10

A Political Economy Survey of Regionalism's Effect on Global Liberalization

Are trade blocs stumbling blocks or building blocks? This chapter reviews many political economy arguments. First come those that suggest that regionalism undermines support for more generalized liberalization. Then we consider arguments that say that regional initiatives help build political momentum for global liberalization. Arguments as to how the adoption of a regional trading area might politically undermine unilateral or multilateral liberalization fall under the following headings: incentives to protect industries, manipulation of the process by special interests, use of scarce negotiator resources, and creation of a political dead end. Other arguments more optimistically offer ways in which regional trading areas might undermine protectionism and reinforce liberalization more generally: by locking in unilateral liberalization, by creating larger units that can negotiate more efficiently, through mobilization of regional solidarity, by building export constituencies to create domestic political momentum, and through competitive liberalization. At the end, we return to the gravity model for a verdict on whether the antiliberalization or proliberalization forces appear to have been dominant among the trading blocs of 1970-92. The evidence seems to tilt in the latter direction—building blocks rather than stumbling blocks.

Although the multilateral system has made large strides toward freer trade, most recently with the conclusion of the Uruguay Round negotiations in December 1993, political constraints inevitably prevent the immediate attainment of the economist's nirvana of worldwide free trade. Since influential producer interest groups in each country typically stand to lose from free trade, full unilateral liberalization rarely occurs, and

the world must instead await the outcome of step-by-step multilateral negotiations. In these negotiations, countries trade concessions in such a way that, at each step, the fraction of the population that stands to gain is sufficiently high to overcome the political opposition.

In this light, the case in favor of regional trading arrangements is usually a second-best argument that takes as given the political impossibility of negotiating further liberalization along most-favored nation (MFN) lines. The uninitiated might assume that free trade economists would under these circumstances necessarily support free trade areas (FTAs). The phrase contains the magic words ''free trade,'' so the general assumption is that economists must be in favor of them. After all, FTAs do entail the removal of barriers to trade. Indeed, many mainstream American economists signed a public letter of support for the North American Free Trade Agreement (NAFTA) during the noisy 1993 debate over its ratification, and virtually none publicly opposed it. But from the standpoint of static economic welfare, trade economists are in fact ambivalent about the desirability of FTAs, as observed in chapter 1.[1] According to the Theory of the Second Best, if the persistence of one distortion is taken as given, thus precluding the first-best solution, then eliminating another distortion does not necessarily yield the second-best solution; it may be only third best, or worse (Lipsey and Lancaster 1956-57). So long as tariffs and other barriers against outsider countries remain in place, the elimination of barriers between two FTA members can as easily intensify distortions as eliminate them.

The familiar classical distinction is between the harmful trade-diverting effects of FTAs and their beneficial trade-creating effects. Although modern theories of trade have gone beyond the diversion/creation distinction, it is still a useful intuitive guide to likely welfare effects. Grossman and Helpman (1995), for example, find in a lobbying model that a free trade area is most likely to be adopted when trade diversion outweighs trade creation, which unfortunately is also when it is most likely to reduce aggregate welfare. Some simulations and econometric estimates also give the answer that regional trading arrangements are harmful with respect to the static effects on worldwide economic welfare. On the grounds of trade-diversion effects and other considerations discussed below, Bhagwati and Panagariya generally oppose regional trading arrangements.[2]

---

1. Those who rejected these economists' cautions most often did so under misunderstandings regarding both the case for free trade and the case for free trade areas: that the argument relies on either a naive assumption that other countries will reduce their barriers as much as the United States or on an altruistic willingness to incur domestic costs for the sake of the trading partners' gain.

2. Bhagwati (1995, 11) and Panagariya (1995, 20, note 8) have confessed that they were prepared to oppose NAFTA publicly in 1993, if asked. They are now skeptical of other ongoing initiatives.

Stumbling Blocks or Building Blocks?

This still leaves what Bhagwati (1993a) calls the dynamic time-path question. Rather than taking trade policies outside trade blocs as given, let us recognize that they are determined endogenously by political economy forces. Can regionalism help build political momentum for more general liberalization? Or is it more likely to undermine multilateral efforts? Can trade blocs be building blocks, or are they more likely to be stumbling blocks on the path to free trade? (Bhagwati invented this catchy phrasing [1991, 77], and Lawrence [1991c] popularized it, followed by the Council of Economic Advisers [1995], and others.) Many authors have made specific arguments on both sides of this question. They range from formal models to historical anecdotes. This chapter attempts to review the literature, with an emphasis on the many contributions of the last few years.

Regardless of the side of the debate on which they fall, the arguments can be categorized according to the unit of decision making and the level of analysis. In one class of models, countries are treated as rational unified actors maximizing their national economic welfare. Under the right conditions, it can even be assumed that if a group of countries forms a customs union, this group then becomes a new rational unified actor in setting its trade policy vis-a-vis the rest of the world. A second set of arguments introduces costs to the process of negotiation per se. A third focuses institutionally on the potential for manipulation of technical aspects of policymaking by narrow special interests. A fourth formally models the process, either as majority voting or as other forms of political influence exerted by firms, via lobbying or campaign contributions. It notes that the ultimate outcome can depend on the sequence in which decisions are made—regional agreements first and multilateral (or unilateral) liberalization second, or the other way around.[3]

After enumerating the arguments on both sides, we evaluate which set of forces has been dominant over the last 30 years. The tentative conclusion is that regionalism has in recent experience been politically consistent with more general liberalization, in the sense that members of trade blocs have tended to increase their trade with nonmembers as they intensify their trade (even more) with each other. This is true more often than not in estimates for the European Community, Mercosur, the Association of Southeast Asian Nations (ASEAN), and East Asia, though not for some

3. Rodrik (1995) and Helpman (1995) survey the political economy models of trade policy in general and their shortcomings. Baldwin (1996) has argued that a model of lobbying by pressure groups, along the lines of Grossman and Helpman (1994, 1995), is more appropriate to analyze the recent regionalist movements than a median-voter model. He notes that majorities of the public have opposed many of the recent moves, and that in many cases they have taken place anyway because government elites and business groups determine the outcomes, not voters.

other groups. While one cannot infer causation from correlation, the evidence is consistent with the hypothesis that, over recent history, the political forces that have given rise to liberalization on a regional basis have also given rise to liberalization on a more general basis.

Stumbling Blocs: Negative Implications of Regionalism for Multilateral Liberalization

Blocs' Market Power and Incentive to Protect

The standard experiment presumes that the level of trade barriers against outsiders remains unchanged when a customs union is established. Kennan and Riezman (1990), Gatsios and Karp (1991), and Krugman (1991a) relax this assumption, assuming instead that any group will set its external tariff so as to maximize its economic welfare. The larger the group, the more the collective monopoly power it has, and the higher the optimal tariff it can profitably command. This is the ''incentive to protect.''

In a world consisting of a few large blocs, each unit will have more monopoly power and thus will be more tempted to seek to shift the terms of trade in its favor by raising tariffs against the other blocs. Collectively, the blocs fail to improve their terms of trade and merely reduce everyone's welfare, as in the classic tariff war of Johnson (1954). But this does not stop them from trying. The blocs are trapped in a harmful ''prisoners' dilemma,'' also known as the Nash noncooperative equilibrium.

This temptation to raise tariffs will be minimized in a world of many small trading blocs because each has little monopoly power to exploit. (A system of MFN is such a world—that is, each country is its own bloc). In a world of a few large blocs, the noncooperative equilibrium features a higher level of interbloc tariffs and a lower level of economic welfare. In Krugman's (1991a) simulation, three turns out to be the worst number of blocs in which the world could be divided.[4]

Froot and Yoffie (1993) point out some implications of foreign direct investment for blocs' incentive to protect. If firms own facilities in each other's countries, there will be less temptation for blocs to raise tariffs against each other.

The Krugman model assumes that members of a trade bloc set their external tariffs together—that is, that the arrangement is a customs union. The ''incentive to protect'' story would be different for a standard free trade area, in which each country can set its tariffs independently with respect to nonmembers.

---

4. As we saw in chapter 8, this result generalizes with intercontinental transport costs and partial preferences, so long as the latter are large relative to the former.

A number of authors have explicitly compared FTAs' and customs unions' compatibility with liberalization toward nonmembers. Sinclair and Vines (1994) argue that in the FTA case, there is actually an incentive for each country to *reduce* its external tariffs, just the opposite of the customs union monopoly-power case. Wonnacott and Wonnacott (1995) give some examples of this.[5] Richardson (1993) derives the same result—that is, the superiority of an FTA to a customs union—in a model in which tariffs are set endogenously. The government objectives are assumed to include the profits of protected industries, in addition to consumer welfare. The FTA member with a comparative disadvantage in a particular good will experience a decline in the political influence of that industry as competition within the FTA diminishes the industry's economic size. As a result, the country will tend to reduce protection for that industry in a way that would not be possible if the country were bound by the common external tariff of a customs union.

Panagariya and Findlay (1994) assume that protection is the endogenous outcome of lobbying and derive the opposite result: the lobby chooses a lower external tariff under a customs union than under an FTA. The customs union is more effective at diluting the power of interest groups.[6]

Bagwell and Staiger (1993a) introduce a temporal dimension. They examine the condition under which plurilateral tariff liberalization is self-enforcing through fear of future retaliation by other blocs. They find that this condition holds only during the early stages of customs union formation, when the threat of future retaliation counts heavily, but not when the customs union is complete and able to exercise its full monopoly power. The sequence is reversed in Bagwell and Staiger (1993b), when free trade areas are considered: members do not set external tariffs jointly and thus cannot exercise collective monopoly power as in a customs union.

In reality, governments in one sense are less capable of national economic optimization than the Krugman or Bagwell-Staiger models presuppose, and in another sense they are more capable. In both respects, large trading blocs are less vulnerable to the incentive to raise tariffs against each other than under Krugman's assumptions. First, governments are less capable of optimization, in that maximum exploitation of the terms of trade (through imposition of the ''optimum tariff'') is in practice one of the *less* prevalent determinants of trade policy. More commonly, one sees political arguments made regarding the need to protect infant industries, protection of scarce factors of production, the need to increase

5. They also recommend a hybrid FTA with a common external tariff in some sectors (where it is most desirable to minimize rules of origin), in order to get the best of both kinds of arrangements.

6. Another paper comparing FTAs with customs unions, pointing out the ability of the latter to exploit monopoly power, is Ludema (1993).

employment, and adjustment costs. Second, governments are *more* capable of optimization in that they have already addressed the prisoners' dilemma by instituting a cooperative international regime, namely the General Agreement on Tariffs and Trade (GATT), as Bergsten (1991) pointed out in his comment on Krugman (1991b).

GATT Article XXIV explicitly rules out Krugman's concern. This provision allows deviations from the MFN principle only for FTAs or customs unions that do not raise the average level of their tariffs against nonmembers. Under the World Trade Organization (WTO), which began operations in 1995, the rules governing regional trading arrangements have been strengthened. For example, Article V of the General Agreement on Trade in Services (GATS) in large measure extends Article XXIV to include services along with goods, making the requirements that much harder for a proposed regional arrangement to meet.

There are several reasons to worry that blocs' ''incentive to protect'' survives despite the existence of Article XXIV. First and most obviously, Article XXIV is often virtually disregarded, as Bhagwati (1992) reminds us. Second, as Bagwell and Staiger (1993a, note 25) point out, when explicit tariff increases are ruled out, exacerbation of the incentive to protect can take the form of ''gray area'' measures in customs unions. Third, the multilateral process is on a path whereby worldwide tariff rates are being gradually reduced through negotiation, so this path is the relevant benchmark, not unchanging tariffs. Regionalization might help this process, or it might hinder it. On the one hand, Bond, Syropoulos, and Winters (1995) show that the elimination of remaining internal tariffs among members of an existing customs union may actually alter bargaining relationships in such a way as to enlarge the set of sustainable tariff-reducing agreements vis-a-vis nonmembers. On the other hand, Bond and Syropoulos (1995) show that arriving at the cooperative equilibrium of an agreement for interbloc liberalization in a repeated game, which it is seen as the WTO's role to facilitate, becomes more difficult as the size of the blocs, and therefore their monopoly power, rises.

Manipulation by Special Interests

The special-interests argument points out that creating a regional trading arrangement abounds in opportunities for trade-sensitive industries to manipulate the process, particularly in those sectors that might be hurt. As Wonnacott and Lutz (1989, 65-66) emphasize, negotiators frequently seek to exclude from regional FTAs precisely those sectors that would be most threatened by welfare-enhancing trade creation. The members of ASEAN, for example, have in the past exempted almost all the important sectors from the system of preferences that they are supposed to grant

each other (Panagariya 1994, 828-29).[7] In a famous example, Indonesia excluded major sectors but offered to liberalize imports of snow ploughs.

Grossman and Helpman (1995) have used their lobbying model to understand how the possibility of such industry exclusions increases the chances of FTAs being adopted. This was the primary reason for another restriction that the GATT Article XXIV places on FTAs: that ''substantially all'' barriers within the region be removed. In practice, FTAs and customs unions have tended to comply less than completely with this provision. Examples include the European Economic Community's treatment of agriculture and, in practice, steel and many other goods.

Even when Article XXIV is obeyed, favored industries can win long, drawn-out periods during which protection against the partner is gradually phased out, as was the case with agriculture under NAFTA (Orden 1996). Grossman and Helpman (1995) show that one would expect to get slow phase-outs of protection of a domestic industry when (1) the domestic industry is strong, (2) the foreign industry is weak, or (3) the foreign country wants the FTA more than does the domestic country. Kowalczyk and Davis (1997) study NAFTA and find that American industries that had been able to win higher duties in the past tended to get longer periods of adjustment in the NAFTA negotiations. The same pattern does not seem to hold for Mexican industries, however. The speed of US phase-out for a sector seems to affect the speed of Mexican phase-out, suggesting reciprocity in the negotiations within narrow categories. The authors conclude that slow phase-outs were a concession that Mexico granted the United States because it wanted the agreement more.

Another example of special-interest involvement is the exploitation of rules of origin (Krueger 1993, 1995). An FTA, unlike a customs union, does not involve the setting of common external tariffs. Richardson (1992), Krueger (1993), and Krishna and Krueger (1993) show how individual industries in the FTA negotiation can augment the protection they receive when their governments institute rules of origin that enable them to capture their FTA partner's market in addition to their own, thus diverting trade from foreign suppliers. Richardson (1992) emphasizes that prices at which producers can sell are equalized within an FTA, even when rules of origin are successful in keeping the consumer price higher in the higher-tariff country. Competition for tariff revenue among the FTA members may then result in an equilibrium in which external tariffs are reduced to zero. Krueger (1995) reaches a different conclusion—that customs unions are always Pareto-superior to FTAs because they have no rules of origin that can be exploited in this way. Nagaoka (1994) develops a model in which the government is committed to preserve a given ''strategic'' monopolistic industry—for example, by manipulating rules of origin.

7. It should be harder to exclude sectors under the new WTO rules that came into effect in 1995.

He finds some effects whereby regional integration can reduce the incentive for protection for that industry and thereby accelerate liberalization vis-a-vis the rest of the world but also finds that the formation of a customs union can exacerbate the incentive to protect.

Bhagwati (1993a, 30-31; 1995, 22) and Panagariya (1995, 16-21) argue that large countries such as the United States may use their overwhelming bargaining power within regional groups to obtain distorting concessions from small countries that they might not obtain in more balanced multilateral negotiations. Whalley and Hamilton (1996) and Perroni and Whalley (1994) point out that small countries have been the supplicants in recent regional agreements, seeking assured access, or ''protection against administered protection.'' While large countries have all the bargaining power on their side, one model (Casella 1995) shows that smaller members stand to gain more from expansion of an existing bloc than larger members, based on economies of scale. The smaller members then have to give up proportionate voting power in the bloc to induce the larger ones to go along.

Lawrence (1996) notes that special interests may capture the decision-making process of the larger country and thereby thwart liberalization of an entire customs union. But he also notes that larger customs unions are more likely to contain countervailing interest groups. Because the power of each interest group is diluted, liberalization can proceed. Although French farmers succeeded in blocking the Uruguay Round for a while, eventually other EU members forced France to go along.

Scarce Negotiator Resources

The argument that negotiating resources are limited underscores the fact that creating trade agreements, ratifying them, and compensating losers are not costless endeavors. If they were, then the world would have already agreed to free trade. If a US trade representative spends all his or her time and political capital with Congress on a regional agreement, then there might be less time or capital left over for multilateral negotiations. As with the incentive-to-protect argument, regional trading arrangements may set back the negotiation of worldwide trade liberalization under the WTO.

Firms' Support for FTAs May Be a Political Dead End

Regional initiatives might prevent multilateral initiatives when the sequence of decisions matters. The forces in favor of liberalization might win out over protectionists if the only choice is between the status quo and multilateral liberalization, but when offered the option of a regional free trade area, the political process might then be diverted to the regional

route to the exclusion of the multilateral route. Bhagwati (1993a) worries that business executives and bureaucrats, having achieved regional integration, might then find the effort involved in multilateral negotiation too difficult. ''Lobbying support and political energies can readily be diverted to preferential trading arrangements such as FTAs. . . . That deprives the multilateral system of the support it needs to survive, let alone be conducive to further trade liberalization'' (Bhagwati 1993b, 162).

Krueger (1995, 22-24) shares these concerns and argues that the diversion of political energies is likely to be worse in the case of FTAs than in the case of customs unions. She reasons from two propositions: First, once trade diversion has taken place as the result of any preferential arrangement, the newly established firms producing for the partner country's market will oppose moves away from the new status quo toward global free trade. Second, trade diversion is more likely to occur under an FTA than a customs union, due to rules of origin. It then follows that it will be harder to muster the political support to move from an FTA to multilateral free trade than from a customs union.

A few authors have sought to model issues of sequence formally. Krishna (1995) assumes that a country will accept proposed changes in trade policy if its firms see a net increase in their profits (across all markets) from the change. He then derives two conclusions: first, member countries are more likely to support more trade-diverting preferential arrangements because the firms' gains come at the expense of nonmembers, and second, these trade-diverting preferential arrangements will reduce the incentives to seek multilateral liberalization. The result is that otherwise-attainable multilateral agreements might be precluded once countries start down the FTA path. The argument is similar to Krueger's except that it does not rely on rules of origin.

Levy (1993) offers what might be called a median-voter dead-end model, in which a bilateral free trade agreement can undermine support for multilateral liberalization because it is a dead end in the political process. The model assumes that trade policy is determined by the median voter: any proposition that benefits at least half the voters will pass. Trade itself is determined in some sectors by differences in factor endowments and in others by considerations of imperfect substitutes. As other authors have argued, the intra-industry sort of trade that is generated in imperfect substitutes is easier to accept politically than the factor-endowment kind of trade. The reason is that adjustment to import competition requires workers only to move from the assembly line for one product variety to the assembly line for another variety of the same product. Trade based on differences in factor endowments is much more difficult to accept politically because it requires workers in previously protected industries to move to different industries (and at lower wages in the case of capital-intensive industrialized countries).

Levy argues that trade policy may thus be a trade-off between the gains afforded by increased varieties and the losses inflicted by a fall in the relative price of the product that is intensive in the scarce factor (labor, in the case of industrialized countries). If liberalization is not attainable, it is because the losses from factor-endowment-based trade dominate. If a vote is held first on whether to join a bilateral free trade area, it is more likely to pass when the potential partner has similar factor endowments. (That is, it is easier politically to achieve a European Union than a Free Trade Area of the Americas, or an FTA, of APEC members.) This is because the gains from increased trade in imperfect substitutes will be large, while the losses from a fall in the relative price of labor-intensive products will be small. But if a vote is subsequently held on multilateral liberalization, it may fail: those key sectors that stand to profit from trade in imperfect substitutes will already have reaped those gains, and there will be fewer political forces to countervail the sectors that lose from the additional factor-endowment-based trade. The FTA can thus turn out to be a dead end, even though the multilateral liberalization might have been accepted if it had been voted on first. In this way, regional free trade agreements in this model undermine the political support for more general liberalization.[8]

In sum, there is no shortage of models and arguments in which regional trading arrangements can undermine multilateral liberalization—that is, in which trade blocs can operate as stumbling blocks rather than building blocks. Fortunately, the effects that they are designed to illustrate are not the only ones that operate.

Building Blocs: Positive Implications of Regionalism for Multilateral Liberalization

Locking In and Mobilizing Regional Solidarity

In the late 1980s, Mexican Presidents Miguel de la Madrid and Carlos Salinas reversed a half-century of Mexican protectionism and imposed sweeping, unilateral liberalization measures. Future presidents of Mexico might seek to reverse this liberalization. Thus, a good argument for NAFTA was that it locked in the reforms in a manner that would be difficult to reverse (e.g., Lawrence 1991c; Tornell and Esquivel 1995).

8. Levy (1995) pursues a further application of the Grossman-Helpman model of trade policymaking based on political contributions. He looks at the effect on trade relations between two large countries when one of them forms an FTA with another group of countries. The paper offers an example that assesses whether the regional agreement will increase or decrease political support from multilateral liberalization by altering the marginal profitability to the lobbies of tariff changes.

Panagariya (1995, 22-26) and others respond that tariff bindings under the WTO are better devices for locking in reforms than are regional agreements. Furthermore, Mexico in fact joined GATT before it joined NAFTA so that one cannot claim that the regional agreement was a steppingstone to acceptance of multilateral obligations. Each of these points brings counterpoints. Regarding tariff bindings, most less developed countries have bound their tariffs at rates well above the rates actually applied so that in practice the bindings are—so to speak—not binding. Regarding the sequence in Mexico's case, a US-Mexico bilateral pact on countervailing duties in 1985 was a steppingstone both to Mexico's GATT accession and to NAFTA. Moreover, NAFTA contains some attributes, such as further protection against nontariff barriers, that the WTO does not assure.

A comparison of the 1982 and 1994 Mexican debt crises illustrates the lock-in point. In the first debt crisis, Mexico raised trade barriers against all partners sharply. In the second crisis, with NAFTA in place, Mexico raised tariffs on some products against partners outside of the Western Hemisphere but continued to cut tariffs on US trade, as called for under the agreement. While the trade-diverting effects of such a differential policy response are regrettable, this response illustrates that the regional agreement bound tariffs more strongly than did WTO membership (and did so to the benefit of the United States).

Elsewhere, such as in the Andean Pact countries, leaders have used political support for regional solidarity to achieve liberalization that would be politically impossible if pursued unilaterally. De Melo, Panagariya, and Rodrik (1993, section 3) model the process whereby governments can adopt rules or institutions in a regional context to insulate themselves from pressure by private-sector lobbies for intervention on their behalf.

Efficiency of Negotiating with Larger Units

The idea in the 1980s was that progress toward free trade could be made more rapidly by regional negotiations than by GATT negotiations, which seemed to drag on interminably. (''GATT is dead,'' went the conventional wisdom.) On the other hand, it was pointed out at the time that even the most successful of the customs unions, the European Community, had taken 35 years to form, from the Treaty of Rome to 1992. Now, with the Uruguay Round successfully completed and a new institution, the WTO, in place, it has become harder to argue that multilateral progress is slow or impossible. In the meantime, however, it has also become harder to argue that progress on the regional front is slow. Such recent FTAs as NAFTA and Mercosur have gone from initial proposal to implementation in a few short years. The question is no longer which route to free trade is the slower, but rather, which route is the faster?

Within the context of multilateral negotiations, it can be slow and awkward to negotiate separately with over 100 small countries. (As of early 1996, there were 112 members of the WTO, with another 27 waiting in line.) Some authors have argued that the costs of negotiation rise with the number of countries involved so that it is easier for a smaller group of countries to negotiate a customs unions first. With a common external trade policy, they can then enter multilateral negotiations as a group (see, e.g., Deardorff and Stern 1992; Kahler 1995b; Krugman 1993; Summers 1991). This is thought to increase the efficiency of the negotiations and to make a satisfactory worldwide agreement more likely. The European Union is certainly the most important example of this. Other groups, such as ASEAN, the Central American Common Market (CACM), and the Caribbean Common Market (Caricom), have also been urged to integrate regionally so they can speak to the larger powers with a unified voice.

Others question the practicality of this approach (Bhagwati 1993a; Winters 1997a; Panagariya 1994, 830-31). Figure 1.1 illustrated how dozens of regional trading arrangements among Western Hemisphere countries have created a spaghetti-like web. If one contemplates some kind of negotiation for every possible group of countries, a set of 10 countries produces 2^{10}, or 1,024 combinations, the Western Hemisphere produces 2^{35}, or 3.44 trillion combinations, and so on.[9] This proliferation could complicate negotiations rather than simplify them.

Kahler (1995a, 125-27) suggests that negotiations among a small number of regional neighbors may allow more efficient treatment of new issues than do global negotiations. Lawrence (1996) singles out intellectual property rights, common standards, and currency links as examples of issues that have made their debut in regional forums. Similarly, Bergsten (1994, 23-24; 1996, 110) and Hufbauer and Schott (1995) have suggested that regional groups such as the Canada-US FTA, NAFTA, and the Australia-New Zealand Closer Economic Relations (CER) pact have served as test laboratories for discussions regarding trade in services, investment, and competition policy and have thereby created models for later multilateral negotiations. The general principle that some public activities are better handled at the local level, some at the national, and some at the regional is known as fiscal federalism in the public finance literature and as subsidiarity in EU-speak. Two other Europhrases, ''variable geometry'' and ''flexible integration,'' refer to frameworks in which member countries can opt in for some dimensions of integration and opt out for others (e.g., labor standards, free migration, exchange rate arrangements). There is

9. We can subtract 11 and 36, respectively, from these numbers, eliminating the groups that constitute only one member each, as well as the single group that includes no members at all. Negotiation costs for these combinations would be zero. But it doesn't help to reduce the total count much.

also the principle of the optimum currency area, which says that a group of countries that are highly integrated with respect to trade and labor mobility should share a common currency, while other countries should keep their monetary independence.

Building Export Constituencies to Create Domestic Political Momentum

As Destler (1995) and others have pointed out, for trade negotiations to succeed (whether multilaterally or bilaterally), the proliberalization interests of the *exporters* in each country usually must outweigh the interests of pro-protection, import-competing firms. Consumers are not much help in this regard. The question for our purposes is whether export interests are likely to support multilateral liberalization more strongly after regional liberalization or less. The ''bicycle theory'' implies that any source of liberalization momentum is good (e.g., Bergsten 1996, 108; Destler 1995, 17-18; Staiger 1996, section 5.4), while the ''dead end'' arguments cited above suggested the reverse.

Milner (1997) argues that certain sectors, such as those in which economies of scale are important, are more supportive of regional agreements than others and finds that FTAs lead to greater liberalization within those sectors than within others. She concludes that, unlike in the 1930s, regionalism in the 1990s is ''a complement to, and/or precursor of, greater multilateral trade liberalization.''

Wei and Frankel (1995a) have made a start at modeling an argument regarding political constituencies (see also Frankel and Wei 1997). We consider the problem of building export constituencies in a model in which a country chooses its trade policies by majority vote, and voters suffer costs in changing jobs from one sector to another. Our hypothesis is that, under certain conditions, leaders might not be able to obtain a majority vote in favor of multilateral liberalization, much less unilateral liberalization, and yet might be able to obtain a majority vote in favor of regional liberalization, which, when completed, shifts the economic incentives so as to produce a majority in favor of wider liberalization. The model can be viewed as a formalization of the argument that, as FTA preferences take effect, less competitive industries will contract and thereby lose the political weight necessary to resist further liberalization (e.g., Hufbauer and Schott 1994, 164; Richardson 1993).[10]

This model is inspired by Fernandez and Rodrik's (1991) analysis of status quo bias. They consider a (nonregional) situation in which the

10. McCulloch and Petri (1994) also point out that a regional trading arrangement can eliminate the high-cost industries that would otherwise operate as a political constituency opposed to general liberalization.

majority in a country votes against unilateral liberalization even though a majority would *ex post* gain from it economically. Divide the population into three groups: those who know they would gain from liberalization because they are confident of their ability to compete on world markets, those who will eventually gain from liberalization because they will turn out to be competitive on world markets but do not know this *ex ante*, and those who will lose from liberalization because of new import competition but do not know this *ex ante*. If the two-thirds of the population who are uncertain have as little as a 49 percent chance of gaining, all of this group—a majority of the entire population—will oppose liberalization *ex ante*. This is true even though a majority of the population gains *ex post* (two-thirds times .49, plus the one-third who are sure gainers). The interesting aspect of the model is that if the leaders are somehow able to push liberalization through anyway and a new vote is taken after the uncertainty is resolved, a majority will then vote in favor of maintaining the new liberalized status quo. In essence, the act of liberalization itself builds a constituency for liberalization, as those who are good at exporting discover their previously unknown talents. Similar conclusions could be reached in a model in which capital and labor move from previously protected sectors to new trade-oriented sectors, although the status quo bias in this case would hold for reforms that did not benefit a majority in addition to those that did.

In the Wei-Frankel (1995a) version, political leaders may be able to obtain a majority vote in favor of regional liberalization because fewer sectors are adversely affected. More firms then discover their export potential, making it possible to obtain a majority support for previously unattainable MFN liberalization. This story is thus a counterexample to the overly strong claim of Levy (1993) that ''bilateral free trade agreements can never increase political support for multilateral free trade.''

Competitive Liberalization

''Competitive liberalization'' is a means of building political momentum for liberalization (Bergsten 1995a). A particular claim is that each APEC member will seek to dismantle trade barriers faster than the next as a way of competing for international capital. An alternative claim is that trade liberalization is more likely in an economy that is part of a region in which other economies are liberalizing because each observes its neighbors' prosperity. Drysdale and Garnaut (1993) and Garnaut (1994) call this the ''prisoners' delight,'' as contrasted with the standard prisoners' dilemma. The connotation of liberalization as a game-theoretic equilibrium is somewhat misleading, however, as the argument depends on learning about economic structure rather than on strategic considerations.

A synonym for competitive liberalization is ''ratcheting up'' of the level of liberalization (Bergsten 1994). The formation of the European Economic Community (EEC), for example, was the explicit US reason for launching the Dillon and Kennedy Rounds of GATT negotiations. It is also claimed that the enlargement of the EEC from six to nine gave the United States the impetus to push for the Tokyo Round (e.g., Bergsten 1991; WTO 1995, 53-54) and that the US pursuit of regional agreements with Israel and Canada in the 1980s helped to convince the European Community to drop its earlier veto against a new multilateral round and thus to allow the meeting in Punta del Este, Uruguay, that launched the Uruguay Round (Bergsten 1996).

A more recent illustration is President Bill Clinton's ''triple play'' of late 1993 (Bergsten 1997, 18-20; Destler 1995, 228; Kahler 1995b). He upgraded the Seattle meeting of APEC ministers that had been scheduled for November 1993 by adding a high-profile leaders' meeting. In this way, the United States signaled to the Europeans that if they continued to allow French farmers to hold up the Uruguay Round, other countries might proceed with other initiatives without them. This message carried credibility because of its fortunate timing, coming as it did on the heels of the hard-fought approval of NAFTA in the US Congress. Thus, the NAFTA outcome demonstrated the political will necessary for meaningful agreements, while the APEC meeting demonstrated the possibility that an agreement would cover a fraction of the world economy that was sufficiently large and dynamic to give the Europeans cause for worry at the prospect of exclusion. German policymakers have reportedly confirmed that this was part of their motive for concluding the Uruguay Round in December 1993. In this episode, at least, it appears that regional initiatives helped bring about multilateral agreement.

Of course, the game need not always come out so well. The trouble with making credible threats is that sometimes they must be carried out. The process that is traditionally feared is competitive regionalization, in which the formation of one regional group puts pressure on other countries to form a bloc of their own rather than to liberalize unilaterally or multilaterally. The usual interpretation of the 1930s, for example, is that competitive regionalization exacerbated the decline in world trade and thus contributed to the Great Depression and perhaps to World War II. This view is held even by those who see regionalism in recent history as benign.[11]

In an important analysis of the political economy of regional blocs, Oye (1992) argues that the expected costs of exclusion from groups change the political dynamics by strengthening the antiprotectionist constituencies

11. Eichengreen and Frankel (1995) compare the current and 1930s encounters with regionalism.

domestically, so as to draw countries into multilateral negotiations. Whereas many authors read the recent experience as one in which regionalism helps build support for multilateral liberalization, Oye (73-74) finds that this was also true of the 1930s experience. The US Tariff Act of 1930 (Smoot-Hawley), which itself was nondiscriminatory, led to discriminatory retaliation on the part European countries over the subsequent four years. Thus regionalism was the response to the collapse of the liberal trading order, not the cause (see also Perroni and Whalley 1996; Whalley and Hamilton 1996). The subsequent American reaction to the European actions was not to escalate the protectionism. Rather, Secretary of State Cordell Hull negotiated a series of new tariff-reduction agreements, which were also discriminatory.

Whatever the reading of history, there is a logic that suggests the danger that bloc formation may be protectionist. The worst situation for a country is to be one of the few that do not belong to any bloc because the terms of trade then turn against it. Thus every country will wish to join a bloc. The question is whether a bloc will allow it to join, and if so, which one. Ever since the influential note by Kemp and Wan (1976, 96), the hope has been that a large existing bloc would continue to add new members: ''. . . an incentive to form and enlarge customs unions persists until the whole world is one big customs union, that is, until world wide free trade prevails.'' The incentive is simply that it is theoretically possible to form a sequence of ever-larger blocs so that each move is Pareto-superior—that is, some countries benefit, and none are harmed.[12]

Why hasn't worldwide free trade already been achieved through this route? To say that a bloc can theoretically expand so as to improve its welfare without hurting nonmembers is not to say that it will actually do so. The original reason to ask whether trade blocs would be building blocks for worldwide free trade was the question of whether bloc members would in fact find it in their interest to admit nonmembers.

Mansfield (1993, 212), for example, argues contrary to the happy Kemp-Wan world in which nobody gets hurt at any step along the beneficent

12. Kemp and Wan were not at all specific about what attributes of the customs union would ensure the Pareto-superior outcome. Grinols (1981) refined the analysis of the transfers among members that would be necessary to make sure that all members gained from the customs union. With regard to nonmembers, the standard interpretation is that tariffs on nonmembers should be cut at least to the point at which trade quantities are the same as before the regional trading agreement; then nonmembers will be no worse off. McMillan (1993) proposes on these grounds actually revising GATT Article XXIV to replace the stipulation that tariffs against nonmembers must not rise with a stipulation that trade quantities with nonmembers must not fall. To make this prescription practical, however, one would want to measure changes in trade with nonmembers *relative to what would have happened in the absence of the regional trading arrangement*. This is another possible use for the gravity model. Kowalczyk and Wonnacott (1992) is an application to the Western Hemisphere.

path toward worldwide free trade. He argues that free trade areas are in practice formed among countries that are allies in a political-military sense, with the deliberate aim of helping their friends economically (through trade creation) and hurting their adversaries (through trade diversion). He points out as a corollary that members of an FTA will not want to include countries that it does not wish to see grow economically.

There is a danger that the world will become stuck in a Nash noncooperative equilibrium of several competing FTAs. Each continent, for example, forms an FTA because, given that the next continent is doing so, it will be hurt if it does not respond in kind. In the resulting equilibrium, all are worse off than they were under the status quo of MFN. We saw this in figure 8.14, which illustrates Stein's (1994, 83-93) model of continental FTAs with intercontinental transport costs. In the simulation, the status quo of MFN features worldwide welfare that falls short of free trade by only about 0.5 percent of GNP (which may not be enough to overcome negotiating costs). Each continent in sequence has an incentive to form an FTA, raising its welfare but lowering that of all the other continents, until all four have done so. In that noncooperative equilibrium, the loss relative to global free trade is about 2.5 percent. Hence the argument for discouraging FTAs altogether.

Even if continents are allowed to choose the level of intrabloc preferences to maximize their individual welfares, rather than being constrained to create full-fledged FTAs as Article XXIV stipulates, in equilibrium they will still choose a level of preference that is so high as to leave everyone worse off, as illustrated in figure 8.15. Each bloc opts for an FTA in the prisoners' dilemma equilibrium. This suggests a surprising conclusion: it would not be enough for Article XXIV to *allow* partial preferences; it would have to *require* them.

On the other hand, since the goal is worldwide free trade, it is not certain that the ultimate political-economy dynamic is bad. Worldwide economic welfare is so reduced by a noncooperative equilibrium of four continental FTAs that it may then become politically possible for them to agree multilaterally to remove the barriers that remain between them and move toward worldwide free trade. This would seem to follow if the chief obstacle to a move from MFN to worldwide free trade is a moderate, fixed resource cost to negotiations (say 1 percent of GDP, to buy off producers that stand to lose). The leap to free trade would be all the more likely to follow if the resource cost to negotiation increases with the number of distinct entities involved.

The Ever-Expanding Free Trade Bloc

When a bloc forms, nonmember countries might respond by seeking to join it rather than by forming blocs of their own. This was certainly

the hope when Kemp and Wan proved their theorem of a hypothetical sequence of ever-expanding FTAs that could improve the welfare of all countries at every stage. Several authors (each with somewhat different specifications of the model) have shown that nonmember countries one by one will find it in their interest to join a given FTA, even without the sort of (often impractical) redistributive side payments that Kemp and Wan had in mind (Bond and Syropoulos 1995; Deardorff and Stern 1992; Saxonhouse 1993; Stein 1994; Yi 1994).

The incentive to join lies chiefly in the fact that the bloc turns the terms of trade in favor of the members and against the nonmembers. Baldwin (1996) models the political economy process. When the number of members in a bloc rises, firms in nonmember countries lose demand and so lobby their representatives to join the bloc. Each new member in turn heightens the incentive for the remaining nonmember countries to join. While the bloc expands, its members gain progressively as the terms of trade are shifted further and further in their favor. Those that continue to be left out lose progressively.

In the model of Deardorff and Stern (1992), the bloc will continue to grow until it encompasses the whole world—the optimal outcome of global free trade. Their model, however, assumes that the bloc at each stage places prohibitive tariffs on outsiders, a rather extreme assumption.

Haveman (1992), Saxonhouse (1993), Stein (1994), and Hamada and Goto (1996) consider the same problem, while allowing trade with nonmembers. As already noted in chapter 8, Haveman uses a model in which trade is determined by differences in factor endowments. When the bloc is constrained from raising the level of its tariffs against nonmembers, due, for example, to Article XXIV, he gets the same result as Deardorff and Stern do: the bloc will continue to add members until all countries are happily included. In this context, the Article XXIV constraint is critical. When the incentive to exploit monopoly power is given free reign, Haveman gets the same result as Krugman (1991b): a bad equilibrium of a few large blocs with high tariffs.

The other three papers each use the imperfect substitutes framework. They find that when the bloc reaches a certain size (20 out of 30 members in Saxonhouse, 16 out of 30 in Stein, and ''about half'' in Hamada and Goto), it will choose not to accept any new members because its own welfare starts to decline after that. To the extent the bloc derives its benefits from shifting the terms of trade in its favor, the benefits would disappear if there were no countries left outside the bloc to be exploited. What makes this story especially alarming from the viewpoint of ultimate multilateral liberalization is that the single bloc is truly a dead end: welfare of the bloc members is higher than it would be under worldwide free trade, so they have an incentive to reject multilateral liberalization that they did not have when the alternative was MFN. At this unfortunate dead-end point, worldwide welfare is close to its minimum, the very low welfare of the nonmembers outweighing the high welfare of the members.

At some point, the nonmembers will presumably wise up and form a bloc of their own. But given two competing blocs, the incentive for individual countries will be to join the larger of the two and thus to share in its monopoly power. A world of two equal blocs is unstable (Bond and Syropoulos 1995). A simulation in Stein (1994, 99-102) shows that the stable equilibrium has 26 out of 30 countries in one large bloc and 4 in the other. Again, the large bloc has no incentive to take mercy on those that are excluded.[13]

Yi (1994, 1996) has proposed a theoretical solution to this difficulty that also has the virtue of defining the ambiguous phrase, popular in the Pacific context, ''open regionalism.'' The authors cited so far assume that an existing bloc will admit additional candidates as members only if it is in the economic interest of those already in the club. One could imagine adding a clause to Article XXIV that would specify that FTAs must let in anyone that wants to join. Yi shows that, unlike in the case of exclusive regionalism, which eventually results in a world divided into one large bloc and one small bloc, this open regionalism rule would result in a customs union that expands to include the whole world. The rule has also been called ''inclusivity.''

Stein (1994, 103-05) has another proposed solution to the same problem: that Article XXIV be amended to state that preferences within a bloc cannot go beyond a specified low level (22 percent is the magic limit, in his simulation). We have already seen that such a restriction—the opposite of the current Article XXIV requirement for 100 percent preferences—would be welfare-improving in a world of equal-sized continental blocs. The same is true when there are no intercontinental transport costs and there is a temptation for countries to join the larger of two blocs. The equilibrium still features one large bloc (24 countries) and one small (6 countries). But with the limit on the margin of preferences in place, the large bloc has nothing to lose by moving to worldwide free trade so that the felicitous outcome is still ultimately attainable. Of course, the members of the large bloc would vote against such a rule in the WTO. However, if the issue is decided before any single incipient group is large enough to know that it will be the dominant bloc, then everything will work out for free trade.

Which Effects Appear to Dominate in the Trade Data?

In sum, there are many possible channels of political causation running from regionalism to multilateralism—some positive and some negative.

13. The theoretical model of Seidman and Winter (1996) gives a different result. The smaller coalition or bloc sues for peace in the tariff war with the larger bloc, thus allowing the attainment of worldwide free trade via the steppingstones of regionalism, where a single

Trading blocs have some building-block attributes and some stumbling-block attributes.

How can one get an idea as to which effects in practice dominate, on average? The gravity framework of chapters 4 to 6 can shed some light on the net effect of political interactions such as the ones we have described, as they have actually played themselves out over the last 30 years. For each group that is believed to have undertaken regionalization, we look at the dummy variable for ''openness.'' This dummy variable indicated the effect when at least one country of the pair is a member of the group in question, not necessarily both countries.

If tariffs and other barriers against imports from nonmembers remain unchanged when a given regional group is formed, then the coefficient on the openness variable should be negative, indicating trade diversion. This was often the case. Trade creation is indicated by a positive coefficient on the standard bloc variable (the dummy variable indicating when both countries in the pair are members of the group in question). If trade diversion is large enough relative to trade creation, then the FTA may reduce economic welfare. If trade diversion is small, the FTA is likely to improve welfare. A third possibility is that adoption of a regional FTA is associated with political momentum in favor of more widespread subsequent liberalization, for any of the reasons enumerated in this chapter. In this case, which is the best outcome from the standpoint of economic welfare, the coefficient on the openness variable would be positive rather than negative.

We saw in chapter 5 that the results for most of the openness variables differ depending on the precise specification. We focus here on estimates that pool the years 1970, 1980, 1990, and 1992, which tell a relatively optimistic story. East Asia and Western Europe both show highly significant openness with respect to the rest of the world. Table 5.1 reports a set of results for formal FTAs.[14] When East Asia is divided into ASEAN countries and others, both show highly significant openness. When Western Europe is divided into the EC countries and others (the old European Free Trade Association), it is the European Community that shows positive openness. The Western Hemisphere shows no significant openness effect when considered in the aggregate. This result, which pools data from 1970-92, masks a pattern of some significant negative openness coefficients in 1965-75, followed by significant positive coefficients in 1985 and 1990, as Latin America entered its era of benign FTAs. Estimates on formal FTAs show this pattern for Mercosur in particular, as noted in chapter 5. When one allows for trends over time, NAFTA and ASEAN

large leap to free trade would have been rejected by countries that hoped to do better in blocs. As in Riezman (1985), in the absence of international transfers, a move to global free trade can be vetoed by a single country that stands to lose.

14. Table 5.1, and the results for the broader groups, are taken from Wei and Frankel (1997).

show statistically significant upward trends in openness (appendix table B5.3). The other FTA that shows some evidence of openness is the Australia-New Zealand CER pact. The group showing the strongest evidence of trade diversion or negative openness is EFTA.

Estimates of openness for each of the countries in our sample (with respect to exports and imports, separately) are reported in appendix table B5.5a-d. The population and bloc variables are omitted, and the GNP coefficients are constrained to 1. Otherwise, however, the framework is that of the gravity equation. The most open countries include the Netherlands, United Kingdom, Belgium, Singapore, and Chile. If one adjusts for its large population, the United States is the most open of the industrialized countries. Adjusting for its high GDP per capita, the United States is also one of the most open in the entire data set, though surpassed by Singapore, Chile, and a few others. Among the most closed is Sudan. The country with the greatest decline in openness over the sample period is Iran. The countries with the greatest increase in openness include Argentina, Brazil, South Korea, Malaysia, Mexico, Paraguay, and Thailand. There may be some correspondence between membership in an important regional trade arrangement and high or increasing level of openness.

Our results suggest that the third possibility enumerated above is often the most relevant one. Some countries have tended to open up with respect to all trading partners at the same time that they have opened with respect to members of their own group. This conclusion matches that of a recent WTO report (1995; also Fishlow and Haggard 1992, 60), to the effect that the recent regional arrangements among its members have not been fortresses but have, to the contrary, sometimes helped to promote freer trade worldwide.

Thus, the net political effect of the removal of regional barriers can be to support liberalization with respect to nonmembers as well. The effect of further liberalization has in many cases been more than enough to offset any trade diversion resulting directly from the original regional arrangements themselves. Trading blocs can be building blocks rather than stumbling blocks. From the economists' viewpoint, this verdict, fragile though it may be, is an encouraging one.