This chapter critically reviews what many scholars (e.g. Caves, 1996, 1998) see as the dominant explanation for the multinational enterprise (MNE), the transaction costs approach. An MNE is fundamentally an organization that extends employment contracts over national boundaries. In other words, the distinguishing characteristics of MNEs are their use of hierarchical methods of coordination (managerial directives) to organize cross-national interdependencies. Market prices are another method of organizing international interdependencies. To understand why MNEs exist, one must therefore explain why individuals located in separate countries are more efficiently coordinated when they are employees of an MNE than if they are independent entrepreneurs responding to market prices. This chapter argues that transaction costs theory provides the best and more parsimonious explanation of why this is the case, and hence the best reason for the existence and development of MNEs.

This chapter does not pretend to be an exhaustive survey of the literature on the transaction costs approach to international business. Rather it shows how transaction costs theory can provide a unifying paradigm to explain the most common forms taken by the expansion of firms across national boundaries. Owing to space limitations, not all applications of transaction costs theory to the MNE are covered, and the analysis focuses mostly on statics. After a brief outline of the transaction costs model, I show how it can account for the main motives behind the international expansion of firms. I then explain how some of the unconventional trade practices used by developing countries are second-best strategies used in response to government limitations on the expansion of MNEs. The third section shows how transaction costs theory can throw light on two of the main choices faced by MNEs entering foreign markets. The conclusion outlines the differences between the approach taken here and some of the other explanations of the MNE.

It is important at the outset to explain the differences between two distinct, yet related, phenomena. An MNE is a firm that has employees abroad. When it finances its activities abroad with funds obtained from its home country, it undertakes foreign direct investment (FDI). FDI is a country’s export and import of long-term capital into investments controlled by its residents (as opposed to portfolio investments), as recorded in the capital account of its balance of
payments. There is no one-to-one correlation between the expansion of firms across national boundaries (the growth of MNEs) and flows of FDI, since MNEs can (and do) obtain the funds necessary for their foreign expansion from the foreign country where they invest. FDI is therefore a rather poor indicator of the growth of MNEs.

Transaction costs as a general theory of the MNE

In a chapter in *Multinational Enterprise in Historical Perspective*, D.K. Fieldhouse writes that

Historical research has shown that the tidy logic of growth of firms theory, the main intellectual foundation for the concept of the MNC [multinational corporation], simply does not fit the unruly variety of corporate motivation. It is critical that most early theorizing was based on recent American experience, mainly in manufacturing and petroleum, and even this was not uniform. Still more variable were the historical reasons for US and European investment in overseas mining, plantations, and utilities. In none of these last did FDI commonly or necessarily flow from the growth patterns of metropolitan firms. In many of them capitalists established new corporate enterprises without a previous home base, specifically to produce a commodity which could only be obtained elsewhere, or to exploit an evident overseas need for public services, such as railways or telecommunications in Latin America. Motives for FDI were therefore infinitely more complex than any unitary theory of the MNC could possibly comprehend, and had no necessary connection with the internalization concept.

(Fieldhouse, 1986: 25)

Are motives for the MNE so disparate that a general theory of the MNE cannot be constructed? This section demonstrates that transaction costs theory can in fact account for the wide variety of forms taken by the expansion of MNEs.

Structural and natural market imperfections

MNEs present a paradox. Operating overseas is usually more costly than operating at home, because a foreigner does not have the same contacts and knowledge of local customs and business practices as indigenous competitors, while being often subject to discrimination by host country governments and private institutions. Hence it is difficult to understand why firms based in one country would do business in another country. If a firm has some unique assets of value overseas, why not sell or rent these assets to local entrepreneurs, who could then combine them with local factors of production at lower costs than those experienced by foreign direct investors?

The answer to this paradox is that there might be circumstances under which using market exchange to coordinate the behavior of agents located in two
Jean-François Hennart

separate countries is less efficient than organizing their interdependence within a multinational firm. When this is the case, a firm located in one country may find it profitable to incur the additional costs of operating in a foreign environment. This idea that MNEs owe their existence to “market imperfections” was first put forward by Hymer (1960, published 1976), Kindleberger (1969), and Caves (1971). The market imperfections they had in mind were, however, “structural” imperfections in markets for final products. Hymer, for example, considered two firms, each a final product monopolist in its own market, isolated from competition by high transportation costs and tariff and non-tariff barriers. A decline in these costs exposed them to each other’s competition and reduced their profits. A combination of the two firms, through merger or acquisition, into an MNE would then maximize their joint income by forcing them to take into account the gains and the losses competition inflicts on them. The transformation of two domestic firms into one MNE thus internalized “pecuniary” externalities and produced a gain for the owners of these two firms, but not necessarily for society, since it redistributed income towards the MNE and away from its customers.

A similar case arose when the technology developed by firms based in one country was valuable abroad. That technology had often few substitutes and the number of potential licensees in any given foreign market was also often limited, thus creating a bilateral monopoly. The consolidation of licensor and licensee within an MNE (by acquisition or merger of the potential licensee or by vertical integration of the innovator into overseas manufacturing) reduced haggling and made it easier to enforce price discrimination schemes across countries (Hymer, 1976: 49–50). This analysis of the reasons behind the emergence of multinational firms led Hymer to take a negative view of MNEs, which he considered “an instrument for restraining competition between firms of different nations” (Hymer, 1970: 443).

Hymer’s world is one where market imperfections are structural, arising from structural deviations from perfect competition in the final product market due to exclusive and permanent control of proprietary technology, privileged access to inputs, scale economies, control of distribution systems, and product differentiation (Bain, 1956), but in their absence markets are perfectly efficient. By contrast, the insight of transaction costs theories of the MNE, simultaneously and independently developed in the 1970s by McManus (1972), Buckley and Casson (1976), Brown (1976), and Hennart (1977, 1982), is that “market imperfections” are inherent attributes of markets, and MNEs are institutions to bypass these imperfections. Markets experience “natural” imperfections, i.e. imperfections that are due to the fact that the implicit neoclassical assumptions of perfect knowledge and perfect enforcement are not realized (Dunning and Rugman, 1985; Teece 1981).

**Market transaction costs**

The dominant method of organization in markets is the price system. If knowledge were perfect and individuals perfectly honest, organizing interdependencies
Transaction costs and the multinational 75

between individuals would be costless; in other words, transaction costs would be nil. The three tasks that must be performed by any system of organization – to inform individuals of the needs of others, to reward them for productive behavior, and to curb bargaining – would be costlessly performed by prices. Through prices everyone would be fully appraised of everyone else's needs, and incited to adapt to those needs so as to maximize social welfare. With a large number of buyers and sellers, prices would be exogenous, thus curbing bargaining. Prices would also reward agents in proportion to their output (measured at market prices): someone who takes the day off would see his income correspondingly reduced.

In practice, opportunism and bounded rationality make prices not fully efficient, and market transaction costs (the sum of information, enforcement, and bargaining costs) positive. This is because humans have "bounded rationality" and a tendency toward opportunism (Williamson, 1975). Bounded rationality means that humans do not have infinite intellectual abilities and are not perfectly knowledgeable. With bounded rationality, the value of the goods and services exchanged will never be perfectly measured; hence prices will provide flawed signals, and a price system will not maximize the social product. Agents will generate non-pecuniary externalities. Opportunism means that individuals cannot always be expected to take actions that increase the welfare of others at their own expense. Because of bounded rationality, transactors cannot always predict who will be opportunistic and who will not. Positive measurement costs joined with opportunism will make it possible for individuals to cheat – i.e. given positive detection costs, they will be able to alter the terms of trade to their advantage within a given range without loss of revenue. Lastly, imperfect information will also segment the market, thus making prices endogenous, and bargaining profitable.

On the enforcement side, it is important to understand that the total costs of using prices to organize transactions are those of measuring outputs plus those of failing to do so, i.e. the level of cheating that can be expected from imperfect measurement. At some point the marginal benefit of spending an additional dollar to measure output will be less than the expected loss due to cheating, and further investment in measuring outputs will no longer make sense.

If it is very costly to measure the value of goods and services, and opportunities for bargaining and cheating are therefore high, it may pay to eliminate these opportunities by aligning the interests of the parties, i.e. by reducing the incentives they have to cheat. This can be achieved by breaking the connection between output and performance. The price system can be replaced by a mode of organization in which buyers and sellers no longer profit from their ability to cheat, but instead are rewarded for following the directives of a central party directing the exchange. These directives, which can better reflect the overall costs and benefits of the activity, will supersede flawed market prices. Such a system of organization is called hierarchy (Hennart, 1982, 1986c, 1993a, 1993b).

MNEs use hierarchy to eliminate market transaction costs. By transforming independent entrepreneurs into employees, they reduce their incentives to cheat. MNEs make it possible to organize some interactions which cannot be organized
by prices, thus bringing gains of trade to the interacting parties and a net gain for society. Transaction cost scholars have shown, as we shall see below, that there are circumstances under which MNEs organize the international transfer of inputs more efficiently than markets. Hence while Hymer and Kindleberger see the MNE as a method of maximizing monopoly power, or, in other words, as a way of internalizing pecuniary externalities, for transaction costs theorists, MNEs arise to reduce transaction costs and internalize non-pecuniary externalities.

The distinction between pecuniary and non-pecuniary externalities is subtle but important. Pecuniary externalities arise from structural imperfections in markets for final products, while non-pecuniary externalities can be explained by natural market imperfections, i.e. imperfections in markets for intermediate products.

**The benefits and costs of hierarchy**

Some of the early attempts to explain MNEs have tended to suggest that the presence of transaction costs in international markets was a sufficient condition for the existence of MNEs. But if opportunism and bounded rationality reduce the efficiency of markets, they can also be expected to affect that of firms. Firms will experience positive organization costs that will have to be subtracted from the gains of exchange and coordination. Sometimes these costs will be higher than those experienced by markets. Even if they are lower, they still could be so high as to absorb all of the gains from exchange and no economic interaction will then take place, either within firms or in markets.

For example, during the first part of the nineteenth century, England had a technological advantage comparable to that enjoyed by US firms in the two decades following the Second World War. Property rights in knowledge were even weaker then than in the postwar period, so a theory of the MNE that solely focuses on the market transaction costs of knowledge transfer would predict that British firms would have chosen to set up captive manufacturing operations abroad as the most efficient way to capitalize on their knowledge. Yet while US firms have exploited since 1945 their technological advantage by vertically integrating into foreign manufacture, this was not how British know-how was transferred abroad in the first half of the nineteenth century. Rather, it was transferred through the migration of skilled artisans, who took that know-how with them and set up their own firms overseas. Why this difference in transfer modes? Because the costs of both market and interfirm exchange were so high in the nineteenth century as to make migration the only feasible alternative (Hennart, 1982: 128–30).

To be complete, a transaction costs theory of economic organization should therefore consider *simultaneously* the costs of conducting market exchange (market transaction costs) and those of effecting exchange within the firm (internal organization costs), and show how firms can, in some cases, reduce organization costs below those incurred in markets and below the potential gains of exchange. In other words, alongside a theory of why markets fail, we need a theory of why firms succeed.
Some scholars have argued that firms can be more efficient than markets because they replace failing market prices with internal ones (Buckley, 1983; Rugman, 1981). Although some firms do use prices for some types of internal coordination, this concept of “internal markets” fails to capture the fundamental reason why firms displace markets. Most employees are not guided and rewarded by market prices, but by directives voiced by their superiors, formalized through company rules, or internalized through indoctrination. Even in the relationship between head office and foreign subsidiaries, interdependencies and measurement problems limit the use of internal prices as control mechanisms (Robbins and Stobaugh, 1973: 511; Shapiro, 1984). If firms can be more efficient than markets when the latter fail, it is not because they replicate what markets do, but rather because they use a method of organization that is radically different from that used in markets (Hennart, 1982, 1986c).

As argued earlier, firms can succeed when markets fail because they use a system of organization, namely hierarchy, which has a very different incentive structure. Under hierarchy, parties to the exchange are no longer rewarded for their ability to change the market terms of trade in their favor but, instead, they are paid for carrying out the directives of a central party, the boss. Those directives can be explicit (in the form of direct orders or bureaucratic procedures) or implicit (when, through socialization, employees can be persuaded to do what their bosses want them to do even in the absence of managerial supervision). What makes hierarchy, in some circumstances, a more efficient method of organization than prices is the fact that it eliminates market transaction costs by breaking the connection between output and performance. This has, however, one unavoidable consequence: it tends to change the incentives of agents. Since they no longer are paid in proportion to the output they generate, but now in function of their obedience to managerial directives, they will tend to exert less spontaneous effort and initiative, a behavior we will call shirking. Now that rewards are no longer proportional to output, we cannot expect agents always to give their best, and management will have to expand resources to motivate employees. Employees will also have fewer incentives to carefully collect information than if they were working for themselves, since they do not directly benefit from it, but instead are supposed to pass it on to the boss. The firm will have to replace price constraints by behavior constraints. It will have to tell employees what to do, and to monitor their behavior. The costs of using hierarchy, “internal organization costs,” are those of motivating employees to fully contribute to the firm’s goals (so as to minimize shirking costs), and to collect information and transmit it faithfully to their superiors.

Hierarchy will therefore be a more efficient coordination method than prices when the costs of controlling shirking and of insuring effective information collection and transfer are lower than those of measuring goods and services and curbing cheating. Hence an MNE does not avoid the market when it takes on (internalizes) a transaction: it merely shifts the transaction from the market for goods and services to that for labor. The expansion of firms (the internalization of a market) will take place when transaction costs in the product market are higher than those experienced in the labor market.4
The preceding analysis offers some interesting implications for research. Keeping market transaction costs constant, whether a particular transaction will be internalized by the firm or not, will depend on the costs incurred by firms in monitoring employees. These costs are likely to vary across activities, across time periods, across countries, and across firms. Observing behavior is easy on machine-paced processes, because in that case an employee’s behavior provides a good clue to his performance. Controlling shirking will be costly, on the other hand, for tasks that are not programmable but require judgment and on-the-spot decisions, and for those that require employees to be dispersed over space. As noted below, this explains why, for example, some activities are franchised while others are operated with employees, and why trademark owners usually operate themselves easily accessible units but franchise dispersed ones.

Another implication is that managing employees is also likely to be more costly, the greater the cultural differences between the MNE’s home and target country. Cultural differences between countries raise the cost of doing business there. One would expect that the cost of market transactions would grow less steeply with cultural distance than that of handling the transaction within the firm, because extending the firm abroad requires closer contacts with locals (employees, governments, and other stakeholders) than entering into market transactions with them (Johanson and Vahlne, 1977). If this is the case, the greater the cultural distance between the home base of the MNE and the target market, the less likely it is that the firm will internalize the transaction. As we shall see later, there is some evidence to support this point.

**Why firms use both price and hierarchy**

One criticism that has been levied against the previous analysis is that “the distinction made between markets and hierarchies is greatly overdrawn” since most real-world institutions do in fact use both methods of organization simultaneously (Perrow, 1988). While Perrow sees this as invalidating transaction costs theory, there is a simple answer to this apparent paradox. The answer (Hennart, 1993) lies in the distinction between methods or organization (prices and hierarchy) on one hand, and institutions (markets and firms) on the other. Institutions make use of the two methods of organization in various proportions, with markets using predominantly prices, but also some hierarchical behavior constraints, while firms use mostly hierarchical methods, albeit with some price constraints as well.

We have seen that firms replace price constraints by managerial directives, but that taking away price incentives reduces the motivation that employees have to contribute to the firm’s goals. As individuals switch from being self-employed to being employees, their incentives to cheat decline with the decrease of the percentage of their output they get to keep, but so does their motivation to work. The firm must therefore increase the behavior constraints it will put on employees. Diminishing returns is a basic law in economics, and there are good reasons it should apply to methods of coordination. As firms increase
Transaction costs and the multinational

their reliance on behavior constraints, the marginal returns to this method of organization will decline. As Ghoshal and Moran (1996) show, over-monitoring and micro-managing employees are likely to destroy morale and to kill initiative. Hence, whenever it is important to elicit initiative and effort from employees, firms may find it advantageous to selectively reintroduce price constraints for some activities. They will pay employees through piecework, and they will set up some activities as profit centers and reward their managers for maximizing the profits of their subunits. In so doing they will end up using a mix of price and hierarchical (behavior) constraints. Hence what distinguishes the firm from the market is the mix or constraints used: firms are institutions that use primarily behavior constraints, while markets use mostly price constraints. The additional assumption of diminishing returns to the exclusive use of a single method of organization (prices or hierarchy) allows transaction costs theory to account for the fact that most real-world institutions use a mixture of price and hierarchical constraints. Nevertheless, it is because firms use hierarchy, i.e. because they replace price by behavior constraints, that they can experience, in certain transactions, lower organization costs than markets. Going back to the exclusive use of prices within firms (setting up a full “internal market”) would reintroduce the problems that the extension of the firm has sought to remedy.

The preceding analysis explains the existence in many MNEs of quasi-independent subsidiaries free to buy and sell at market prices to whatever internal or external customer they choose. While prices may be used in such MNEs to organize most interactions between parents and subsidiaries and between the subsidiaries themselves, some interactions, for example know-how, will remain organized through hierarchy. If all interactions were organized through market prices, subsidiaries would be independent firms (Hennart, 1993b).

Inversely, the increasing biases due to imperfect output measurement force firms to supplement price with behavior constraints. This is the case, for example, in the relationships between Japanese car assemblers and their suppliers, and those between franchisers and their franchisees. In this case, as we shall see below, the use of prices to organize the goodwill interactions between franchisors and franchisees would require a mechanism that would instantly price the impact of the behavior of each franchisee on the reputation of the brand, and would automatically credit or debit the account of the franchisees with the gains or losses their behavior inflicts on the overall reputation of the chain. Since such a price solution is presently technically infeasible, the impact of franchisee behavior on brand reputation is more efficiently controlled by a series of contractual rules (e.g. McDonald’s QSC standards). The results are networks of franchisee firms linked to franchisors by franchising contracts, a mix of price and hierarchical constraints (Hennart, 1993a).

To sum up, MNEs are firms that expand across the boundaries of nation states. To explain why they do, one must explain why the method of organization that distinguishes them from markets, viz. hierarchy, can in some cases be a more efficient coordination mechanism than prices. This is the case whenever behavior constraints are more efficient than output constraints. This happens
when the characteristics of outputs are poorly known and hard to measure. Inversely, behavior constraints are efficient when bosses have a good knowledge of the employee production function, and when observing behavior gives them good clues on performance. Because of diminishing returns to the exclusive use of a single method of organization, price or hierarchy, most real-world solutions are a mix of both, although the boundaries of firms are generally unambiguous. The following pages show that such a model can explain all the forms taken by the international expansion of firms.

**Why do multinational firms expand abroad?**

As we have seen, there are two basic (and non-mutually exclusive) reasons why firms may expand abroad. The first one, to which we now turn, is to internalize the pecuniary externalities firms inflict on one another when they compete on the market for final products. MNEs then arise to reduce competition. We will later discuss how MNEs internalize the second type of externalities, the non-pecuniary ones that result from "natural" market imperfections.

**Structural market imperfections**

Recall that pecuniary externalities are those that competitors impose on one another through the impact of their actions on the prices they face. One particular instance is competition. Consider a homogeneous good produced by single-plant monopolists located in a number of different countries. Competition between these producers will reduce their income. Competitors can, however, maximize their joint income if they agree to merge and concentrate production in the lowest cost plant. For Hymer, the MNE is the vehicle through which such collusion is organized.

This argument, nevertheless, only provides a partial explanation for the existence of MNEs since MNEs can also arise to internalize non-pecuniary externalities. Moreover, Hymer’s argument is incomplete. Extension of the firm is not the only method available to reach collusion. Competitors can also coordinate their behavior through contract by taking part in a cartel (Casson, 1985). The main problem with a cartel is that of free-riding: all members would like to sell more in the high price market, but by doing so they lower prices. Consolidating all firms into an MNE eliminates incentives to cheat, but raises management costs and may be opposed by host governments. This suggests that, without legal restrictions on cartels, the propensity to internalize pecuniary externalities by cartel rather than through an MNE will be greater the easier it is to detect cheating and the lower the need to adapt to changing conditions. Hence cartels should be more prevalent in industries producing homogeneous products and characterized by slow growth and static technology (Casson, 1985).

Pecuniary externalities are also generated in the licensing process. Consider a patent owner licensing one producer in each country. Because of differences in the elasticity of demand, the optimal price for the product is likely to vary across
maximizations of rents (and hence of royalties received by the licensor) requires that licensees be prevented from invading each other’s markets. In some cases, high transportation costs, tariffs, or government regulations segment markets. When barriers to trade are low, the licensor must explicitly forbid licensees from exporting products to other markets. In practice, such territorial restrictions are often illegal. A firm that integrates into foreign manufacture is, however, better able to prevent competition between plants producing the licensed product because it needs no explicit contract to eliminate competition between its subunits (Casson, 1979). Hence transaction costs theory explains when the internalization of pecuniary externalities will be achieved by the establishment of an MNE rather than by contractual means.

Natural market imperfections

Figure 4.1 sketches the decision-tree of a firm with proprietary intangible assets valued abroad. The firm becomes an MNE when it ends up having employees overseas. These employees can be in sales subsidiaries, when the firm manufactures at home but handles its exports with its own employees, or they can be in production subsidiaries when the firm manufactures overseas. According to our definition, the firm is not an MNE if it exports but uses independent distributors, or if it decides to transfer its intangibles by contract (license, franchise, management contract, etc.). Whether the firm is an MNE or not thus depends on whether it extends hierarchical control across national boundaries, through sales or production subsidiaries.

Figure 4.1 shows that the extension of an MNE overseas occurs as a result of a location and of a governance decision. The location decision is based on a comparison of delivered cost, itself a function of the relative production cost of a domestic and foreign location, of transportation costs, of tariff and non-tariff barriers to trade, and of political risk. A second, largely independent decision, is whether hierarchy is more efficient than price to organize the interdependencies between agents located abroad and at home. The combination of location and governance decisions determines whether a firm will be multinational (Dunning, 1977, 1979, 1981).

A similar analysis can be used for a firm that requires inputs from abroad (Figure 4.2). The firm must choose between procuring its inputs from home or foreign locations (the location decision). It must also choose between integrating into the production of the inputs it needs or procuring them at arm’s length from independent suppliers (the governance decision).

Figures 4.1 and 4.2 thus show that the expansion of firms abroad can take very diverse forms, from setting up sales subsidiaries to developing mines and plantations (all cases that correspond to the expansion of firms abroad are in italics). Yet, once a foreign location is optimal, whether international coordination will take place within the firm (whether an MNE will emerge) or through the market can be explained by the relative costs of using firm or market governance to organize that specific interdependency.
International interdependencies can be of many types. Knowledge and reputation developed in one country can have productive uses in another. Products manufactured in one country may require complementary marketing services in another. Raw materials and components most efficiently produced abroad may be needed at home. Lastly, financial capital, abundant in one country, may be scarce in another. Table 4.1 presents the hypothesis that MNEs exist to organize all these international interdependencies. As we shall see in the following pages, all these motives for the existence of MNEs can be explained by the same set of theoretical tools.

*Knowledge: licensing vs. foreign production*

Most applications of transaction costs theory to the MNE have focused on the international exchange of one particular type of input – knowledge (Buckley and
Table 4.1 The transaction costs theory of the MNE

<table>
<thead>
<tr>
<th>Type of MNE</th>
<th>Market internalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizontal integration by R&amp;D-intensive firms</td>
<td>Technological and managerial know-how</td>
</tr>
<tr>
<td>Horizontal integration by advertising-intensive firms</td>
<td>Reputation; managerial skills in quality control and marketing</td>
</tr>
<tr>
<td>Vertical integration into distribution</td>
<td>Distribution and marketing services</td>
</tr>
<tr>
<td>Vertical integration into raw materials and components</td>
<td>Raw materials and components</td>
</tr>
<tr>
<td>Freestanding firms</td>
<td>Capital</td>
</tr>
</tbody>
</table>

Casson, 1976; Cantwell, 1989; Casson, 1979; Caves, 1982; Hennart, 1977; Magee, 1977; Teece, 1981). Know-how developed in one country is often useful in others. When and why will the international transfer of such know-how be organized within a firm, and when will it be organized through arm’s length or contractual exchange? Microeconomic theory tells us that markets are efficient when buyers and sellers have good knowledge of what is being sold. This suggests that knowledge itself will have a particular problem when it is the commodity being transacted. As Arrow (1962) first argued, the buyer of know-how does not, by definition, know its exact characteristics and is thus likely to underpay for it. If the seller were to provide that information, he would be revealing it to the buyer, and thus transferring it free of charge. The basic problem of the transfer of knowledge is therefore one of information asymmetry between buyer and seller. Information asymmetry will be lower the older the technology and the smaller the differences between the technological capabilities of the transactors.

The patent system offers one solution to this information asymmetry problem. By giving owners of knowledge a monopoly in its use, patents encourage them to disclose it, thus reducing information asymmetry between buyer and seller. However, the efficiency of a patent system crucially depends on (a) the ability of patents fully to describe the invention and (b) the power and the willingness of public authorities to establish and enforce monopoly rights in knowledge. By taking a patent, inventors are disclosing their know-how to potential buyers (licensees), but also to potential imitators. It is therefore crucial that inventors be protected against infringement. Inventors who fear that their rights will not be protected will keep their inventions secret and will exploit them themselves, i.e. they will internalize the market for their knowledge by vertically integrating into the manufacture of products incorporating their know-how. On the other hand, when patent rights are well protected, knowledge holders will be able to sell or rent (license) the rights to their know-how to domestic and foreign manufacturers. The degree to which patents provide protection depends on technological factors (such as the extent to which knowledge can be clearly defined and described on
Jean-François Hennart

paper and the difficulty of designing around the patent), and on government policies, such as the length of the patent grant, the efficiency of the court system, and the level of penalties for infringement. Arm's length transfer of knowledge through licensing will be more prevalent whenever patent rights are easy to establish and to defend, while transfer within the firm will be chosen for knowledge that is difficult to codify into patents and easy to copy (Caves et al., 1982; Hennart, 1982). Since governments vary in the extent to which they provide patent protection, the efficiency of transferring knowledge through licensing will also depend on the country being entered.

All in all, the limited available research shows that patents are not a very effective method of knowledge appropriation because they do not provide patentors with solid rights of exclusive use and because the rights to transfer patents to others (licensing) are subject to governmental limits. The research and development (R&D) managers of US firms surveyed by Levin et al. (1987) reported that patents were generally a less effective method of knowledge appropriation than vertical integration into manufacture. On a scale of 1 to 7, with 7 being “very effective,” they rated process patents to secure royalty income at 3.52, and product patents at 4.33. Other methods of appropriation that were associated with vertical integration, such as being the first to manufacture a product, or having a strong sales and service effort, were rated more highly. Being the first to sell a product rated 5.11 for processes and 5.41 for products. As we would expect, there were substantial differences across industries in the level of patent protection, with chemicals and drugs enjoying better protection than computers and communication equipment.

Given those differences in the protection given by patents, it is not surprising to see major differences across industries in the mix between international licensing and vertical integration into manufacturing as strategies to exploit knowledge abroad. Table 4.2, for example, shows the ratio of licensed production relative to total UK foreign production (the sum of sales by foreign licensees of UK firms and sales of foreign subsidiaries of UK concerns); this provides an index of the extent to which technology was transferred through contract. This ratio varies from a high of 71 percent in the shipbuilding industry to a low of 4 percent in food, drink, and tobacco.

International licensing is also often subject to limitations imposed by host governments. Host governments have disallowed the patenting of otherwise patentable knowledge, and they have outlawed patent clauses, such as grant-backs (which provide licensees with access to improvements made by licensees) and territorial restrictions (which make it possible to price-discriminate between markets). These government limitations further reduce the profitability of patenting and make it a second-best method of exploiting one's know-how (Contractor, 1984). One consequence is that licensing is primarily used to exploit “peripheral technology” in “peripheral markets.” Because imperfect property rights in knowledge (imperfect patent protection and restrictions on licensing) prevent licensors from getting as high a return on their technology as vertical integration into manufacture, and expose them to potential technology leakages to
the other operations of licensees, firms prefer to exploit internally technology that they see as central to their future. Licensing is also used primarily in small, peripheral markets where the risks and the consequences of building up competitors are small and the limited market size does not warrant the high fixed costs of establishing production subsidiaries (Caves et al., 1982).

There are few empirical tests of these propositions because of the difficulty of obtaining detailed information on large samples. Davidson and McFetridge (1984) tested the propositions that older technology was easier to license because it was better known, and the risk a firm runs in having its know-how build up competitors was lower if the technology it transferred was not central to its survival. Analyzing 1,382 cases of technology transfer undertaken by 32 US MNEs, they found, as expected, that the probability of licensing an innovation was greater the older the technology, the more peripheral it was to the innovator’s business, the smaller the investment in R&D necessary to develop it, and the greater the innovator’s experience in international licensing.

Another implication of transaction costs theory is that, keeping the transaction costs of licensing constant, the costs of transferring knowledge within the firm should be lower the smaller the cultural distance between the innovator’s home country and the target country. This was also supported by Davidson and McFetridge (1985), who found that the probability that US firms would transfer an innovation internally (as opposed to licensing it) increased if the target country was geographically and culturally close to the United States. Buckley and Davies (1981) estimated the share of foreign subsidiary sales to total foreign sales of British firms (the sum of foreign subsidiary sales of British MNEs and sales by licensees of British firms). This measure of the extent to which British firms used hierarchical over market transfer of knowledge was higher in the former sterling area countries, which are culturally close to the UK, than in other countries.

### Table 4.2

<table>
<thead>
<tr>
<th>Sector</th>
<th>Share %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food, drink, tobacco</td>
<td>4.0</td>
</tr>
<tr>
<td>Chemicals and allied industries</td>
<td>25.8</td>
</tr>
<tr>
<td>Metal manufactures</td>
<td>12.3</td>
</tr>
<tr>
<td>Mechanical and instrument engineering</td>
<td>21.5</td>
</tr>
<tr>
<td>Electrical engineering</td>
<td>7.6</td>
</tr>
<tr>
<td>Shipbuilding</td>
<td>71.1</td>
</tr>
<tr>
<td>Motor vehicles</td>
<td>45.1</td>
</tr>
<tr>
<td>Textile, leather, clothing, footwear</td>
<td>5.0</td>
</tr>
<tr>
<td>Paper, printing and publishing</td>
<td>13.6</td>
</tr>
<tr>
<td>Rubber</td>
<td>34.8</td>
</tr>
<tr>
<td>Other manufacturing</td>
<td>21.2</td>
</tr>
<tr>
<td>All manufacturing</td>
<td>17.5</td>
</tr>
</tbody>
</table>

*Source: Adapted from Buckley and Prescott (1989)*
While much of the literature has tended to focus on the problem of a firm seeking to exploit technology, the theory sketched above predicts that the presence of high transaction cost in the market for knowledge will also lead firms to expand abroad to acquire technology. The same transactional difficulties that make it difficult for a domestic firm to license some types of knowledge to foreign firms make it difficult for a foreign firm to license similar types of knowledge from domestic firms. The desire to acquire technology will therefore lead firms to expand abroad through acquisitions, joint ventures, and greenfield investments.

Reputation: franchising vs. foreign production

Trademarks are to reputation what patents are to knowledge: they are the legal instrument establishing property rights in reputation. Trademarks are valuable because they reduce customer search costs. Tourists and businesspeople are likely to patronize abroad businesses bearing trademarks of their own country. Hence a firm’s home country trademarks are sometimes valuable abroad. The ability of a firm to exploit its reputation abroad depends on the extent to which its trademarks are protected from unauthorized imitation (counterfeiting). Counterfeiting will destroy the value of a trademark since counterfeiters have incentives to run down the reputation of the brand by selling lower-quality products if they sell to non-repeat customers. A firm that owns a trademark can exploit it by itself producing goods and services bearing its trademark, or by drawing franchising contracts to rent the use of its trademark to local entrepreneurs. The efficiency of franchising depends on the extent to which trademarks are protected from counterfeiting, since no one will rent what can be acquired for free.

Franchising is also subject to free-riding. A trademark becomes a public good to all those who are using it, in the sense that the quality of goods and services supplied by anyone using the trademark will affect all those who share the trademark. Franchisees can maximize their income by reducing the quality of the goods sold under the trademark. The greater the proportion of non-repeat customers, and the higher the cost of detecting free-riding, the greater the payoff to franchisees who reduce quality, and the losses they will inflict on all the other users of the trademark. Consequently, franchisors write contracts that carefully stipulate minimum quality standards franchisees must follow and spend significant resources enforcing them.

One way of reducing a franchisee’s incentive to free-ride is to transform him into an employee. As an employee, the erstwhile franchisee is paid a straight salary, unrelated to the outlet’s profits, and gains nothing from reducing quality. But breaking the link between output and reward has an unfortunate, but unavoidable side-effect: the former franchisee, now an employee, has little motivation to exert any effort which cannot be specified and measured by his employer. Hence the choice between franchising independent owners and establishing company-owned outlets will depend on the comparison of two types of cost: that of monitoring employees to guarantee their level of effort vs. that of specifying...
Figure 4.3 Determinants of franchising

and enforcing a minimum level of quality by contract to prevent free-riding on quality (see Figure 4.3). Everything else being constant, franchising will be efficient when (a) it is relatively easy to write rules, which, if followed, will guarantee a minimum level of quality, and whose violation can be easily detected and proven to third parties and (b) it is relatively costly to ensure that employees will work hard and take initiatives. Trademark owners will operate outlets with their own employees in the reverse case. No international exchange of reputation will take place when it is difficult to monitor employees and contractually to stipulate quality.

These considerations explain why franchising is commonly used for the international exchange of goodwill in some industries, but not in others. In fast food, hotel, and car rental, drawing up and enforcing contractual quality standards is relatively easy, whereas the dispersion of outlets raises the cost of monitoring employees. Franchising is therefore common. The reverse seems to be true in banking, insurance, advertising, and management consulting. In these activities minimum quality levels are difficult to stipulate by contract in a manner that is enforceable in court, while the relatively small number of outlets necessary to reach customers makes monitoring employees relatively easy (Hennart, 1982: 89–93). In these industries goodwill is exploited in outlets staffed by employees of the trademark owner.

The relative levels of market transaction costs and internal organization costs also vary within a given firm. In the fast food industry, high-volume outlets in concentrated locations, which are easy to monitor, tend to be operated by the trademark owner, while small, dispersed outlets are franchised (Brickly and Dark, 1987; Hennart, 1982). Dunning and McQueen (1981) uncovered a similar pattern in the international hotel industry. Worldwide, more than half of the rooms in hotels operated by MNEs are run under franchise or management contracts, but the proportion is much greater (82 percent) in developing countries, where political risk and cultural distance raise the cost of managing hotels over that of franchising them to local operators.
Raw materials and components: spot purchases and long-term contracts vs. foreign production

One early and persistent reason why firms have expanded abroad has been to secure minerals and agricultural produce necessary for their downstream activities. The foreign investments of integrated oil companies into crude oil, those of steel firms in iron ore, and those of rubber manufacturers into natural rubber plantations are but a few examples (Wilkins, 1970, 1974). Transaction costs theory suggests that such “backward vertical integration” will be chosen whenever markets for raw materials and intermediate inputs are characterized by high transaction costs. These costs arise when (a) the number of parties to the exchange is small (small-number conditions) and/or (b) parties differ in the amount of information they have on the transaction (information asymmetry).

SMALL NUMBER CONDITIONS

Small-number conditions result from economies of scale, high transportation costs, government barriers, and physical asset specificity. Physical asset specificity arises when one or both parties to the transaction invest in equipment specially designed to carry out the transaction, and which has lower value in other uses (Williamson, 1985). When these conditions are present, spot markets are likely to fail, because a party making transaction-specific investments, and consequently for whom the costs of switching partners are high, will fear that the more flexible party will opportunistically renegotiate the terms of trade after investments have been made. One possible way for parties to protect themselves is to write a contract fixing the terms and conditions of the trade over a period of time corresponding to the life of the plant.

Contracts reduce enforcement costs by specifying ex ante the terms and conditions of the trade and the compensation to be paid in case of breach, relying on third parties (the courts or private arbitrators) to determine whether a violation has taken place, to decide on the compensation to be paid in each case, and to enforce payment of such compensation. Contracts suffer, however, from a number of limitations. They are more costly than spot markets, for while a large number of traders typically share the costs of running spot markets, the costs of writing contracts must be paid by the contracting parties themselves. The additional cost of writing a contract can be justified only if trades are long-lived. Contracts can be risky because they rely on enforcement by third parties (the courts or arbitrators). Breach, while obvious to those involved, may be difficult to prove to others. The aggrieved party has also limited control on the amount and form of the compensation he will receive for breach.

The most severe drawback of contracts is that they often fail under conditions of high uncertainty. As the degree of uncertainty increases, specifying ex ante all possible contingencies and the contractual changes to be made in each case becomes increasingly difficult. Yet leaving contracts incomplete allows parties to exploit each other. Hence contracts are efficient when uncertainty is low or when it can be indexed.
Here as elsewhere, an alternative to contracts is to have buyers and sellers of raw materials become employees of the same firm, i.e. vertical integration. Vertical integration transforms one of the parties into an employee of the other party. As an employee, the erstwhile trader is no longer rewarded for his opportunism, but instead for obeying the directives of his boss. Vertical integration makes it possible to reduce opportunism by aligning the incentives of both parties (Williamson, 1979).

This theoretical framework provides a convincing explanation of the pattern of vertical integration found in many domestic industries, for example that between coal mines and electric power plants (Joskow, 1985), between automobile assemblers and parts manufacturers (Monteverde and Teece, 1982; Walker and Weber, 1984), between aerospace firms and their component suppliers (Masten, 1984), and between wood processors and timber growers (Globerman and Schwindt, 1986). The same logic can be applied to foreign backward investments by MNEs, since they are vertical investments that cross borders.

The aluminum industry provides an interesting example (Stuckey, 1983). There are three stages in the production of aluminum: bauxite is mined, then shipped to alumina plants where it is refined into alumina, and alumina is then smelted into aluminum. In the late 1970s, close to 90 percent of transfers of bauxite, alumina, and primary aluminum was organized within MNEs. The reason is that markets for bauxite, and to a lesser extent for alumina, are narrow. Bauxite refining has high asset specificity because bauxites are heterogeneous and significant cost savings can be obtained when refineries are built to process a single type of bauxite. Once an alumina refinery is built, switching costs are high. Because bauxite is of low grade, transporting it long distances is costly. Lastly, high economies of scale in bauxite refining and smelting further reduce the number of potential buyers and sellers of bauxite and alumina. To organize what is often a bilateral relationship through spot markets would be hazardous, because, after investments have been made, the owner of the mine could exploit the owner of the alumina plant (or vice versa) by unilaterally changing the price of bauxite.

Using contracts also entails serious risks. Because both bauxite mining and refining require significant investments of long economic life, contracts must cover very long periods – typically 20 to 25 years. Over such a long timespan they cannot effectively protect the parties against changes in the environment which impact their profit stream. The long-term contracts signed in the 1960s by Japanese bauxite purchasers specified prices that were denominated in dollars, with adjustments based on the quality of the bauxite actually shipped. Shortly after they were signed, the end of the Bretton Woods system of fixed exchange rates radically altered the profitability of the contracts by changing the local currency equivalent of the dollar prices, while the two oil shocks made the Japanese aluminum industry, which relied for its electricity on oil-fired power plants, increasingly uncompetitive. This led to acrimonious bargaining, and to a shift from contracts to vertical integration (Stuckey, 1983).

We have seen that the high degree of asset specificity in bauxites, and to a lesser degree in alumina and aluminum, has made vertical integration the
overwhelming method to effect the transfer of these commodities. In contrast, coordination between stages in the case of alluvial tin is efficiently performed by spot markets. There is no asset specificity in the smelting of alluvial tin ores since these ores are very homogeneous (they are nearly pure tin). Tin ores are also of high value, so their transportation costs are low relative to their value, and they can be transported long distances. The result is an efficient market for tin ores, eliminating the need for vertical integration between mining and smelting.\textsuperscript{15}

Tin ores obtained from lode mining, however, tend to carry various impurities, and the smelters handling these ores must be specifically designed to treat those impurities. Consequently, the lode sector of the tin industry is characterized by greater vertical integration (Hennart, 1986, 1988).

The considerations outlined above explain the need for MNEs to own suppliers of other intermediate inputs, such as parts or subassemblies. In most cases MNEs will own their foreign suppliers when the components they manufacture are specific to the purchaser, while independent suppliers will be used for standard parts, which are sold in a relatively broad market.\textsuperscript{16}

QUALITY CONTROL

Another reason for vertical backward investments is quality control. Quality control problems arise in situations of information asymmetry. If buyers cannot distinguish \textit{ex ante} between good and bad quality, they will tend to reduce their offer price to reflect this risk. Sellers of high-quality products may not be able to persuade a buyer that the goods offered are of high quality, and will therefore avoid the market. Markets will fail in the sense that they will be used to sell goods of increasingly lower quality (Ackerlof, 1970). In that case, sellers and buyers have incentives to integrate.

The banana industry offers an interesting example of this motive for vertical integration. Bananas are certainly an unsophisticated product, so it is surprising that vertically integrated MNEs dominate their international trade.\textsuperscript{17} But bananas are highly perishable, as they spoil 21 days after cutting. Their quality also depends on careful handling and proper ripening conditions. Improper handling and ripening are difficult to detect \textit{ex ante}: damage incurred at the cutting and shipping stages will be revealed only once the banana reaches retailers. Hence it is difficult to achieve consistent high quality if grower, shipper, and distributor are separate concerns (Litvak and Maule, 1977; Read, 1986). Consistent quality is better assured by vertical integration because it reduces incentives to cheat at each stage (Casson, 1982). This explains why US banana distributors integrated early on into banana plantations (Wilkins, 1970).

\textit{Distribution and marketing services: independent distribution vs. company sales force}

The international expansion of firms to take over the distribution of their products can be explained by the same general factors that lead to the integration of buyers
and sellers of raw materials and components described above, i.e. the difficulty of coordinating the behavior of buyers and sellers of distribution services when markets for these services are narrow and when their quality is difficult to measure (Chandler, 1977; Hennart, 1982; Nicholas, 1983; Williamson, 1981).

SMALL NUMBER CONDITIONS

To distribute a product efficiently requires investments both in physical assets (warehouses, stocks, transportation networks, repair facilities, offices, or retail stores) and in knowledge. The distributor must learn how to demonstrate and service the product, how to price it, and how to adapt it to local tastes and conditions of use. These physical and intellectual investments vary in size and especially in specificity. In some cases, they are “general purpose” and can be used to sell the products of a number of manufacturers; in others, they are specific to a single supplier and have little or no value in other uses.

As in the case of backward vertical integration, there are two main, often reinforcing reasons why the market for distribution services is often narrow. First, distribution is often subject to high economies of scale or scope. A single distributor may be dealing with a single manufacturer; joining both in an MNE may solve the resulting bargaining stalemates. Second, effective distribution sometimes requires substantial manufacturer-specific investments. A distributor may be reluctant to make such investments, fearing that, after they are made, the manufacturer will opportunistically renegotiate the margins by threatening to sign a new contract with another distributor. This fear may cause the distributor to commit fewer resources to distribution than would be optimal (Nicholas, 1983).

Contracts may provide solutions. A manufacturer can persuade a distributor to make manufacturer-specific investments by offering exclusive distribution rights. Here again, the more uncertain the environment and the longer the time needed to recover its transaction-specific investments, the greater the chances that such a long-term contract will break down. In these cases, vertical integration of manufacturers into distribution may be the most efficient option. Integration into distribution will thus be observed in the distribution of products requiring specialized facilities or specific demonstration and repair.

This model is supported by some recent studies of the factors that determine whether firms will use independent agents or employees to handle their exports. Anderson and Coughlan (1987) found that the probability that US manufacturers of semiconductors would use their own employees to sell abroad increased with the level of product knowledge that they needed to sell the product effectively. Klein et al. (1990) analyzed the choice of Canadian exporters between independent distributors and a company sales force. They found that a company sales force was chosen when the effective distribution of the product required manufacturer-specific investments.
PERFORMANCE INSEPARABILITY

Another problem with independent distributors arises when consumers cannot separate the contribution that manufacturers and distributors make to the satisfaction they derive from the product-plus-service bundle (Chen and Hennart, 1998). For example, poor control by the retailer of the temperature in the freezers containing branded ice cream will affect its texture. A consumer will not be able to know whether the poor texture is due to poor manufacture or improper storage, thus allowing the manufacturer and the distributor to inflict uncompensated damages on each other (inversely, there is no easy mechanism to allocate rewards for investments made by either party to increase ice cream quality). The problem is similar to that experienced in franchising and arises when the distributor and the manufacturer can both affect the quality of the good and services as perceived by the consumer. One solution is for manufacturers contractually to negotiate the imposition of behavior constraints on retailers, of the type franchisors put on franchisees. Another is for manufacturers to integrate into distribution or for retailers to integrate into manufacturing. Vertical integration will be chosen if manufacturers cannot easily define and enforce contractual rules to prevent distributors from debasing quality but can cheaply and effectively monitor the behavior of distributors if they become employees (Caves and Murphy, 1976; Hennart, 1982).18

The observed pattern of vertical integration into distribution reflects a tradeoff between the need to have distributors make the requisite level of investments and keep quality at an agreed level, on one hand, and the cost of managing company-owned distribution facilities on the other. That cost rises if there are scope economies in distribution and if it is difficult to monitor the behavior of employees.

Inter-industry differences in the level of integration between manufacturers and distributors seem consistent with such a model. Here again, there is a striking parallel between domestic and international integration. Williamson (1981) has shown how the extent of domestic forward integration by US manufacturers varied at the turn of the century from none in the case of hardware, jewelry, liquor, and dry goods items, to full integration into both wholesaling and retailing in the case of new, complex, high-priced machines requiring specialized demonstration and repair. Vertical forward integration by MNEs abroad had then a similar pattern: no integration for dry goods, integration into wholesaling for products that required specialized handling, and integration into retailing in the case of new, sophisticated products which required demonstration, installation, and after-sales service and whose quality could be damaged by improper handling (Chandler, 1959, 1977; Hennart, 1982; Nicholas, 1983; Wilkins, 1970). In short, the observed pattern of vertical forward integration abroad, as well as that of backward vertical integration, is consistent with the view that MNEs expand abroad to bring in-house activities which are subject to high market transaction costs and relatively low internal organization costs.
Financial capital: loans vs. equity

Financial capital is, alongside know-how, reputation, and raw materials and components, a factor of production that can often be profitably transferred overseas since its return often differs across countries. As in the case of the other intermediate products considered above, financial capital can either be transferred across countries by contract (bonds and loans) between a lender and a borrower/investor, or its international exchange can be organized by hierarchy. In the latter case, the market for financial capital is internalized, and the lender and the borrower are joined within an MNE.

Capital can be transferred on international markets through loans and corporate bonds. This mode of transfer gives the lender no right to the profits of the venture and no “general and discretionary right” to direct the behavior of the borrower. Lending involves making funds available to the borrower, to be paid back later with interest. There are many reasons why international lending is subject to high transaction costs. First, in contrast to many other transactions, the two parts of a lending transaction are not simultaneous. Money is lent today and repaid later. Unlike other physical assets, money is fungible, and can be used for purposes other than those approved by the lender. Finally, incentives in a loan transaction are often not symmetrical. Borrowers who invest in risky projects will capture, if they are successful, all of the returns of their investment net of interest payments and repayment of principal. If they fail, lenders will shoulder all the costs (Jensen and Meckling, 1976). While a borrower’s loss of reputation is supposed to curb such opportunism, such reputation effects are weak in international lending, as it is often difficult for lenders to know whether the borrower’s default was due to dishonesty, ineptitude, or just bad luck.

Default would not arise if lenders had perfect information on borrowers and projects and if the enforcement of loan contracts were costless. Given positive transaction costs, lenders will have to use a number of second-best strategies to protect themselves against default. One consists in concentrating on projects and borrowers known to them. Restricting the use of funds to purposes specifically authorized by them is another. However, this requires lenders to know what expenditures are needed to carry out the proposed business. A third strategy is to ask the borrower to pledge a collateral to be forfeited in case of default. Usually that collateral will consist of some of the assets financed through the loan (Hennart, 1994).

Attempts by lenders to protect themselves against default by using these three strategies will cause inefficiencies in the international market for financial capital. First, lenders will tend to screen out borrowers not personally known to them, those without long track records, and those undertaking projects with which they are not familiar. Second, lenders will try to avoid projects that do not yield good collateral. Good collateral are assets that are not specific to a particular user and have good resale value, for example general purpose trucks or airplanes. Because the value of collateral to a lender varies systematically across projects, so will the transaction costs of lending. Investments in mining or R&D offer poor collateral,
because capital sunk into unsuccessful mines or R&D does not yield salable assets, and therefore will tend not to be financed by loans.

The market for financial capital is therefore likely to fail if prospective borrowers cannot show a good track record, if they are not known to lenders, or if they want to invest in projects that do not yield good collateral. One solution in that case may be to organize the market for financial capital within a firm, i.e. to have the lender become full or part owner of the borrower's business (or vice versa). This will reduce transaction costs for three main reasons. First, equity links provide lenders with much greater control than loan contracts, as equity owners have the right to review decisions ex ante and have easier access to internal documents. By contrast, lenders are strictly limited in the quantity, quality, and timeliness of the information they can obtain on their borrowers. Equity control is also more flexible than loans, because it allows greater discretion to preserve the value of a going concern when problems occur (Williamson, 1988). Lastly, transforming borrowers into employees reduces their incentives to take excessive risks because their reward as employees is now more independent of the success of the project. This, of course, has also the unintended consequence of decreasing the borrower's incentive to exert effort.

The presence of transaction costs in the market for financial capital explains the respective roles played by lending (bank lending or corporate bonds) and by intra-firm transfers in the international movement of financial capital. Historically, banks have provided only short-term international lending, mostly on receivables (except during the 1970s and 1980s, when banks lent long term to newly developing countries, with disastrous consequences). In the nineteenth century, long-term international lending was through bonds, mostly issued by governments and railways, and through mortgages. In all these cases, the transaction costs of lending were relatively low. Governments have low default risk because of their ability to raise taxes to repay loans. Land is also a good collateral, especially in fast-growing economies, and in the nineteenth century considerable sums were lent by European investors to farmers in North America, Australia, and Argentina (Wilkins, 1989). The construction of railroads was also financed through bonds. In Europe and Russia governments guaranteed the bonds of railroad companies. Railroads were private in the US, but were given huge tracks of land, which they used as collateral.

Outside the transfers to developing countries made in the postwar period by bilateral or multilateral aid agencies, the bulk of long-term international financial transfers has been effected through MNEs. This transfer can take place in two ways. First, MNEs can emit shares in capital-rich countries and use the proceeds to finance projects in capital-poor ones, thus intermediating the international transfer of capital. The import of capital need not, however, be intermediated. For example, firms in capital-poor countries (or firms in capital-rich countries investing in capital-poor countries) may sell directly shares on the stock markets of capital-rich countries. The best example of this last type of transfer is that effected by the so-called "freestanding company" (Wilkins, 1988: Wilkins and Schroeter, 1998).
Freestanding companies were the dominant type of British MNEs until 1914, and were also important in The Netherlands, Belgium, and France (Wilkins and Schroter, 1998). While they were overshadowed by the more traditional type of MNEs after 1914, they continued to be formed after that time and some are still being established today (Hennart, 1998).

Freestanding firms do not fit at all the traditional model of MNEs developed in the literature and they seem to support Fieldhouse’s contention that no single theoretical model can explain all the forms taken by MNEs. Freestanding firms were domiciled in the dominant equity markets of the time (in London, but also in Brussels, Paris, Amsterdam, Boston, etc.) and they raised equity there by selling shares directly to the public to finance investments that were located exclusively abroad. The traditional theory of the MNE argues that firms expand abroad to exploit the firm-specific advantages (such as proprietary technology) they have developed in their home market (Dunning, 1981). Besides a skeletal head office, freestanding firms did not, however, conduct any business in their country of registration, as all of their actual operations (plants, mines, plantations) were abroad, typically in the least developed part of Europe (Spain, Italy, Russia), in the regions of recent settlement (Australia, Canada, the United States, Argentina), and in developing countries, both under colonial rule (India, Malaysia, the Dutch East Indies, Africa) and outside it (Persia, Siam, Latin America). While they were engaged in a wide variety of businesses, from services (hotels, utilities, docks, newspapers, banks) to manufacturing (breweries, jute mills, fish canning, flour milling), the majority of freestanding firms were in two sectors, namely agriculture (ranches and plantations – natural rubber, tea, sugarcane, cinchona) and mining and petroleum.

Without any domestic business, it is hard to see how the raison d’être of freestanding firms could have been the exploitation of firm-specific advantages. The apparent paradox of freestanding firms is compounded by the fact that the technology they used abroad was generally unknown in their country of registration. A large number of British and Dutch freestanding firms were in tropical agriculture. While the British were active in tin mining at home, the mining techniques they used in the Far East came from California and New Zealand (Yip, 1969).

Here as elsewhere, the raison d’être of MNEs must not be sought in the internalization of firm-specific advantages, but in the internalization of markets, in this case, for financial capital. Freestanding firms allowed lenders, located in capital-rich countries, to exercise direct management control over the projects that made use of their funds. This made it possible to finance projects which, because of their characteristics (they were in areas poorly known by savers and had poor collateral) could not be financed through loans.

Freestanding firms usually started when affluent individuals recognized an investment opportunity and pooled their personal funds into “syndicates” to take advantage of them. After these syndicates had sufficiently developed the project, they obtained additional finance (and cashed in on their entrepreneurial skills) by floating shares on the principal stock market of the time. There is a wealth of
evidence to support the view that the main reason for freestanding firms was the internalization of the market for financial capital.

First, case histories show how local firms registered in London or other capital-rich countries with an efficient stock market when they were unable to obtain sufficient local capital (Cushman, 1986; Drabble, 1973; Hennart, 1987, 1998). Second, freestanding companies had usually bigger capitalization than local firms. For example, in the Dutch East Indies, where Netherlands-registered freestanding companies worked side by side with Dutch-operated but locally registered firms, the Netherlands-registered companies had higher capital (Gales and Sluyterman, 1998). Third, as we have seen, freestanding firms were domiciled in capital-exporting countries with liquid capital markets, while their operations were concentrated in the rapidly developing but capital-poor countries of the time. Freestanding companies located their headquarters wherever they could efficiently access financial capital. While they were heavily concentrated in London prior to 1914, the imposition of capital export controls in the UK after the First World War led to a relative decline of London to the benefit of New York, where a number of freestanding firms were then formed to operate in Latin America (Hennart, 1998).

Fourth, the theory suggests that setting up a new venture as a freestanding firm would be efficient as long as equity capital could be raised more efficiently abroad than locally, and as long as the characteristics of the projects led to high transaction costs on the loan market. Freestanding firms created before 1914 survived longer in countries with underdeveloped capital markets. By the 1920s, UK-based freestanding firms had disappeared in the United States, whereas they survived in Malaysia and in Nigeria until the 1960s (Hennart, 1986b, 1987; Wilkins, 1988).

Fifth, as predicted by our theory, the activities typically financed by freestanding firms, namely speculative agriculture and mining, were those that, yesterday as well as today, are difficult to finance through loans. Mining is risky, and much of the investment is in diggings and site-specific infrastructure, yielding poor collateral. As a result, new mining ventures cannot generally obtain debt financing, and have to be financed through equity. Freestanding firms were also the dominant method used to finance, around the turn of the century, rubber and cinchona in South East Asia, and ranches in the United States and Argentina. All of these investments were substantial and then highly speculative, and used previously undeveloped land with little alternative uses, and hence little collateral value (Allen and Donnithorne, 1954; Jackson, 1968; Stillson, 1971).

To sum up, freestanding firms were the type of MNEs used to transfer financial capital internationally whenever lending would have been subject to high transaction costs and when the optimal scale of the investment was small. Freestanding firms were single-project firms (single mine, single plantation, single ranch, and so on), although they were often partly owned by trading companies, mining finance houses, or equipment sellers (Hennart, 1998). MNEs today perform the same function, since they also transfer financial capital across boundaries, but they tend to intermediate the transfer, in the sense that they raise money to make
investments in a variety of projects that are not necessarily defined \textit{ex ante}. In both cases, firms will expand abroad if hierarchy is a more efficient way than loans or bonds to link lenders and borrowers located in different countries.

\textbf{Conclusion}

The aim of this section has been to show that, contrary to what is sometimes asserted, transaction costs theory can account for all the major types of MNEs. As shown in Table 4.1, different types of MNE result from the internalization of various types of markets (naturally, in many cases the MNE will simultaneously be internalizing a number of markets). Hence transaction costs theory is a powerful tool to reveal the fundamental features of MNEs which are hidden behind the wide diversity of their forms.

\textbf{New forms of investment and countertrade}

One implication of the transaction costs argument that MNEs reduce the cost of organizing cross-border interdependencies is that attempts by governments to ban or limit their operations should, if binding, have generally negative consequences for the efficiency of the international organization of interdependencies involving knowledge, reputation, distribution services, raw materials and components, and capital. In this respect, transaction costs theory leads to conclusions and recommendations that are diametrically opposed to those made by the proponents of the “new forms of investment.” Indeed, transaction costs scholars have shown that the development of nontraditional forms of trade, such as countertrade, can be seen as a second-best attempt by firms located in countries where governments discouraged the establishment of hierarchical links between home firms and foreign partners to recreate some of the incentives inherent in MNEs.

In the mid-1980s a number of scholars (Buckley, 1985; Oman, 1984) suggested that developing countries, rather than obtaining the inputs they needed from foreign firms as a bundle and giving the profits of the venture to the foreign investor (as happened with the expansion in their country of foreign-based MNEs), should instead purchase these inputs in unbundled forms through a wide range of contractual substitutes (such as licensing, franchising, turnkey and management contracts). These “new forms” of investment would have the advantage of leaving the profits in local hands, thus protecting developing countries from exploitation by foreign MNEs. Some of these scholars saw the trend toward these new forms as inevitable, and applauded their use as providing a solution to the conflicts between MNEs and host countries.

Transaction costs analysis suggests that preventing the transfer through MNEs of the inputs supplied by foreign investors (such as technology, management skills, and access to foreign markets) has negative efficiency consequences. MNEs align incentives between seller and buyer of inputs, and this property cannot always be fully emulated by contractual arrangements, no matter how sophisticated. Hence there are cases where the \textit{mandated} replacement of MNEs...
by contracts will reduce the efficiency of transferring capabilities held by foreign firms to developing countries (Hennart, 1989a).

To better understand why, consider Figure 4.4. Efficient production overseas requires the combination of inputs held by A, a developing country firm, and B, a developed country firm. Sometimes the intermediate inputs supplied by A are sold in relatively efficient markets (for example, commodity chemicals), sometimes in inefficient ones (such is the case with tacit knowledge of local conditions). The same is true for developed country firm B. Sometimes its know-how is easy to license, and sometimes not. When the know-how held by B is easy to license, but the input contributed by A is difficult to measure, the arrangement that minimizes transaction costs consists in A taking a license from B, and holding title to the residual (cell 2). This is because A is both more incited and better able to cheat in the transfer of its input to B than B is in the licensing of its know-how to A. Giving a right to the residual to the party most likely to cheat reduces total transaction costs (Grossman and Hart, 1986). The “new forms” are in this case an efficient way to obtain foreign inputs. If, on the other hand, the inputs held by B are difficult to transact on markets, but A’s inputs are sold on efficient markets, then efficiency demands that B keep full title to the profits, that it become a direct investor, buying A’s input on the market (cell 3). Imposing the use of “new forms” is inefficient in this case, for it increases the level of total transaction costs incurred in combining inputs.

Davies (1977) provides some evidence to support this argument. He compared the technology transfer provisions of joint venture and licensing contracts between British and Indian firms (the Indian government prohibited then foreign firms from entering India with wholly owned subsidiaries). He found that the type and extent of knowledge transferred by joint venture agreements differed significantly from that in licensing contracts, with licensors providing only technical knowledge, while joint venturers were more likely to add tacit marketing and management advice. Because of the imperfections in the market for knowledge described above, some types of tacit know-how could only be transferred through

<table>
<thead>
<tr>
<th>Host country factors held by firm A</th>
<th>Marketable know-how</th>
<th>Non-marketable know-how</th>
</tr>
</thead>
<tbody>
<tr>
<td>Home country factors held by firm B</td>
<td>Marketable know-how</td>
<td>1) Indeterminate</td>
</tr>
<tr>
<td>Non-marketable know-how</td>
<td>3) A licenses B who enters with a wholly owned affiliate</td>
<td>4) A and B joint venture</td>
</tr>
</tbody>
</table>

*Figure 4.4 Institutional modes for market entry*
MNEs. Restricting the entry of MNEs will therefore deprive the country of that know-how.

**Countertrade**

Consider the situation of a firm in a developing country whose government has banned the entry of foreign MNEs and which sees itself incapable of expanding abroad. It needs the tacit knowledge that is most efficiently transferred through hierarchy within MNEs. What can it do to recreate the incentives provided by hierarchy? One solution suggested by Williamson (1985) is to attempt to build reciprocity into its transactions. Reciprocity increases the enforceability of contracts by creating hostages that will be sacrificed in case of breach. This scenario provides the best explanation for the emergence of the most common forms of a particularly misunderstood type of trading arrangements known as “countertrade.”

The term “countertrade” describes a variety of trade practices developed in the 1960s by Soviet-bloc countries to trade with Western firms, and which spread in the 1980s to developing countries. Between 1975 and 1984 the number of countries mandating countertrade rose from 15 to 88, and by the mid-1980s the consensus estimate was that trade under some sort of countertrade arrangement made up between 15 and 20 percent of world trade.

Countertrade has been generally described as a form of barter. As such, its increasing popularity among developing countries has been seen as a puzzling return to archaic practices (de Miramon, 1985). However, a careful look at the structure of three common forms of countertrade contracts (counterpurchase, buy-backs, and offsets) shows that they do not include barter clauses (i.e. the swapping of goods for other goods) but that they consist instead of two separate money-for-goods contracts, with imports made conditional on the exporter purchasing goods and services from the importer. Hence the essence of most countertrade contracts is not barter, but reciprocity (Hennart, 1989b).

Take the case of buy-back contracts. Buy-backs are long-term arrangements made up of two separate but linked contracts. The first contract is for the sale of a plant in exchange for hard currency. In the second contract the plant seller agrees to buy back part or all of the output of the plant in exchange for hard currency. How do we explain these clauses?

Buyers of plant and equipment, especially those located in developing countries, are often considerably less knowledgeable than sellers, and therefore exposed to their opportunism. The economic value of plants also usually depends on continued technical support from vendors. Once the plant is sold, the vendor is usually in a strong bargaining position vis-à-vis the plant buyer. Note that neither of these problems arises when plant sellers integrate vertically into foreign manufacture, since in this case they get their returns not from the sale of the plant but from the output generated by the plant. Consequently, they have no incentives to cheat plant buyers and strong incentives to provide them with technical assistance, since they now all belong to the same firm.
The turnkey contract is one solution to these problems that has been touted by the proponents of the “new forms.” Under these contracts, full payment for the plant is contingent upon the satisfactory completion of trial runs. There is evidence, however, that such contracts do not protect uninformed buyers from opportunism by plant sellers (Abdallah-Kodja, 1984). A buy-back contract under which the plant seller has to purchase some of the output of the plant is one way of reducing transaction costs in the sale of plants by establishing reciprocal commitments. A plant seller that has to take back some of the output of the plant has incentives to provide after-sales service at reasonable cost because interrupted supplies may have adverse consequences. Similarly, the price at which the plant seller agrees to take back the plant’s output reveals whether the plant is outdated or whether there is demand for the plant’s output. Buy-backs are not perfect substitutes for vertical integration into manufacture by the plant seller, however, since they do not protect plant buyers against misrepresentations of operating costs.

Counterpurchase, the other major type of countertrade, can be explained in similar terms. Counterpurchase also involves two parallel goods-for-hard-currency contracts, but in this case the goods taken back are not those produced by the equipment sold, but typically those that lack a ready market. Through counterpurchase, a firm that has difficulty marketing its exports can use the threat of not buying to motivate their suppliers to market their products. This solution is likely to be sought if effective distribution of the product taken back requires manufacturer-specific investments. In that case, a simple distribution contract may not be enough to persuade the distributor to make such investments. One solution in this case is to have manufacturers and distributors hierarchically coordinated within an MNE. When this is not possible because of government restrictions, counterpurchase is the next best alternative.

In both cases, countertrade provides better protection against opportunism than simple contracts when markets for plants or marketing services are subject to high transaction costs, but that protection is not as good as that offered by hierarchical coordination. This suggests that countertrade contracts will mostly be used whenever there are barriers to having a foreign firm expand into local production or a domestic firm into foreign sales. Such is the case for firms located in countries that have banned or severely restricted both inward and outward foreign direct investment (Hennart, 1989b, 1990; Kogut, 1986; Mirus and Yeung, 1986; Murrell, 1982). This hypothesis was investigated by Hennart and Anderson (1993). They reasoned that if countertrade was a second-best alternative to MNEs, then a country’s propensity to countertrade should be greater the higher the barriers it puts on entry by MNEs. Countertrade propensity should also be high when the host country might have wished to attract MNEs, but was prevented from doing so by its high level of political risk. The alternative hypothesis that a country’s countertrade was driven by its inability to obtain foreign exchange to pay for its imports implied that its propensity to countertrade would be higher the more stringent its exchange control (the more overvalued its domestic currency) and the lower its international credit rating. After controlling
for a country's export volume, the results showed no statistically significant relationship between a country's countertrade activity and the stringency of its exchange controls. Furthermore, countries with good credit ratings tended in fact to do more countertrade than those with poor ratings. On the other hand, the higher a country's level of political risk and the limits it put on incoming foreign direct investment, the higher its level of countertrade. This supports the view that countertrade contracts are a second-best attempt to replicate the benefits of hierarchical organization when entry by MNEs is banned.26

Modes of market entry
Transaction costs theory can also throw light on two decisions that MNEs face when entering foreign markets: should they enter (a) with joint ventures or with wholly owned affiliates and (b) with acquisitions or with greenfield (de novo) investments?

Equity joint ventures
One international business institution that is supposedly gaining increasing importance is the equity joint venture (Dunning, 1997). There is no shortage of explanations why firms choose equity joint ventures, but most of the reasons given (pooling complementary inputs, pooling similar inputs) apply as well to other methods of combining the services of assets, such as contracts, greenfield investments, or mergers and acquisitions. Much of the literature thus fails to state the necessary and sufficient conditions under which equity joint ventures will be preferred to the above-named alternatives.

Equity joint ventures are shared equity arrangements.27 The first step in the explanation of joint ventures is to explore the choice between equity and non-equity arrangements. While the literature has often lumped both types into theoretically vague categories such as "cooperative ventures" or "strategic alliances," it is important to distinguish between equity and non-equity ventures because they differ in their incentive structures, and hence in their suitability in organizing a particular type of transaction. Parties to non-equity arrangements (which, for the sake of simplicity, I shall call "contracts") are paid up front for their contribution and hence have the possibility to cheat if measurement costs are high. Equity arrangements, on the other hand, pay input suppliers from the profits of the venture to which they are contributing. Input sellers who deliver inferior inputs will shoulder part of the cost of their dishonesty, since supplying the venture with inferior inputs lowers the profits of the venture and imposes a loss on the input sellers/owners which is proportional to their equity stake. Equity arrangements are therefore more efficient than contracts when the market for the inputs to be obtained is subject to high transactions costs. As we have seen, some types of know-how, reputation, and distribution services fall into this category.

Equity joint ventures are a particular type of equity arrangement, in the sense that equity in the venture is shared between two or more input contributors/
owners. A simple explanation of this is that equity joint ventures arise when the markets for the intermediate inputs supplied by two or more parties are both subject to high transaction costs. Otherwise, as shown in Figure 4.4, the party with the more marketable input could gain by transferring this input by contract to the party with the less marketable input. Assume that the production of a particular good or service requires the combination of two types of knowledge, \( a \) and \( b \), held by firms A and B, respectively. If \( a \) can be easily licensed, but not \( b \), \( a \) and \( b \) will be combined by B, with A licensing \( a \) to B. If the reverse is true, B will license \( b \) to A. For a joint venture to be the chosen method to combine \( a \) and \( b \), the market exchange of both \( a \) and \( b \) must be subject to high transaction costs (Hennart, 1988b).

Figure 4.4 shows the conditions that make equity joint ventures the optimal way to combine complementary inputs (link joint ventures). A similar logic can be used to explain cases where joint ventures will be used to pool similar inputs (scale joint ventures). Scale joint ventures will be optimal when (a) high transaction costs in the market for the inputs make vertical integration between one production stage and another desirable and (b) differences in minimum efficient scale (MES) between stages are such that sharing the equity of one of the two stages is desirable (Hennart, 1988b). In the aluminum industry, higher MES at the alumina than at the aluminum stage, joined to high transaction costs in the market for bauxite and alumina, have led aluminum companies to form joint ventures to operate captive, but efficiently sized, alumina refineries (Stuckey, 1983).

The presence of at least two simultaneously “hard to sell” inputs held by at least two separate parties is not sufficient, however, for joint ventures to emerge. Inputs that are difficult to exchange on the market could be combined if either firm bought out the other, or if they merged, or if either firm were able to replicate the other firm’s assets by purchasing the necessary inputs on the market (the replication option). Further conditions for the existence of joint ventures must therefore be that (1) replicating the assets is more expensive than sharing them, and that (2) mergers and acquisitions of the firms owning the complementary assets are more costly than pooling the services of the assets in a joint venture. Condition (1) obtains when the desired assets have high fixed costs and low marginal costs. Besides the obvious case of governments restricting mergers and acquisitions, (2) will obtain when the assets that yield the desired services are a small and inseparable part of the total assets held by both potential partners or when a merger or a total acquisition would significantly increase management costs (Buckley and Casson, 1987; Hennart, 1988b; Kay et al., 1987).

While joint ventures constitute an efficient way to combine poorly marketable inputs held by two or more firms, they also have offsetting costs. The incentives that suppliers of inputs have to cheat are not totally eliminated, since each joint venture partner can only claim a fraction of the residual value of the business. Each partner may therefore still find it advantageous to maximize his gain at the expense of the venture by, for example, supplying fewer inputs to the venture than contractually agreed when it was formed. This contrasts with full equity
control, where the parent, having full rights to the residual of the venture, is incited to maximize it. Another problem is that the joint venture partners may transfer to their wholly owned operations some of the inputs contributed to the venture by the other partners, and then use these inputs usurped from their partner to compete with them (Hamel, 1991). As a result, the efficiency of a joint venture hinges on the convergence of the goals of parties to the agreement, or, failing this, on the degree to which opportunism by the partners can be controlled by contractual means or by the appropriate design of the venture. Whenever partners have conflicting goals that cannot be reconciled by contracts, their actions will lower the profits available for sharing, and the joint venture mode of organization will prove to be very costly for one or both parties.

**Joint ventures vs. wholly owned affiliates**

This transaction costs model of the joint venture is consistent with the findings of the pioneering studies of Franko (1971) and Stopford and Wells (1972). These authors examined the choice made by US MNEs between wholly owned subsidiaries and joint ventures and concluded that MNEs resisted entering into joint ventures (a) when they already held or could acquire on the market all the assets necessary to operate abroad; and (b) when the market for the assets they were contributing to the venture was characterized by high transaction costs. Hence parents supplying their affiliates (or buying from them) intermediate products which had no market prices were likely to insist on full ownership. Similarly, parents who exploited types of knowledge and goodwill that were difficult to protect through contracts were less likely to form a joint venture (Stopford and Wells thought this was well proxied by the parent’s R&D-to-sales and advertising-to-sales ratios). On the other hand, parents tended to choose joint ventures when they needed complementary resources that they could not easily acquire on the market. The need to joint venture was particularly strong (a) when the foreign affiliate represented a diversification move for the parent, and hence the parent needed industry-specific knowledge or distribution facilities; (b) when the MNE had little experience of the market entered, and hence needed country-specific knowledge; and (c) when it needed resources controlled by local firms. Similar findings are reported by Stopford and Haberich (1978) in the case of British MNEs, and by Yoshino (1976), Yoshihara (1984), and Tsurumi (1976) in the case of Japanese MNEs.

Econometric studies provide further support. Gatignon and Anderson (1988) found that the probability that US parents fully owned their overseas subsidiaries varied positively with the ratios of R&D expenditures to sales and advertising expenditures to sales of the industries of the parents, and with the firm’s international experience (as proxied by the number of previous foreign investments). Joint ventures were more likely in countries with cultures radically different from the US, controlling for legal restrictions on incoming investment and political risk. A study by Gomes-Casseres (1987), using the same database of pre-1975 affiliates of US MNEs, also showed that advertising intensity, international experience, a
high percentage of intrasystem sales within the MNE, and familiarity with the host country all tended to lead to full ownership. Affiliates that were in an industry different from the parent’s and which operated in resource-intensive industries, were more likely to be joint ventures. R&D intensity was insignificant, but an interaction term between diversification and R&D intensity had a significantly negative coefficient, indicating that when the subsidiary was outside the parent’s core business, R&D intensity encouraged joint ventures, whereas when the subsidiary was active in the parent’s main product line, R&D intensity led to full ownership of subsidiaries.

Hennart (1991) looked at the choice made by Japanese MNEs between partial and full ownership of their US subsidiaries. As in the case of US MNEs, Japanese investors were more likely to enter joint ventures when they had little experience in the US market or when their US subsidiary was either in a different industry from that of the parent or in a natural resource industry, indicating the need for country- and industry-specific knowledge and for access to natural resources. Unlike the other two studies, R&D and advertising intensities of the Japanese parents had, however, no impact on the choice between full and partial ownership of their subsidiaries. Hennart’s explanation is that some R&D-intensive Japanese firms might have come to the United States to source knowledge via joint ventures, while others had entered to exploit it through wholly owned affiliates.

Hennart used the parent’s R&D intensity, an improvement over Gomes-Casseres and Gatignon and Anderson’s use of the R&D intensity of the industry of the parent. The correct operationalization of this variable should be the subjective evaluation by the parent of the risk of leakage of their knowledge if they chose joint ventures. This is what Bell (1996) used in his study of the determinants of the choice made by Dutch MNEs between full and partial ownership of their foreign affiliates. He found that parents who transferred crucial proprietary know-how to their foreign affiliates and who thought that know-how could easily leak out to third parties tended to choose wholly owned affiliates.

Hennart and Larimo (1998) looked at whether Japanese firms behaved differently from Finnish firms when choosing between wholly owned affiliates and joint ventures to enter the United States. They argued that because Japan is culturally further from the United States than Finland, the need to acquire the knowledge of how to operate in the United States would be higher, \textit{ceteris paribus}, for Japanese than for Finnish firms. Joint ventures are a good choice to enter an unfamiliar country, since the management of the venture can be delegated to the local partner (Kogut and Singh, 1988). Hence Japanese investors should be more likely to opt for the joint venture. Controlling for all other factors that affect the choice between these two alternatives, Japanese investors were indeed more likely than their Finnish counterparts to choose joint ventures over wholly owned affiliates.

\textit{Joint ventures vs. acquisitions}

As seen above, transaction costs theory suggests that joint ventures will be chosen over acquisitions in three cases:
Transaction costs and the multinational 105

1 in the presence of political, institutional, or cultural restrictions on acquisitions/merger;
2 whenever the foreign investor is worried about the impact of a full acquisition on the motivation of the top management team of the acquired firm; and
3 whenever the assets that provide the desired services are tied up with other, non-desired assets.

One implication of conditions (1) and (2) is that the ratio of mergers and acquisitions to joint ventures should be higher between firms domiciled in a given country than between firms domiciled in two different countries, since cultural and institutional barriers to acquisitions are higher for cross-national than for domestic mergers. Using European Community data on joint ventures and acquisitions/mergers between 1983 and 1992, Kay et al., (1996) found that the ratio of acquisitions/mergers to joint ventures was highest between EU firms based in the same country, intermediate between EU firms based in different EU countries, and lowest between EU and non-EU firms.

Hennart and Reddy (1997) investigated condition (3). They looked at the choice made by Japanese investors in the United States between acquiring a US firm or joint venturing with one. They hypothesized that Japanese investors would be more likely to acquire American firms if they were “digestible,” i.e. if they did not include unneeded assets. Digestible firms, they argued, were firms that were small, or, if large, that were divisionalized, so that Japanese investors could acquire only the relevant parts of the firm. The Japanese would, on the other hand, enter into a joint venture with large, non-divisionalized American firms, because through joint ventures they could access the services of desired assets without having to purchase undesired ones. As predicted, Hennart and Reddy found that the “digestibility” of the American partner had a statistically significant influence on the choice between joint ventures and acquisitions.

Acquisitions

Another major decision facing a firm entering a foreign market is whether to establish a new business from scratch (a greenfield entry) or whether to acquire an existing business (an acquisition). Transaction costs theory can throw light on this choice as well, although the theory here is somewhat less developed. Making an acquisition means purchasing a going business. Setting up a greenfield entry means purchasing local inputs in disembodied form and combining them with inputs held by the investor to form a productive unit. This suggests that the choice between greenfield entry and acquisitions should be determined, at least in part, by the characteristics of the contributions made by local and foreign factors of production, and how they interact.

Consider the following cases. In the 1960s American food companies thought that they knew better than their European counterparts how to advertise and distribute their products. At the same time, they did not know much about how to
Jean-François Hennart

operate in Europe. Marketing know-how, tacit and not protected by patents, could not be profitably licensed. The most efficient way to leverage it abroad was to acquire European food companies with reputable trade names and quality products, and to provide these acquisitions with superior Yankee marketing know-how (Horst, 1974). Here acquisitions made sense since the American investors had little knowledge of European food tastes and production conditions, and US marketing know-how could efficiently be merged into an existing firm.

Japanese automobile assemblers, on the other hand, have entered the United States through greenfield investments. The Japanese automaker's competitive advantage has been superior quality, obtained through sophisticated plant design and labor management practices, such as work rotation, work teams, and quality circles (Florida and Kenney, 1991; Pucik and Hatvani, 1983). Because the successful implementation of many of these techniques requires a change in attitudes and values, they are very difficult to incorporate into an existing American firm (Brannen et al., 1998). Instead it is much easier to inculcate them into a greenfield plant, since this makes it possible to select a location and a labor force that will facilitate the transfer.

The preceding analysis suggests that, ceteris paribus, acquisitions will be chosen over greenfield investments (1) the easier it is to merge the foreign investor's contribution with an existing firm and (2) the less knowledgeable investors are with the market entered. Diversified firms are firms where quasi-independent subsidiaries share general management or marketing know-how generated at headquarters. Hence we would expect that investors who are diversified, and/or who invest in industries unrelated to their domestic business, would tend to prefer acquisitions to greenfields. Acquisitions are also indicated when the foreign investor wants to combine and rationalize existing physical assets (economies of scale), or needs access to existing distribution channels (economies of scope) (Hassan and Jemison, 1991). Greenfields will be chosen when the foreign investor wants to transfer socially embedded processes.29

The empirical literature provides some support. Caves and Mehra (1986), in the case of American firms, Zejan (1990) in the case of Swedish firms, and Barkema and Vermeulen (1998) in the case of Dutch ones, found a significant correlation between a parent's overall product diversification and its preference for acquisitions over greenfields. Hennart and Park (1993) looked at Japanese entries into the United States. Controlling for financial factors, such as the investor's leverage and the relative levels of stock prices and exchange rates, they found that R&D-intensive Japanese parents, whose advantage consisted in work practices that were socially embedded, favored greenfields over acquisitions. Acquisitions were chosen, on the other hand, by Japanese firms effecting a diversification. Financial factors were not significant. Andersson and Svensson (1994) analyzed the choice of Swedish MNEs between acquisitions and greenfields and confirmed Hennart and Park's findings that a parent's R&D intensity encouraged greenfield entry.
Conclusions

As Kay (1991: 143) has pointed out, “it is misleading to talk of ‘the’ transaction costs approach when there are in fact a number of variants of transaction costs reasoning.” Transaction costs theory is an approach more than a set theory. The richness and the diversity of the literature this approach has stimulated is a sign of its fruitfulness. Given the subtle differences in the way various authors have used transaction costs theory to explain the MNE, it might be useful to highlight what is idiosyncratic in the version developed in this chapter.

First, I have argued that a full understanding of the expansion of MNEs requires a comparison of the costs of organizing interdependencies within firms and those of doing it in markets. A theory of why markets fail is not sufficient to explain the MNE; one must also have a theory of why firms succeed. The preceding pages argue that firms succeed not because they use “internal prices,” but because they use a method of organization called hierarchy which extinguishes market transaction costs by changing the incentives of the parties. The unavoidable consequence of using hierarchy is the generation in firms of a different type of cost, which arises from the decrease in the motivation of employees who are no longer paid in strict proportion to their output measured at market prices. The choice of institutional form involves therefore a tradeoff between the cost of measuring intermediate outputs (market transaction costs) and that of motivating and directing employees who are no longer paid in strict proportion to their output (internal organization costs). Taking full equity is inherently costly, because it automatically dulls the incentives of agents. This simple model explains why firms will sometimes use non-equity forms (such as franchising) to exploit abroad their reputation, or will willingly choose to share the ownership of their foreign affiliates.

Second, the analysis of this chapter has focused on the organization of international interdependencies, not on the internalization of “firm-specific advantages.” As the preceding pages show, there are many cases when firms expand abroad without any firm-specific advantages: for example when they make backward investments to procure raw materials and intermediate products. When US steel makers invest into the mining of Liberian iron ore, they do not do so to exploit abroad a monopolistic advantage in a superior proprietary technology. In fact, many steel firms have so little competence in the mining of iron ore that they have their captive mines managed by specialist firms (Cleveland Cliffs and Hanna Mining, for example). Similarly, many freestanding firms did not have any firm-specific advantages to exploit abroad because they had no domestic business. Focusing on the internalization of advantages also makes it difficult to explain why some firms invest abroad to acquire technology. There are many cases of Japanese foreign direct investors acquiring a share in American high-technology firms in exchange for fresh capital. In this case the Japanese investor does not have any technological advantages to exploit (Christelow, 1995).

The raison d’être of MNEs is more generally that they are more efficient than markets in organizing international interdependencies. In other words, MNEs
exist because they internalize markets, including, but not exclusively, markets for advantages. Hence US steel makers invest abroad when organizing the coordination between mine and blast furnace through hierarchical processes which is less costly than procuring iron ore through long-term contracts. Focusing on the internalization of markets also offers a simple and straightforward explanation for technology-seeking investments: an equity link between a Japanese firm seeking access to technology and the high-technology American firm which possesses it is made necessary by the presence of high transaction costs in the market for knowledge, while high transaction costs in the international market for financial capital may explain why the American firm will need to offer an equity stake to its Japanese lender.

This brings us to our third point. The right question to ask to understand international business institutions is not so much what arrangements the MNE needs to use to safeguard its “monopolistic advantages,” but rather, what is the most efficient way to combine the local and foreign inputs needed to operate in a foreign country? In all cases an MNE which operates abroad must obtain access to local inputs. It will keep full ownership of its foreign affiliate if it can access these inputs on the market. If this is not the case, and the necessary inputs are held by a local firm, the foreign investor will form a joint venture with that firm. But in all cases the most efficient solution will be the one that minimizes the costs of organizing the international interdependence between factors of production located in two separate countries.

Thinking in terms of interdependence also throws light on an apparent puzzle in the transaction costs theory of the MNE. The theory as it stands predicts that an inefficient market will be internalized, but it does not predict by whom. Recall our example of the banana industry. Transaction costs theory explains why plantations and distribution facilities ought to be integrated, but not why banana distributors have integrated into banana plantations rather than the reverse. As we have argued above, firms experience internal organization costs when they internalize a market. These costs arise because reducing cheating costs is achieved by reducing the motivation of employees. Whether the distributor will take over the banana plantation, or whether the reverse will take place, should depend on the costs experienced by firms of each country in motivating employees of the other country. Hence if the banana distributor is more skillful at monitoring banana plantation employees than the banana plantation is at monitoring employees of the distributor, the distributor will find it profitable to take over the banana plantation firm. In other words, the foreign direct investor will be the interacting party who has the better management skills, and therefore experiences lower cost in managing the other party’s employees (Grossman and Hart, 1986). This explains why the pattern of vertical integration in the banana industry has been one of American shippers and wholesalers integrating into Central American plantations, and not one of Central American growers integrating into US distribution. It also explains why Middle East petroleum producers have been slow at integrating into European distribution in order to recreate the vertical links between crude oil production and refining that had been
established by Western oil companies, but broken with the nationalization of their Middle East holdings.⁴⁰

Acknowledgments

I wish to thank Jean Boddewyn, Peter Buckley, Mark Casson, Geoffrey Jones, Roger Sugden, Mira Wilkins, Bernard Yeung, and especially Kathleen Saul, for many valuable suggestions on an earlier version of this chapter.

Notes

1 Transaction costs theory differs in subtle ways from its close relative, internalization theory (Rugman, 1981). Internalization theory focuses on the internalization of advantages, while transaction costs theory considers the internalization of markets. The transaction costs theory version outlined here also presents a different explanation of why firms can be more efficient than markets and considers both home country and host country factors of production when considering the forms taken by expansion abroad. All these points are covered below.

2 In 1990, for example, more than 50 percent of the financing of the foreign affiliates of US firms was obtained locally (through reinvested profits or local borrowing) (Delapierre and Milleti, 1995: 18).

3 Contrary to what is sometimes asserted by its critics, transaction costs theory does not assume opportunism to be constant across space and time, and so what individuals rationally do to protect against opportunism will vary according to local conditions. One always locks one's car in New York except in special circumstances; one never locks one's car in Nova Scotia, except in special circumstances.

4 Note that, in contrast to principal-agent theory, transaction costs theory can accommodate any assumption about the risk preference of agents and principals, and is compatible with assumptions that both principals and agents are risk neutral.

5 This is true whether the firm operates the subsidiary through expatriates or through local personnel: the cross-cultural contact, with its potential for misunderstandings and deception (Root, 1987), must be made in all cases, whether between HQ and the subsidiary manager, between the subsidiary manager and his or her employees, or between the employees and local suppliers or customers.

6 The fact that there are decreasing returns to monitoring is perfectly compatible with transaction costs theory, contrary to what Ghoshal and Moran (1996) argue.

7 One can make the parallel assumption that the opportunities for fraud rise more than proportionally when one substitutes price for behavior constraints. Hence the presence of behavior constraints in market transactions (i.e. contracts).

8 Note that the imposition of tariffs per se is not a sufficient reason for foreign production, unless MNEs can minimize tariff payments by undervaluing intra-firm trade.

9 Tacit technology is difficult to define in a patent, and hence will be difficult to license (Hennart, 1982).

10 If its consumers are repeat customers, a franchisor who would reduce quality would bear the full monetary consequences of his action. However, the greater the percentage of repeat customers, the less valuable a franchised trademark, since repeat customers do not have high search costs and hence have less need for trademarked products. Hence we can expect the free-riding problem to be significant in the case of most franchised trademarks.

11 Franchisors also shirk when they fail to monitor the quality of the franchisees’ output. One of the reasons for McDonald’s success was its founder’s policy of not granting
area franchises, but instead of franchising one outlet at a time. At the time, franchisors granted such area franchises in exchange for up-front cash payments, and then withdrew from the business (Love, 1987). One additional reason why franchisors may own outlets is that they can use such outlets to persuade potential franchisees that they will stay in the business.

In February 1996, for example, the average cash flow of an American franchised McDonald's unit was $15,340, or 13.7 percent of sales of $112,000. Company-owned stores had a lower cash flow of $12,690, equal to 10.4 percent of sales of $122,460. Hence franchised units were more efficiently run than company-owned ones (Gibson, 1996).

An efficiently sized bauxite mine costs half a billion dollars and a refinery between 500 million and a billion.

Similar problems arose in the iron ore industry following the slump in demand in the 1970s (Franz et al., 1986).

There are some cases of vertical integration between mines and smelters in the lode sector, but they reflect political motives. For further details see Hennart (1986a, 1988a).

Fiscal considerations may also be relevant here. Vertical integration transforms arm's length trades into internal ones. This makes it possible to alter the nominal price at which transactions take place without affecting revenues. By altering nominal prices, the MNE has the possibility of shifting accounting profits to low tax jurisdictions, to reduce \textit{ad valorem} tariff duties, and to repatriate earnings disguised as expenditures. This possibility is limited by the presence of world prices for the goods shipped and by the sophistication of tax authorities.

In 1980, 60 percent of the banana export trade was handled by three vertically integrated MNEs (Casson, 1986: 51).

Inversely, retailers will vertically integrate into manufacturing if it is more difficult for them to control the quality of the manufacturers' goods than it is to control shirking by manufacturers if they are employees of retailers. Note that a third solution is to have retailers put their brand on the manufacturer's product (Chen and Hennart, 1998).

In large corporations this right is exercised through the board of directors.

"Just a brass nameplate in the City" is how Nicholas (1982) described it.

About three-fourths of the companies registered between 1862 and 1914 in Edinburgh to do business abroad were in mining and agriculture (Schmitz, 1993). Forty percent of the 2,544 British freestanding firms operating outside the US and Canada listed in a 1916 US Federal Trade Commission report (US FTC, 1916) were in mining, and another 28 percent were in agriculture. Sixty-nine percent of the Dutch freestanding companies identified by Sluyterman as active in 1913 were in tropical agriculture and in mining and petroleum extraction (Sluyterman, 1994).

As late as 1967 UK-registered freestanding firms were producing over 75 percent of Nigeria's tin output (Schatzli, 1971).

Today shares of small companies developing North and South American mines are sold directly to the public on the penny stock markets of Spokane, Denver, and Vancouver (Mikesell and Whitney, 1987), as they were in London in the heyday of the freestanding firm.

Here as well, there are performance inseparability problems, since failure of a turnkey plant may be due to defective technology or to mistakes made by local operators. Ascertaining who is responsible is often very difficult.

The traditional argument is that countertrade is barter. Barter makes it possible to bypass exchange control regulations by swapping exports for imports without having to use foreign exchange. A low credit rating makes it difficult to borrow foreign exchange, and thus encourages importers to resort to barter.

For similar findings at the firm level, see Marin and Schnitzer (1995).
Some authors restrict the term “equity joint ventures” to shared-equity greenfield ventures, and exclude partial acquisitions. I know of no theoretical reasons to exclude partial acquisitions, and, unless specifically noted, the term “joint venture” in the following paragraphs refers to both greenfield joint ventures and partial acquisitions.

Benito (1996) had similar results in the case of Norwegian firms investing abroad. Larimo (1993), on the other hand, did not find any impact of cultural distance on ownership preferences.

That acquired subsidiaries tend to be less integrated into the parent’s network is suggested by the fact that Japanese manufacturing subsidiaries in the United States that had been acquired had, ceteris paribus, a statistically higher probability of being sold off between 1980 and 1991 than those that had been established as greenfields (Hennart et al., 1998).

Alongside increasing managerial sophistication in some Arab countries (notably Kuwait), asymmetries between European countries and Middle East countries in restrictions on inward foreign direct investment are also a factor.

References


Cantwell, J. (1989a) "Theories of International Production." University of Reading Discussion Paper in International Investment and Business Studies.


Transaction costs and the multinational


Jean-François Hennart


Transaction costs and the multinational


Klein, S., G. Frazier, and V. Roth (1990) “A Transaction Cost Analysis Model of


Transaction costs and the multinational


