

Regional Agreements and Trade in Services: Policy Issues

Aaditya Mattoo and Carsten Fink*

Abstract: Every major regional trade agreement now has a services dimension. Is trade in services so different that we need to modify the conclusions on preferential agreements reached so far in the realm of goods? This paper examines, first, the implications of unilateral policy choices in a particular services market. It then explores the economics of international cooperation and identifies the circumstances in which a country is more likely to benefit from cooperation in a regional rather than multilateral forum.

Keywords: Regional integration, trade in services

JEL Classification: F13, F15

World Bank Policy Research Working Paper 2852, June 2002

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the view of the World Bank, its Executive Directors, or the countries they represent. Policy Research Working Papers are available online at <http://econ.worldbank.org>.

* Development Research Group, World Bank, 1818 H Street, NW, Washington, DC 20433. Email: amattoo@worldbank.org and cfink@worldbank.org. This paper has benefited from discussions with Simon Evenett, Bernard Hoekman, Marcelo Olarreaga, G.V. Rao, David Tarr, and, in particular, Maurice Schiff.

Regional Agreements and Trade in Services: Policy Issues

Non-technical Summary

There is a large literature on the costs and benefits of integration agreements on trade in goods, and hardly any analysis of the implications of such agreements in services. This is surprising because nearly every major regional agreement now has a services dimension. The question arises: is trade in services so different that we need to modify the conclusions reached so far in the realm of goods? In particular, what would happen if a country liberalized services trade faster in the regional context than at the multilateral level? And if a country were to obtain preferential access to foreign goods markets, would the benefits justify granting preferential access in services to its home market?

The first part of this paper is concerned with the efficiency effects of unilateral policy choice in a particular services market, and addresses two questions: in its independent choice of services policy, is a country likely to improve upon the status quo by liberalizing on a preferential basis? And between preferential and non-preferential liberalization, which is likely to produce larger welfare gains? The main conclusions are as follows:

- Compared to the *status quo*, a country is likely to gain from preferential liberalization of services trade at a particular point of time—as distinct from the more ambiguous conclusions emerging for goods trade. The main reason is that barriers are often prohibitive and not revenue generating, so there are few costs of trade diversion.
- As in the case of goods trade, the scope for increased competition and exploitation of scale economies, as well as the possibility of inducing knowledge spillovers, strengthens the presumption that a country would gain from a preferential agreement in services.
- *Non-preferential* liberalization is likely to produce larger gains than preferential liberalization *ceteris paribus*, because it does not in any way bias consumer choice, and allows consumers to import from the most competitive source.

The second part examines the economics of international cooperation and addresses the question: are there circumstances in which a country is more likely to benefit from cooperation in a plurilateral forum than in a multilateral forum? We find three main arguments in favor of a plurilateral approach:

- Participants in a plurilateral agreement may gain at the expense of the rest of the world either through improved terms-of trade in competitive markets or, more

likely in services, by shifting rents towards participants' firms in oligopolistic markets - unless excluded countries retaliate by concluding similar agreements.

- More efficient bargaining may be possible in a plurilateral context than in the multilateral context: there is less concern that outsiders will be able to free-ride on the reciprocal exchange of concessions than if there were a general MFN obligation.
- Regulatory cooperation may be more feasible and in many cases more desirable among a subset of countries than globally.

But there is an important general caveat to each of these three arguments:

- The sequence of liberalization matters more in services trade than in the case of goods trade. In particular, the benefits of eventual non-preferential liberalization may be different if it is *preceded* by preferential liberalization. This is because location-specific sunk costs of production are important in many services, so even temporary privileged access for an inferior supplier can translate into a long-term advantage in the market. Thus, while the elimination of preferences may lead to a relatively painless switch to more efficient sources of goods supply, the entry of more efficient service providers may be durably deterred if their competitive advantage does not offset the advantages conferred by incumbency. These considerations are particularly relevant for the large number of countries that export mainly goods and import many services.

Regional Agreements and Trade in Services: Policy Issues

Introduction

There is a large literature on the costs and benefits of integration agreements on trade in goods, and hardly any analysis of the implications of such agreements in services.¹ This is surprising because nearly every major regional agreement now has a services dimension. The question arises: is trade in services so different that we need to modify the conclusions reached so far in the realm of goods? In particular, what would happen if a country liberalized services trade faster in the regional context than at the multilateral level? And if a country were to obtain preferential access to foreign goods markets, would the benefits justify granting preferential access in services to its home market?

We recognize that the choice of integration strategy may be determined primarily by political considerations (World Bank, 2000). There is, nevertheless, a need for an assessment of the economic benefits and costs of alternative approaches to services liberalization. How such an assessment might be undertaken is the subject of this paper. We do not seek to provide a comprehensive analysis of regional integration, but to highlight, for the most part, why the implications may be different in agreements involving services. We proceed in two steps.

The first part is concerned with the efficiency effects of unilateral policy choice in a particular services market, and addresses two questions: in its independent choice of services policy, is a country likely to improve upon the status quo by liberalizing on a preferential basis? And between preferential and non-preferential liberalization, which is likely to produce larger welfare gains?

The main conclusions are as follows:

- Compared to the *status quo*, a country is likely to gain from preferential liberalization of services trade at a particular point of time—as distinct from the more ambiguous conclusions emerging for goods trade. The main reason is that barriers are often prohibitive and not revenue generating, so there are few costs of trade diversion.
- As in the case of goods trade, the scope for increased competition and exploitation of scale economies, as well as the possibility of inducing knowledge spillovers,

¹ One recent paper, Baier and Bergstrand (2001), does seek to examine the implications of a free trade agreement (FTA) in services but the assumptions limit the value of the results. Services are assumed to differ from goods because they have higher or prohibitive transport costs – but strangely, only when transported across continents while transport costs are zero between countries on the same continent. Not surprisingly, continental FTAs in services are found to be desirable—a conclusion similar to the finding by Frankel, Stein and Wei (1995) for trade in goods. The only other difference examined between trade in goods and services is that the latter face a higher level of protection.

strengthens the presumption that a country would gain from a preferential agreement in services.

- *Non-preferential* liberalization is likely to produce larger gains than preferential liberalization *ceteris paribus*. Non-preferential liberalization is superior because it does not in any way bias consumer choice, and allows consumers to import from the most competitive source.

The second part examines the economics of international cooperation and addresses the question: are there circumstances in which a country is more likely to benefit from cooperation in a plurilateral (or regional) forum than in a multilateral forum?

We find three main arguments (which are not necessarily specific to services) in favour of a plurilateral approach:

- Participants in a plurilateral agreement may gain at the expense of the rest of the world either through improved terms-of trade in competitive markets or, more likely in services, by shifting rents towards participants' firms in oligopolistic markets - unless excluded countries retaliate by concluding similar agreements.
- More efficient bargaining may be possible in a plurilateral context than in the multilateral context: there is less concern that outsiders will be able to free-ride on the reciprocal exchange of concessions than if there were a general MFN obligation. However, departures from the MFN principle may create inefficiencies due to increased uncertainty about the value of concessions.
- Regulatory cooperation may be more desirable among a subset of countries than globally. Ideally, countries choose their partners spontaneously sector by sector, depending on the costs and benefits of regulatory harmonization. But under certain circumstances, it may be desirable to choose partners *ex ante* in macro-agreements and then seek to deepen integration across sectors.

But there is an important general caveat to each of these three arguments:

- The sequence of liberalization matters more in services trade than in the case of goods trade. In particular, the benefits of eventual non-preferential liberalization may be different if it is *preceded* by preferential liberalization. This is because location-specific sunk costs of production are important in many services, so even temporary privileged access for an inferior supplier can translate into a long-term advantage in the market. Thus, while the elimination of preferences may lead to a relatively painless switch to more efficient sources of goods supply, the entry of more efficient service providers may be durably deterred if their competitive advantage does not offset the advantages conferred by incumbency. These considerations are particularly relevant for the large number of countries that export mainly goods and import many services.

1. The Economics of Unilateral Policy Choice

1.1 Standard economics of preferences

The conventional analysis of regional agreements focuses on goods trade and emphasizes two main types of effects.² The first are '*trade and location*' effects. The preferential reduction in tariffs within a regional agreement will induce purchasers to switch demand towards supply from partner countries, at the expense of both domestic production and imports from non-members. This is trade creation and trade diversion. The former is beneficial, but the latter may be costly. In particular, governments will lose tariff revenue, and the overall effect on national income may be positive or negative, depending on the costs of alternative sources of supply and on trade policy towards non-member countries.

Furthermore, changes in trade flows induce changes in the location of production between member countries of a regional agreement. These relocations are determined by the comparative advantage of member countries, and by agglomeration or clustering effects. In some circumstances they can be a force for convergence of income levels between countries. For example, labor intensive production activities may move towards lower wage countries, raising wages there. In other circumstances they can be a force for divergence. For example, industry may be pulled towards a country with a head-start or some natural advantage, driving up incomes while other countries lag.

The second source of economic change are '*scale and competition*' effects. Removal of trade barriers is like a market enlargement, as separate national markets move towards integration in a regional market. This allows firms to benefit from greater scale, and attracts investment projects for which market size is important, including foreign direct investment. Removing barriers also forces firms from different member countries into closer competition with each other, possibly inducing them to make efficiency improvements. In sum, enlarging the market shifts the trade-off between scale and competition, and it becomes possible to have both larger firms and more competition.

1.2 Economics of preferences in services

The analysis of preferential agreements in services requires an extension of conventional theory in two ways. First, since services trade often requires proximity between the supplier and the consumer, we need to consider preferences extended not just to cross-border trade, but also to foreign direct investment and foreign individual service providers. Secondly, preferential treatment could be granted not through tariffs (which are rare in services trade), but through discriminatory restrictions on the movement of labor and capital (e.g. in terms of quantity or share of foreign ownership), and a variety of domestic regulations, such as technical standards, licensing and qualification requirements. The consequences of preferential tariff arrangements are well understood. Do preferences through alternative instruments, impacting both on product and factor mobility, raise new issues?

² This section draws on World Bank (2000). A review of the standard economics of preferential agreements and new developments is to be found in Panagariya (2000).

The implications of preferential liberalization on factor mobility depend, first of all, on whether it is temporary, i.e. only for the fulfillment of a particular service contract, or relatively permanent. At one extreme, temporary preferential access for foreign consultants or construction companies is analogous to preferential liberalization of trade in products and can be expected to have similar effects. It is as if the service product were carried to the consumer by the supplier, after which the supplier returns home.

At the other extreme, an integration agreement could imply full integration of product and factor markets, as in the case of the European Union.³ In between, there may be a limited extent of permanent movements of individual suppliers (through migration) and capital (through foreign direct investment). Such movement would imply a change in the factor endowments of participating countries. The positive impact will depend on the specificity of the factors that move and the normative impact on the extent to which incomes are repatriated. Most of the integration agreements that exist or are being considered are as ambitious as the European Union as far as the product market is concerned, but less ambitious in terms of the implied liberalization of certain types of factor mobility, particularly relating to labor. Our discussion will, therefore, focus on the more limited types of agreements.

The manner in which privileged access can be granted by a country to some suppliers depends on the instrument of protection that it has in place. For instance:

- Where a country imposes a quantitative restriction on services output or on the number of service providers, it could allocate a larger proportion of the quota to a preferred source. Examples of the former can be found in air, road and maritime transport, where many countries allocate freight and passenger quotas on a preferential basis, and in audiovisual services, where preferential quotas exist on airtime allocated to foreign programs. Examples of the latter are restrictions on the number of telecommunications firms, banks, and professionals that are allowed to operate.
- A country could also relax restrictions on foreign ownership, type of legal entity, branching rights, etc., on a preferential basis. For example, under NAFTA, Mexico eliminated ownership restrictions on financial institutions established in Canada and the United States, but for a certain period maintained restrictions on financial institutions based outside these countries.⁴ Preferential treatment with regard to local incorporation and branching rights were an issue in the WTO

³ While services trade policy at the multilateral level is likely to see a persistence of the asymmetric attitude to the two factors, i.e. temporary movement of individuals combined with relatively permanent movements of capital, a more symmetric attitude may be feasible at a regional level – witness, for instance, the experience of the European Union and Caribbean Community.

⁴ The effective preference granted by Mexico to the United States and Canadian banks was limited because European banks based in the United States were also able to benefit from the NAFTA opening. In fact, several Spanish and Dutch banks first established a presence in Mexico through their subsidiaries in the United States. This illustrates how a liberal "rule of origin" can limit the scope for trade diversion.

financial services negotiations and eventually led to the so-called grandfather provisions - whereby certain existing firms were allowed to operate under more favorable conditions.

- There could be discrimination through taxes and subsidies. In many countries, all firms established in a country are assured of national treatment, i.e. there is no discrimination against such firms even if they are foreign-owned - so there is limited scope for preferential treatment of some foreign providers post-establishment. Nonetheless, in some countries, foreign providers are subject to different tax rates and do not have access to subsidies.
- Far more feasible is preferential treatment through domestic regulations pertaining to technical regulations, licensing and qualification requirements. Many countries today impose qualification and licensing requirements on foreign providers that are not necessary to achieve regulatory objectives. Where these are waived only for some of the foreign providers who deserve the benefit, de facto preferences result. Regulatory preferences also arise in other sectors, ranging from transport to financial services. For instance, owing to the reciprocal recognition of the proof of solvency between the European Union and Switzerland, insurance companies that have their principal place of business in the territory of one of the contracting parties are not obliged to localize funds to a significant extent. The United States agreement with Canada eliminates the need for chartered accountants trained in these countries to duplicate all steps in the licensing process, and provides for abbreviated examination requirements.

Table 1: Measures affecting trade in services

		Generating domestic rents	Not generating domestic rents
Measures which increase variable costs of foreign providers	(1)	(2)
	... fixed costs of foreign providers	(3)	(4)
Quantitative restrictions on final sales	(5)	(6)
	... number of providers	(7)	(8)

Table 1 provides a classification of the measures affecting trade in services. Traditional trade theory has focused on the impact of preferences when barriers are of type (1), i.e. tariff-like instruments, or (5) and (6), i.e. quantitative restrictions on sales, essentially on

cross border trade. We wish to highlight three forms of discrimination that seem particularly relevant to trade in services:

- Through variable cost-increasing protectionist measures that do not generate rents (“frictional barriers”) (2)– relevant for all modes but easiest to analyze for cross-border trade.
- Through measures that affect the fixed cost of supply (3) and (4)– most relevant for commercial presence and easiest to analyze when cross-border trade is not feasible.
- Through quantitative restrictions on the number of service providers (7) and (8).

1.3 Preferential access and frictional barriers

A large variety of measures that a country maintains can have the effect of increasing the variable costs of operation without generating (equivalent) rents. The problem is that it would not usually be correct to treat all the additional costs imposed on foreign services or service suppliers of conforming to local recognitions as a form of protection. It is necessary to distinguish between the regulatory burden imposed on the foreign supplier that is necessary to ensure the desired quality of the service and that which is excessive. For instance, the requirement that foreign financial service providers incorporate locally (rather than enter as branches) obliges all entities to fulfill local capital and reserve requirements, which has the effect of increasing their costs of operation. It may be that the imposition of some but not all of these costs is justified – e.g. a part of the capital requirement could be fulfilled by the parent institution.

A variety of other measures can also have the effect of increasing variable costs of operation. One example are the excessive border formalities that impose a burden on international transport service providers. Another example of cost-increasing measures are local content requirements, such as the stipulation that foreign firms use a certain proportion of local employees - if the foreigner would not freely do so. Restrictions on foreign ownership may also translate into a higher variable cost if such restrictions dampen the incentive of the foreign provider to improve performance, e.g. by transferring technology or improving management.

We illustrate the implications of preferential access by considering the relatively complex case where the preference is a consequence of selective recognition of foreign regulations pertaining to standards, qualifications or licensing. We can think of any standard as made up of two parts: a "universal" element, consisting of u units of quality, which is identical between countries, and a country-specific element equal to v_i units – reflecting either the preference for a higher quality or a different variety. In some sectors, for instance in construction, financial and transport services, it is reasonable to presume that there is a high universal component to standards though there is usually also a country specific element. Within certain professional services, like medicine and engineering, the universal component is also likely to be high, whereas in other professions, like law and accountancy, the country-specific component is likely to be high.

Let us also assume that the cost of meeting a unit of the standard in country i is constant at c_i . The variations in c_i are meant to capture inherent advantages that certain countries have in certain areas. If a foreign provider wished to provide a service in country i , it would necessarily have to accept an increase in costs by $c_i v_i$. But it is possible that the first country also refuses to acknowledge the equivalence of the universal part of the standard, and insists on full re-qualification, implying costs $c_i(u + v_i)$. In this case, $c_i u$ would be a measure of the excessive regulatory burden. More subtle forms of protection could involve understatement of the universal element u , and exaggeration of the country specific element v_i . Preferential recognition agreements may exempt certain suppliers from incurring whole or part of these costs.

In situations where the impact of regulation is on variable costs, as is assumed here, the analysis of discriminatory regulation is to some extent analogous to the analysis of tariffs (see Annex A). As we know, when tariffs are the instruments of protection, the costs of trade diversion can be an important disincentive to conclude preferential liberalization agreements. The displacement of high-tariff imports from third countries by low or no-tariff imports from preferential sources implies lost revenue, which may offset the gains in consumers' surplus from any liberalization. The same reasoning also applies to other regulations which imply a transfer from foreign suppliers to domestic interest groups. However, the situation is different when the protectionist instrument is a regulatory barrier that imposes a cost on the exporter without yielding a corresponding revenue for the government or other domestic entity. There is then no cost to granting preferential access because there is no revenue to lose. Therefore, preferential liberalization would necessarily be welfare-enhancing.

However, countries outside the preferential arrangement may lose. The exemption from a wasteful regulation implies reduced costs for a class of suppliers and hence a decline in prices in the importing country. This decline in prices hurts third country suppliers who suffer reduced sales and a decline in producers' surplus. Interestingly, preferential exemptions are likely to increase global welfare even though excluded suppliers lose. The gain to consumers from any decline in price is necessarily greater than the loss to a subset of suppliers. This makes intuitive sense: eliminating wasteful duplication should enhance global welfare. Though, of course, a non-preferential recognition agreement would enhance national and world welfare even more, because the service would be supplied by producers from the most efficient locations.

The analysis of discriminatory regulation is also relevant to quantitative restrictions on the sales of services. In the case of goods, the quota rents can be appropriated by domestic intermediaries like the importer rather than the foreign exporter. However, in many services, intermediation is difficult because the service is not storable and directly supplied by producers to consumers. Rents are, therefore, usually appropriated by exporters rather than domestic importers. As in the case of frictional measures, there can be no cost of trade diversion to the preference-granting country.

Policy implication:

- Where a country maintains regulations that impose a cost on foreign providers, without generating any benefit (such as improved quality) or revenue for the government or other domestic entities, welfare would necessarily be enhanced by preferential liberalization. However, *non-preferential* liberalization would lead to an even greater increase in welfare nationally and globally because the service would then be supplied by the most efficient locations.

1.4 Preferential access and the fixed costs of entry or establishment

A number of measures that countries maintain can have the effect of increasing the fixed costs of entry or establishment. For instance, the requirement to establish a local presence; license fees for entry into the market; the need to requalify for foreign professionals. Again some of these costs may be justified by the regulatory objective.

To analyze the impact of fixed costs, we need to move away from the perfectly competitive model presented in the previous section. Consider a service industry where firms face constant marginal costs and two types of fixed costs, a firm-specific fixed cost of setting up production (unrelated to policy) and a fixed cost of selling to each market (related to policy).⁵ The three countries – home (X), partner (Y) and rest of world (Z), are assumed *ex ante* identical to reduce complexity. Marginal production costs are also initially assumed to be identical. The three markets are assumed to be segmented – which is plausible where cross-border delivery is not feasible. For simplicity, we assume that the part of the fixed cost that is necessary is incorporated into the “technologically” given fixed cost, and so the part that we observe is unnecessary.

We assume that each market has its own norm and complying with these costs F in each market. We also assume that there are no other restrictions on entry and exit in any of the markets, so the number of firms can change as policy changes. This is likely to be a reasonable description of a wide range of services sectors ranging from professional to financial services, when explicit barriers to entry are eliminated but local qualification or licensing requirements remain. In equilibrium, each firm's profits will just cover its fixed costs.

Initially the same fixed-cost regulation applies on a non-discriminatory basis. Any firm that wants to sell in more than one market must incur a fixed cost of F in each. So a firm from any country (X, Y or Z) that wishes to sell in all three markets has to pay a total of $3F$, and in any two markets (say X and Y) a total of $2F$. Say liberalization takes the form of mutual recognition between X and Y, and a service/services supplier that complies with *either* X's or Y's norms, can sell freely in both markets. That is, X-firms and Y-firms have to pay only a total of F to access both the X and Y markets.

This mutual-recognition privilege may or may not extend to Z-firms. Say it does not, and only X and Y firms benefit. In practice, this exclusion is enforced by restrictive rules of origin. That is, X and Y firms only need to meet the norm in their home market and then they can sell in both markets, but Z firms must meet the norm in each market separately.

⁵ The model, drawn from Baldwin (2000), is described in more detail in Annex B.

This means that X and Y firms pay only F to access the combined X-Y market, but Z-firms must continue to pay $2F$. This results in improved profitability for X and Y firms, which will lead to new entry by, and an increase in the number of, X and Y firms. The increased competition will lead to a decline in prices, and a drop in Z firms. Thus, exclusionary mutual recognition will imply that new X-based and Y-based firms crowd out Z-based firms.

But consider the situation where Z-firms too are allowed to sell to both X and Y markets after certifying their product in either. Then the fixed cost liberalization benefits all firms equally. This is because now any firm that wishes to sell in the X-Y market need only incur a fixed cost of F – i.e. Z firms are no longer disadvantaged. And any firm that wishes to sell in all three markets must incur a fixed cost of $2F$ – non-exclusionary recognition has reduced the costs of selling to all three markets for each firm from $3F$ to $2F$. This raises profits and attracts new entrants in all countries. Given symmetry, the number of firms in each market rise equally.

It is evident that even mutual recognition among a sub-set of countries improves the market outcome by leading to increased competition. But does it matter for country X whether recognition is open or exclusionary? Even if firms from each country are identical but there are certain technological fixed costs of entry that limit the total number of firms, then the answer is yes. With exclusionary recognition, Z firms get crowded out of the X-Y market leading to less competition in each market than if Z firms had been allowed to benefit from the mutual recognition agreement.

If the excluded Z firms have a lower marginal cost of production, then their displacement would lead to an increase in average marginal costs of the firms operating in the market. This increase could even offset the benefits of increased entry and hence competition, created by reducing the fixed costs for X and Y firms. Hence, discrimination through fixed costs would be particularly costly if it were directed against the more efficient provider (see also the section on knowledge spillovers).

Finally, consider the situation where all three countries participate in the recognition agreement. Now a firm in any country must incur a cost of only F in order to sell in all three markets. This raises profits even more and leads to greater entry and competition than in any of the previous cases. Furthermore, there is now no risk of crowding out the more efficient firm. Clearly mutual recognition among all countries is the most desirable outcome.⁶ Its desirability is even greater if we allow for product differentiation across countries and the fact that consumers of services benefit from greater diversity.

⁶ In this section, we have not addressed the issue of whether fixed costs of entry generate rents for the host country. If all firms have identical costs and there are no other barriers to entry, then even preferential liberalization of fixed cost-increasing barriers are likely to increase welfare. Consider the less obvious case when measures generate rents, as in the case of license fees. Recall that if there are no restrictions on entry then in equilibrium, aggregate industry profits are equal to total fixed costs. If a reduction in fixed costs induces new entry, then the increased competition implies a lower price. The loss in rents from liberalization is equal to the decline in aggregate industry profits which is necessarily lower than the increase in consumer surplus.

Policy implications:

- A country is likely to benefit from eliminating, even on a preferential basis, any *excessive* fixed costs of entry imposed on foreign providers - e.g. by removing unnecessary qualification, licensing and local-establishment requirements in professional and financial services.
- The gains from a particular preferential agreement leading to the elimination of fixed-costs of entry depend on the competitiveness of the partner countries' service providers.
- Regardless of the chosen partners, the presumption that a country will benefit from such initiatives is greater if agreements are not exclusionary – i.e. they do not apply restrictive rules of origin. This means, for example, that a Brazilian dentist who has qualified in Portugal would benefit from recognition in Germany in the same way that a Portuguese dentist would.
- The greatest benefits arise if recognition agreements include all countries that have comparable regulation. That is, if Brazil has basically the same educational and training system for dentists as Portugal, then it should also be made party to the mutual recognition agreement. The benefits come from both increased competition and greater diversity of services.

1.5 Preferential access and quantitative restrictions on the number of suppliers

In the previous section, it was assumed that there were no restrictions on entry other than technological or regulatory fixed costs. However, as noted above, in many countries there are restrictions in a wide range of services sectors on the number of firms that operate. While there are sometimes good reasons to limit entry, such as the existence of significant economies of scale, it is not clear that the observed restrictions are motivated by these considerations.

In such situations, it matters how entry is allowed, i.e. by acquisition (as in financial services) or greenfield (as in telecommunications). Interestingly, allowing limited new entry by foreign firms (i.e. stopping short of removing all barriers to entry), irrespective of whether this is done preferentially or MFN, may not be welfare-enhancing. The main reason is that even though consumers benefit from the increased competition, this gain may be offset by the transfer of rents from domestic oligopolists to foreign oligopolists.⁷ These considerations may affect the preferred mode of entry: entry through acquisition implies less competition than greenfield entry, but it allows domestic firms to extract some of the potential rents through the sale price.⁸

⁷ Restrictions on greenfield entry are also imposed to direct new foreign capital into weak domestic financial institutions to help the restructuring process.

⁸ In these circumstances, price regulation could help prevent the emergence of super-normal profits.

We can only be confident that there will be gains from any form of liberalization if all barriers to entry are removed.⁹ But if only limited entry were allowed, then open non-discriminatory access – e.g. through global auctions of licenses - would dominate preferential access. This would ensure that the new entrants are the most efficient suppliers in the world. In contrast, preferences may lead to entry or acquisition by inferior suppliers. In either case, the downside of preferences may be higher prices for consumers, lower takeover prices for domestic firms or lower license fees for the government.

Policy implication:

- Allowing limited foreign entry in concentrated markets can lead to a decline in welfare. In any case, non-discriminatory allocation of entry quotas is better than preferential allocation.

1.6 Sunk costs and the sequence of liberalization

Sunk costs are important in both goods and services production. However, location-specific sunk costs, i.e. those incurred in supplying a particular market, are arguably more important in a number of services sectors, because their provision requires proximity between the supplier and the consumer.¹⁰ One consequence is that preferential liberalization may have more durable consequences than in the case of goods. For instance, concluding an agreement that allows inferior providers to establish may mean that a country could be stuck with such providers even when it subsequently liberalizes on an MFN basis.

Sunk costs matter because they have commitment value and can be used strategically by those who are allowed to enter the market first (Tirole, 1988). A firm that establishes a telecommunications or transport network today signals that it will be around tomorrow if it cannot easily resell the equipment. The commitment value is stronger the more slowly

⁹ However, the existence of significant economies of scale implies that free entry need not lead to a socially optimal outcome (Tirole, 1988).

¹⁰ Quantitative evidence on the importance of sunk costs in services is not easy to find because they are hard to estimate. Nevertheless, some studies illustrate their importance in widely differing services sectors. Breshnan and Reiss (1993) find that sunk costs are important even in professional services. Their study of US rural counties reveals that dentists sink significant costs. Furthermore, dentists do not usually compete much on prices and entry also does not foster price competition because patients face substantial switching costs. A single incumbent exists in markets with 800 people or less. The Whinston and Collins (1990) event study of entry suggests that even deregulated airline markets – often cited as an example of low sunk costs - are not contestable, and one reason could be the significance of sunk costs. Beesley (1997) finds that the entry of London Express Aviation into the London-Singapore route in the mid 80s involved non-renewable committed costs (sunk costs) of 400,000-450,000 pounds. The authors argue that even though monetary outlays in the form of recognizable committed costs were negligible (not more than about 1.6 percent of the total cost outlays to be incurred in the same year), the problems of which these costs were symptoms, such as the risk of large capital investments, were not. Gual's (1999) analysis of market integration policies in European banking argues that endogenous sunk costs in banking include the development of a brand image, the investment in electronic banking, or the development of a strong capital base. The study predicts that if competition focuses on such sunk costs, then concentration is likely to increase substantially with the integration of the European market.

capital depreciates and the more specific it is to the firm. Then if some firms are allowed to enter the market early, these incumbents may accumulate a quantity of “capital” sufficient to limit the entry of other firms.

Capital need not necessarily take a physical form. A firm may be able to develop a clientele through advertising and promotional campaigns that pre-empt demand. The more imperfect the consumers’ information and the more important the costs of switching suppliers, the greater the clientele effect. Consumers are often reluctant to switch banks and telecommunications suppliers even when new entrants offer better terms. Such incumbency effects may be stronger in services with network externalities, like telecommunications, where new entrants’ technical standards must be the same as those of the incumbent. The incumbent may also succeed in assuring itself of the services of the most capable franchisees by selecting them initially and imposing exclusivity on them. Each of these forms of “capital accumulation” enhance the first-mover advantages and allow the established firms to restrict or prevent competition.

Because of the importance of sunk costs, sequential entry can produce very different results from simultaneous entry. A market outcome where one firm enters first is not necessarily worse than one where all firms enter at the same time, but it may well be for several reasons. First, if entry is costly, then the incumbent may be able to completely deter entry so that the outcome is a much more concentrated market structure.¹¹ Second, and from our point of view more important, the first-mover advantage may be conferred on an inferior supplier who may nevertheless use it to establish a position of market dominance. How durable such a position is depends on the importance of sunk costs relative to differences in costs and quality.

Two qualifications to this argument based on sunk costs are important. First, entry by the more efficient firm could take place through acquisition circumventing some of the problems of first-mover advantage. But this would require no asymmetry of information about the value of assets and no direct costs of transferring assets. Secondly, in certain services sectors, firms could learn by doing: the experience acquired by the established banks during the previous period reduces their current costs, enhancing their competitiveness and discourages others from entering. This form of entry deterrence may well promote welfare.

Policy implication:

- Location-specific sunk costs are important in a large number of services sectors, ranging from professional to telecommunications and financial services.

¹¹ In situations of network externalities, entry deterrence could also be through the choice of a standard that is incompatible with that of potential entrants.

Therefore, a country needs to carefully evaluate not just the static costs of granting preferential access to a particular country, but also how the eventual benefits from multilateral liberalization are likely to be affected.

1.7 Scale, competition, agglomeration, knowledge creation and spillovers

Competition and scale

As noted above in Section 1.1, it has been recognized that the combining of markets through a regional integration agreement can lead to gains arising from a combination of scale effects and changes in the intensity of competition. In a market of a given size, there is a trade-off between scale economies and competition: if firms are larger, there are fewer of them and the market is less competitive. Enlarging the market shifts this trade-off, as it becomes possible to have both larger firms and more competition (World Bank, 2000).

The gains from preferential agreements are likely to be substantial in areas where there is scope for fuller realization of economies of scale, as in certain international transport services, and increased competition, as in business services. In principle, these gains can also be realized through MFN liberalization; but in practice, the full integration of markets requires a deeper integration of regulations which might be more feasible and desirable in a regional context, as discussed below.

Liberalization as an inducement to FDI

Apart from changing the organization of local industry, if regional agreements create large markets and do not impose stringent ownership-related rules of origin, they may also assist in attracting foreign investment when economies of scale matter. For example, a foreign transport service provider might not find it worthwhile to establish in Latin America if each country market were segmented, but might find Latin America attractive with a continent-wide integrated market.¹²

Regionalism and Learning-by-doing

One rationale for a regional agreement is a variant of the infant-industry argument. South-south regional agreements, in particular, are seen as a form of gradual liberalization. Exposure to competition first in the relatively protected regional market could help prepare firms for global competition. This approach improves on traditional infant-industry protection because some degree of international competition is created. There is also the possibility that firms that have become competitive in a regional context are less likely to resist broader-based liberalization. In this sense, regional agreements are seen as a “building block” towards multilateral liberalization (Bhagwati, 1993).

¹² However, if the FDI is induced by high levels of protection against cross-border trade, then welfare may decline – because the private benefits to the foreign investor may outweigh the social benefit from his presence.

However, a regional agreement is only justified on these grounds if the eventual benefits from learning-by-doing at the regional level offset the immediate costs for consumers, in terms of the higher prices they pay when a country chooses regional rather than multilateral liberalization. There is also another concern. In the national context, infant industry protection once granted has proved difficult to eliminate –either because of the continued weakness of domestic industry or the strength of vested interests. Similarly, regionalism may create a new constellation of vested interests that would resist further liberalization – raising the concern that regionalism can become a stumbling block to further multilateral opening (Bhagwati 1993, Krishna, 1998). To an extent, it may be possible to address these concerns by credibly committing to future multilateral liberalization, e.g. by making commitments under the GATS to eliminate barriers at a future date.

Agglomeration

One force that drives the relocation of activity in a RIA is comparative advantage. But as economic centers start to develop, so cumulative causation mechanisms come into effect, leading to the clustering of economic activity in certain locations (World Bank, 2000). How might preferential liberalization of services trade affect the interaction between “centripetal” forces, encouraging firms to locate close to each other, and “centrifugal” forces, encouraging them to spread out? At this stage, it is possible to make only a few superficial observations.

Consider first the impact on the location of goods production. Certain services, ranging from telecommunications to transport, have a critical impact on the cost of distance and this would be reduced by the liberalization of these services, even on a preferential basis. The incentive to locate production close to areas where consumers or inputs are concentrated would be dampened.

There are, of course, likely to be centripetal forces operating in the services sectors themselves, making it attractive for firms to locate close to each other. These could result from knowledge spillovers or other beneficial technological externalities, or labor market pooling effects, which encourage firms to locate where they can benefit from readily available labor skills. The elimination of barriers to trade in services and factor mobility may encourage the production of certain services to gravitate to particular locations. For instance, global advertising service producers might gravitate to New York if there were no barriers to trade in such services. On the one hand, if these services could be easily supplied long-distance, then there need not be an effect on the location of goods production. On the other hand, if these services required proximity between the supplier and consumer, then that would set off further agglomeration forces as the production of goods which rely on these services moved to the same locations.¹³ More research is needed to improve our understanding of the impact of services liberalization on agglomeration.

¹³ It is of course, conceivable, that if goods production is already locked into certain locations, that the producers of face-to-face services are obliged to move to the same locations – regardless of the strength of centripetal forces operating within services sectors. See also Markusen (1987).

Knowledge flows

Preferential agreements may also promote knowledge flows between member countries. A growing body of work argues that trade flows provide a means for the transfer of technology between countries. For instance, it has been found that access to foreign knowledge is a statistically significant determinant of the rate of growth of total factor productivity.¹⁴ There is some evidence to suggest that developing countries benefit from foreign knowledge, first, according to how open they are, and, second, according to whether they are open to those countries that have the largest knowledge stocks.

It would seem that such effects are even stronger when we consider not just cross-border trade but also the movement of factors associated with international transactions in services. In particular, opening trade with countries which are well-endowed with knowledge may lead to beneficial transfers of technology. And agreements which cause trade to be diverted away from such countries, adversely affect growth. While MFN liberalization would not divert trade away from technologically attractive partners, non-discrimination is not optimal – it would actually be optimal to subsidize entry by firms which generate larger positive externalities. However, there are two reasons why committing to MFN treatment may nevertheless be a sound strategy. First, the government may find it difficult to prejudge the sources from which firms are likely to generate the greatest spillovers. Second, it is likely that the most competitive suppliers (in terms of costs and/or quality) also generate the greatest positive externalities.¹⁵

Policy implication:

- The gains from increased competition and exploitation of scale economies, as well as the possibilities of inducing knowledge spillovers, strengthen the presumption that a country would gain from a preferential agreement in services. However, each of these arguments provides even stronger support for non-discriminatory liberalization.

2. The Economics of International Cooperation

2.1 Gains at the expense of the rest of the world

There are certain circumstances in which a single country or a group of countries may derive greater benefits from a preferential arrangement than from a multilateral

¹⁴ Coe, Helpman and Hoffmaister (1997). See also Lumenga-Neso, Olarreaga and Schiff (2001).

¹⁵ However, the correlation of quality of a service from which the consumer benefits and the positive spillovers that are generated may create a dilemma for a poor country. Because consumers may prefer the cheaper low quality service even though the more expensive high quality service is socially preferable.

arrangement. First of all, any one country may benefit if its trading partner offers more profitable reciprocal preferential access than is available multilaterally. Thus for instance, if the European Union were willing to offer a developing country unrestricted access to its protected agricultural market in return for preferential access to that country's services market, then this could be more attractive for this country than a multilateral arrangement with more general reductions in European agricultural protection.

However, it is more difficult to think of circumstances where each country in a preferential arrangement would be better off than if each were to choose non-discriminatory liberalization. One possibility that could arise in competitive markets is that participant countries realize terms of trade gains vis-à-vis the rest of the world.¹⁶ In imperfectly competitive markets, there may be collective gains if the arrangement makes it possible to shift rents away from third countries. Finally, as noted in Section 1.7, there may be gains over time, if there are dynamic economies – for instance, a larger protected regional market offers the opportunity to learn by doing and develop comparative advantage in new areas.

2.2 Efficient bargaining

There is inherently a tension between reciprocity-based bargaining and the most favored nation (MFN) principle. In certain circumstances, a regional approach may be preferable because it facilitates more efficient bargaining. This argument has been made in the context of goods negotiations.¹⁷ Reciprocity-based bargaining has so far played less of a role in services negotiations, but in principle the argument is also relevant to services.

The MFN principle is not simply a rule that constrains trade discrimination, but also a rule that influences the multilateral bargaining process. Say a country is negotiating bilaterally with the European Union in the WTO. It knows that if it grants improved access to its market to the EU, then the MFN rule would oblige it to offer the same level of access to all other countries without obtaining anything in return. At the same time, the country also knows that any improved access it obtains from the European Union will automatically be available to other exporters who have not given the EU anything in return. The temptation to be one of those other countries would be strong.

More generally, the problem is that countries may try and free ride on each other during bargaining. Each of the beneficiaries of a concession from a trading partner may be tempted to understate their willingness to pay for it, hoping that offers of reciprocal concessions from other countries will be sufficient to induce the concession. If each country behaves in this way, the result could be that mutually beneficial deals will not be struck. But it is not clear how well-founded these fears of free-riding are in practice.

¹⁶ See Panagariya (2000), Winters (1996) and the recent papers by Chang and Winters (forthcoming) and Schiff and Chang (forthcoming).

¹⁷ This section draws upon Schwartz and Sykes (1996). But see also Wonnacot and Wonnacott (1981) for an early depiction of the rationale for preferential agreements, and Levy (1994) and Krishna (1998) for a political economy -based analysis.

Nevertheless, in so far as free-riding is a real problem, two possible solutions exist: for beneficiaries to form a coalition to work out a collective offer for a desired concession. For example, consider agriculture. Production subsidies play a significant role, and the benefits of any reductions in such subsidies are necessarily extended on an MFN basis. This may be one reason for the emergence of the Cairns group of agricultural exporting countries. In services, the European Union and the United States could be a possible coalition. Together they account for the bulk of world services exports.¹⁸ The second solution to the free-riding problem is to abandon product-by-product negotiations and agree instead that every country will reduce barriers according to a mutually agreed formula. This approach may, however, be difficult in services since the instruments of protection do not lend themselves to quantification.¹⁹

However, because there is no complete solution to the free-rider problem, it is likely that the MFN principle prevents the parties in large multilateral trade negotiations from exhausting all joint political gains and makes the optimum bargain unattainable. *Ceteris paribus*, the desirability of discriminatory arrangements would be greater, the more serious the free-rider problem for non-discriminatory arrangements. Under these circumstances, preferential trading arrangements may make some political sense as a second best alternative.

However, apart from the efficiency costs of departing from the MFN principle described in the previous section, there is another concern raised by preferential arrangements: they tend to undermine the security of trade concessions. A country's willingness to pay for a concession may be undermined if it fears that its partner might subsequently grant preferred access to someone else. Faced with this possibility it would offer less for the concession in the first place and fewer mutually beneficial deals would be struck. The MFN obligation protects the value of any negotiated concessions against future erosion through discrimination and thus brings benefits for the bargaining process.²⁰

Policy implication:

¹⁸ In the year 2000, the shares of the EU and US in world exports of commercial services on a balance of payments (BOP) basis were around 40 percent and 20 per cent, respectively. Their shares are probably even greater once FDI is taken into account.

¹⁹ Sapir (1998). Some have tried to develop formulae for services negotiations but it is doubtful that they will be used (Thompson, 2000).

²⁰ A blanket MFN provision is of course not the only solution to the problem. Each party to a trade agreement could secure in conjunction with the concession additional promises limiting the future concessions that a trading partner could subsequently make in the same domain to other trading partners. However, the transaction and information costs of proceeding in this fashion would be high. Alternatively, members could confine all concessions to well defined multilateral negotiating rounds - as is generally the case, with the exception of regional initiatives. No nation would commit to anything until it could review the entire schedule of proposed discriminatory concessions by each of its trading partners. These approaches offer the prospect of more focused reciprocity-based bargaining. But they compare unfavourably with the simple MFN rule in terms of ease of administration and the maximization of joint surplus.

- As a country undertakes negotiations in multiple fora, it must assess the costs and benefits of negotiating in the WTO in the knowledge that all concessions will be multilateralized; of negotiating outside the WTO separately with its main trading partners in the knowledge that concessions can be preferential; and of doing both as at present. In this assessment, it must weigh the advantages of the MFN principle in ensuring economic efficiency and safeguarding the value of concessions, against its disadvantages in terms of inhibiting mutually beneficial bargains because of the free-rider problem.

2.3 Regulatory cooperation: The role of international rules

The economic case for regulation in services, as in the case of goods, arises essentially from market failure attributable to three kinds of problems, asymmetric information, externalities, and natural monopoly or oligopoly. In the first two cases, national remedial measures can themselves become an impediment to trade; in the third case, it is the absence of national regulation that can create trade problems. In order to ensure that domestic regulations at home and abroad support trade, a country must decide on the appropriate forum (multilateral, regional) and the approach (international rules, mutual recognition or harmonization) to pursue in each service sector.

First of all, in dealing with trade impeding domestic regulations (like standards, qualification, and licensing requirements), international rules could deepen the basic non-discrimination obligation by creating a variant of the so-called “necessity test”.²¹ While this test has been given differing interpretations, it may be possible to provide an economic interpretation.²² Economic principles can provide meaningful rules for the choice of instruments. Ideally, state intervention is meant to remedy distortions. The “first-best” instrument is the one which attacks the divergence between the private and the social cost at the source. For instance, in the case of professionals like doctors, a requirement to re-qualify would be judged unnecessary, since the basic problem, inadequate information about whether they possess the required skills, could be remedied by a less burdensome test of competence. In terms of the earlier model, the role of international rules could be to ensure that (a) the v component (where a difference in quality or incompatibility is the issue) is not exaggerated at the expense of the u component (which reflects a certain universal requirement); and (b) that the measures necessary to ensure compliance with the u and v components are not unduly trade restrictive.

The second way in which international rules have taken sought to deepen the principle of non-discrimination is by creating *pro-competitive regulatory principles* where monopolistic control of essential facilities may impede market access. These might pertain to regulating interconnection prices to telecommunications networks (as under the GATS Telecom Reference Paper) or to the allocation of airport slots or to administering

²¹ See Mattoo (2000).

²² For instance, in the WTO context, more emphasis has been placed on WTO-consistency or “least trade-restrictiveness” of measures rather than on the economic efficiency of measures.

deposit insurance schemes. Note that international differences in regulatory standards need not inhibit trade if they conform to certain broad principles. The fact that country A's approach to the regulation of interconnection differs from that of country B does not by itself limit the ability of an operator from B to supply services in A. However, greater similarity of regulatory approaches is likely to make life easier for international traders and investors.

However, there are limits to what can be accomplished through international rules. First, national regulators may object to what they regard as excessive international scrutiny of their judgments regarding whether certain universal requirements (the *u* component) have been fulfilled. And more seriously, there is little that international rules can do to address differences across jurisdictions in country-specific requirements (the *v* component) pertaining to: *vertical* standards, where differences between countries arise because of different quality standards, such as for instance, in the training of doctors or supervision of banks; and *horizontal* standards, where differences between countries arise because of incompatibility, e.g. because of differences in their legal systems or network standards in communication and transport services. Differences may reflect national preferences for certain levels of quality or particular varieties, or simply be a legacy of history.

Policy implication:

- The institution of some variant of a necessity test in international trade agreements would enable exporters, e.g. of professional and construction services, to challenge unnecessary regulatory barriers abroad. At the same time, the ability of foreign suppliers to challenge excessively burdensome domestic regulations, e.g. in professional services, would ensure that domestic regulations serve legitimate objectives rather than protectionist interests, and hence create benefits for domestic consumers. International rules on access to essential facilities, such as ports and telecommunications networks, provide a similar dual benefit. But there are limits to what can be accomplished through international rule-making, and there will be need for further regulatory cooperation.

2.4 Regulatory cooperation: harmonization and mutual recognition

As noted above, international rules can do little to address impediments to trade arising from fundamental differences across countries in standards. In such circumstances, two approaches are usually proposed: harmonization and mutual recognition. Even though these approaches are sometimes presented as alternatives, the former is either a precondition or a result of the latter. Where differences in mandatory quality standards matter, recognition can only happen once there is a certain degree of harmonization (to establish mutually acceptable minimum standards). Where differences in mandatory standards are so narrow that they do not matter, recognition can be granted and then harmonization happens via a race to the minimum standard.²³ A similar logic applies to

²³ Of course differences in quality standards may well persist as firms produce different qualities to suit different individual tastes, but there would be little logic in maintaining different *mandatory* quality standards once there is mutual recognition.

compatibility standards, but there may be no alternative to full harmonization if differences matter, as for instance in the case of differences in railway gauges and legal procedures.

There is little, if any, empirical guidance on the payoffs to regulatory cooperation. What are the costs and benefits of deeper harmonization of regulatory standards and/or the establishment of mutual recognition agreements? The lack of empirical evidence complicates the task of deciding on the scope and depth, as well as the geographical reach and the institutional form of cooperation. Nonetheless, several conceptual considerations may assist a country in formulating a strategy for regulatory cooperation.

First of all, if national standards are not optimal, then international harmonization can be a way of improving national standards – as has happened with the Basle accord on financial regulation. In such situations, the best partners for regulatory cooperation are those with the soundest regulatory framework. Secondly, sometimes standards are captured by protectionist interests, in which case harmonization can serve as a purely liberalizing device. Third, if standards are separable, in the sense that it is possible to have one standard for export and another for domestic consumption, then regulatory cooperation is less of a challenge. But standards may not be separable: economies of scale may make it prohibitively expensive to have separate standards for export production; or a trading partner may make the adoption of country-wide higher standards (of financial regulation, privacy) a condition for trade.²⁴

If national standards optimally serve national objectives and are not separable between markets, there is a trade-off between the gains from integrated markets and the costs of transition and of departing from nationally optimal standards. For instance, a poor country may prefer to maintain a low mandatory standard for certain services, because that reflects the socially optimal trade-off between price and quality whereas the socially optimal trade-off in the rich country leads to a higher standard. Harmonization of standards would create benefits in terms of increased competition in integrated markets (as discussed earlier), but would necessarily impose a social cost in at least one market.

The aggregate adjustment cost of harmonization depends on the distance between the policy-related standards of the countries. The costs are likely to be smallest when foreign regulatory preferences are similar and regulatory institutions are compatible. The benefits of eliminating policy differences through harmonization depend on the prospects of creating a truly integrated market, which depends on the natural ties between countries, and that in turn depends on geographic proximity, legal systems and language, etc. We can conceive of an *optimum harmonization area* that defines the set of countries for which aggregate welfare would be maximized by regulatory harmonization.²⁵ However,

²⁴ Mattoo (2000). A similar concern can also arise in the design of pro-competitive regulatory principles. International trade and investment may be facilitated by the detailed harmonization of regulatory regimes, yet there may be a cost if the regulatory measures set out in the international agreement do not coincide with what a country would have chosen in a purely domestic context.

²⁵ In the definition of an optimal harmonization area, it must also be recognized that cooperation forums can be a vehicle to exchange information on different experiences with regulatory reform and to identify

whether an individual country benefits from harmonization, and its willingness to participate in such an area, depends on where the standard is set - which determines who will bear the costs of transition.²⁶

This raises an important question: should regulatory cooperation take place in each sector among spontaneous sets of self-selected countries; or should the group of countries be determined ex ante, and then this group pursues deeper integration across each sector? E.g. should a regional agreement pursue deep integration internally in all sectors, or should sectoral regulatory cooperation be pursued on a cross-cutting basis among countries with like regulations? Intuitively, the latter would seem to be the preferable course. It is certainly true for countries that have identical regulations or whose regulations differ in a way that does not matter. And an optimum harmonization area is likely to differ across services sectors, suggesting that it may be desirable for a country to cooperate with certain countries on a narrower set of quality standards compared to other countries.

But it is conceivable that an optimum harmonization area in a particular services sector would not satisfy the participation constraint for each country: i.e. group welfare would be maximized by setting the standard at a level at which it is not attractive for some countries to participate. For example, say countries in a particular region as a whole would benefit from the adoption of the same educational system, but a particular country would lose from moving to the ideal harmonized standard. And in some other sector, say transport services, that same country would gain a lot from harmonized regulations but some other countries would lose. Therefore, in the absence of a sectoral compensation mechanism, it might be easier if a number of such sectoral negotiations were bundled together within a group whose composition was fixed in advance, i.e. a horizontal-agreement.

Policy implication:

- There are gains for a country from regulatory cooperation, but also costs. The former will dominate where national regulation can be improved, as in the case of financial services, or is excessively burdensome in all countries, as in the case of professional services. Once national regulations are optimal, the benefits of international harmonization in terms of greater competition in integrated markets must be weighed against the costs of departing from nationally desirable regulations. Finally, a country would gain from pursuing regulatory cooperation on a sectoral basis as widely as feasible, but also exploit the scope within cross-sectoral agreements to overcome sector-specific constraints to cooperation.

good regulatory practices. This form of cooperation can be especially useful for regulating new services in sectors with continuous technical change. Developing countries may then have an interest in cooperating with advanced industrial countries that have the longest experience with regulatory reform and/or where the newest technologies are often introduced first.

²⁶ We can think of this issue being the outcome of bargaining. The incentive to harmonize may depend on the relative market size, with the small country having more to gain.

2.5 A caveat: The role of sunk costs

An important question is whether the sequence in which regional and multilateral liberalization occur matters. At least three insights have been presented in the literature, each related to the importance of market-specific sunk costs, discussed in Section 1.6. One is that the prospect of a regional integration agreement prompts investments geared towards supplying the markets of partners in the regional agreements. These investments have a “lock-in” effect and for a small country can substantially reduce bargaining power in the subsequent negotiations (McLaren, 1997). Secondly, once such investments are made, multilateral liberalization may become less attractive for a country than regional liberalization (McLaren, 1997, 1999). These arguments are less relevant when regional and multilateral negotiations are proceeding in parallel, and, therefore, current investment decisions are unlikely to be based on expectations of one or the other outcome.

The third insight is that in oligopolistic markets, a regional route to multilateral liberalization may offer greater benefits to participants than direct multilateral liberalization (Freund, 2000). The intuition is that initial privileged access to each others markets provides firms with a first-mover advantage and an opportunity to commit to high bilateral exports by creating larger distribution networks (a sunk cost) so as to limit future exports by outside firms. As a result, member nations earn higher profits in each others markets from the regional path than the multilateral path – and are no worse off in third markets.²⁷

However, this model is symmetric in terms of the importance of market-specific sunk costs. Consider a somewhat different and perhaps more relevant scenario: one country has a comparative advantage in goods, and the other in services. Both can choose to liberalize MFN, or to sequence liberalization: preferential followed by MFN.

As argued in Section 1.6, location-specific sunk costs are arguably more important in services like professional, telecommunications, finance and distribution, than in goods, so even temporary privileged access for an inferior supplier can translate into a long-term advantage in the market. Thus, while the elimination of preferences may lead to a relatively painless switch to more efficient sources of goods supply, the entry of more efficient service providers may be durably deterred if their competitive advantage does not offset the advantages conferred by incumbency. This is an important consideration for a country, which is likely to export goods and import services. The benefits of obtaining preferential access for that country may be small and temporary, while the costs of granting preferential access could be large and durable. How much this matters depends on whether the provider who benefits from preferences is the most efficient provider globally.

Finally, regulatory harmonization itself could involve sunk costs of transition. The sequence in which a country chooses to harmonize its regulations with different trading partners is not irrelevant. One reason is that the sequence of harmonization may influence

²⁷ In the model, world welfare turns out to be higher because the gains to the original members are greater than the loss to the excluded countries.

the bargaining power of different country groupings in the negotiation over where the harmonized standard should be set.

Policy implication:

- If a country's main export interest is in agricultural and manufactured goods, where location-specific sunk costs are relatively small, and it is likely to import services, where sunk costs are relatively important, then it must also consider the implications of current preferences on the gains from eventual multilateral liberalization.

3. Conclusion

The welfare implications of regional liberalization in goods trade are well-understood—benefits from greater intra-regional competition may be offset by so-called trade diversion if policies discriminate against more efficient non-regional suppliers. These welfare implications cannot be straightforwardly extended to the realm of services.

Compared to the status quo, a country is likely to gain from preferential liberalization of services trade at any particular point of time—in contrast with the more ambiguous conclusions emerging for goods trade. In goods trade, where tariffs are the main instrument of protection, the trade-creating benefits of preferential liberalization could be offset by the costs of trade diversion due to the loss in tariff revenue. However, in services, trade barriers do not generate revenue for the government - import tariffs are rarely used to restrict trade - and barriers are often prohibitive. Consequently, there are few costs of trade diversion. Furthermore, as in the case of goods trade, the scope for increased competition and exploitation of scale economies, as well as the possibility of inducing knowledge spillovers, strengthens the presumption that a country would gain from a preferential agreement in services.

But, other things equal, non-preferential liberalization is likely to produce larger gains than regional agreements that imply preferential liberalization. Nondiscriminatory liberalization is superior because it does not bias consumer choice, allowing imports from the most competitive source. It also leads to a less complex policy regime than a preferential arrangement and therefore implies lower administration costs for government agencies and lower transactions costs for the private sector.

Also, if a country initially takes the regional route, the benefits of eventual multilateral liberalization may be lower. This is because location-specific sunk costs of production are important in many services that a country is likely to import, so even temporary privileged access for an inferior supplier can translate into a long-term advantage in the market. While the elimination of preferences for trade in goods may lead to a relatively painless switch to more efficient sources of supply, the entry of more efficient service providers may be durably deterred if their competitive advantage does not offset the advantages conferred by incumbency.

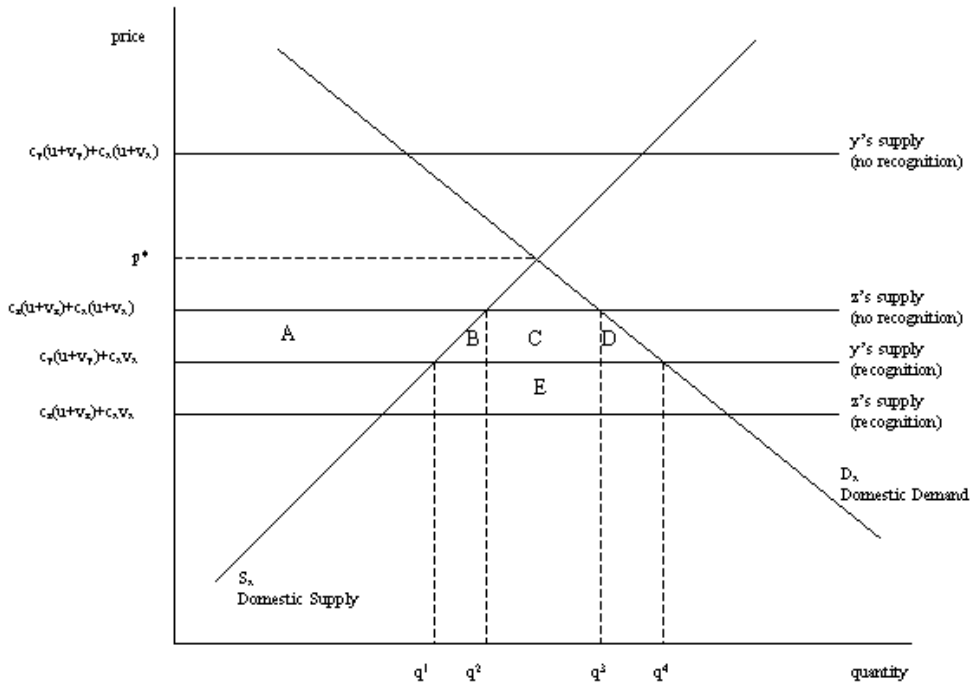
A case can, nevertheless, be made for engaging at the regional level because there is a greater possibility of striking mutually beneficial bargains. More efficient bargaining may be possible in a plurilateral setting than in the multilateral context; because there is less concern that outsiders will be able to free-ride on the reciprocal exchange of concessions than if there were a general obligation not to discriminate between all trading partners. Political economy pressures may also lead to better access to foreign markets in a preferential context. For example, if exporters gain more from preferential access to markets abroad than from multilateral liberalization, then they may be a stronger countervailing force against protectionist interests at home.

Regulatory cooperation is the one area where it is possible to support a regional approach without qualification. Deeper regulatory cooperation, in the form of harmonization and mutual recognition, naturally takes place among a subset of similar countries. Ideally, countries choose their partners spontaneously sector by sector, depending on the costs and benefits of regulatory harmonization. But in certain circumstances, it may be desirable to choose partners *ex ante* in horizontal agreements and then seek to deepen integration across sectors.

Annexes

A. Preferential recognition: Variable cost-increasing measures

Assume that there are three countries: country X is an importing country, facing an upward sloping domestic supply of a particular service (say because of increasing opportunity costs) and countries Y and Z are potential exporters. Let us say that $c_x > c_y > c_z$, and that in the absence of any trade the prevailing price in country X is p^* . Assume now that country X recognizes the equivalence of the universal component of the standard obtained in country Y, but not in country Z. There are several possibilities, but we consider only two:



No trade prior to recognition: Prior to recognition, service providers from both countries Y and Z were required to meet the full standards in country X but neither found it worthwhile to do so, i.e. $c_y(u + v_y) + c_x(u + v_x) > p^*$ and $c_z(u + v_z) + c_x(u + v_x) > p^*$.²⁸ But $c_y(u + v_y) + c_x v_x < p^*$, i.e. when the universal

²⁸ It may be asked why some individuals from countries Y and Z do not directly meet the standard in country X instead of first qualifying at home. One reason is that providers are often allowed to enter foreign markets only on a temporary basis (as under the GATS), so they need to be qualified also to serve the home market. Where longer term movement of providers is allowed, we need to assume that the individual elements of qualifications are not separable, and that there is a part, say w , which is universally recognized as equivalent. The incentive to meet the standard at home arises as long as $c_y w$ and $c_z w$ are sufficiently smaller than $c_x w$. The w element is suppressed here to keep the notation simple. Non-separability is indeed an aspect of many qualifications: a student can usually not switch institutions after doing only part of a course.

component of standards in country Y is recognized as equivalent to that in country X, then suppliers from Y find it worthwhile to export to country X, and the price in X will fall. Hence, if all foreign suppliers had been completely deterred from supplying to country X by the absence of recognition, then any recognition agreement is necessarily trade creating.

Country Z exports to country X prior to any recognition: In this case, $c_z(u + v_z) + c_x(u + v_x) < p^* < c_y(u + v_y) + c_x(u + v_x)$, i.e. when all foreign suppliers were required to requalify, only those from the third country were willing to supply country 1. But $c_y(u + v_y) + c_x v_x < c_z(u + v_z) + c_x(u + v_x)$, i.e. once the second country suppliers are exempted from the basic qualification requirement, they gain a competitive edge over the third country. That is, if suppliers from the third country were already present in the first country, then the recognition of second country suppliers would put them at a competitive disadvantage and could lead to trade diversion.

This situation is depicted in Figure 1.²⁹ The pre-recognition situation involves domestic output q_2 , consumption q_3 , and imports from Z, $q_3 - q_2$. After country X recognizes qualifications in country Y, domestic output declines to q_1 , consumption increases to q_4 , and imports from Y, $q_4 - q_1$, displace imports from Z. The welfare effects are straightforward. Consumer surplus rises by $A + B + C + D$ as consumption expands from q_3 to q_4 . The area A is a gain at the expense of domestic suppliers, whose surplus falls with their output. Area D is the gain from the better allocation of consumption expenditures; area B the gain from the resources released as inefficient domestic supply contracts; and area C the gain arising from the elimination of wasteful requalification.

The area $C + E$ is of crucial significance, and can be interpreted in several different ways. It helps to recall the analysis of preferential arrangements when tariffs are the instruments of protection. In that case, area $C + E$ would be the loss in government revenue because preferential imports displace high tariff imports. While C is gained by consumers, E is completely lost because supply comes from the more expensive source, and is the loss due to trade diversion. The net gain to the country is only $B + D - E$, and could be positive or negative. In the example here, $C + E$ were the costs of requalification for country Z suppliers when they supplied country X. If these costs were completely dissipated, then they do not enter the welfare calculus of country X. That is, *there is no cost of trade diversion for the importing country* and the net gain to the importing country from the recognition agreement is $B + C + D$. If, however, part of these requalification costs (say a fraction α) were appropriated by country X, perhaps as the producer surplus of its qualification granting industry or as some form of regulatory rent, then they would be foregone with trade diversion and would need to be taken into account: the net gain to the importing country from the agreement would be $B + C + D - \alpha(C + E)$.

B. Preferential recognition: Fixed cost-increasing measures

Following Baldwin (2001), we consider an industry marked by Cournot oligopolistic competition with firms facing constant marginal costs and two types of fixed costs, a firm-specific fixed cost of setting up production (unrelated to policy) and a fixed cost of selling to each market (related to policy). The three countries – home (X), partner (Y) and rest of world (Z), are assumed ex ante identical to reduce complexity. The inverse demand function of a typical nation is $p_j = 1 - Q_j$, where p_j and Q_j are the price and total sales in market j ($j = X, Y$ or Z). The total quantity in market j is $(n_j q_{ij})$ where q_{ij} is sales of a typical i -based firm in market j ; n_j is the corresponding number of i -based firms. Marginal production costs are initially assumed to be identical. The three markets are assumed to be segmented – which is plausible where cross-border delivery is not feasible. Domestic regulations impose an additional fixed cost on non-local firms. To capture this simply, we assume that each market has its own norm and complying with these costs F in each market.

²⁹ The initial part of the following discussion resembles closely the discussion in Pelkmans and Winters (1988).

The equilibrium price in market X, p_X , equals the sum of two terms.¹⁶ The first term, $1/(1+n_i)$, reflects the level of overall competition. The second term, $(n_i c_{iX})/(1+n_i)$, reflects the average marginal cost of firms active in the market (c_{iX} is the market-specific marginal cost). It is assumed that the number of X and Y firms are identical, n , and n^* is the number of Z firms. The number of firms operating in the domestic market is allowed to vary with policy. Let us say that there are no other restrictions on new entry in the country. Then, in equilibrium each firm's profits will just cover its fixed costs. It is intuitively obvious (and easy to show) (i) that the equilibrium number of X and Y firms, i.e. n , falls as their fixed cost F rises, and (ii) that n rises as the Z-firms' fixed cost F^* rises. Similarly, the equilibrium number of Z firms, i.e. n^* , falls as F^* rises and as F falls.

¹⁶ The equilibrium is found by solving the nine segmented market first order conditions for the nine levels of sales. The model's linearity allows us to find solutions for prices, consumption, welfare, the numbers of firms, trade flows, etc.

References

- Baier, S. L., and Bergstrand, J.H. (2001). International Trade in Services, Free Trade Agreements, and the WTO, Chapter 7 in R. M. Stern (ed.), *Services in the International Economy*, University of Michigan Press, Michigan.
- Baldwin, R.. (2000). Regulatory protectionism, developing nations and a two-tier world trade system. Centre For Economic Policy Research (CEPR) Discussion Paper Series (October) 2574:1-39. London: CEPR.
- Beesley, M. E. (1997) "Privatization, regulation and deregulation", Institute of Economic affairs, (2nd edition), (Ch. 10 Entry to the airline industry), p. 168-187.
- Bhagwati, J. (1993). "Regionalism and Multilateralism: An Overview", in A. Panagariya and J. de Melo (eds.), *New Dimensions in Regional Integration*, Washington, D.C., World Bank.
- Breshnan, T.F. and Reiss, P.C., "Measuring the importance of sunk costs", *Annales d'economie et de statistique*, No.31, 1993.
- Chang, W. and Winters, L.A. (forthcoming). "How Regional Blocs Affected Excluded Countries: The Price Effects of MERCOSUL," *American Economic Review*.
- Coe, D.T., E. Helpman, and A.W. Hoffmaister. (1997). "North-South R&D Spillovers," *Economic Journal*, 107.
- Frankel, J., E. Stein and S. Wei. (1995). Trading Blocs and the Americas: The Natural, the Unnatural and the Supernatural, *Journal of Development Economics*, 47, 61-96.
- Freund, C. (2000). "Different Paths to Free Trade: The Gains from Regionalism." *Quarterly Journal of Economics*, pp.1317-1341.
- Gual, J. (1999). "Deregulation, Integration, and market structure in European Banking," *Journal of the Japanese and International Economies* 13, p. 372-396, 1999.
- Krishna, P. (1998). "Regionalism and multilateralism a political economy approach." *Quarterly Journal Of Economics* (February) 113:227-51.
- Levy, P. I. (1994). "A political-economic analysis of free trade agreements." *American Economic Review* (September) 87:506-19.
- Lumenga Neso, Olivier, Marcelo Olarreaga and Maurice Schiff. (2001). "On Indirect Trade-Related Research and Development Spillovers," Policy Research Working Paper #2580. Washington, D.C.: The World Bank.

- Markusen, J.R. (1987). Trade in producer services and in other specialized intermediate inputs. *American Economic Review* 79:85-95.
- McLaren, J. (1997). Size, Sunk Costs, and Judge Bowker's Objection to Free Trade. *American Economic Review* (June) 87:400-420.
- McLaren, J. (1999). A Theory of Insidious Regionalism. Department of Economics, Columbia University. June, 1999. Mimeo.
- Panagariya, A. (2000). Preferential Trade Liberalization: The Traditional Theory and New Developments, *Journal of Economic Literature* 38(2): 287-331.
- Pelkmans, J. and Winters, A. (1988), Europe's Domestic Market, Chatham House Papers 43, Royal Institute of International Affairs, Routledge, London.
- Sapir, A. (1998). GATS 1994-2000, *Journal of World Trade*, 33, 51-66.
- Schiff, M. and W. Chang. (forthcoming). "Market Presence, Contestability, and the Terms-of-Trade Effects of Regional Integration," *Journal of International Economics*.
- Schartz, W.F. and Sykes, A.O. (1996), Toward a Positive Theory of the Most Favoured Nation Obligation and Its Exceptions in the WTO/GATT System, *International Review of Law and Economics*, 16, 27-51..
- Thompson, Rachel. (2000). "Formula Approaches to Improving GATS Commitments", in Sauvé, Pierre Stern, Robert M. (2000), 473-486.
- Tirole, J. (1998). "*The Theory of Industrial Organization*." MIT Press, Cambridge, MA.
- Whinston, M. D. and S. Collins. (1990). Entry Contestability and Deregulated Airline Markets: An Event Study Analysis of People Express" by, and, Working Paper No. 33187, NBER, April 1990.
- Winters, L.A. (1996). Regionalism versus Multilateralism, World Bank Policy Research Working Paper, 1687, World Bank, Washington, D.C.
- Wonnacott, P. and R. (1981). Is Unilateral Tariff Reduction Preferable to a Customs Union? The Curious Case of the Missing Foreign Tariffs. *The American Economic Review* (September) 71(4): 704-714.
- World Bank. (2000). *Trade Blocs*, Oxford University Press, New York.