1′) $P_E^1 E^0 \geq P_M^1 M^0$

where E^0 and M^0 stand for outside countries' vectors of exports and imports in the pre-union equilibrium and P_E^1 and P_M^1 are corresponding price vectors in the post-union equilibrium. Thus, $p^1 e^0 \equiv P_E^1 E^0 - P_M^1 M^0$. Inequality (1′) says that, at post-union prices, the pre-union export bundle of nonmembers should be able to buy at least their pre-union import bundle.

In principle, (1′) can serve as the simple test sought by McMillan to sort out the desirable FTAs and CUs from undesirable ones. But, in practice, it suffers from two key limitations. First, its ability to guarantee no harm to outside countries is based on the assumption of no distortions and balanced trade in the pre-union equilibrium in nonmembers. Second, and more importantly, it assumes that all changes in the 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, the union as a whole has market power and it holds external trade barriers at their original level, the terms of trade

of outside countries are highly likely to deteriorate with the removal of within-union barriers. Since substitutability is not an especially strong assumption 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.

7.   Concluding Remarks

As this review demonstrates, trade theorists have responded quickly to the challenges thrown by the current wave of regionalism as they did to the first wave. 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.

There, nevertheless, remain major gaps. One area in which theory remains almost non-existent is the relationship among regional, multilateral and unilateral liberalization in trade in services. Formal models 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. The modelling of trade preferences in services has received little attention from theorists.

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

Finally, there are frequent claims in the policy debate that regional arrangements offer commitment mechanisms that are superior to the available alternatives. Once again, neither theory nor empirical work has shed much light on this relationship. Much of the debate has rested on the judgements of the participants in the debate.

References

Anderson, K. and R. 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, K. and Staiger, R., 1997a, "Multilateral Tariff Cooperation During the Formation of Free Trade Areas," *International Economic Review* 38(2), May, 291-319.

Bagwell, K. and Staiger, R., 1997b, "Multilateral Tariff Cooperation During the Formation of Customs Unions," *Journal of International Economics* 42, 91-123.

Baldwin, Richard, 1995, "A Domino Theory of Regionalism." In Baldwin, Richard, Haaparnata, P. and Kiander, J., eds., *Expanding Membership of the European Union*, Cambridge, U.K.: Cambridge University Press.

Baldwin, Richard and Venables, Anthony, 1995, "Regional Economic Integration," in Grossman, Gene and Rogoff, Ken, *Handbook of International Economics Volume III*, North Holland, Amsterdam.

Berglas, Eitan, 1979, "Preferential Trading: The n Commodity Case," *Journal of Political Economy* 87, No. 21, 315-331.

Berglas, Eitan, 1983, "The Case for Unilateral Tariff Reductions: Foreign Tariffs Rediscovered," *American Economic Review* 73, December, 1142-43.

Bhagwati, Jagdish, 1971, "Trade-Diverting Customs Unions and Welfare Improvement: A Clarification," *Economic Journal* 81, 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 Melo and Panagariya, eds., 22-51.

Bhagwati, Jagdish, David Greenaway and Arvind Panagariya, 1998, "Trading Preferentially: Theory and

Policy," *Economic Journal* 108(449), 1128-1148.

Bhagwati, Jagdish, Krishna, Pravin and Panagariya, Arvind, *Trading Blocs: Alternative Approaches to Analyzing Preferential Trade Agreements*, Cambridge, MA: MIT Press, February.

Bhagwati, Jagdish and Panagariya, Arvind, 1996a, "Preferential Trading Areas and Multilateralism: Strangers, Friends or Foes?" in Jagdish Bhagwati and Arvind Panagariya, eds., *The Economics of Preferential Trade Agreements*, AEI Press, Washington, D.C.

Bhagwati, Jagdish and Panagariya, Arvind, 1996b, "The Theory of Preferential Trade Agreements:Historical Evolution and Current Trends." *American Economic Review* 86, May, 82-87.

Bond, Eric W. and Syropoulos, Constantinos, 1996, "The Size of Trading Blocs, Market Power and World Welfare Effects," *Journal of International Economics* 40, 411-437.

Brecher, Richard and Bhagwati, Jagdish. "Foreign Ownership and the Theory of Trade and Welfare." Journal of Political Economy, June 1981, BI(3), pp. 497-51 1.

Cadot, O, Melo, Jaime de and Olarreaga, J., 1996, "Regional Integration and Lobbying for Tariffs Against Non-Members," mimeo, University of Geneva.

Davis, Donald, 1997, "Common External Tariffs of a Customs Unions: Alternative Approaches. Comment," *Japan and the World Economy* 9(4).

Deardorff, Alan and Stern, Robert, 1994, "Multilateral Trade Negotiations and Preferential Trading Arrangements," in Deardorff, A. and Stern, R., eds., *Analytical and Negotiating Issues in the Global Trading System*, Ann Arbor, MI: University of Michigan Press.

de la Torre, A. and Kelly, M., 1992, "The Regional Trading Arrangements," Occasional Paper No. 93, Washington: International Monetary Fund.

Ethier, W., 1998, "Regionalism in a Multilateral World, *Journal of Political Economy* 106(6), 1214-45.

Frankel, J., 1997, *Regional Trading Blocs in the World Trading System*, Washington, D.C.: Institute for International Economics.

Frankel, J., ed., 1998, *The Regionalization of the World Economy*, Chicago: University of Chicago Press.

Frankel, J., Stein, E. and Wei, S., 1995, "Trading Blocs and the Americas: The Natural, the Unnatural and the Supernatural," *Journal of Development Economics* 47, 61-96.

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.

Gehrels, Franz, 1956-57, "Customs Union from a Single-Country Viewpoint," *Review of Economic Studies* 24(63), 61-64.

Gros, Daniel, "A Note on the Optimal Tariff, Retaliation and the welfare Loss from Tariff Wars in a Framework with Intra-Industry trade," *Journal of International Economics* 23 (1987): 357-367.

Grossman, Gene and Helpman, Elhanan, 1994, "Protection for Sale," *American Economic Review* 84(4), September, 835-50.

Grossman, Gene and Helpman, Elhanan, 1995, "The Politics of Free Trade Agreements," *American Economic Review* 85(4), September, 667-690.

Helpman, E. and Krugman, P.R., *Market Structure and Foreign Trade. Increasing Returns, Imperfect Competition, and the International Economy*, (MIT Press, 1985).

Kemp, Murray, C., 1964, *The Pure Theory of International Trade*, Prentice-Hall, Englewood Cliffs, N.J., 176-177.

Kemp, M. and Wan, Henry Jr., 1976, "An Elementary Proposition Concerning the Formation of Customs Unions", *Journal of International Economics* 6 (February), 95-98.

Kowalczyk and Sjostrom, 1994, "Bringing GATT into the Core," *Economica* 61, 301-317.

Krishna, K. and A. O. Krueger, 1995, "Implementing Free Trade Areas: Rules of Origin and Hidden Protection," in Deardorff, A, Levinsohn, J. and Stern, R., eds., *New Directions in Trade Theory*, Ann Arbor: University of Michigan Press, 149-187.

Krishna, Pravin, 1998, "Regionalism and Multilateralism: A Political Economy Approach," *Quarterly Journal*

of Economics 113(1), February, 227-251.

Krishna, Pravin and Jagdish Bhagwati, 1997, "Necessarily Welfare-enhancing Customs Unions with Industrialization Constraints," *Japan and the World Economy* 9(4).

Krueger, Anne O., 1993, "Free Trade Agreements as Protectionist devices: Rules of Origin," NBER Working Paper No. 4352.

Krugman, P., 1980, "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review* 70, no. 5, December, 950-959.

Krugman, P., 1991a, "Is Bilateralism Bad?" in E. Helpman and A. Razin (eds.), *International Trade and Trade Policy*, Cambridge, Mass.: MIT Press.

Krugman, P., 1991b, "The Move to Free Trade Zones," Symposium Sponsored by the Federal Reserve Bank of Kansas City, *Policy Implications of Trade and Currency Zones.*

Krugman, P., 1993, "Regionalism versus Multilateralism: Analytical Notes," in Melo and Panagariya (eds.)

Levy, Philip. 1997, "A Political-Economic Analysis of Free-Trade Agreements," *American Economic Review* 87(4), September, 506-519.

Lipsey, Richard, 1957, "The Theory of Customs Unions: Trade Diversion and Welfare," *Economica* 24, February, 40-46.

Lloyd, Peter J., 1982, "3x3 Theory of Customs Unions," *Journal of International Economics* 12, 41-63.

Ludema, R., 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 Kym Anderson and Richard Blackhurst, eds., *Regional Integration and the Global Trading System*, New York: St. Martin's Press.

McMillan, John and McCann, Ewen, 1981, "Welfare Effects in Customs Union," *Economic Journal* 91, 697-703, September.

Meade, James, E., 1955, *The Theory of Customs Unions*, Amsterdam: North-Holland.

Melo, Jaime de and Panagariya, Arvind (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*, 1-13.

Ohyama, M., 1972, "Trade and Welfare in General Equilibrium," *Keio Economic Studies* 9, 37-73.

Panagariya, Arvind, 1996, 'The Free Trade Area of the Americas: Good for Latin America?' *World Economy* 19, no. 5, September, 485-515.

Panagariya, Arvind, 1997a, "Preferential Trading and the Myth of Natural Trading Partners," *Japan and the World Economy* 9(4): 471-489.

Panagariya, Arvind, 1997b, "The Meade Model of Preferential Trading: History, Analytics and Policy Implications," in Cohen, B.J., ed., *International Trade and Finance: New Frontiers for Research. Essays in Honor of Peter B. Kenen* (Cambridge University Press, New York) 57-88.

Panagariya, Arvind, 1998a, "The Regionalism Debate: An Overview," Center for International Economics Working Paper No. 40, Department of Economics, University of Maryland; (forthcoming in the *World Economy*)

Panagariya, Arvind, 1998b, "Do Transport Costs Justify *Regional* Preferential Trade Arrangements? No," *Weltwirtschaftliches Archiv* 134(2), 230-301.

Panagariya, Arvind, 1999, "Some Simple Welfare Results on Customs Unions and Free Trade Areas," in progress.

Panagariya, Arvind and Ronald Findlay, 1996, "A Political Economy Analysis of Free Trade Areas and Customs Unions," Robert Feenstra, Douglas Irwin and Gene Grossman, eds., *The Political Economy*

of Trade Reform, Essays in Honor of jagdish Bhagwati, MIT Press.

Panagariya, Arvind and Pravin Krishna, 1997, "On the Existence of Necessarily Welfare-Enhancing Free Trade Areas," Center for International Economics, Department of Economics, University of Maryland, mimeo.

Richardson, Martin, 1993, "Endogenous Protection and Trade Diversion, " *Journal of International Economics* 34(3-4), 309-24.

Richardson, M. 1994. "Why a Free Trade Area? The Tariff Also Rises," *Economics & Politics* 6, no. 1, March, 79-95.

Riezman, Raymond, 1979, "A 3x3 Model of Customs Unions, *Journal of International Economics* 9(4), 341-354.

Srinivasan, T.N., 1993, "Regionalism versus Multilateralism: Analytical Notes. Comment," in Melo and Panagariya, eds.

Srinivasan, T. N., 1997, "Common External Tariffs of a Customs Unions: Alternative Approaches," *Japan and the World Economy* 9, 447-70.

Summers, L., 1991, "Regionalism and the World Trading System," Symposium Sponsored by the Federal Reserve Bank of Kansas City, *Policy Implications of Trade and Currency Zones.*

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, World Bank, Washington, D.C.

Winters, Alan, 1997, "Regionalism and the Rest of the World: The Irrelevance of the Kemp-Wan Theorem," *Oxford Economic Papers* 49, 228-234.

Wonnacott, Paul and Lutz, Mark, 1989, "Is There a Case for Free Trade Areas?" in Schott, Jeffrey, *Free Trade Areas and U.S. Trade Policy*, Washington, D.C.: Institute for International Economics, 59-84.

Wonnacott, Paul and Wonnacott, Ronald J., 1981, "Is Unilateral Tariff Reduction Preferable to a Customs

    Union? The Curious Case of the Missing Foreign Tariff," *American Economic Review* 71, September,

    704-14.

Yi, Sang-Seung, 1996, "Endogenous Formation of Customs Unions under Imperfect Competition: Open

    Regionalism is Good," *Journal of International Economics* 41, 153-77.