2. Regional integration and foreign direct investment: theory and lessons from NAFTA

Lorraine Eden

INTRODUCTION

One of the strongest worldwide trends in trade policy in the past 20 years has been the rebirth of regional integration schemes, or preferential trading agreements (PTAs). These agreements first became fashionable in the 1960s, after the 1957 Treaty of Rome created the European Economic Community (EEC). Many of the ‘first wave’ of regional integration schemes were started in Latin and South America, but languished in the 1970s. In the mid-1980s, the desire to deepen European integration with the EC1992 project regenerated interest in a ‘second wave’ of regional integration schemes (Dunning 1997b; Serra and Kallab 1997). By 1996, there were in excess of 100 PTAs around the world, more than double the number in 1985 (Bhagwati and Panagariya 1996). The ‘first wave’ PTAs (for example, CARICOM) were revitalized; others (for example, MERCOSUR) were newly created.

Regional integration is a topic that has attracted interest from many disciplines: international economics, international politics, international trade law, and more recently, international business. There is a large and rich literature on the economics of regional integration that stretches back to the early 1950s and the work of international trade economists such as Jacob Viner (1950), Richard Lipsey (1960) and Bela Balassa (1961). Much of the early international trade literature focused on types of preferential trading arrangements and their static and dynamic effects on international trade and national welfare.¹

International business (IB) scholars first became interested in regional integration by examining the impacts of the formation of the EEC on foreign direct investment (FDI) from the United States (Dunning 1988). More generally, IB scholars have been interested in the way that multinational enterprises (MNEs) have responded to the formation of PTAs, and, in turn, how their strategies have influenced the nature and pace of regional integration.² Interestingly, the research on regional integration by international economics and
international business scholars has proceeded on parallel but separate tracks, possibly because IB researchers have been more focused on firm strategies and FDI patterns, whereas international economics scholars have been more focused on trade patterns.

Where is the IB literature today on MNEs and regional integration? What lessons have we learned? What issues remain unexplored or controversial? In this chapter, we first briefly review the international economics literature on regional integration theory and then move to the IB literature on the strategic responses of MNEs. A review of empirical work on MNEs and regional integration, focusing on North America, follows. The chapter concludes with an outline of several possible new directions in IB research on MNEs and regional integration.

THE THEORY OF REGIONAL INTEGRATION

Regional integration schemes, in theory, have four general economic effects. The first set of effects are the short-run welfare gains that come from improved specialization of resources and greater opportunities for exchange within the region. These are known as the static gains from trade. The second set of effects are the long-run welfare gains, the dynamic gains from trade, that come from exploiting region-based economies of scale and scope, attracting inflows of foreign direct investment and technology transfers, and greater competition among firms in national markets. The third set of effects are the transitional costs that fall, in the short run, on inefficient sectors and immobile factors as firms rationalize and reallocate their activities throughout the region as they respond to regional integration. Lastly, greater economic interdependence within the region is likely to occur in response to rising interregional linkages created by trade and investment flows. Greater interdependence means more sensitivity and vulnerability to instabilities, but also creates additional potential gains from the multiplier effects of economic linkages with other member countries.

The size of the effects depends on several factors, the most important of which are probably the scope of the PTA in terms of number of member countries, industries and products covered, the degree of liberalization of tariff and nontariff barriers among the member countries, and the current and potential economic complementarity of the member countries relative to non-members. We explore these below.

Shallow versus Deep Integration

Preferential trading arrangements vary significantly in terms of how shallow or deep is the integration process among the member countries (UNCTC
1993). *Shallow integration* schemes normally involve little more than lowering or removal of tariff barriers among PTA members, whereas *deep integration* schemes involve significant removal and/or harmonization of nontariff barriers in addition to tariffs (UNCTAD 1993: 35). At the 'shallow end' of the PTAs are *free trade areas* (FTAs) where tariffs are removed against member countries, but members keep their own external tariffs against non-members and no removal or harmonization of nontariff barriers is required. Rules of origin, which define the necessary amounts of regional content required to qualify for duty-free entry, are used to prevent non-member country products from coming in through the lowest-tariff country and moving freely inside the FTA afterwards.

Since more coordination among member countries is required for a *customs union*, where a common external tariff against non-member countries is added to the zero internal tariff requirements, a customs union is deeper than a free trade area. A common market is deeper again, since removal and/or harmonization of nontariff barriers that restrict factor mobility (in particular, mobility of labor and capital) is an added requirement. The last regional integration scheme before the 'deep end' (full political union, as in the United States of America) is an economic union, where the member countries, in addition to the requirements of a common market, adopt common monetary and fiscal policies and a common currency.

Effective deep integration removes all intra-regional barriers that discourage the efficient allocation of international production within the PTA. This includes elimination of barriers to trade in business services, right of establishment and fair treatment for foreign direct investment (FDI), and protection of intellectual property. UNCTC (1993) argues that governments press for shallow integration but deep integration comes from the pressures of multinational enterprises to remove intra-regional impediments to the flow of goods, services, intangibles, capital, and people.

Deep integration occurs in two ways; first, through the extension of the GATT norm of *national treatment* (foreign activities performed within a country’s borders receive the same treatment as activities of nationals), to intra-regional flows of investment, services, and intellectual property. National treatment means that a country treats foreign activities performed within its borders in the same way as it treats domestic activities. Foreign goods, services, and investments are treated the same as domestic goods, services, and investments, once they have cleared customs and become part of a country’s internal market.

However, deep integration requires that countries go further than simply national treatment. Greater *policy coordination and harmonization* in specific areas takes place as governments harmonize and coordinate a variety of domestic policies and adopt common standards in various fields that are not
directly trade related but do affect multinational enterprises. Removal of internal barriers facilitates the exploitation of economies of scale and scope at a regional level through the location of MNE plants where they are most efficiently located.

Macro-regions and Micro-regions

Ethier (1998: 1150) argues that the ‘second wave’ of regional integration schemes in the 1980s and 1990s is a ‘new regionalism’ with specific characteristics that differentiate it from old regional schemes such as the EEC. First, the new regionalism typically involves differences in country size, as one or more small countries joins with a large country; NAFTA (North American Free Trade Area) is an example. Second, before entry, the small countries have unilaterally reformed their trade regimes, and, after entry, liberalization is primarily by the small countries. Third, regional integration involves neighboring countries, creating what Dunning (2000) calls ‘macro-regions’. Lastly, regional integration often involves deep integration as members harmonize and liberalize nontariff barriers among themselves.

The spread of regional integration schemes in the 1980s and 1990s has meant that cross-border barriers to trade and FDI flows within a PTA have fallen relative to barriers against non-member countries. As a result, the movement of capital, people, goods, and services within regions is probably freer than it has been since the late 1800s. Over the same period, international institutions have strengthened their regulation of cross-border flows, most notably international trade flows, with the lowering of tariff and nontariff barriers (NTBs) and the creation of the World Trade Organization in the Uruguay Round. The rapid spread of bilateral investment treaties and tax treaties also suggests that an international regime for investment is forming around the national treatment standard (Eden 1996).

Dunning (2000) argues that the liberalization of trade and FDI regulations at the regional level is creating ‘macro-regions’ as spatial entities. Some authors (for example, Ohmae 1995) have suggested that the region is replacing the nation state as the key spatial economic unit, while others (Kobrin 1995) see globalization as an offsetting force to regionalism. At the same time, ‘micro-regions’ are developing within countries as clustering activities by firms create ‘sticky places within slippery space’ (Markusen 1996). How PTAs affect firms’ location decisions, in terms of both macro-regions and micro-regions, has attracted recent and substantial attention in the IB literature, which I explore later in this chapter.
Trade Creation and Diversion Effects

International trade economists have long studied the welfare impacts of PTAs, generally focusing on the customs union case, where the member countries erect a common external tariff (Lipsey 1960; Bhagwati et al. 1998). Because the benefits of PTAs are restricted to member countries, their formation has two conflicting effects on trade flows. Trade creation occurs when the reduction of trade barriers within the PTA causes a shift from higher cost producers to lower cost producers within the PTA. In this case, differences in comparative costs cause shifts in trade, production, and investment patterns that favor the lower cost producers and improve economic efficiency within the PTA. For example, if the lowest cost producer (country Z) of steel plate is a PTA member, then country X will benefit by joining a PTA with Z, even though X’s steel plate producers will suffer dislocations and their production will shrink. The gains to consumers in X outweigh the losses to X’s producers, so that trade creation causes a net welfare gain for country X.

Trade diversion, on the other hand, occurs when the PTA causes a shift to higher cost internal producers from lower cost external producers because the products of the external producers have become uncompetitive in the internal market. Before the PTA, both inside and outside countries faced the same tariff barriers; after the PTA, only the outside countries face the tariff barriers. Removal of the tariff barriers against the inside countries may give them a competitive advantage that diverts trade and production away from the most efficient producers (the outsiders) towards the less efficient – but inside and therefore advantaged – countries.

These effects work against one another: regional integration creates welfare losses since trade is not with the lowest cost producer, but provides welfare gains because the average level of tariffs has fallen. Whether or not the country contemplating the PTA is actually worse off, in welfare terms, depends on the cost disadvantage between the member and non-member countries as compared to the tariff savings from removal of the tariff within the PTA. Since trade creation and trade diversion effects will vary by product and industry, the net impact of the formation of a PTA on the welfare of member and non-member countries depends on many factors. The general presumption is that the more trade expands between two countries after the formation of a PTA and the less the negative impact on trade with non-member countries, the more likely that trade creation effects have dominated trade diversion effects.

Bhagwati et al. (1998: 1130) argue that 'trade diversion is not necessarily a negligible phenomenon in current PTAs'. Several empirical studies have found significant estimates of trade diversion. In addition, PTAs can lead to
endogenous trade diversion as member countries raise trade barriers against non-members. The authors also argue that the proliferation of PTAs has created a ‘spaghetti bowl phenomenon’ (1998: 1138) whereby numerous crisscrossing PTAs and tariff rates create a ‘who is whose’ problem that can also raise protectionism and reduce welfare levels.

MNE RESPONSES TO REGIONAL INTEGRATION

Macro-regions and FDI

The formation of a preferential trading area is a policy shock, a change in the environment facing firms inside and outside the PTA (Eaton et al. 1994a). We therefore expect firm strategies to change in response to this policy shock. Dunning (1993) identifies four basic motivations for foreign direct investment: market access, resource seeking, efficiency seeking, and strategic asset seeking FDI. The formation of a preferential trading area should lead to all four types of FDI as firms take advantage of the lower tariff and nontariff barriers within the PTA.

Rugman and Gestrin (1993b) argue that the FDI response depends initially on the country-specific advantages (CSAs) of each of the member countries and the region-specific advantages (RSAs) that will be available once the PTA is fully phased in. They use internalization theory and the concept of CSAs and RSAs to predict how MNEs would react to NAFTA, but their arguments apply to any preferential trading area. They argue that the ability of MNEs to internalize their own firm specific advantages (FSAs) and take advantage of the country and region-specific advantages determines MNE profitability, market share, and growth. Rugman and Gestrin argue that regional integration reduces the transactions costs (for example, tariffs) associated with intra-regional trade and creates more certainty for investment. How regional integration affects CSAs and RSAs, and their interaction with FSAs, should determine trade and investment responses. Where trade and FDI are complements, they predict that regional integration should increase flows in both directions. However, where transactions costs were the main factor driving FDI (that is, tariff jumping FDI), MNEs should substitute trade for investment, closing tariff factories, and centralizing production in the home market. Where strategic considerations, other than tariffs, are the main driver of FDI, the impact should be largely neutral.

Eden (1998: 166–8) argues that the likely reactions of firms to regional integration depend on the type of firm and its activities before the formation of the preferential trading area. Three types of firms are critical:
- **Insiders** – multinationals headquartered inside the region with significant investments across the region.
- **Outsiders** – foreign firms headquartered outside the region. These outsiders may be *traders* that export or import in the region or *investors* that have foreign investments inside the region.
- **Domestics** – local firms that are primarily focused domestically on their local/national market; they may be exporting to or importing from other member countries but do not have any foreign investments across the region.

Firms that are already established within the region and have affiliates in the member countries should view regional integration primarily as an opportunity. Insider firms should see benefits from lower intra-regional barriers and respond by rationalizing product lines (horizontal integration) and/or production processes (vertical integration) better to exploit economies of scale and scope across the region. That is, the primary response of insiders should be *efficiency seeking* FDI. There should be both a short-run response as MNEs engage in locational reshufflings in response to the falling trade barriers, and a long-run response where insiders locate, close and/or expand their plants with the whole regional market in mind. The result should be reduced numbers of product lines in various plants and increased horizontal trade among plants. MNEs are also likely to segment their production process among plants so that more vertical intra-firm trade takes place. Certain product lines, industry segments, and plant functions should shift within the region to the lowest cost location, causing job losses and plant closures in high cost locations.

Because regional integration normally leaves trade and FDI barriers against non-member countries unchanged, outsider firms exporting into the PTA may face trade diversion as insider firms receive preferential treatment. For defensive reasons, outsiders are likely to respond by investing in the region in order to protect their market access; that is, their defensive response is *market seeking (or protecting)* FDI. If privileged access depends on firm nationality, or if insiders hold complementary assets needed by the foreign firms, we also expect *strategic asset seeking (or asset augmenting)* FDI. Asset augmenting FDI is likely to be in the form of joint ventures or strategic alliances with insider firms. Outsiders that already have transplant operations in the region are likely to behave as the insiders, expanding and rationalizing their investments to take advantage of the larger regional market. Where just-in-time production is critical, these locational reshufflings will induce subsequent investments by upstream suppliers.

For local firms without established links to other PTA member countries, regional integration is both an opportunity (new markets, access to lower cost
inputs) and a threat (more competition). These firms, with encouragement, may start or increase their exports within the region and possibly open up distributors or offshore plants where market size or costs warrant. They will, however, have to face the difficult task of breaking into established distribution networks of domestics and MNEs already located in the other PTA countries. The key question is whether to ‘go regional’ and branch outside the home country into other parts of the region, or stay at home and become less competitive. The FDI strategies of domestics are therefore likely to be either market seeking or resource seeking (de novo FDI). If the domestics lack the complementary assets needed to penetrate other member country markets (for example, distribution networks), we also expect strategic asset seeking FDI.

When the PTA consists of countries with very different market sizes (in terms of population and income), another useful distinction can be made between firms headquartered in the hub economy (the largest market in the region) and in the spoke (small) economies (Eden and Molot 1993). If high trade barriers separate these countries prior to the formation of a PTA, one would expect firms to engage in tariff jumping FDI, setting up plants in each country to supply the local market. These so-called tariff factories were common in North America in the 1960s and 1970s, and in Mexico before NAFTA. Regional integration, in this case, is likely to lead MNEs headquartered in the hub economy to engage in efficiency seeking FDI, closing tariff factories in the smaller markets and supplying them with exports from large-scale hub plants. The larger, regional market should also be more attractive as an investment location for market seeking FDI. MNEs located in the spoke economies are also likely to expand, both for offensive and defensive reasons, although they may need to change their market focus from the national to the regional market, adapting to what Rugman and D’Cruz (1991) call the ‘double diamond’ model of competitive advantage. On the other hand, firms located in the spoke economies that have not invested in the hub market (the domestics) are likely to respond to regional integration by moving closer to the border. The move to relocate to border areas and port cities should be more pronounced for firms in small countries due to the relatively stronger pull of the larger market.

Lastly, if the PTA uses rules of origin to determine duty-free status, outsider firms that are unlikely to meet the rules of origin tests are more likely to cluster in the largest country, thus reducing their intra-PTA tariff duties. Large differences in sizes of member countries therefore suggest that the hub economy may reap a disproportionate share of regional FDI flows.
Micro-regions and FDI

Until Krugman published *Geography and Trade* (1991), little attention had been paid to the spatial effects of international trade. Since then, there has been a virtual explosion of work among international trade and international business scholars on clustering. The key tension identified by these authors lies between the pull to centralize activities in one location and the pull to disperse activities closer to factor and product markets.

For example, in deciding where to locate a plant, the firm must first choose whether to centralize the activity by locating it at home (with the parent firm) or decentralize the activity to a foreign location. Eaton et al.’s (1994b) argument explains how the interaction of economies of scale together with transportation and communication costs produce clustering. Essentially, these authors argue that, ‘the degree of agglomeration is determined jointly by the interaction of economies of scale at the plant level, which work to create agglomeration by encouraging fewer larger plants, and transportation and communication costs, which work to limit agglomeration by encouraging smaller plants’ (1994b: 82). Economies of scale at the plant level encourage centralization of production in one location with exports being used as the mode to supply foreign markets. Transportation and communication costs, on the other hand, raise the costs of exporting and encourage decentralization. High tariffs act like transport costs, discouraging foreign firms from locating production in one central place and exporting to a variety of foreign markets. Tariff jumping FDI is likely to occur as foreign firms set up domestic plants in order to supply the local market, particularly where the market is large and attractive.

While Eaton et al. (1994b) focused on the individual firm’s decision to centralize or decentralize its activities, the unit of analysis for other authors such as Markusen (1996), Enright (1998), Porter (1998a,b) and Dunning (2000) has been the formation of spatial clusters, or ‘sticky places within slippery space’ (Markusen 1996). These authors identify several different types of clusters or micro-regions, depending on their scope, nature of activities, growth potential, innovatory capacity, and governance structures (Enright 1998). Markusen (1996) identifies four types of clusters: (1) Marshallian new industrial district based on flexible specialization, as exemplified by the north Italy garment district; (2) hub-and-spoke district where satellite firms locate around flagship firms in one or a few industries (for example, the Toyota complex near Tokyo); (3) the satellite industrial platform, such as export processing zones, where branch plants cluster together to take advantage of low wages and/or available resources; and (4) state-centered district where a major government institution (for example, military or research facility) anchors the regional economy.
Where firms are likely to locate is partly dependent on the advantages of firms clustering in one location; that is, whether or not to locate near other firms (upstream suppliers, downstream customers, or competitors). External economies and the benefits of information sharing encourage clustering of firms, particularly in knowledge intensive sectors. Access to natural resources or specialized assets and infrastructure will also encourage resource-seeking firms to cluster in one location. The benefits of labor pooling can similarly encourage clustering, for example, where highly specialized workers are needed. *Horizontal clusters*, firms engaged in similar lines of activity, are likely to form under these circumstances.

The shift from mass to lean production methods can also cause clustering. New process technologies have reduced the importance of labor, transport, and communications costs; shortened the minimum efficient scale of production (thus reducing the importance of economies of scale at the plant level); and increased the need for supplier firms to locate close to their downstream customers in order to use just-in-time production and delivery methods. Lean production therefore encourages the formation of *vertical clusters* of suppliers and buyers. In many cases, these clusters may form around one or more flagship firms, with upstream and downstream firms as satellites around them. Head et al. (1995) perform an empirical study testing whether industry-level agglomeration economies influence location decisions. By examining the location decision of Japanese manufacturing plants in the United States, they found that the Japanese ventures do not mimic the geographical pattern of US establishments; instead, they follow initial investments by other Japanese firms in the same industry or industrial group. Therefore, the authors suggest that locational choice supports the theory on agglomeration externalities rather than the theory on differences in factor endowments.

Regional integration can affect the location of economic activity inside countries. Puga and Venables (1997) examine the effects of preferential trading agreements on industrial location. Free trade areas have the pulling effect of attracting industry into the integrating countries. Moreover, when trade barriers fall, agglomeration economies favor centralization of production so some member countries may gain industry at the expense of others. A hub-and-spoke arrangement will favor location in the hub, with better access to the spoke countries.

Eden and Monteils (2000) explore the impacts of regional integration on clustering. They argue that the type of firm – insider, outsider, domestic – influences the firm’s location decisions and therefore the creation of micro-regions. For insider firms, as tariff rates fall in a PTA, where plant-level economies of scale are important, firms that had tariff-jumping factories may close down smaller plants and shift production to the largest, most efficient plant, relying on exports to reach the smaller markets. If the PTA
is also accompanied by a decline in transportation costs (for example, liberalization of cross-border transport routes is part of the PTA package), this also encourages centralization of production. This suggests that insider MNEs with investments throughout the PTA are more likely to rationalize production by closing inefficient plants and centralizing production, where economies of scale gains are significant and transport costs low. Alternatively, rationalization of product lines between plants and increased intra-industry horizontal trade is an alternative solution for differentiated product industries.

Outsider firms may be induced by the PTA to locate inside the region. Their locational patterns may also cluster if they are following downstream producers (particularly where lean production techniques are prevalent), choose to locate their market seeking FDI in urban centers, or are attracted to knowledge-based clusters for their external economies and information sharing. Such firms may be more likely to see the region as a whole and make decisions from a regionally efficient perspective, thus increasing their competitiveness relative to member firms. On the other hand, outsider firms may be less well equipped to take advantage of clustering, particularly where the advantages are based on knowledge spillovers (Enright 1996: 204).

For domestics, a PTA expands the set of markets available to firms, if they had not previously engaged in exports or FDI to the member countries. In order to access these new markets, these firms are likely to move to locations with good access to the other markets, such as border areas and port cities. As firms move to border locations, a self-reinforcing movement may occur due to agglomeration economies, creating new clusters. Thus, some clusters could expand with a PTA while other micro-regions shrink.

Hanson (1998) argues that border clusters may be encouraged by the formation of a PTA, as firms move to the border in order to be able to access adjoining markets. Small cities along the border may develop into transportation and wholesale trade hubs, facilitating cross-border flows of goods and services liberalized under the PTA. Large cities, on the other hand, may develop into full-sized regional production sharing networks, where firms from both countries specialize their value adding activities along the value chain, engage in sophisticated subcontracting strategies, and establish cross-border alliances.

**LESSONS FROM NAFTA**

Any review of the IB literature on MNEs and regional integration is incomplete if it only focuses on theoretical contributions to the literature. In this section, we review empirical work on regional integration in North America.
NAFTA shares three characteristics that distinguish the 'new regionalism' from older PTAs. First, NAFTA allows free movement of goods, services, and capital between two developed market economies and a developing country. Opening trade and investment between countries with very different institutional, legal, political, social, and economic profiles not surprisingly should be more difficult than creating a PTA between two rich countries (for example, the Canada–US Free Trade Agreement (CAFTA)). The extension of CAFTA to Mexico had problems similar to those experienced by West Germany in its amalgamation with East Germany, and transaction costs were, and continue to be, large. Second, the major adjustments in terms of reducing tariffs and nontariff barriers fell on Mexico, as the smaller partner with the highest tariff barriers and the largest amount of necessary adjustment. On the other hand, Mexico was expected to make the largest gains in the long run. Third, country sizes are very different with Canada having the smallest population (30 million), followed by Mexico (100 million) compared to the United States (230 million). In addition, North American trade and investment patterns are dominated by US multinationals. North America can therefore be seen as a hub-and-spoke relationship, where the US hub is linked through trade and investment to two spokes: Canada and Mexico (Eden and Molot 1993).

In this section, we examine the impacts of the formation of a macro-region – NAFTA – on MNE trade and FDI patterns, separating our review into macro-region and micro-region (clustering) decisions.

**Macro-regions, Trade and FDI**

Because NAFTA has only been in place for six years, much of the work in the 1990s has looked at the earlier Canada–US Free Trade Agreement (CAFTA). Rugman (1990) surveyed MNEs in Canada and the United States about their anticipated reactions to the 1989 Canada–US Free Trade Agreement (CAFTA). He found that MNEs in both countries supported free trade, anticipated few adjustment problems, and were particularly attuned to competitiveness. Another study, done two years later by the Conference Board of Canada (Krajewski 1992), also posed questions to its members about their reactions to CAFTA. The Conference Board study looked at two groups: Canadian parents with US subsidiaries and Canadian subsidiaries of US parents. CAFTA was seen by the respondents as a primary driver, in addition to globalization of markets in general, pushing multinationals in Canada to rationalize their production and sales for the North American market. The firms had a sense of new opportunities and/or felt the necessity to compete globally to survive. CAFTA provided new business opportunities through a more open door to the US market.
Schwanen (1997), in a C.D. Howe Institute study, compared the trade and FDI growth rates in sectors liberalized by CAFTA, relative to those that were already barrier free, between 1988 and 1995. He found that Canada–US bilateral trade grew more quickly in liberalized sectors than in tariff-free sectors. Intra-industry specialization, as evidenced by rapid growth in two-way trade, occurred in several sectors. On the other hand, North America’s share of global FDI fell over the period, as did the importance of Canada and the US in each other’s FDI portfolio. Schwanen concluded that external events (for example, liberalization and privatization in South America, rapid growth in Asia) had attracted FDI outside of North America.

In a later study looking at NAFTA, Rugman et al. (1997) argued that NAFTA posed three challenges for Canadian firms. First, NAFTA did not provide secure access to the US market because US anti-dumping and countervailing duties could still be used to harass Canadian firms. Second, Canadian firms were ethnocentric and needed to develop national responsiveness to the whole North American market, not just Canada. Third, the high degree of foreign ownership in certain sectors complicated business-government relations in Canada. Because Canadian firms depend so heavily on the US market, one way to develop a ‘North American mindset’ was to establish business networks to maintain competitive advantages (D’Cruz and Rugman 1993). Using the chemical industry as an example, Rugman et al. (1997) argued that a business network arrangement, where key partners were coordinated by a flagship firm, provided an effective strategic response to external environmental changes, such as NAFTA.

Eden (1994b) theorized about the locational strategies of US multinationals after NAFTA. She provided detailed statistics about US MNEs in Canada and Mexico in 1990, and argued that US MNEs were best placed to take advantage of the opportunities created by North American economic integration. A follow-up Conference Board study looking at NAFTA (Blank et al. 1994, 1995) focused on the restructuring of US firms and their Canadian subsidiaries and confirmed Eden’s hypotheses. Large US MNEs were quickly moving to adopt North American strategies and structures. Canadian subsidiaries were being rapidly integrated into a continental production system, and more rapidly than their Mexican sister affiliates. Driving this reorganization was what the authors termed the emerging ‘architecture’ of North America – a North American economic space – as well as intensified global competition, the early 1990s recession and technological change. Many Canadian subsidiaries expected a rise in intra-firm trade as fewer goods were produced in Canadian plants, a decrease in subsidiary autonomy, a loss of production capacity and jobs, redefinition of their role within the corporate network, and growing intra-firm competition for product and marketing mandates (Blank et al. 1995).
Johnson et al. (1995) surveyed senior operations executives at 139 North American manufacturers in 1993. The managers were asked to assess how they thought NAFTA would affect their operations strategies and to outline the responses they had undertaken in the past two years. Canadian firms did not expect to meet global competition through low manufacturing costs, but through superior customer service, dependable deliveries and high quality. Their response was to improve capacity in these areas, seek more international customers, and avoid markets where fast deliveries and product proliferation were critical. In the new North American economic space, both Canadian and Mexican firms saw their greatest potential in the US market, but also saw their US competitors as their greatest threat.

The US International Trade Commission (ITC) conducted a major three-year review of the impacts of NAFTA on the US economy in nearly 200 industrial sectors (ITC 1997; see also the analysis in GAO 1997). The ITC concluded that NAFTA has minimal impacts on the US economy in terms of trade, employment or hourly earnings. Maquiladora-related trade expanded sharply, leading the ITC to conclude that production sharing along the US-Mexico border would continue to expand due to the complementaries of the US and Mexican economies. Intra-industry trade, both Canada-US and Mexico-US, increased in sectors characterized by product differentiation and a high percentage of manufactured components. US-Mexico integration is perhaps proceeding fastest in the auto sector, where high Mexican trade barriers (tariffs on autos, domestic content regulations, trade balancing requirements) are being dismantled and the gains from continental integration are large (USTR 1998).

In a later study of NAFTA and MNE strategies, Blank and Haar (1998) surveyed senior managers of US MNEs with Mexican operation in 1994–96. Incorporating the results of two earlier Conference Board studies of US and Canadian managers, the authors analyze MNE strategies from the perspective of all three countries. They find that ‘cross-border corporate integration has been deeper and more far-reaching ... than governments ... seem to realize’ (1998: 2). Their interviews with North American managers led Blank and Haar to conclude that an integrated corporate system in North America is emerging where MNEs view North America as a single spatial unit when making their configuration and coordination strategies.

Micro-regions, Trade and FDI

So far, there has been little work asking whether regional integration in North America has encouraged clustering, either within a country or along the borders between member countries. In one of the early statistical analyses of Canada-US trade flows post-CAFTA, Little (1996) found that both US and
Canadian firms tended to rely on trade rather than FDI to serve the US-Canadian market. She focused on changes in the industrial composition of trade at the regional level, which were often obscured when national-level data were employed. For example, inward FDI from Canada and geographic shifts in US industry activity significantly affected New England's export performance, because trade activity shifted to the south and west with the implementation of CAFTA.

A second exception is Gordon Hanson who, across several papers, has examined the impact of NAFTA on US-Mexican trade and FDI patterns. Industry responses to NAFTA were expected to be most evident in terms of US-Mexico trade and FDI flows because the major effect of NAFTA was to bring Mexico into CAFTA. Studying economic activity in the US-Mexico border cities, Hanson (1996) found evidence that export manufacturing expansion in these border cities has increased manufacturing employment in the US border cities, suggesting that NAFTA can positively influence the relocation of US manufacturing production to the US-Mexico border region, especially when transport costs are an important consideration for industry location. As is evidenced by such relocation from interior US region to the border cities, Hanson (1996) suggests that this negates the prevailing view that the smaller Mexican economy would not have any significant effects on the US economy.

Parallel to the US manufacturing relocation, there is a similar pattern of relocation in Mexico (Hanson 1998). Manufacturing employment has increased in northern Mexico and decreased in central Mexico, also suggesting manufacturing relocation to the border cities with the US. The author suggests that NAFTA is likely to have more impact on Mexico in industrial location compared to the two larger economies, the US and Canada. This view is in broad agreement with Rugman and Gestrin (1993a,b) who suggested that the impact of NAFTA on the US and Canada would be more neutral, as most effects should already have taken place since CAFTA, while Mexico would also benefit from significant investment diversion away from other LDCs. Therefore, the most significant post-NAFTA changes were likely to take place in Mexico.

Eden and Monteils (2000) develop a theoretical model of MNE strategic responses to regional integration and then provide some evidence of these responses, looking at FDI patterns in North America over 1985-97. They found that North America became less attractive to inward FDI relative to other regions, especially East Asia and the former Soviet Union, over this period. CAFTA and NAFTA did not appear to have resulted in significant increased inward FDI. Although the dollar value of the FDI stock increased among all the NAFTA partners and in both directions, relatively more investments were directed outside of North America.
Looking specifically at inward FDI entries to the United States, they found that Canadian MNEs invested more frequently at the beginning of the CAFTA and engaged in relatively more new FDI entries, primarily through mergers and acquisitions. Insider FDI entries, for Canada, while more clustered in geography and industry than FDI entries from non-NAFTA countries, were overall similar in terms of the top ten destinations for inward FDI. Mexican investors, reflecting Mexico's joining the PTA only in 1994, its higher trade barriers and developing country status, engaged in small numbers of new investments in the US market which were geographically and industrially clustered. Mexico, in particular, invested more heavily along the US–Mexico border and in manufacturing industries relative to all home country entries. Eden and Monteils concluded that North American firms were making their locational decisions from a macro-regional perspective, but that in the 1990s this meant primarily locational reshufflings as firms rationalized investments on a continental basis. Over the longer term, once NAFTA is fully phased in, they expected new investment decisions to be made treating the North American macro-region as the 'home base'.

What is evident from these studies is that, in response to NAFTA, multinationals are engaged in locational reshuffling, as Vernon (1994) predicted, designed to integrate Mexican and Canadian industry into regional production networks. This is proceeding fastest in the automotive, electronic equipment, and textile sectors, as evidenced by the rapid growth in two-way trade in components and finished manufactured goods and the movement of firms to the US–Canada and US–Mexico border regions.

NEW DIRECTIONS IN IB RESEARCH

Based on our outline of the IB literature on MNEs and regional integration, we suggest the following as potential areas for research. First, there continues to be room for statistical analyses of the impact of regional integration schemes on trade and FDI flows, both extra-region and intra-region. For example, the definitive study of the empirical impacts of NAFTA on FDI patterns, at the national and industry levels, has not been completed. Dunning (1997a,b) is a tour de force on FDI patterns in response to first wave and second wave regional integration within the European Union. Similar studies for other regions remain to be explored.

Second, the dynamics of preferential trading arrangements, as they weaken or strengthen over time, pose opportunities and threats for domestic and multinational firms. Change in the structure and strength of PTAs is likely to occur in two dimensions: breadth and depth. The breadth of a regional integration scheme increases when the PTA takes on new member countries,
causing further rounds of trade and FDI creation and diversion effects. For example, the 1989 bilateral CAFTA expanded into the 1994 trilateral NAFTA. While Congress did not approve of Chile as a fourth NAFTA member, the governments in North and South America have been discussing a Free Trade Area of the Americas (FTAA) linking both hemispheres, starting in 2005. Our analysis suggests that insider MNEs, with investments in both hemispheres, are most likely to be the beneficiaries of an FTAA. PTAs can also be deepened through the broader application of national treatment legislation and/or the increased harmonization and coordination of internal policies. The adoption of a common currency within the European Union is one such example.

Third, one can study the impacts of PTA deepening or broadening on the cross-border configuration and coordination decisions of multinational firms. The reactions of different types of firms – insiders, outsiders and domestics – to broadening and deepening PTAs is an under-explored topic. Comparative studies exploring firm responses to different types of PTAs (for example, free trade areas versus customs unions, PTAs with similar sized members versus hub-and-spoke PTAs, PTAs with rich country members versus PTAs that include developing countries) have not been done.

Fourth, although little attention (due to space constraints) has been paid in this chapter to the political economy of MNEs and regional integration, an area that begs for further exploration is how firms can affect the dynamic path of regional integration. Two-level bargaining games should be visible as MNEs negotiate with their domestic governments and with the regional apparatus. Different types of PTAs should lead to different configurations of MNE responses, but no work has been done in this area.

Fifth, another area that was not explored in this chapter due to space constraints, and is under-explored in the literature, is the impact of regional integration schemes on the mode of entry decision into member country markets. We briefly discussed the demand for strategic alliances as a way to obtain strategic assets and attain insider status within the PTA.

Lastly, the life cycle of micro-regions or clusters within countries and along border regions within a larger PTA is a new topic in the IB literature. How different types of micro-regions are formed and how they foster firm-level and national competitiveness is not well understood. Some work has been done on follow-the-leader FDI as suppliers follow flagship firms and create vertical clusters; other work has focused on knowledge clusters; however, this remains a relatively new topic in the IB literature.
CONCLUSIONS

Regional integration schemes have been with us since the late 1950s. Early work on understanding these preferential trading arrangements was conducted primarily by economics and political scientists. In this chapter, we have reviewed the international business literature on multinational enterprises and regional integration, focusing on ‘the new regionalism’ and its impact on MNE location decisions at both the macro-region and micro-region levels. Much work remains to be done, however. IB researchers need to inform their own studies of MNEs and regional integration with parallel research being undertaken in international economics and political science.

NOTES

1. For recent overviews of this literature see Baldwin and Venables (1996), Bhagwati and Panagariya (1996), De Melo and Panagariya (1996), El-Agraa (1997) and Bhagwati et al. (1998).

2. We can distinguish at least two separate strands of IB literature on regional integration. The first strand takes a political economy approach, focusing on MNE attitudes towards regional integration and how firms can affect the nature and speed of regional integration (for example, Dunning 1988, Eden and Molot 1993, Milner 1997; Rugman 1994a,b). The second strand examines the impacts of regional integration on inward and outward FDI patterns and MNE location decisions. In this paper, we focus on the second strand of the literature.

3. For example, Mexico responded to the peso crisis in 1995 by raising more than 500 tariffs against non-member countries while leaving those against its NAFTA partners unchanged.

4. See also Vernon (1994) and Eden and Monteils (2000).


6. For example, Dunning argues that US multinationals were the major beneficiaries from Mark I regional integration in the European Community because they ‘were able to take advantage of the removal of tariff barriers, and surmount the transactions costs of the remaining non-tariff barriers better than their EC equivalents’ (1997a: 5).

7. In this paper, we focus on regional integration within North America. On European integration, see for example, Dunning (1994, 1997a, b).


REFERENCES

Balassa, B. (1961), The Theory of Economic Integration, Homewood, II: Richard D. Irwin, Inc.


The macro-environment


# Contents

**List of figures** vii  
**List of tables** viii  
**List of boxes** ix  
**List of contributors** x  
**Foreword**  M. Moshe Porat xvii  
**Foreword**  Arvind Phatak xviii  
**Preface** xx

1. International business research: from functional to issue-based focus  
   *Masaaki Kotabe and Preet S. Aulakh*  1

## PART I  THE MACRO-ENVIRONMENT

2. Regional integration and foreign direct investment: theory and lessons from NAFTA  
   *Lorraine Eden*  15

3. Intellectual property rights and international business  
   *Subhash C. Jain*  37

4. Global financial markets and global firms: implications for international business research  
   *Jongmoo Jay Choi*  65

5. Cultural Balkanization and hybridization in an era of globalization: implications for international business research  
   *Bryan W. Husted*  81

## PART II  INTERFACES BETWEEN BUSINESS AND INSTITUTIONS

6. Emerging issues in MNC–host government relations in developing countries  
   *Ravi Ramamurti*  99

7. National export promotion: A statement of issues, changes, and opportunities  
   *Michael R. Czinkota*  123