The Boundaries of Multinational Enterprises and the Theory of International Trade

James R. Markusen

The Boundaries of Multinational Enterprises and the Theory of International Trade

James R. Markusen

The so-called "new trade theory" and more recently the literature on "geography and trade" have greatly enriched economists' understanding of trade. In the new trade theory, trade and gains from trade can arise independently of any pattern of comparative advantage (as traditionally understood) as firms exploit economies of scale and pursue strategies of product differentiation in an imperfectly competitive environment. The literature on geography and trade is a natural extension of this line of research, focussing on how industry agglomeration and regional differentiation can arise endogenously as a consequence of transport costs, market sizes, and the trade policy regime.

These newer streams of literature are very limited in their treatment of firms. In these models, a firm is generally synonymous with a plant or production facility; that is, a firm is an independent organization that produces one good in one location. Multiplant and multiproduct production, whether horizontal or vertical, are generally excluded from the analysis. This is potentially troubling. After all, industries characterized by scale economies and imperfect competition are often dominated by multinationals. As a result, the policy and normative analysis that comes out of the new trade theory may be significantly off base. For example, conclusions of the "strategic trade policy" literature are fundamentally bound up with the notion of clearly defined national firms competing via trade with the national champions of other

■ *James R. Markusen is Professor of Economics, University of Colorado, Boulder, Colorado, and Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts.*

countries. Substantial foreign ownership of domestic production facilities radically alters the policy implications (Dick, 1993).

This paper provides an update on recent research on the theory of the multinational enterprise. Multinationals are firms that engage in direct foreign investment, defined as investments in which the firm acquires a substantial controlling interest in a foreign firm or sets up a subsidiary in a foreign country. The terms "multinational enterprise" and "direct foreign investment" will be used fairly interchangeably.

Since it is necessary to circumscribe the article, I will focus on research and models produced by international trade economists, whose research is motivated by two sets of questions. What circumstances lead a firm to serve a foreign market by exports versus foreign production? In the latter case, why might the firm chose direct investment versus some type of alternative mode of entry, such as a joint venture or a licensing arrangement? I will limit the discussion of these alternatives to simply the licensing option, both because the literature is better developed than that on joint ventures and other modes of entry, and because consideration of the licensing option is sufficient to illustrate some of the key tradeoffs.

Second, I will focus on horizontal direct investment, meaning the foreign production (not just investment in distribution, wholesaling, and servicing) of products and services roughly similar to those the firm produces for its home market. Horizontal direct investment is more important quantitatively than vertical investment (which in this context means fragmenting the production process geographically, by stages of production) and relates closely to the issues raised in the literature on new trade theory and strategic trade policy.

An Empirical Background

Before plunging into the theory, a real-world background is needed to provide a context within which to evaluate the theory and indeed to understand its origins. Much of the recent theory is fairly closely tied to the evidence, or at least consistent with it. I first offer six macro facts from the aggregate data, and then six micro facts found in analyses of the industry and firm-level data. The citations given in this abbreviated discussion should be taken only as illustrative of what I believe to be a broadly agreed-upon finding.

Macro Facts
First, direct foreign investment has grown rapidly throughout the world, with a particularly strong surge in the late 1980s.[1]

---

[1] Documentation of macro facts 1–3 can be found in a number of sources. All are demonstrated in various articles found in Froot (1993). See in particular articles by Krugman and Graham, and by Lipsey in that volume. See also Hummels and Stern (1994), the UNCTAD World Development Report (1993), and Markusen and Venables (1995).

Second, the developed countries not only account for the overwhelming proportion of outward direct foreign investment, but they are also the major recipients of direct foreign investment. Hummels and Stern (1994) report that in 1985 the developed countries were the source of 97 percent of direct investment flows and the recipient of 75 percent.

Third, there is a great deal of two-way direct foreign investment flows between pairs of developed countries, even at the industry level. Julius (1990) reports that the share of all direct investment outflows generated by G-5 countries absorbed by other G-5 countries has been rising and amounted to 70 percent by 1988.[2]

Fourth, most direct foreign investment in production facilities seems to be "horizontal," in the sense that most of the output of foreign production affiliates is sold in the foreign country. For example, Brainard (1993b) reports that foreign affiliates owned by U.S. multinationals export only 13 percent of their overseas production to the United States, while the U.S. affiliates of foreign multinationals export 2 percent of their U.S. production to their parents.

Fifth, a significant percentage of world trade, about 30 percent, is now intra-firm trade (UNCTAD, 1993; Brainard, 1993b). There is some evidence of complementarity between exports and overseas production (Blomstrom, Lipsey, and Kulchycky, 1988; Denekamp and Ferrantino, 1992).

Sixth, there is little evidence that direct foreign investment is related to differences in factor endowments across countries (Brainard, 1993b), or to differences in the general return to capital. There is little support for the idea that risk diversification or tax avoidance are important motives for direct foreign investment (Morck and Yeung, 1991; Wheeler and Mody, 1992). Apparently, most firms first choose foreign production locations, and then instruct their tax departments to minimize taxes.

In short, direct investment has been growing rapidly, and the bulk of it is horizontal direct investment among countries with similar per capita incomes, similar relative factor endowments, and relatively low trade barriers.

Micro Facts

First, there are large differences across industries in the degree to which production and sales are accounted for by multinational firms (Brainard, 1993b).

---

[2]Brainard (1993b) considers an index of intra-industry affiliate sales, measuring the degree of international cross investment in a particular industry, and an index of intra-industry trade. She shows that intra-industry affiliate sales indices are somewhat lower than intra-industry trade indices, but they are still significant. For U.S. inward and outward direct foreign investment, the Grubel-Lloyd index is 0.27 averaged across manufacturing industries versus 0.40 for trade flows. If the sample were restricted to U.S.-Europe-Japan, the affiliate sales index would likely be much higher.

Second, multinationals tend to be important in industries and firms with four characteristics: high levels of R&D relative to sales; a large share of professional and technical workers in their workforces; products that are new and/or technically complex; and high levels of product differentiation and advertising. These characteristics appear in many studies, and I have never seen any of them contradicted in any study.[3]

Third, multinationals tend to be firms in which the value of the firm's intangible assets—roughly, market value less the value of tangible assets such as plant and equipment—is large relative to its market value (Morck and Yeung, 1991).

Fourth, limited evidence suggests that plant-level scale economies are negatively associated with multinationality (Brainard, 1993c; Beaudreau, 1986).

Fifth, there seems to be a threshold size for multinationals, but above that level corporate size is not important. Corporate age is highly correlated with multinationality (Blomstrom and Lipsey, 1991; Morck and Yeung, 1991; Beaudreau, 1986).

Sixth, there is mixed evidence as to whether or not the *level* of direct foreign investment is positively related to the existence of trade barriers or transport costs. New evidence by Brainard (1993c) convincingly demonstrates that the *share* of foreign affiliate sales in the sum of exports and affiliate sales is positively related to trade barriers and transport costs.[4] Thus, trade barriers and transport costs do cause a substitution effect toward direct investment, although they may also reduce the levels of both investment and trade.

From this micro perspective, multinationals are important in industries in which intangible, firm-specific assets are important. These assets can generally be characterized as "knowledge capital," ranging from proprietary product or process know-how to reputations and trademarks. The amount of direct foreign investment increases relative to trade (but not necessarily absolutely) as tariffs and transport costs increase.

---

[3]Much discussion, data, and many references are found in Caves (1982). Buckley and Casson (1976) remains an important study on this question. For more recent evidence, see Morck and Yeung (1991, 1992), Brainard (1993b, c), Grubaugh (1987), and Beaudreau (1986). For events in which firms do transfer technology abroad, articles by Davidson and McFetridge (1984), Mansfield and Romeo (1980), Teece (1986), and Wilson (1977) show technology is more likely to be transferred internally within the firm by R&D-intensive firms producing new and technically complex products. Blomstrom and Zejan (1991) get similar results with respect to joint ventures: firms are less likely to seek a foreign partner when intangible assets are important.

[4]Regression coefficients on tariffs and transport costs or distance have often been insignificant and/or had the wrong sign in the equations with some measure of multinationality as the dependent variable; for example, Beaudreau (1986) using extensive firm-specific data. Brainard (1993c) has mixed results for equations explaining the level of affiliate sales abroad. Part of the explanation seems to be that many firms have substantial imported content in their foreign production and export modest amounts (on average as noted above) back to their parent. In these respects, tariffs and transport costs discourage affiliate production just like they discourage exports. However, using share equations, the *share* of affiliate sales in the total of affiliate sales and exports is increasing and significant in both freight charges and tariffs.

Dunning's OLI Framework

If foreign multinational enterprises are exactly identical to domestic firms, they will not find it profitable to enter the domestic market. After all, there are added costs of doing business in another country, including communications and transport costs, higher costs of stationing personnel abroad, barriers due to language, customs, and being outside the local business and government networks. The multinational enterprise must, therefore, arise due to the fact that it possesses some special advantage such as superior technology or lower costs due to scale economies. This point is found in Hymer's 1960 dissertation (published in Hymer, 1976), and the logic of the argument remains persuasive. It implies that a multinational enterprise brings inherent advantages, such as technology, that potentially constitute an important gain for the host country. There may, however, be offsetting costs such as increased monopoly power resulting in the transfer of rents away from indigenous host-country firms (Hymer, 1976; Kindleberger, 1969, 1984).[5]

Because of the inherent disadvantages and higher costs of foreign production, it is necessary to identify the advantages and conditions under which direct investment will occur as just noted. One organizing framework was proposed by Dunning (1977, 1981), who suggested that three conditions all need to be present for a firm to have a strong motive to undertake direct investment. This has become known as the OLI framework: ownership, location, and internalization.

A firm's *ownership advantage* could be a product or a production process to which other firms do not have access, such as a patent, blueprint, or trade secret. It could also be something intangible, like a trademark or reputation for quality. Whatever its form, the ownership advantage confers some valuable market power or cost advantage on the firm sufficient to outweigh the disadvantages of doing business abroad.

In addition, the foreign market must offer a *location advantage* that makes it profitable to produce the product in the foreign country rather than simply produce it at home and export it to the foreign market. Although tariffs, quotas, transport costs, and cheap factor prices are the most obvious sources of location advantages, factors such as access to customers can also be important. Indeed, many multinationals are in service industries (for example, hotels) in which on-site provision of the services is an inherent part of the companies' business.[6]

Finally, the multinational enterprise must have an *internalization advantage*. This condition is the most abstract of the three. If a company has a proprietary

---

[5] In fact, Hymer (1976) and Kindleberger (1969, 1984) concentrate rather heavily on the market power dimensions of multinationals, without inquiring very deeply as to the efficiency advantages that might go hand in hand with market power.

[6] Of course, it is possible to think of such services as simply being characterized by very high transport costs!

product or production process and if, due to tariffs and transport costs, it is advantageous to produce the product abroad rather than export it, it is still not obvious that the company should set up a foreign subsidiary. One of several alternatives is to license a foreign firm to produce the product or use the production process. Why not just sell the blueprints to a foreign firm rather than go through the costly and difficult process of setting up a foreign production facility? Reasons for wishing to do so are referred to as internalization advantages; that is, the product or process is exploited internally within the firm rather than at arm's length through markets.[7]

Ownership Advantages, Firm-Specific Assets, and Knowledge Capital

Ownership advantages come in many possible forms, and a good approach to identifying them is to seek guidance from the micro facts about direct foreign investment. Remember, the evidence shows that an industry tends to have a greater proportion of multinational enterprises when the output of that industry is characterized by R&D, marketing expenditures, scientific and technical workers, product newness and complexity, and product differentiation. At a broader level, multinational enterprises are identified with a high ratio of intangible assets of the firm to its total market value. These explanatory variables give rise to the concept of knowledge-based, firm-specific assets. These proprietary assets of the firm are embodied in such things as the human capital of the employees, patents or other exclusive technical knowledge, copyrights or trademarks, or even more intangible assets such as management, "know-how" or the reputation of the firm.

There are two good reasons why these knowledge-based assets are more likely to give rise to direct foreign investment than physical capital assets. First, knowledge-based assets can be transferred easily back and forth across space at low cost. An engineer or manager can visit many separate production facilities at a relatively low cost. Second, knowledge often has a joint character, like a public good, in that it can be supplied to additional production facilities at very low cost. Blueprints, chemical formulae and pharmaceuticals, trademarks, and other marketing devices all have this characteristic—but assets based on physical capital such as machinery usually do not. That is, physical capital usually cannot yield a flow of services in one location without reducing its productivity in others.

In turn, the joint-input characteristic of knowledge-based assets has implications for the efficiency of the firm and for market structure. These implica-

---

[7]Attachment to the OLI framework is not universal, although it has been very appealing to trade economists. Rugman (1981, 1985, 1986) in particular argues that internalization is really the only thing that matters to understanding the multinational. OLI is also limited in that it only considers the conditions necessary for direct investment. It has little to offer about the choice among alternatives, such as licensing versus joint venture versus exporting.

tions are encapsulated in the notion of economies of multiplant production. Such economies arise because a single two-plant firm has a cost efficiency over two independent single-plant firms. The multiplant firm (that is, the multinational enterprise) need only make a single investment in R&D, for example, while two independent firms must each make the investment. Cost efficiency then dictates that multinational enterprises (multiplant firms) arise as the equilibrium market structure in industries where firm-specific assets are important, which is consistent with the empirical evidence.[8]

The converse proposition also deserves emphasis. Scale economies based on physical capital intensity do not by themselves lead to foreign direct investment, an argument supported by some evidence (Brainard, 1993c; Beaudreau, 1986). This type of scale economy implies the cost efficiency of centralized production rather than geographically dispersed production. Of course, some industries with high physical capital intensity may also be industries in which firm-specific assets are important (like automobiles).

What then is being traded when we observe multinational production? Basically, multinational enterprises in this framework are exporters of the services of firm-specific assets. These include management, engineering, marketing, and financial services, many of which are based on human capital. They also include the "services" of patents and trademarks, which are other knowledge-based assets. Subsidiaries import these services in exchange for repatriated profits, royalties, fees, or output.

Combining Ownership and Location Advantages

A small number of authors working from the international trade perspective have constructed models in which multinationals arise endogenously in equilibrium. These authors have combined elements of ownership and location advantages, generally leaving aside the question of internalization. Early papers by Helpman (1984) and Markusen (1984) allowed for a headquarters or firm-level activity such as R&D, which could be separated from production. Helpman's model was constructed such that firms have a single production facility, which could be in a different country than the headquarters. The absence of tariffs or transport costs means that the firm will never open more than one production facility, so the model is really one of a vertically integrated firm. In Markusen's model, the multinational enterprise would choose production facilities in both countries, becoming a horizontally integrated

---

[8]R&D, advertising, and technical and scientific workers are often used as proxies for firm-specific assets, and hence multinationality is highly correlated with firm-specific assets using these proxies (for citations, see the studies listed in footnote 3). Alternatively, firm-specific assets (intangible assets) are proxied as the market value of the firm minus the value of tangible assets (Morck and Yeung, 1991). In this case, firm-specific assets are *defined* as a residual, and this residual is highly correlated with multinationality. Care needs to be taken lest the argument become tautological: multinational enterprises tend to be firms with big residual values (unobserved intangible assets), and these residuals are firm-specific assets by definition.

multinational enterprise. The headquarter's activity is modelled as a joint input (a non-rival input) such that adding additional plants does not reduce the value of the input to existing plants. The respective approaches are extended in Helpman (1985), Helpman and Krugman (1985), and Horstmann and Markusen (1987a).

More recently, Brainard (1993a) and Horstmann and Markusen (1992) have produced models in which horizontal multinationals arise endogenously and in which two-way investment, a characteristic of the North Atlantic economy, can arise in equilibrium. The three key elements of both papers are firm-level activities (like R&D) that are joint inputs across plants, plant-level scale economies, and tariffs or transport costs between countries. Although Brainard models firms as producing differentiated products whereas goods are homogeneous in the Horstmann and Markusen model, the results are strikingly similar. Multinationals are supported in equilibrium when firm-level fixed costs and tariffs and transport costs are large relative to plant-level scale economies. Multinationals are more likely to exist in equilibrium when the countries are large (both papers) and when the countries have similar relative factor endowments (Brainard). These results fit well with the empirical evidence noted above.

It may be useful to offer an outline of these newer models. The model sketched here assumes homogenous goods, but it is clear from Brainard (1993a) that a differentiated-good model generates similar conclusions. The model begins with two countries (h and f) producing two goods (X and Y), using the factors "land" and "labor" (R and L). Factors are immobile between countries.

The two goods X and Y have the following characteristics. Y is a homogeneous good produced with constant returns to scale by a competitive industry. Y production uses all of the land (R) and some of the labor (L).[9] X is a homogeneous good produced with increasing returns to scale by Cournot firms. Markets are segmented (arbitrage conditions need not hold). X uses labor as its single factor of production. The costs for producers of X, which differ across countries, can be measured in units of labor. The costs can be divided into four types: firm-specific fixed cost (F); plant-specific fixed cost (one G per plant); constant marginal cost (c); and unit shipping cost (t) between markets, assumed symmetric in both directions.

The model employs three firm types, with free entry and exit into and out of firm types. Type-m firms are multinationals that maintain plants in both countries.[10] Type-h firms are national firms that maintain a single plant in country h. Type-h firms may or may not export to country f. Finally, type-f

---

[9] The existence of the specific factor R in Y produces a general-equilibrium effect: the wage rate in terms of Y rises as the X sector expands, drawing more labor from Y. This effect "convexifies" the model and tends to limit the concentration of the X sector into one country.

[10] Assume that multinational firms, when they exist in equilibrium, draw their labor for firm-specific fixed costs evenly between countries, so that we make no attempt in this minimal model to associate multinationals with particular countries.

firms are national firms that maintain a single plant in country f. Type-f firms may or may not export to country h.

In the context of this model, consider first two countries absolutely identical in technologies, preferences, and endowments. If transport costs were zero, then there would exist only national firms exporting to each other's markets, since no firm would incur the fixed costs of a second plant. If transport costs were very high, there would exist only multinational (two-plant) firms: in this case, a multinational has lower fixed costs per market and therefore outcompetes national firms, which face prohibitive export costs. At intermediate levels of transport costs, multinational firms exist if firm-specific fixed costs and transport costs are large relative to plant-specific fixed costs (plant-level scale economies).[11] Thus, this model predicts that we should find multinationals concentrated in industries that fit at least one of three conditions: firm-level activities or intangible assets are important; plant scale economies are not particularly important; and tariffs and transport costs are high but barriers to direct investment are relatively low.

The empirical evidence is consistent with these results, but it also indicates that multinationals are of greater importance between countries that are relatively similar in size, per capita income, and relative factor endowments, like western Europe and the United States. It is interesting that the simple model does a good job of capturing the association of direct foreign investment and multinational enterprises with the *similarity* of countries.

Consider the situation where countries differ in size, factor endowments, and technologies. The three diagrams in Figure 1 present simulation results in this spirit, drawn from Markusen and Venables (1995). The symbols in the figures represent various combinations of the three different types of firm types active in equilibrium. The horizontal axis of each diagram is identical; it shows transport costs (symmetric in both directions) expressed as a proportion of marginal production costs. In each of the diagrams, the vertical axis measures a characteristic that can differ across the two countries. In each case, the difference between countries is maximized at the bottom row of the graph, while the two countries are identical in the top row. Similarity of countries increases as we move up a column.

Figure 1a gives the relative country sizes on the vertical axis, with country f arbitrarily chosen as the smaller country. As we move up the graph along a column, the country sizes converge.

Figure 1b gives the factor endowment ratio R/L of country f on the vertical axis, normalized so that the "world" endowment ratio is equal to one. Factors are transferred between the two countries in opposite directions, so as to preserve approximately their relative income levels. In the bottom row, country f is relatively well endowed with land (R), the factor used only in Y, and country h is relatively well endowed with labor. Since good X uses only

[11]This last statement is what Brainard (1993c) refers to as the "proximity-concentration hypothesis."

labor in production, country h, where labor is relatively abundant, will have a clear Heckscher-Ohlin advantage in producing X.

Figure 1c gives the cost in units of labor of producing X in country f relative to country h.[12] In the bottom row of Figure 1c, country h has a strong Ricardian comparative advantage in producing X.

The changes in the set of active firm types as we move up a column in any of the diagrams capture what Markusen and Venables (1995) refer to as the "convergence hypothesis." As countries become more similar in size, relative factor endowments, and technical efficiency (the last two determining per capita income), international economic activity will become increasingly dominated by multinationals, which displace trade, provided that transport costs are not very small.

It should be emphasized that the displacement of trade by direct foreign investment is not trivially true due to the fact that trade would disappear as the countries converge, even without direct foreign investment. In this type of industrial-organization model, interindustry trade (Y for X) diminishes as we move up a column in any of the three diagrams, but intra-industry trade rises (X for X). In fact, if multinational production is ruled out—as it is in so many of the "new trade theory" models—intra-industry trade may rise faster than trade in Y diminishes. For example, as we move up the column for $t = .03$ in Figure 1a, which has no multinational activity, the overall volume of trade increases.
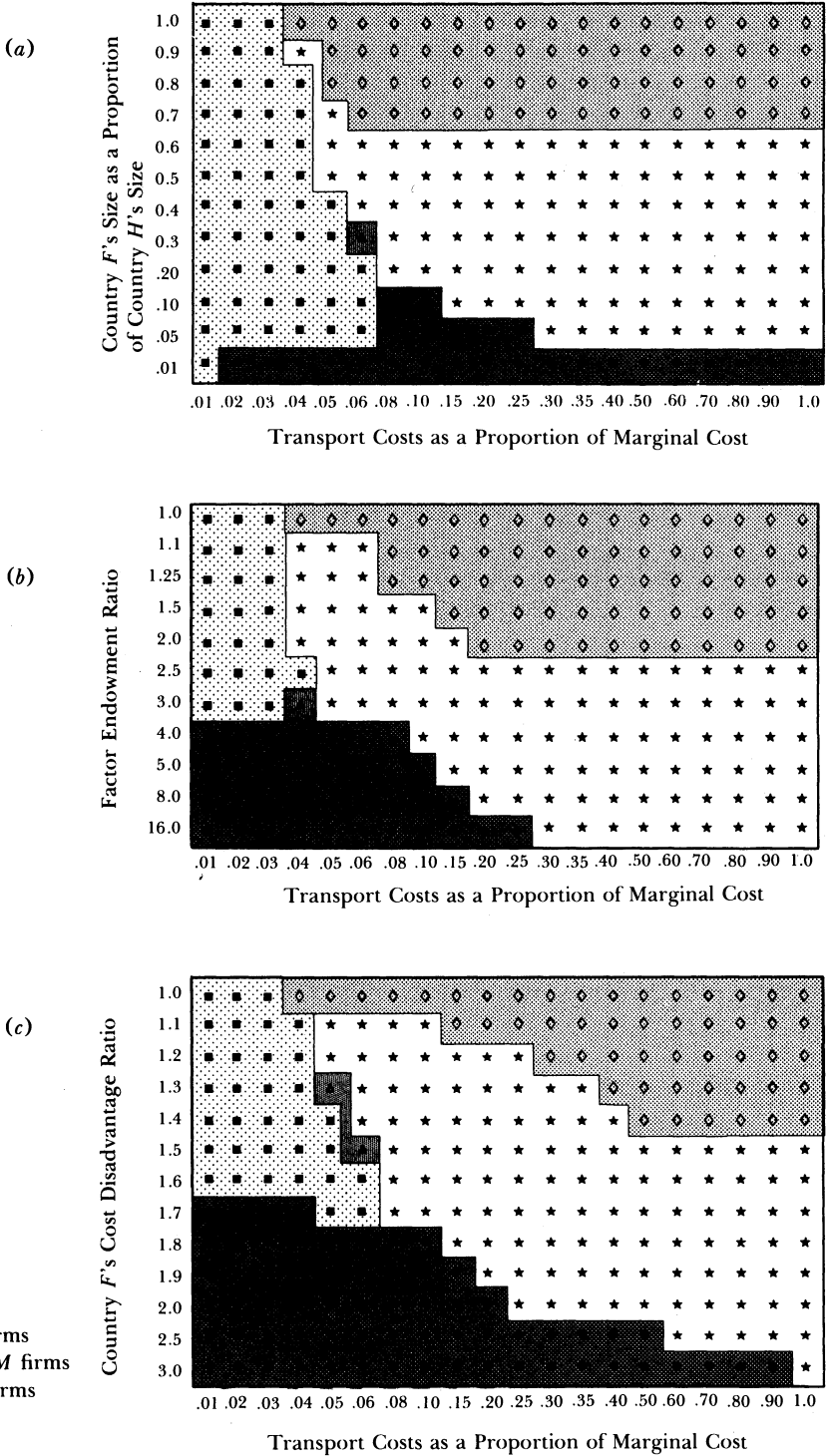
Now consider the column corresponding to $t = 0.6$ in Figure 1a, beginning in the top row in which the countries are identical. Parameter values are chosen such that only type-m firms exist in the top row at this level of t; roughly speaking, F and t (and market sizes) are large relative to G. Now reduce the size (proportionately decrease the factor endowment) of country f, moving down the column.

The fall in the size of country f reduces the profits of type-m firms, causing the exit of some firms and an increase in the price of X in both countries (the latter due to higher markups due to higher concentration). But the decreased size of country f has a smaller effect on the profits of *potential* type-h firms, since their sales are concentrated in the larger market, country h. As prices rise due to the exit of some type-m firms, eventually some type-h firms can enter the market. The set of active firm types shifts to type-m and type-h firms coexisting.

Continuing the movement down a column such as $t = .06$ in Figure 1a, the next shift in regime is due essentially to two effects. As production of X by type-m firms falls in country f, there is an increasing price difference $p_f > p_h$, where p_i is the price of X in terms of Y in country i. Second, the fall in the demand for labor to produce X in country f causes a fall in the wage in

---

[12]To be more precise, firm-specific fixed costs, plant-specific fixed costs, and constant marginal costs are all inflated equally in country f as we move down a column.

Figure 1
When Countries Differ in Size, Factor Endowment, and Technology

(a)

Country F's Size as a Proportion of Country H's Size

Transport Costs as a Proportion of Marginal Cost

(b)

Factor Endowment Ratio

Transport Costs as a Proportion of Marginal Cost

(c)

Country F's Cost Disadvantage Ratio

● H firms
▪ H and F firms
△ H, F, and M firms
★ H and M firms
◊ M firms

Transport Costs as a Proportion of Marginal Cost

country *f*. Eventually, type-*f* firms can enter, their highest domestic price and lower wage compensating for the fact that their low marginal cost domestic market is small.

A further fall in the size of country *f*, however, eventually means that type-*f* firms cannot exist, and all world production is by type-*h* firms, which export a small amount to country *f*. Intuitively, as the size of country *f* goes to zero, all of the sales of both type-*h* and type-*f* firms would have to be in country *h*. But type-*f* firms are high marginal cost suppliers due to the transport cost, and hence type-*f* firms will not exist unless the general-equilibrium effect on the wage in country *f* is very strong.

The stories in Figures 1b and 1c are similar for moderate levels of transport costs. When the countries are very similar in relative endowments or technologies, multinational firms dominate. When there is a moderate degree of difference, type-*m* firms coexist with national firms of the "advantaged" country *h* (advantaged in terms of relative factor endowment or lower unit labor costs in *X*). When the degree of difference is very large, only type-*h* firms are supported in equilibrium. Figure 1 illustrates the convergence hypothesis of Markusen and Venables (1995) mentioned above; multinationals displace national firms and trade as countries become more similar in size, technology, and relative factor endowments. The simplest intuition seems to be that, when the countries are very different in any or all of these characteristics, multinational firms cannot complete against single-plant national firms in the "advantaged" country, which serve the "disadvantaged" countries by exports. These single-plant firms derive their advantage from the fact that their production is concentrated in the country in which sales are larger, factor costs are lower, and/ or real factor productivity is higher. When the countries are quite different, the multinationals derive their disadvantage from having to locate costly additional "capacity" in the small and/ or costly market.

This model is appealing in its consistency with both the micro and macro empirical facts. In addition, it has some implications that contrast rather sharply with the national firm-based models of the new trade theory and the geography and trade literature. In particular, this model with multinational firms has differing implications for the observed volume of intra-industry trade between two nations as they converge in size, endowments, and technologies. The national firm models, characteristic of the writings of Helpman, Krugman, and Venables, suggest a growing volume of intra-industry trade as the countries converge. The multinational models of Brainard, Horstmann and Markusen, and Markusen and Venables suggest that the volume of intra-industry trade at first rises (for example, move up to any column from $t = 0.04$ to 0.1 in Figure 1a) and then falls to zero as multinationals displace trade.

To the best of my knowledge, these contrasting predictions have not as yet been tested, but some data on U.S.-European trade and investment lends support to at least pursuing these ideas. In the 26-year period from 1966–1991, two-way trade between the U.S. and Europe (developed market economies)

rose 12.8 times in nominal value, while the value of direct foreign investment stocks (U.S.-owned in Europe and European-owned in the United States) rose 20.3 times in value. The value of European exports to the United States and the value of European direct foreign investment stocks in the United States were approximately equal as late as 1980, but the value of European direct foreign investment stocks in the United States was 2.5 times the value of European exports to the United States by 1991 (Markusen and Venables, 1995). Although far from decisive, such evidence may indicate that a process of multinationals displacing trade has begun.

Internalization

Even if multinational production makes economic sense, in the terms of the models laid out in the previous section, there is a further question to tackle. A firm might be able to realize many of the advantages of multinational production, while shielding itself from the costs, by signing a licensing agreement with a firm in the foreign country. Thus, a complementary part of the argument must explain why firms choose direct investment, rather than licensing.

There is a small literature on whether a firm's decision to transfer a firm-specific asset (or the services thereof) should happen within a particular firm or through an arm's length licensing agreement with an independent foreign firm. As one might expect, the optimal scope of a firm is determined by factors like the form of corporate governance, the cost of internal transactions versus those in arm's length markets, and the specific characteristics of the knowledge and information to be transferred, along with resulting market failures involving concepts like bounded rationality, agent opportunism, and asset specificity (Williamson, 1975, 1981).

I will limit my discussion here to some fairly specific, formal models produced by international trade economists, using the broader notions developed by Williamson (1975, 1981), Casson (1987), Rugman (1986), Teece (1977, 1986), and others. The set of ideas I will discuss is quite compatible and complementary with the ownership-location model developed above, in addition to fitting well with most of the micro empirical facts laid out earlier.

Many or most of the reasons to transfer assets internally arise from the basic property that knowledge capital can be a joint input to a number of plants; this same property, the reader will recall, was used in the ownership-location models and is consistent with the association of multinational enterprises with R&D, advertising, and product newness and complexity. A number of papers show quite convincingly that transfers tend to be internal, rather than arm's length, when the products are new, complex, have no prior commercial application, and are produced by R&D-intensive firms (Davidson and McFetridge, 1984; Mansfield and Romeo, 1980; Teece, 1977; Wilson 1977).

Thus the same features that create multiplant economies of scale may be responsible for creating advantages of internalization.

But although models of internalization do share an underlying commonality with the ownership-location models discussed in the previous section, internalization models tend to be somewhat different. Because they focus on characteristics of knowledge capital like non-excludability, asymmetric information, moral hazard, adverse selection, and incomplete contracting, the models are often partial equilibrium in nature and bring to bear quite different tools of analysis. In what follows, I will outline some of the ideas that have been advanced in formal models of internalization and present a simple analytical example at the end, much the same as I did in the section on ownership and location.

All of the models generally share a point of departure discussed earlier: firms would like to license due to the costs of doing business abroad, but licensing carries costs as well. These models can be categorized according to how they draw the link from information issues to difficulties in licensing and thus why direct foreign investment occurs.

A first problem is that because of the non-excludability property of new knowledge, a firm may not want to reveal its process or product technology to a potential licensee. After all, the licensee could reject a deal and go and copy the technology at little cost. Conversely, the licensee is not going to deal without knowing exactly what it is getting, which requires revelation on the part of the seller. No licensing deal can be reached under these circumstances, so the technology is transferred instead to an owned subsidiary (Ethier, 1986). More general discussions of buyer uncertainty of this type can be found in Teece (1986) and Rugman (1986).

A second problem involves another sort of informational asymmetry between the firm and the potential licensee, one which particularly affects new or complex products. The firm may know a great deal about the product or the process (such as the quality of the product), but the licensee recognizes that the firm may not have an incentive to reveal the product's quality truthfully. Of course, there exist the standard mechanism-design arguments to motivate full revelation, but the necessary contingent contracts might be difficult to write, particularly when there are multiple dimensions to the uncertainty. Internalization may be preferred to costly and/or incomplete contracts (Ethier, 1986). More general discussions of contracting costs and agent opportunism may be found in the already mentioned writings of Buckley, Casson, Rugman, Teece, and Williamson.

A third informational asymmetry associated with newness focuses on the case where the potential licensee has superior information, usually about how the product will sell in its local market. The multinational enterprise is reluctant to build a foreign plant without information about whether sales will be high or low, information that could be provided by the foreign agent. But the agent knows that if it reveals demand to be high, the firm may decide to

produce directly, or a large share of the rents will be extracted from the foreign agent in subsequent periods. Thus, agent's incentives can cause sales to be low even when demand is high. The multinational enterprise can avoid having to share rents with the licensee by direct investment (Horstmann and Markusen, 1995). In fact, many firms do set up foreign wholesaling and servicing subsidiaries, possibly to deal with this sort of problem (Nicholas, 1983; Zeile, 1993).

A fourth problem is that the same property that makes knowledge easy to transfer internationally may mean that it is easily learned by new employees. If a firm licenses a technology to a foreign producer, the managers and workers may learn the technology quickly and be able to "defect," starting a new domestic firm in competition with the multinational enterprise. While this problem would exist to some extent within a firm as well, it is argued that a firm may more credibly commit than can a licensee to sharing the rents from a string of (uncertain) future products with the employees (Ethier and Markusen, 1993). Although many writers have discussed this problem in general terms, Rugman (1985, 1986) in particular views this as a cornerstone of internalization theory.

A fifth problem focusses on the costs of transferring technology. Certain aspects of a knowledge-intensive technology are bound up in the human capital of a firm's employees and even in the "company culture" (Teece, 1977, 1986). Such technology is costly to transfer arm's length, which does not contradict the possibility mentioned in the previous paragraph that the technology's value could be easily dissipated once the transfer does take place.

A sixth potential problem for licensing arises when the firm's intangible asset is a reputation for product quality. Product quality may only be observed after the product is purchased and used by the buyer. In this situation, the multinational enterprise cannot extract all rents from a licensee because, if it attempts to do so, the licensee can skimp on quality by producing an inferior substitute product for one period and earn positive single-period rents. To avoid this problem, it may be profitable to produce and sell through an owned subsidiary despite the added direct costs (Horstmann and Markusen, 1987b). This problem arises especially in franchising, where the firm wants a uniform level of quality across outlets. Each outlet manager (licensee) has an incentive to free ride on the reputation of the whole (Caves and Murphy, 1976).

Finally, when a firm employs licensees, it must be concerned about the problem of moral hazard. For example, licensees may divert selling effort to competing products of other firms or simply shirk (Nicholas, 1983; Mathewson and Winter, 1985). Of course, these problems can occur within firms as well. Carlos (1994) and Carlos and Nicholas (1990) document how private trading on the part of agents caused difficulties for the Hudson's Bay and Royal African Companies, and how the Hudson's Bay Company was able to create an internal structure and company culture to mitigate the moral hazard and attendant losses. On the other hand, the Royal African Company went bankrupt. Of course, intensive monitoring is one way to deal with licensees, but if a firm is

going to monitor licensees with great care, it may be easier simply to own the foreign operation outright.

To add some concreteness to the idea of asset dissipation, we can outline a highly simplified version of Ethier and Markusen (1993). Consider a simple two-period model in which the multinational wishes to exploit a technology in a foreign market by licensing a foreign firm or by setting up a subsidiary (we will ignore exporting in this example). Because of the costs of doing business abroad, a licensing arrangement generates the most potential rents. The licensee masters the technology in the first period and can defect to start a rival firm in the second period. Similarly, the multinational can "defect" by issuing a license to a second firm in the second period. In other words, we make the strong assumption that no binding contracts can be written to prevent either firm from undertaking such a defection. We will assume here, with no justification or discussion, that defection cannot occur from within a subsidiary: that is, a part of a subsidiary will not split off to form another competitor. (This assumption is relaxed in Ethier and Markusen, 1993.) A subsidiary is thus (by definition!) costly but "secure."

At the beginning of the second period, the multinational and the licensee make simultaneous moves, choosing whether to continue their original relationship. If both the multinational and the licensee defect, then the original licensee and the new licensee will compete as duopolists in the second period. For a two-period licensing contract to be self-enforcing, neither the multinational enterprise nor the licensee must wish to defect in the second period.

For the sake of illustration, and with some loss of generality, let us make some assumptions about the rents available in these different scenarios. If the licensee continues for both periods, let us refer to the total rents as $2R - F$, where R is the rents available in each period and F is the physical capital cost that the licensee (or multinational) must incur to start production. (For simplicity, this example assumes no discounting.) If the multinational sets up a subsidiary, then the rents will be $2M - F$, where M represents the rents received when the subsidiary operates on its own. We will assume that $2R - F > 2M - F$, which just means that the rents are larger if the licensing agreement continues. This assumption that $R > M$ captures the idea discussed earlier that there are costs to establishing a business abroad.

The third situation is where the one-period license is followed by duopoly. In this case, the rents will be $R + D - 2F$, where D represents the total rents for both members of the duopoly in one period, and the capital costs F must be doubled because with two separate producers, the start-up costs must be incurred twice. For the purposes of this example, we posit that the rents from the duopoly option are lowest of these three scenarios; that is, $(2R - F) > (2M - F) > (R + D - 2F)$.

As a final piece of notation, consider the licensing fee, which we will refer to as L_1 in period 1, and L_2 in period 2.

In this setting, what conditions must hold so that the licensing arrangement continues through both periods? For the license to continue, it must be better than the alternative both from the point of view of the multinational and from the point of view of the subsidiary. Let us posit that if one partner defects, that partner must incur the additional costs of F, the non-defecting partner retaining the original F.

For the licensee not to defect and start up production on its own, its second-period earnings ($R - L_2$, with no additional start-up costs) must be at least equal to ($R - F$), its payoff from defecting. For the multinational enterprise, its licensing fee L_2 must be at least equal to ($R - F$), its payoff from defecting. Combining these two inequalities, licensing will be continued in the second period if $R < 2F$; that is, if the rent is no greater than twice the fixed costs.

Furthermore, if the $R < 2F$ condition holds, then the multinational can extract all rents from its licensee. In the second period, the fee $L_2 = F$ is the largest fee that the multinational can charge without causing the licensee to defect, and such a fee will also lead the firm to honor the agreement. The fee $L_2 = F$ leaves the licensee with rents ($R - L_2$) = ($R - F$) in the second period. The multinational can extract these with a fee $L_1 = 2R - F$ in the first period. In other words, the fee schedule $L_1 = 2R - F$ and $L_2 = F$ satisfies the (incentive compatibility) condition that neither partner will wish to defect in the second period and the (participation or individual rationality) condition that the licensee earns nonnegative profits (exactly zero in this case). Notice that $L_1 + L_2 = 2R$, which is to say that all the rents are collected by the multinational through the license fees. To sum up, if the condition $R < 2F$ holds, then the multinational will license, and it will earn all of the rents.

If the $R < 2F$ condition fails to hold—that is, if the rents are greater than twice the fixed costs—then both the firm and the licensee will defect in the second period. In this case, a duopoly game will result in the second period between the original and a second licensee. Assume that ownership of the original fixed cost F remains with the multinational. Then, the original licensee, now on its own, generates a net second-period income of $D/2 - F$ while the second licensee generates $D/2$ (using the original capital stock F). Knowing that defection is coming in the second period, the multinational is limited in what it can charge in the first period. All it can do is charge the first licensee a first-period fee of $L_1 = R + D/2 - F$, which just means that the most the multinational can demand is the second-period profits of the prospective defector. For the same reason, the multinational can charge the second licensee a second-period fee of $L_2 = D/2$. In this case, the total two-period profit for the multinational is ($L_1 + L_2 - F$) = ($R + D - 2F$). Both licensees earn zero profits under this fee schedule, but while the multinational captures all rents, additional fixed costs are incurred, and some rents are dissipated by the duopoly competition. Thus, if the licensing condition fails to hold, the

multinational will seek to avoid this duopoly outcome, and instead will set up a subsidiary. Remember, our earlier assumption was that the rents of a subsidiary arrangement are $2M - F$, which exceeds the rents of duopoly $R + D - 2F$.

Finally, consider the situation where $F = 0$. This can be interpreted as the case of a "pure" knowledge-capital technology; that is, when $F = 0$, the licensee can costlessly enter production in the second period after one period of learning by doing. Under the assumption that $F = 0$, it is clear that $R < 2F$ will fail to hold, and licensing will not sustain itself. As a result, the multinational chooses a costly subsidiary over a rent-dissipating licensing contract. We thus have a result that is consistent with both the theoretical ideas developed here and with some of the micro facts listed earlier. Direct investment in a subsidiary is more likely in cases where the technology has the joint-input characteristic of knowledge capital.

Future Research

The framework presented in this paper may seem quite tidy. The notion of knowledge capital, with its jointness characteristic and attendant risks of dissipation, offers a basis for both ownership and internalization advantages. In addition, it fits well with many of the micro facts regarding characteristics of multinational firms, and generates predictions consistent with many of the macro facts, as discussed in connection with Figure 1.

However, much remains to be done. First, we would like to have more direct empirical evidence rather than relying on proxy variables such as R&D to test the theory. Case studies would be valuable in helping to understand exactly what services parent firms are supplying to subsidiaries. Second, much more work needs to be done on vertical investments in which the production process is geographically decomposed into stages. This type of direct investment is growing rapidly as many developing countries liberalize their trade and investment laws. Why is it necessary for a firm to own the separate stages of production? What are the sources of ownership and internalization advantages for an offshore assembly plant? Third, joint ventures need attention. What market alternatives are being rejected in favor of joint ventures? What problems are being internalized? Do technological and information advantages outweigh possible anticompetitive consequences?

Finally, a good deal of normative and policy analysis needs to be done, qualifying and modifying the strategic trade policy literature. Foreign ownership obviously qualifies rent-shifting arguments, but we need to move beyond this point to consider making endogenous the two-way causality between policy and the existence of foreign ownership. Too much of the taxation literature simply assumes the (exogenous) existence of foreign plants and considers the effect of taxes on marginal price and output decisions. Tariffs and other trade barriers can induce inward direct investment, or domestic taxation discourage

it, with nonobvious welfare consequences. Tax competition between regions for footloose investment needs to be examined. Conversely, direct investment can modify the political economy context within which trade policy is made by sharply altering the incentives facing domestic firms.

References

Beaudreau, Bernard C., "Managers, Learning and the Multinational Firm: Theory and Evidence," unpublished Ph.D dissertation, University of Western Ontario, 1986.

Blomstrom, Magnus, and Robert Lipsey, "Firm Size and Foreign Operations of Multinationals," *Scandinavian Journal of Economics*, 1991, *93*:1, 101–7.

Blomstrom, Magnus, Robert Lipsey, and Ksenia Kulchycky, "U.S. and Swedish Direct Investment and Exports." In Baldwin, Robert E., ed., *Trade Policy Issues and Empirical Analysis*. Chicago: University of Chicago Press, 1988, pp. 259–97.

Blomstrom, Magnus, and Mario Zejan, "Why Do Multinational Firms Seek Out Joint Ventures," *Journal of International Development*, 1991, *3*:1, 53–63.

Brainard, S. Lael, "A Simple Theory of Multinational Corporations and Trade with a Trade-off between Proximity and Concentration." NBER Working Paper No. 4269, February 1993a.

Brainard, S. Lael, "An Empirical Assessment of the Factor Proportions Explanation of Multinationals Sales." NBER Working Paper No. 4580, December 1993b.

Brainard, S. Lael, "An Empirical Assessment of the Proximity-Concentration Tradeoff between Multinational Sales and Trade." NBER Working Paper No. 4583, December 1993c.

Buckley, Peter J., and Mark Casson, *The Future of the Multinational Enterprise*. London: Macmillan, 1976.

Buckley, Peter J., and Mark Casson, *The Economic Theory of the Multinational Enterprise*. London: Macmillan, 1985.

Carlos, Ann, "Bonding and the Agency Problem: Evidence from the Royal African Company 1672–1691," *Explorations in Economic History*, July 1994, *31*, 313–35.

Carlos, Ann, and Stephen Nicholas, "Agency Problems in Early Chartered Companies: The Case of the Hudson's Bay Company," *Journal of Economic History*, December 1990, *50*:4, 853–75.

Casson, Mark, *The Firm and the Market: Studies in Multinational Enterprises and the Scope of the Firm*. Oxford: Blackwell, and Cambridge: MIT Press, 1987.

Caves, Richard E., *Multinational Enterprise and Economic Analysis*. London: Cambridge University Press, 1982.

Caves, Richard E., and William F. Murphy II, "Franchising: Firms, Markets, and Intangible Assets," *Southern Economic Journal*, April 1976, *42*, 572–86.

Davidson, William H., and Donald G. McFetridge, "International Technology Transactions and the Theory of the Firm," *Journal of Industrial Economics*, March 1984, *32*, 253–64.

Denekamp, Johannes, and Michael J. Ferrantino, "Substitution and Complementarity of U.S. Exports and Foreign Based Affiliate Sales in a Demand Based Gravity System," working paper, 1992.

Dick, Andrew R., "Strategic Trade Policy and Welfare: The Empirical Consequences of Cross-Ownership," *Journal of International Economics*, November 1993, *35*, 227–49.

Dunning, John H., *International Production and the Multinational Enterprise*. London: George Allen and Unwin, 1981.

Dunning, John H., "Trade, Location of Economic Activity and MNE: A Search for an Eclectic Approach." In Ohlin, B., P. O. Hesselborn, and P. M. Wijkman, eds., *The International Allocation of Economic Activity*. London: Macmillan, 1977, pp. 395–418.

Ethier, Wilfred J., "The Multinational Firm," *Quarterly Journal of Economics*, November 1986, *101*, 805–33.

Ethier, Wilfred J., "Multinational Firms in the Theory of International Trade." In Bacha, E., ed., *Development, Trade and the Environment*. London: Macmillan, forthcoming 1995.

Ethier, Wilfred J., and James R. Markusen, "Multinational Firms, Technology Diffusion and Trade," NBER working paper, 1993.

Froot, Kenneth A., *Foreign Direct Investment*. Chicago: University of Chicago Press, 1993.

Graham, Edward M., and Paul R. Krugman, "The Surge in Foreign Direct Investment in the 1980s." In Froot, Kenneth A., ed., *Foreign Direct Investment*. Chicago: University of Chicago Press, 1993, pp. 13–36.

Grubaugh, S., "Determinants of Direct Foreign Investment," *Review of Economics and Statistics*, February 1987, *69*, 149–51.

Helpman, Elhanan, "A Simple Theory of Trade with Multinational Corporations," *Journal of Political Economy*, June 1984, *92*, 451–71.

Helpman, Elhanan, "Multinational Corporations and Trade Structure," *Review of Economic Studies*, July 1985, *52*, 443–57.

Helpman, Elhanan, and Paul R. Krugman, *Market Structure and Foreign Trade*. Cambridge: MIT Press, 1985.

Horstmann, Ignatius J., and James R. Markusen, "Strategic Investments and the Development of Multinationals," *International Economic Review*, February 1987a, *28*, 109–21.

Horstmann, Ignatius J., and James R. Markusen, "Licensing Versus Direct Investment: A Model of Internalization by the Multinational Enterprise," *Canadian Journal of Economics*, August 1987b, *20*, 464–81.

Horstmann, Ignatius J., and James R. Markusen, "Endogenous Market Structures in International Trade," *Journal of International Economics*, February 1992, *32*, 109–29.

Horstmann, Ignatius J., and James R. Markusen, "Exploring New Markets: Direct Investment, Contractual Relationships, and the Multinational Enterprise," *International Economic Review*, forthcoming 1995.

Hummels, David L, and Robert M. Stern, "Evolving Patterns of North American Merchandise Trade and Foreign Direct Investment, 1960–1990," *The World Economy*, January 1994, *17*:1, 5–29.

Hymer, Stephen H., *The International Operations of National Firms: A Study of Direct Foreign Investment*. Cambridge: MIT Press, 1976.

Julius, DeAnne, *Global Companies and Public Policy*. London: Royal Institute of International Affairs, 1990.

Kindleberger, Charles P., *American Business Abroad*. New Haven: Yale University Press, 1969.

Kindleberger, Charles P., *Multinational Excursions*. Cambridge: MIT Press, 1984.

Levinsohn, James A., "Strategic Trade Policy when Firms can Invest Abroad: When are Tariffs and Quotas Equivalent?," *Journal of International Economics*, August 1989, *27*:1/2, 129–46.

Lipsey, Robert E., "Foreign Direct Investment in the United States: Changes over Three Decades." In Froot, Kenneth A., ed., *Foreign Direct Investment*. Chicago: University of Chicago Press, 1993, pp. 113–72.

Lipsey, Robert E., "Outward Direct Investment and the U.S. Economy." NBER Working Paper No. 4691, March 1994.

Mansfield, Edwin, and Anthony Romeo, "Technology Transfer to Overseas Subsidiaries by U.S.-Based Firms," *Quarterly Journal of Economics*, December 1980, *94*, 737–50.

Markusen, James R., "Multinationals, Multi-Plant Economics, and the Gains from Trade," *Journal of International Economics*, May 1984, *16*, 205–26.

Markusen, James, R., Thomas F. Rutherford, and Linda Hunter, "Trade Liberalization in a Multinational-Dominated Industry," *Journal of International Economics*, forthcoming 1995.

Markusen, James R., and Anthony J. Venables, "The Increased Importance of Multinationals in North American Economic Relationships: A Convergence Hypothesis." In Canzoneri, Matthew W., Wilfred J. Ethier, and

Vitoria Grilli, eds., *The New Transatlantic Economy*. London: Cambridge University Press, forthcoming 1995.

Mathewson, G. Frank, and Ralph A. Winter, "The Economics of Franchise Contracts," *Journal of Law and Economics*, October 1985, *28*, 503–26.

Morck, Randall, and Bernard Yeung, "Why Investors Value Multinationality," *Journal of Business*, 1991, *64*:2, 165–87.

Morck, Randall, and Bernard Yeung, "Internalization: An Event Study Test," *Journal of International Economics*, August 1992, *33*:1/2, 41–56.

Motta, Massimo, "Multinational Firms and the Tariff-Jumping Argument," *European Economic Review*, December 1992, *36*, 1557–71.

Nicholas, Stephen, "Agency Contracts, Institutional Modes, and the Transition to Foreign Direct Investment by British Manufacturing Multinationals Before 1939," *Journal of Economic History*, September 1983, *43*, 675–86.

Rugman, Alan M., *Inside the Multinationals: The Economics of Internal Markets*. London: Croom Helm, and New York: Columbia University Press, 1981.

Rugman, Alan M., "Internalization is Still a General Theory of Foreign Direct Investment," *Welwirtschaftliches Archiv*, 1985.

Rugman, Alan M., "New Theories of the Multinational Enterprise: an Assessment of Internalization Theory," *Bulletin of Economic Research*, 1986, *38*:2, 101–18.

Smith, Alasdair, "Strategic Investment, Multinational Corporations, and Trade Policy." Centre for Economic Policy Research Discussion Paper No. 137, September 1986.

Teece, David, *The Multinational Corporation and the Resource Cost of International Technology Transfer*. Cambridge: Ballinger, 1986.

Teece, David, "Technology Transfer By Multinational Firms: The Resource Cost of Transferring Technological Know-How," *Economic Journal*, June 1977, *87*, 242–61.

UNCTAD, *World Investment Report 1993: Transnational Corporations and Integrated International Production*. New York: United Nations, 1993.

Wheeler, David, and Ashoka Mody, "International Investment Location Decisions," *Journal of International Economics*, August 1992, *33*, 57–76.

Williamson, Oliver E., *Markets and Hierarchies, Analysis and Antitrust Implications: A Study in the Economics of Internal Organization*, New York: Free Press, 1975.

Williamson, Oliver E., "The Modern Corporation: Origins, Evolution, Attributes," *Journal of Economic Literature*, 1981, *19*:4, 1537–68.

Wilson, Robert W., "The Effect of Technological Environment and Product Rivalry on R&D Effort and Licensing of Inventions," *Review of Economics and Statistics*, May 1977, *59*, 171–78.

Zeile, William J., "Merchandise Trade of U.S. Affiliates of Foreign Companies," *Survey of Current Business*, October 1993, *73*, 52–65.