

1. *The Future of the Multinational Enterprise* in historical perspective

Mark Casson

1 INTRODUCTION

This chapter examines the evolution of the internalization theory of the multinational enterprise over the past 40 years. It looks through the lens of a book that was published (at the time of writing this chapter), 40 years ago: *The Future of the Multinational Enterprise*, co-authored by Peter Buckley and myself. This book has been widely cited. A new edition was prepared for its 25th anniversary. The 30th anniversary was marked by the publication of *The Multinational Enterprise Revisited*, which was a compilation of the author's joint papers. So why carry out another retrospective on the 40th anniversary? One reason is that a 40th anniversary is more interesting than a 30th. There is a longer time to look back on.

Longevity implies survival. This raises interesting questions:

- Has the theory been refuted by new evidence? No, although some have claimed that it has (e.g., emerging market multinationals).
- Has the theory been superseded by new and better theory? No – though some have claimed that it has (e.g., resource-based theory of the firm).
- Has it been forgotten? No – it is regularly cited.

In the light of this, can internalization theory be judged a success? My answer is 'Yes' and 'No'. The 'Yes' is hardly surprising, given the authors' personal bias. The 'No' part is the most interesting. This chapter claims that the theory has failed to develop its full potential. The development of the theory has diverged along two very different paths. Broadly speaking, Peter and I have continued to follow the 'economics' path we started down, but many other scholars – particularly younger scholars – have pursued a 'strategic management' path instead. The theory has certainly not developed along the lines that Peter and I expected. Some of the economic content of the theory has been lost. How much has been gained is a matter for debate.

The 'good news' is that there is exciting work to be done by exploring further down the economics path. The bad news is that many international business scholars may be unwilling to follow because they are already committed to the other path. I hope that readers of this chapter will join us on the economics path once they begin to see its enormous potential.

In this chapter I propose to look at internalization theory at three dates: 1976 (the past), 2016 (the present at the time of writing) and 2026 (the future). The past is examined in sections 2, 3 and 4. The objectives of internalization theory are set out and the logical structure of the theory is explained. Recent advances down the economics path are set out in section 5. I call the present state of the economic theory 'Internalization Plus'. Compared to the original, it is more explicit about the general principles that underpin it, and addresses a wider range of issues. Section 6 examines why it has taken so long for this theory to develop along these lines. I argue that academic incentives have distorted theoretical development by diverting resources excessively into the strategic management path. Section 7 concludes by spelling out a research agenda, and setting goals that ought to be achieved by 2026.

2 THE ORIGIN AND EARLY CONTEXT OF INTERNALIZATION THEORY

Internalization is often presented as a theory of international business (IB), but the theory was developed at a time when IB had only recently emerged as a recognized field of study. It was conceived mainly as a contribution to economics (Buckley and Casson, 1976). It was only later that it became viewed as a theory of IB. In 1976 the *Journal of International Business Studies* (JIBS) was only recently established, and was publishing papers in economics, finance and administration rather than business strategy. Stephen Hymer (1976), John McManus (1973, 1975), John Dunning (1977), Alan Rugman (1981), Jean-François Hennart (1982), Peter Buckley and myself all trained as economists. Internalization theory, therefore, was parented by economists and then adopted by IB. This explains quite a lot about its subsequent evolution, as I suggest below.

There are two main ways of looking at the origins of internalization theory, and both are correct. One is that it was a response to a pressing practical problem, and the other is that it set out a general theory of the firm. First, the problem. As a branch of economics, internalization theory was concerned with the global economic system and the place of multinational enterprises (MNEs) and foreign direct investment (FDI) within it. The structure of this system changed after World War II through

exogenous changes driven by new technologies and policy shifts towards free trade. The endogenous reaction was the growth of MNEs. This elicited a political backlash, reminiscent of present times, which presented FDI as a threat to national sovereignty (Vernon, 1971). But the nature of this threat was not entirely clear because the rationale for FDI was not fully understood. Because economics could not explain FDI convincingly, economists lacked credibility in the policy debate. Second, the theory. The claim of *The Future* was that economics lacked a theory, and not that IB needed one. As far as economists were concerned, IB was simply a novel field of applied economics. The problem was that industrial economists had failed to develop an adequate theory to explain the MNE because they implicitly identified a plant with a firm. This meant that firms were visualized as single-plant operations; since an MNE is necessarily multiplant, this implicitly assumed away the MNE. The failure of industrial economics opened the door for international trade theorists. But they could not do any better either. Indeed, their efforts were a total failure.

Trade economists Donald MacDougall (1960) and Murray Kemp (1970) developed a theory of FDI by extending neoclassical theory. Capital mobility, they maintained, was the key to understanding FDI. An MNE was simply a firm that carried out FDI. Unfortunately, however, they failed to distinguish between portfolio investment and direct investment. Capital, they said, was homogeneous and flowed in only one direction at a time, from low-profit countries to high-profit countries. This applied to all industries. It was a disequilibrium phenomenon of a transitory nature and would cease once profit rates were equalized. They could not have been more wrong:

- FDI flowed in both directions at once.
- It flowed mainly from high-profit countries (e.g., the USA) to low-profit countries (e.g., Europe) and not the other way round (as predicted by neoclassical theory).
- It was concentrated in high-technology and advertising-intensive industries (Dunning, 1958).
- It showed no sign of abating; flows were sustained and, indeed, growing.

Here was an anomaly that could not be ignored.

Internalization theory explained everything – the anomalies and paradoxes – at a single stroke. It showed that the transfer of knowledge, and not the transfer of capital, was key. Markets for knowledge were highly imperfect, unlike markets for financial capital. When markets were imperfect, ownership was crucial in appropriating rents, and ownership involved

direct rather than portfolio investment. Trade theory was relevant to the location of foreign production plants, but foreign ownership could only be understood through a theory of the knowledge-based multiplant firm. The theory applied to all forms of proprietary knowledge, including technology, product design, and brands.

The gist of internalization theory, as expressed at the time, may be paraphrased in the following question:

If a firm headquartered in one country possesses knowledge useful to another country, and a firm in the other country has information on local conditions that the knowledge-owner does not possess, why not sell the knowledge to the firm that has the local knowledge so that they can exploit it in their local market? Why not replicate this procedure for every local market to which knowledge can be transferred? This leaves the knowledge-owner free to do what they do best, which is to continue developing knowledge, and the local firms to do what they do best, which is to exploit their local expertise.

The answer was:

This idea would be great if you could trust your knowledge-partner in the foreign market. But often you cannot, and so you need to keep control of the knowledge yourself. Instead of selling the knowledge, you sell the product that embodies the knowledge. If you produce that product abroad you will become an MNE.

The Future created a stir on two main counts. First, it helped to promote globalization by changing policy attitudes worldwide. After the end of World War II the USA pursued a free trade agenda to encourage foreign countries to open up their markets to US products. The so-called 'dollar shortage' of the 1950s meant that many US products had to be produced abroad, and so US business and government sought access for investment as well. But many countries resisted this. They believed that foreign capital siphoned off domestic profits, thereby impairing local development and increasing 'dependency' on the USA. But if investment was just a vehicle for knowledge transfer, as internalization theory suggested, then the balance of advantage might lie with the host country instead. FDI could boost 'catch-up' industrialization, particularly if knowledge spilled over for free to indigenous businesses. John Dunning (1977) was influential in persuading the United Nations to endorse this policy change. Second, *The Future* offered a new perspective on FDI. It highlighted alternative contractual arrangements such as licensing, franchising and subcontracting. *The Future* did *not* argue that internalization is always the best strategy. On the contrary, it drew attention to the advantages of the external market at a time when internalization was implicitly assumed to be the norm. In the

1970s, no management gurus took the licensing option seriously, and we were even criticized for raising the issue. *The Future* actually predicted that licensing would become more attractive as property rights became more secure, and that, as a result, the global dominance of monopolistic megamultinationals, predicted at that time, would not materialize. The promotion of contractual ‘alternatives to the multinational’ became an important theme of our subsequent work, and a theme of United Nations’ policy too (UNCTAD, 2011).

3 THE LOGICAL STRUCTURE OF INTERNALIZATION THEORY

As far as the authors were concerned *The Future* represented the beginning of a research agenda and not just the solution of a practical problem (Buckley and Casson, 2009). Neoclassical economics was a dominant force in the 1970s, and was taken up as a policy instrument by Margaret Thatcher, Ronald Reagan and other politicians. But since it failed to explain FDI and the MNE, it must surely be fundamentally flawed, we believed. What was needed was not a ‘quick fix’ but a radical overhaul of neoclassical theory. A diagnostic investigation was required into where its problems lay. Rejecting the entire theory on ideological grounds was popular in some intellectual circles (Hymer, for example, converted to Marxism), but we decided to follow a different approach. Our approach was more clinical: to identify and then isolate specific assumptions that were causing problems, and replace them with more plausible assumptions instead.

We sought a theory in which the MNE emerged as a natural form of economic organization, and not as some anomaly that required a special explanation. The key was to develop a theory of the multiplant firm. A multiplant firm operates several plants in different locations. National borders are often arbitrary; for example, they are often drawn up to consolidate territory gained or lost through war. For reasons of security they are often drawn along rivers and other natural features; but rivers are natural conduits of trade and it makes no economic sense to create political barriers along them. If we could identify the economic logic of the multiplant firm then it would be easy to account for the multinational on the basis of the illogicality of national borders. Multiplant firms would naturally be multinational if the ‘costs of doing business abroad’ (or the ‘liability of foreignness’ in modern terminology) were not too high.

It is often said that the easiest way to solve a problem is to embed it in a bigger problem and treat it as a special case. This was our approach. If we could understand the logic of the multiplant firm then we would have

a theory of the multinational as a special case. The unination firm would be another special case. Because unination firms are very common – there are millions of them – it might seem that they must represent the general case. While that may be true statistically, it is not true analytically, because the unination firm is just a special case of an N -plant firm where $N = 1$. If we could embed the theory of the multinational within a theory of the N -plant firm then it would not only explain the difference between a multinational and a unination firm, but also between a regular multi-plant firm and a uniplant firm.

The key intellectual challenge, therefore, was to create a theory of the multiplant firm. By the 1970s, empirical evidence clearly showed the importance of multiplant firms in both a domestic and international context. Chain stores such as Woolworths were common and in many industries national champion firms owned multiple plants, whether they were multinational or not.

The Future, therefore, set out the elements of a general theory of the firm in space. The theory had two distinctive features. (1) It drew on institutional thinking and not just neoclassical economics. (2) Because of the multinational aspect, the spatial analysis was very thorough. It included transport costs, communication costs and costs of foreignness. It thereby addressed not only ‘geographical distance’ but ‘cultural’ and ‘political’ distance too.

The theory rejected four main tenets of neoclassical orthodoxy. Two of these tenets were explicit: (1) markets work perfectly; (2) knowledge is free to all. Two were implicit in the way that the theory was applied: (1) FDI involves a flow of homogeneous capital; and (2) each firm owns just a single plant.

The Future presents a synthesis of five main components of theory:

- institutions;
- knowledge;
- monopoly;
- classical economics of production and trade;
- economic geography.

The comparison with neoclassical theory is summarized schematically in Figures 1.1 and 1.2.

The first two components were key. They were taken from Coase (1937) and Arrow (1962) respectively. The other three were already familiar. Monopoly theory was standard and had previously been used in Hymer’s 1960 doctoral dissertation (published as Hymer, 1976). The classical theory of production and trade incorporated elements of neoclassical trade theory, but shorn of its objectionable features.

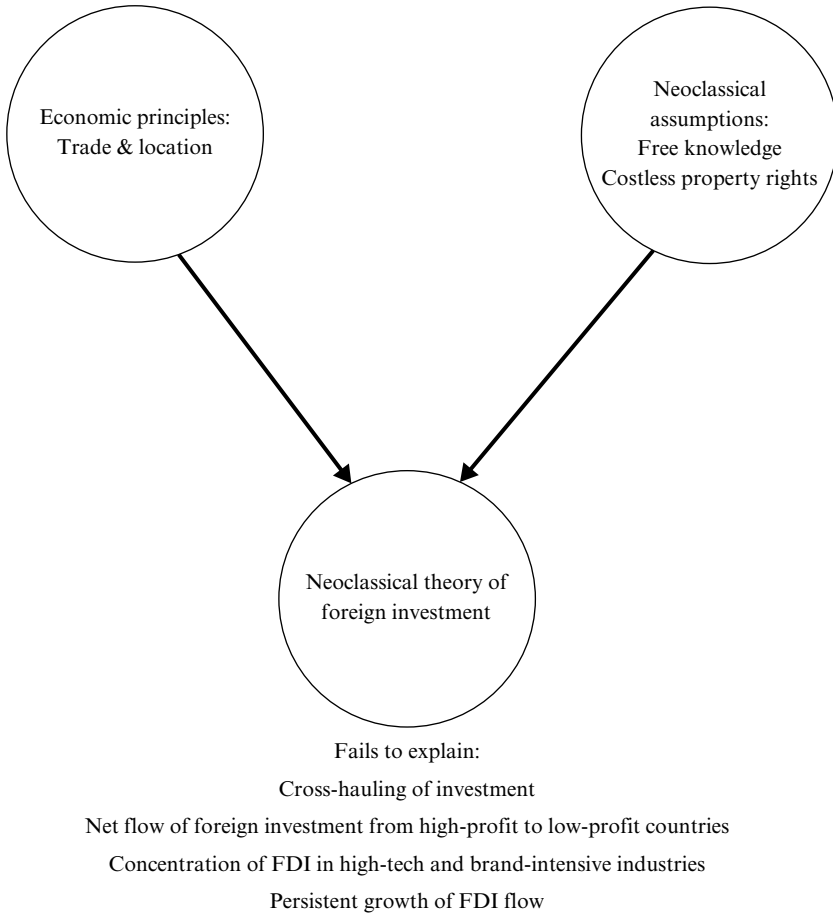


Figure 1.1 Key elements of neoclassical foreign investment theory in 1976

Institutional theory Ronald Coase noted that, because of market imperfections, some markets work better than others and some do not even work at all. In certain cases it may be better to replace markets with a planning mechanism. This is done by bringing interdependent activities under common ownership and control. Final product markets cannot normally be internalized by firms, but intermediate product markets can. Both the buyer and seller become part of the same firm; the market for the intermediate product is internalized within the firm.

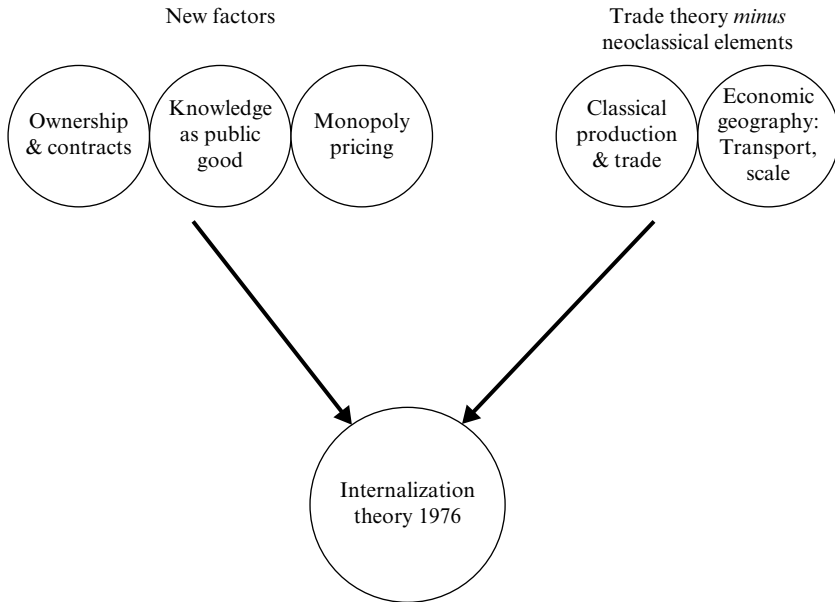


Figure 1.2 Selective synthesis of components in internalization theory 1976

Knowledge The knowledge used by firms is an intermediate product, linking R&D with production and marketing. If the market for knowledge is imperfect then it can pay to internalize it. Knowledge is an intangible asset. The market for it is highly imperfect. The buyer may be uncertain about the knowledge, but if the seller clarifies it they may inadvertently give it away. If the seller keeps it secret, the buyer may suspect that the knowledge is flawed, and so offer less than it is really worth. This is the ‘buyer uncertainty’ problem (Vaitsos, 1974). Patents and copyrights can mitigate this problem, but they cannot resolve it fully.

Knowledge is also a public good. This does not mean that it is free – just that it can be shared. Sharing knowledge is costly, but usually cheaper than replicating its discovery.

Knowledge is a global public good. Some forms of knowledge are more global than others, however. Scientific knowledge can be globally applied, and spread through professional scientific networks; local knowledge can be broadcast globally, but is difficult to interpret outside its local context.

The commercial exploitation of global knowledge normally requires a synthesis of global knowledge and local knowledge. If it is easier to buy local knowledge than to sell global knowledge then it pays to internalize

knowledge use by hiring local managers with local knowledge instead of selling the global knowledge to a local licensee. In some cases, though, it may be easier to buy global knowledge than to sell local knowledge, in which case licensing may be more efficient instead.

Monopoly and cartels The standard theory of monopoly, developed by Chamberlin (1933) and Robinson (1933), can be applied to knowledge. If R&D is privately funded then knowledge must be exploited through a monopoly, for otherwise competition from imitators will eliminate profit. If global knowledge is exploited through local licensees then these local licensees must not compete in each other's territories, because this would drive down prices and erode profits. To control harmful competition between licensees, internalization may be required.

Classical economics Classical economics analyses the division of labour, the modularization of production tasks, and specialization based on comparative advantage. These are all aspects of the theories of production and trade developed by Adam Smith (1776 [1976]) and David Ricardo (1817 [1951]), well before neoclassical economics appeared.

Economic geography Economic geography analyses 'distance' effects, and in particular, the interplay of economies of scale, transport costs and tariffs. Economies of scale encourage the concentration of production in a single plant, while high transport costs and tariffs encourage the dispersal of production across separate plants (Weber, 1929; Iammarino and McCann, 2013).

Internalization theory synthesized these five carefully selected components of theory. Logic demanded that a common set of fundamental principles was used throughout. These converted the theory from an eclectic mix of ideas into a coherent and consistent whole.

Because *The Future* was written for economists, these principles were not made explicit. Although they are mentioned in the literature cited in the book, they were not discussed in detail. It was assumed that economists who read the book would share a common understanding of them, while non-economists would not be interested in them. To extend the theory, however, it is important that to make them explicit, because consistency must be retained as the theory evolves. The principles are therefore presented in updated form in the Appendix. These principles guide the further development of the theory as explained in the remaining part of the chapter, and indeed, underpin the analysis in the remaining chapters in this book.

4 MODELLING THE PRODUCTION SYSTEM

The basic unit of analysis in *The Future* is not the firm but rather the global economy. Unlike the firm, which is dependent on external links to individual owners, customers and workers, the global economy is a self-contained unit, and is therefore an ideal unit of analysis. The theory analyses the impact of the global economy on the behaviour of the firm, rather than studying the firm as an autonomous unit. This makes a lot of sense in both theoretical and practical terms.

Because the global economy is highly complex, it is useful to make some judicious simplifications. Internalization theory therefore focuses on a single industry. The firm is embedded in the industry and the industry in the economy. This was the traditional approach in industrial economics. Conventional industrial economics, however, focused on the national economy rather than the global economy. It also focused, as noted above, on the single-plant firm. To overcome these limitations, internalization theory focused on the production system as a whole and, in particular, the production system in a global industry. This system was conceptualized as a modular system in which different specialized types of plant were connected to each other by intermediate product flows. Each plant has a specific location. National borders partition space.

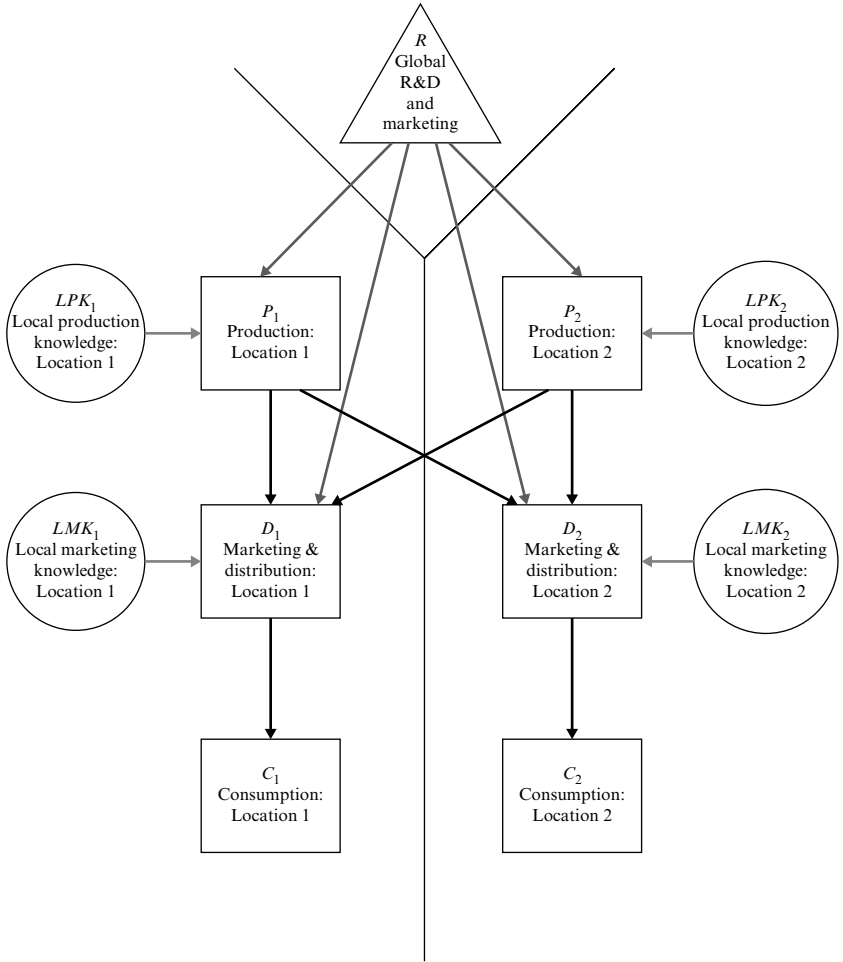
Internalization is often analysed in terms of just two activities, usually production and R&D. It is said that R&D generates proprietary knowledge, and feeds this knowledge into production. In fact, *The Future* distinguishes four activities. Production is split into upstream production and downstream production, and marketing is introduced as well. There are therefore two production-based activities and two knowledge-based activities (marketing and R&D). Moving away from a strictly dyadic approach enriches the theory.

In the dyadic model the MNE serves a global market from production plants in different locations. In the four-activity model there are other possibilities. The MNE may integrate two stages of production in different countries; for example, a metal-manufacturing MNE may mine in a mineral-rich country but refine in an energy-rich country. A confectionery-manufacturing MNE may exploit technology or marketing expertise by serving different markets through local production in different countries. A motor-manufacturing MNE may serve local markets through offshore production in third-country locations. These distinctions correspond to Dunning and Lundan's (2008) classification of resource-seeking, market-seeking and efficiency-seeking FDI.

The four-activity model was summarized on page 34 of *The Future*. More activities means more linkages. It is the number and nature of the

linkages that are crucial to internalization theory. These linkages all involve intermediate product flows.

A simplified version of the four-activity model is presented in Figure 1.3. In this version, the second stage of production is identified with local



Note: *R*: research and development; *P*: production; *D*: distribution; *C*: customer; *LPK*: local production knowledge; *LMK*: local marketing knowledge; box: physical activity; triangle: generating knowledge; circle: generating local information; lines: black: product, dark grey: knowledge, light grey: information; arrows: direction of principal flow.

Figure 1.3 Interaction of production, marketing and R&D: Potential intermodular linkages involving flows of intermediate products

marketing and distribution. R&D includes product design and brand development as well as technology development. The figure distinguishes between global knowledge generated by development activity, which applies across all locations, and local knowledge of production conditions and markets, generated by practical experience, which is specific to particular locations. Successful innovation, the figure indicates, requires a synthesis of global and local knowledge, as explained above. The figure is confined to the case of just two countries, but the same principles apply however many countries are involved.

The links between production, marketing and consumption, illustrated in the figure, constitute a rudimentary supply chain. This product supply chain is supported by a knowledge supply chain that feeds in general knowledge from development and specific knowledge from local sources.

This system needs to be coordinated. Production plans for individual plants need to be harmonized, because otherwise there will be supply shortages, or excess inventory, both of which involve unnecessary waste. This process is known as coordination (see Appendix). Product flows need to be synchronized. Synchronization requires transport planning, inventory management and sales forecasting (Buckley and Casson, 1985). The transfer of knowledge from R&D to production and marketing also needs to be coordinated.

Coordination is effected through contractual arrangements, and these arrangements need to cover each linkage in the system. The system contains many potential linkages, and location decisions determine which of these linkages actually operate. Internalization determines how each operational linkage is coordinated.

Location decisions are, in general, interdependent, for example, location of R&D may influence the location of production. Internalization decisions are also interdependent, for example, if R&D and production are owned by the same firm, and production and marketing are also owned by the same firm, then R&D and marketing are necessarily owned by the same firm as well. The benefits of internalizing the linkage between R&D and marketing may therefore influence decisions to internalize production and marketing or production and R&D.

Crucially, location and internalization decisions are interdependent too. The costs of doing business abroad mean that the costs of internalization may vary between locations (Hymer, 1976; Zaheer, 1995). Thus, location can influence internalization (e.g., domestic production may be internalized when foreign production is not) and internalization can influence location (exporting may replace FDI if a foreign government prohibits FDI).

Many different combinations of ownership and location are possible.

The whole system could be coordinated by a single firm. If this firm carries out activities in both countries it will become an MNE. There are various configurations of MNE; for example, the firm could control production but franchise local distribution, or control distribution but subcontract production, or it could do both. It could pursue different strategies in different countries. Domestic strategies often differ from foreign strategies, and strategies in different foreign markets may differ too.

Different configurations involve different numbers and different types of firms. There are several possibilities, even in the two-country case. If the MNE subcontracts to local firms in both countries while retaining control of local marketing then there are three firms in the system altogether: the MNE and the two subcontractors. On the other hand, if the MNE subcontracts to the same firm in both countries then there are two MNEs: the MNE that developed the product and the MNE that produces it. The MNE that developed the product could still be an MNE because it would control distribution in both markets. A licensee can become an MNE while the technology owner is not. If the technology owner grants a single global licence then the licensee may operate a global network of plants while the licensor does not produce at all. The licensee could even offshore production to the licensor's headquarters country.

Internalization theory therefore generates many possible configurations for both firm and industry. Which configuration prevails at any given time depends, in general, on the state of the global economy and the state of each global industry in particular.

Different industries may exhibit different configurations and there may be differences between firms within each industry too (Spender's 1989 'industry recipes'). Internalization explains the nature and direction of these changes. Internalization theory is not a theory of any particular type of MNE. In the context of the theory 'US', 'European', 'Chinese' and 'emerging market' MNEs simply represent different outcomes of the internalization process. Which type of MNE emerges is contingent on the headquarters location and the nature of the industry. Internalization is a general theory, as noted above and it is unnecessary to develop special theories every time circumstances change.

5 'INTERNALIZATION PLUS'

If *The Future* were being written today, it would be easy to extend the scope of the theory. Additional components could be included. R&D location and headquarters location could be made endogenous, augmenting the set of location strategies. Pricing policies could be analysed.

Combining information on revenues and costs would make it possible to calculate profits.

It would also be possible to incorporate theoretical developments that have been made over the past 40 years. Oligopolistic competition, for example, was mentioned in the book but not incorporated into the formal analysis (Rowthorn and Hymer, 1971; Knickerbocker, 1973; Graham, 1974; Flowers, 1976). Recent developments in game theory now make it possible to integrate this component into the theory (see, e.g., Petit, Sanna-Randaccio and Sestini, 2012).

Including these components produces a powerful new general theory that includes the original theory as a special case. It may be termed 'Internalization Plus'. This is the 2016 vintage of the theory. The basic framework has already been developed. The process of development is charted in this book (Chapters 2 and 5–7). Only a brief summary is presented here.

The basic unit of analysis is the global industry. This industry interacts with the rest of the economy. Firms in the industry recruit labour and raw materials at wages and prices set in the economy as a whole, and sell products to customers at prices the firms set themselves (because they enjoy a degree of monopoly power). Customers derive income (wages and profits) from both inside and outside the industry concerned.

The theory involves multiple firms serving multiple local markets. Entrepreneurs in the industry scan the technological and marketing possibilities in order to develop potential projects. Each entrepreneur establishes a knowledge-owning firm. All knowledge-owning firms in a given industry are rivals. They each exploit different knowledge, embodied in different products, but consumers regard their products as interchangeable. In economic terms their products are perfect substitutes, so that the firms are in direct competition with each other.

The production system is modelled along the lines explained in section 4 above. Each knowledge-owner can serve any national market from production plants in any country, subject to transport costs and other barriers to trade. Products cannot be arbitrated by independent resellers, however, so that price competition takes place separately in each national market. Competitive price determination is described in Box 1.1.

Each knowledge-owner decides whether or not to exploit their knowledge. If they innovate a product they incur fixed costs of R&D and fixed costs of headquarters. All other costs are directly proportional to the amount of product produced (i.e., there are constant returns to scale in production). If they decide not to innovate they incur no costs, earn no revenue and therefore make zero profit. Firms maximize profit, and so they do not innovate unless it is profitable.

BOX 1.1 DETERMINATION OF PRICE IN A MARKET WITH *N* COMPETING SUPPLIERS

Suppose that in each country individual consumers purchase products at a uniform price that is set by competition between firms. All firms that innovate are potential competitors in every market. Consumers are indifferent between the products of different firms; they choose only on price. They are sufficiently rational that they always purchase from the lowest-price supplier. Tastes, incomes and population differ between countries. Each market has its own demand schedule.

In each country the number of units demanded, q , is a declining linear function of price. Inverting this demand schedule, price, p , may be expressed as a declining function of demand:

$$p = a - bq$$

where a and b are fixed parameters; a measures the intensity of demand (i.e., 'willingness to pay'), whilst b measures the price-sensitivity of demand. The values of these parameters vary between countries.

Consider a given scenario. Unless there is a 'tie' between two or more competitors, only one firm will serve the market, namely the firm with the lowest unit cost of supply. This cost is achieved by optimizing its production location and internalization strategies. Let c^* be the lowest cost of supply. Let c^{**} be the second-lowest cost of supply, that is, the unit cost of the firm's closest rival in the local market. The maximum price that the least-cost supplier can charge is c^{**} , because if they charge more their rival can undercut them and take the whole market; c^{**} is the 'limit price' (Milgrom and Roberts, 1982).

The least-cost supplier may find it profitable to charge less than the limit price, however. In the absence of a limit price constraint the supplier will set the 'monopoly price':

$$p_m = (a + c^*)/2$$

A profit-maximizing least-cost supplier will charge the minimum of the limit price and the monopoly price:

$$p^* = \min[c^{**}, p_m]$$

and the corresponding quantity sold will be

$$q^* = (a - p^*)/2b.$$

The profit margin is:

$$m^* = p^* - c^*$$

and the quantity sold is

$$q^* = a - bp^*$$

The profit accruing from the market is

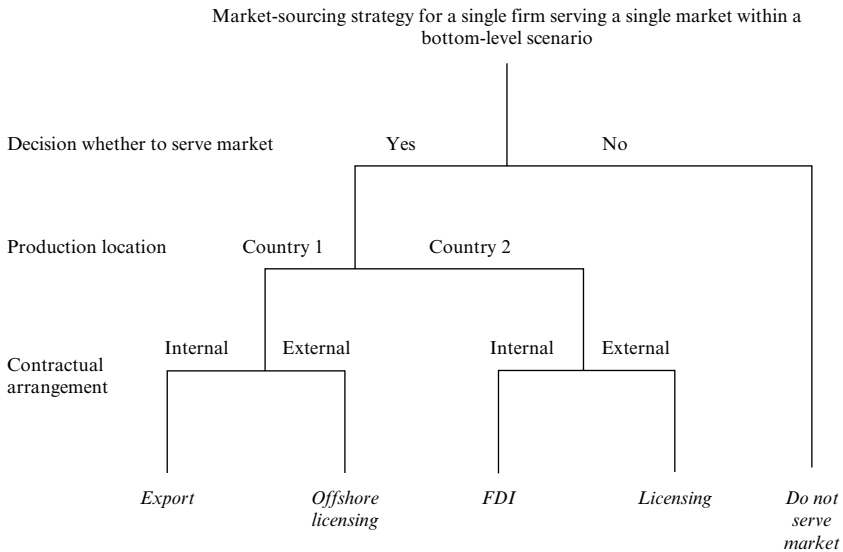
$$\pi^* = m^*q^*$$

They can choose where to locate their R&D and their headquarters; these are independent decisions and co-location of headquarters and R&D is just a special case.

Because of rivalry, each firm's innovation decision depends on the decisions of the other firms. Their location decisions for R&D and headquarters are interdependent too. Innovation decisions for different firms may be sequential or simultaneous. The outcome determines which firms innovate and which do not; this in turn determines the number of potential competitors in each local market. With sequential decisions the industry outcome is unique, but may vary according to the sequence of the moves. With simultaneous decisions there may be no equilibrium (which is unlikely) or multiple equilibria (which is quite likely) (Fudenberg and Tirole, 1991).

The resulting model has a simple logical structure that means that it is easy to solve (Casson, 2016). There are three main stages to the solution:

1. The first (bottom-level) stage involves the familiar market-sourcing decision. It is illustrated by the decision tree shown in Figure 1.4. This problem must be solved by every firm for every market, and each solution is independent of the others. Licensees and franchisees are headquartered in the host market, and subcontractors are headquartered in the country in which they produce.
2. The second (middle-level) stage is to determine the price in each market, which depends on which particular firms decide to innovate and where they locate their headquarters and R&D. The mechanism of price determination is explained in Box 1.1 above. This stage also determines which firms will supply which markets when a given set of firms decide to innovate.
3. The third (top-level) stage determines each firm's location of headquarters and R&D. It also determines whether they innovate or not. This stage is illustrated by the decision tree in Figure 1.5. Unlike bottom-level decisions, which are independent of other firms, these top-level decisions depend on what other firms decide to do. For each location strategy it is possible to calculate the firm's overall profit conditional on the decisions of other firms. These calculations populate a table in which each cell reports the profit of each firm on the basis of the individual strategies of all the other firms. Firms innovate if their profit is positive and otherwise they do not. Using this figure it is possible to determine the outcome under alternative sequences of moves, and under simultaneous moves. Imposing a particular schedule of moves solves the model.



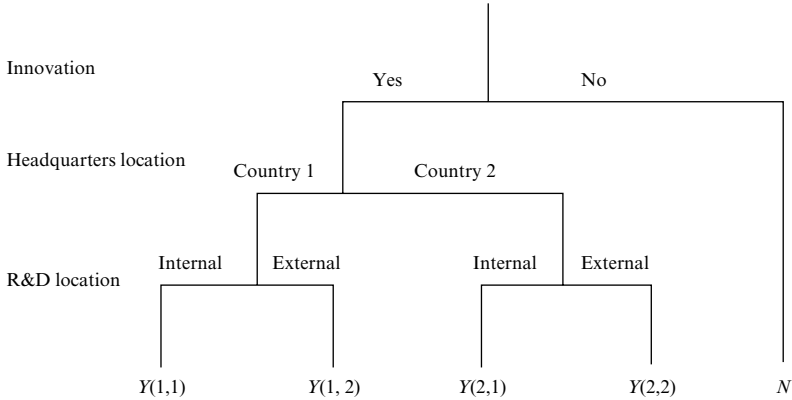
Note: The vertical lines illustrate various options open to the decision-maker. The horizontal lines indicate the choices between the alternative options that are available at each stage. The decision-maker takes the top decision first and works downwards so that they make the bottom decision last. However, the decisions are optimized by working from the bottom up; the decision-maker decides what would be their best response at the final stage to each decision that they could make at the middle stage and, in light of this, they decide what would be the best response at the middle stage to any decisions they might make at the first stage. They are then in a position to optimize the first-stage decision. Once they have made the first-stage decisions they know from their previous calculations what the other decisions should be.

Figure 1.4 Decision trees for bottom-level decisions

The solution of the model indicates which firms innovate, and which do not, where they are headquartered and where they carry out their R&D. But it does much more than this. It provides a complete profile of the global industry. It determines how many knowledge-owners innovate, and how many licensees, franchisees and subcontractors they use. It tells us where these other firms are headquartered, and whether they are multi-nationals or not. It tells us which knowledge-owners serve which market, and indicates whether they are ‘global’ or ‘regional’, and even whether they are multinational at all. This underlines the point that internalization is really a theory of the industry and not just a theory of the individual firm.

It is not only the number and nature of the firms that is endogenous but their boundaries too. Internalization theory is sometimes described as a

(a) Top-level decisions for a single firm (to interpret this diagram, see note to Figure 1.4)



(b) Interaction of top-level decisions for two firms (set of $5 \times 5 = 25$ scenarios under which bottom-level market-sourcing decisions can be played out)

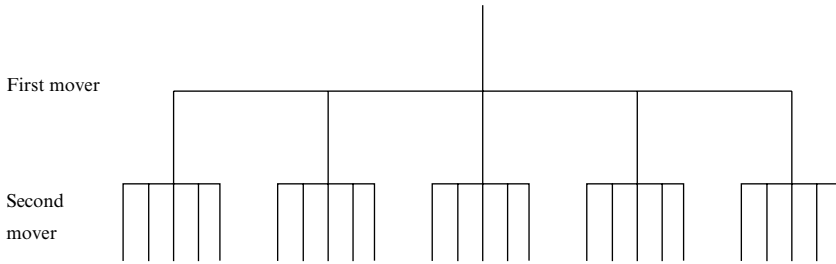


Figure 1.5 Decision trees for top-level decisions

theory of the boundaries of firms, but many applications of the theory are so trivial that the significance of this proposition is lost. The boundaries of firms are set where active linkages are externalized. The theory above specifies a full set of possible linkages, determines which linkages are active, and whether they are internal or external. Which linkages are active depends on which firms innovate and the location strategies they choose. Which of these linkages they internalize is determined by the internalization decision itself. The solution, therefore, makes it possible to ‘picture’ the boundaries of firms and to visualize the firms as clusters of interdependent internalized activities.

The solution also identifies the factors that govern each aspect of the industry outcome. These are represented by the exogenous parameters of the model (see Figure 1.6). In principle, each exogenous parameter can

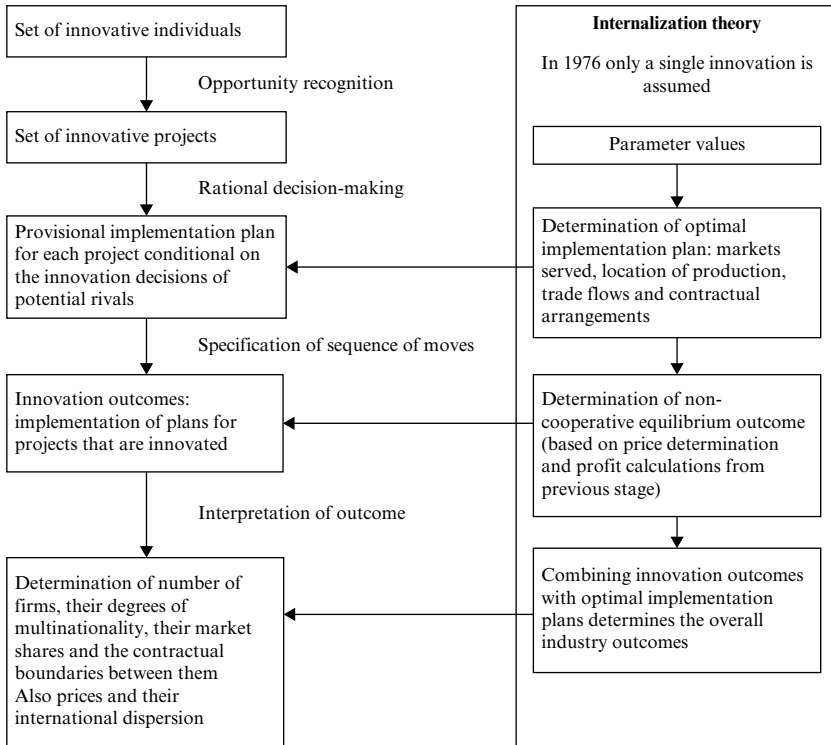


Figure 1.6 *Internalization Plus: Determination of outcomes conditional on parameter values*

influence every endogenous aspect of the outcome, but in practice certain parameters influence certain outcomes more than others.

The parameters reflect spatial characteristics of the global economy, the nature of the knowledge that is generated, the tastes of consumers and the skills of the workers at each location (see Figure 1.7). The impact of exogenous parameter changes on endogenous firm behaviour may be summarized as follows:

- The number of MNEs depends crucially on the innovative potential of the economy, as represented by the technological opportunities available and the costs of R&D. It also depends on the ‘costs of distance’; low costs of distance will encourage global monopoly, whereas high costs will encourage multiple ‘regional’ monopolies, as described in Rugman (2005).

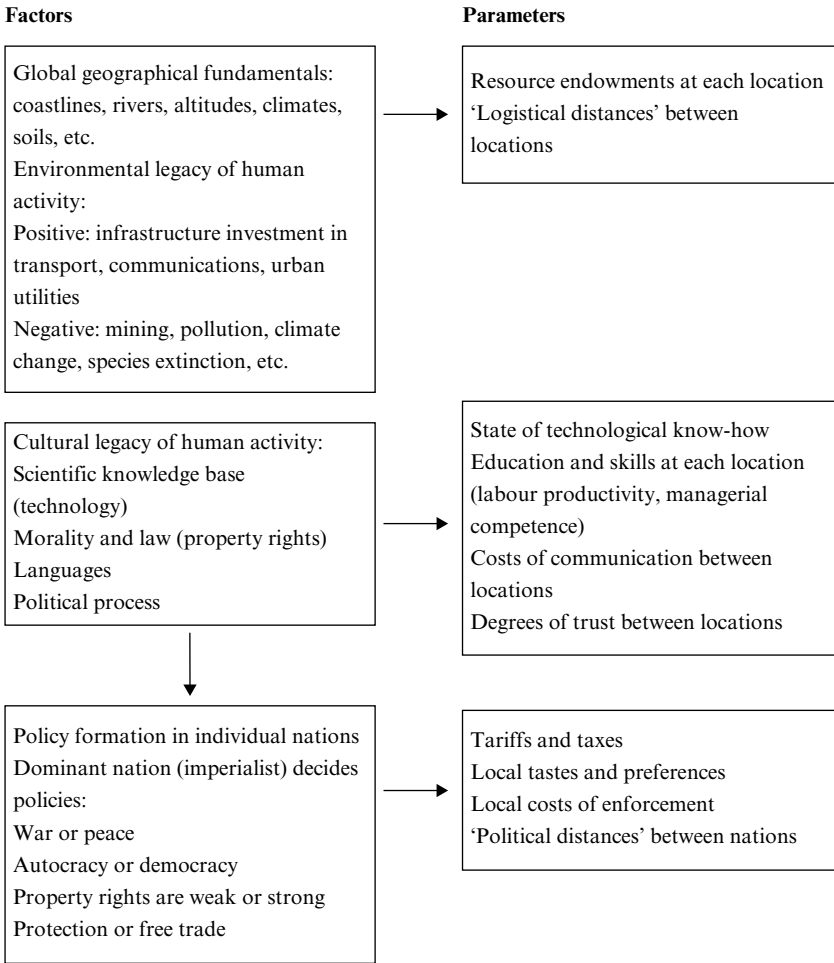


Figure 1.7 Determination of basic parameters

- The locations at which each product is produced depend on the nature of the product (which determines transport costs) and the variation between locations in terms of tastes, incomes and labour productivity.
- The boundaries of MNEs are determined by internalization decisions, which will, in principle, vary according to the characteristics of the product and, for any given product, according to the characteristics of individual locations. High costs of licensing will encourage the emergence of highly integrated MNEs that control a large number of subsidiaries, whereas low costs will encourage the

establishment of networks of independent licensees, as exemplified by 'flagship' firms (Rugman and D'Cruz, 2001). In the first case the global industry will comprise a small number of very large firms and a few smaller firms, and in the second case a small number of medium-size firms and a larger number of smaller firms. Overall, higher costs of licensing lead to fewer firms, higher average firm size, and higher 'firm concentration' of production.

As the parameters of the global economy change, so the number of firms will change and their boundaries will change as well. This will have implications for the average number of firms, the average size of firms, and the range of activities encompassed by each firm.

6 THE COSTS OF SURVIVAL: OVERSIMPLIFICATION

The transition from internalization theory 1976 to Internalization Plus has taken 40 years. Why has it taken so long? Much of the progress has occurred only in the last few years. My explanation is that many internalization scholars have been more concerned with the survival of the theory than with its development. Over the past 40 years economics has become an increasingly technical and mathematical subject, and has consequently been marginalized within many business schools. Not only do students not learn the subject, but many younger faculty have never had an opportunity to learn it either. They have learned strategic management instead. In a changing environment, such as modern academia, survival requires adaptation, and internalization theory has survived by repackaging itself as a theory of international business strategy (Dunning, 1993; Verbeke, 2013).

Adaptation often comes at a cost, however. In order to survive the theory has changed in three main ways:

- *The unit of analysis has changed from the global industry, to the firm.* As a result, the perspective of the theorist has narrowed from 'the big picture' to the boardroom view. The concept of an 'industry equilibrium' or 'industry recipe' has been lost. The role of internalization theory has been reduced to the analysis of the market-sourcing decision. The IB profession has lost influence with governments and international organizations as a result. Policy advice to governments requires the bigger picture; the narrow view is useful only for management consultancy.

- *References to fundamental economic principles have been stripped out.* The IB system is inherently complex: ownership and location interact; internalization trades off benefits and costs; and so on. That is why IB is such a good 'laboratory' for testing theory; it pushes theory to the limit and tests it to destruction, as happened with neoclassical theory. That is why internalization 1976 involved a synthesis of five strands of theory, as explained above. Fundamental economic principles were the 'glue' that held this intellectual synthesis together. As these principles have been purged from the theory, so the glue has dissolved away. The subtle ideas that linked key concepts have been lost. The presentation of the theory is increasingly reduced to a list of simple bullet points.
- *Little use is made of formal models.* Complex systems require models to analyse them. Formal models recognize complexity but seek to simplify it. While IB modelling was an accepted methodology in the 1970s, it is no longer so today. Without a model it is tempting to pretend that reality is simple even when it is not. This is positively dangerous. Presenting a complex system as simple merely distorts the picture. It suggests that simple remedies may work when they will actually make the situation worse.

In the 1980s, IB theory was exported to other disciplines, including economics, where it influenced mainstream trade theory (Markusen, 2002). Today IB exports relatively few ideas to other disciplines and, like many other fields of study, is in danger of becoming a silo (Rugman, 2014).

My conclusion is that the theory has become oversimplified. There is a confusion between simplicity and clarity. It is often said that a good theory simplifies reality by abstracting from unnecessary detail. It is possible to oversimplify, however, as any encounter with politicians and the media will show. Clarity is best achieved by constructing models that respect the complexity of a system. The global economic system is complex: models clarify it; oversimplification distorts it.

Oversimplification retards theoretical progress. Once the intellectual glue has been lost, and formal modelling abandoned, the foundations become too weak to support further theory-building. As one oversimplification is proved inadequate, another is introduced. As simplifications proliferate, confusion develops. Instead of one sophisticated theory that works, there are multiple theories that fail. Theory, rather than reality, becomes the subject of investigation, as researchers clarify the relationships between competing theories.

Simple theories often involve a single concept, which is hailed as an intellectual innovation, and treated as the private property of the person

who gave it a name. Hymer, for example, introduced ‘monopolistic advantage’ in 1960. Then ‘ownership advantage’ appeared in 1977, ‘competitive advantage’ in 1980, and ‘firm-specific advantage’ around 1990. There is also ‘absolute advantage’ and ‘strategic advantage’. Many papers have been written about the relations between these concepts, but what have we learned from the process? One answer is ‘almost nothing’, because they are all just different words for the same thing. If you ask ‘Advantage relative to what?’ the answer is always ‘Advantage relative to a potential competitor’. But the potential competitor itself is almost never examined. If you ask what the advantage consists of, the answer is usually ‘monopolistic control of a knowledge-based asset’ – in other words, Hymer seems to have got it ‘right first time’. Students learning IB waste a lot of time trying to work out how these concepts relate to each other. They would be much better off using that time to learn how economic models work instead.

7 CONCLUSION

The contribution of *The Future* was not to create a special theory of the MNE but rather to embed the MNE within a general theory that would explain other things as well. Internalization is best understood as a general organizing principle within a global economic system. It transcends the field of IB; IB is just a special (but important) case. The future of internalization theory lies in the generality of the internalization model rather than the particularity of the historical anomaly it addressed.

The firm is an important specialized part of the global economic system in which individuals associate together for productive purposes (see the Appendix). But a firm-centred view cannot do full justice to the interdependencies within the global system. Firms are created to exploit the plans of entrepreneurs. These plans involve systems of production, which may involve links between production, marketing and R&D. The implementation of parts of these plans may be delegated to independent firms such as licensees, franchisees and subcontractors. A single production system may therefore involve several different firms. Internalization theory rightly focuses on the production system rather than individual firm. Focusing on the production system explains the rationale of the firm – namely the coordination of the system – while focusing exclusively on the firm explains almost nothing at all.

Different entrepreneurs may have different plans for meeting the same customer needs. These plans involve different technologies and product designs that have different types of production system. Competition between these plans determines which production systems prevail, and therefore what types of independent firm emerge. Competition is a

two-stage process: rivalry between entrepreneurs determines which firms innovate and which do not, while competition within individual markets determines which innovator supplies which market, and hence where production for each market is located. Without a full appreciation of competition it is impossible to understand why certain firms survive and prosper and others do not. Competitive survival implies that a firm is marginally better than its potential competitors, while prosperity based on monopoly profits implies that the firm is significantly better than its competitors.

Internalization Plus is a significant advance, but it does not complete the research agenda; it is just a milepost along the way. There are many opportunities to refine and develop the theory further. Other issues that can be addressed are:

- the policy framework (including tariffs and taxes);
- economies of scale in production;
- management capacity constraints;
- economies of agglomeration, for example, industrial districts;
- international joint ventures and other alliances.

The policy environment may be analysed as the outcome of a political game involving governments and firms (see Figure 1.8). Economies of scale can be analysed by introducing additional location strategies in which all production is concentrated at a few locations. Management capacity constraints can be invoked to set a limit on the sizes of firms; they generate a trade-off for each firm between the number of countries that are served and the degree of internalization that is used. Economies of agglomeration can be analysed as a game played out between firms at particular locations where the players seek either to co-locate or to keep their distance from each other. The study of international joint ventures fits very naturally into

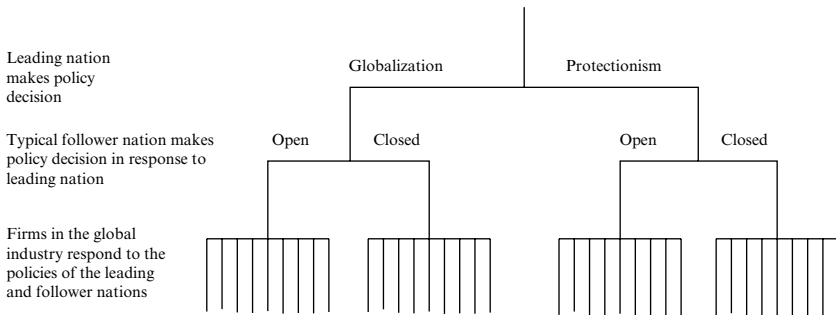


Figure 1.8 Policy context

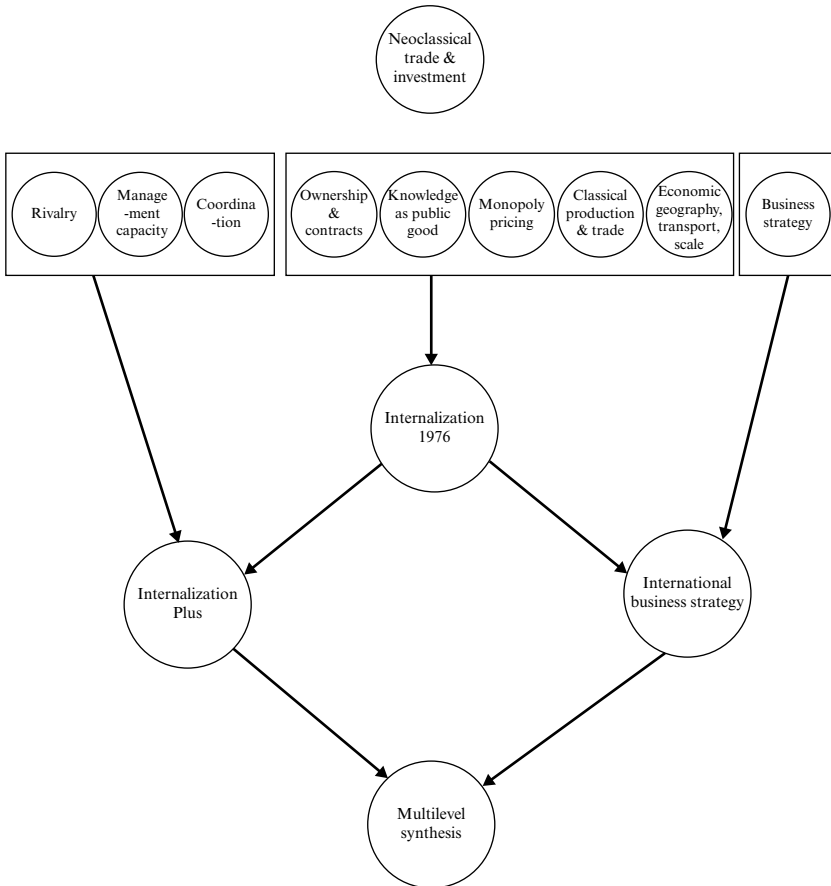


Figure 1.9 Future developments in theory

the framework of Internalization Plus because it allows for cooperation as well as competition between innovating firms.

All of these issues have been addressed before from a firm-centred view, but within a global industry view they appear in a different light. Internalization Plus, for example, makes explicit an important distinction between R&D joint ventures and production joint ventures, and shows that they have very different implications for innovation, market entry, pricing and profitability.

The further development of Internalization Plus will require resources. Over the past 40 years considerable effort has gone into developing the firm-centred view of IB. The firm-centred agenda has the potential to

inform management decision-making, but it cannot provide a ‘big picture’ view of the global economy. While the two agendas are rivals for resources, they are not rivals in intellectual terms because their levels of analysis differ. One is focused on the firm and the other on the industry and the global economy. In principle, they complement each other (see Figure 1.9). The development of Internalization Plus is not a threat to firm-level research, but it does require some redirection of intellectual effort.

REFERENCES

- Arrow, K.J. (1962). Economic welfare and the allocation of resources for invention. In J. Lerner and S. Stern (eds), *The Rate and Direction of Inventive Activity*. Princeton, NJ: NBER, pp. 609–26.
- Buckley, P.J. and M.C. Casson (1976). *The Future of the Multinational Enterprise*. London: Macmillan [40th Anniversary Edition, 2016].
- Buckley, P.J. and M.C. Casson (1985). *Economic Theory of the Multinational Enterprise*. London: Macmillan.
- Buckley, P.J. and M.C. Casson (2009). The internalization theory of the multinational enterprise: A review of the progress of a research agenda after 30 years. *Journal of International Business Studies*, **40**: 1563–80.
- Casson, M.C. (1997). *Information and Organization*. Oxford: Oxford University Press.
- Casson, M.C. (2016). *The Theory of International Business: Economic Models and Methods*. Cham, Switzerland: Palgrave Macmillan.
- Chamberlin, E.H. (1933). *The Theory of Monopolistic Competition*. Cambridge, MA: Harvard University Press.
- Coase, R.H. (1937). The nature of the firm. *Economica* (New Series), **4**: 387–405.
- Dunning, J.H. (1958). *American Investment in British Manufacturing Industry*. London: Allen and Unwin.
- Dunning, J.H. (1977). Trade, location of economic activity and multinational enterprise: A search for an eclectic approach. In B. Ohlin, P.O. Hesselborn and P.M. Wijkman (eds), *The International Allocation of Economic Activity*. London: Macmillan, pp. 395–418.
- Dunning, J.H. (1993). Internationalizing Porter’s Diamond. *Management International Review*, **33**(2): 7–15.
- Dunning, J.H. and S.M. Lundan (2008). *Multinational Enterprises and the Global Economy*. 2nd edition. Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.
- Flowers, E.B. (1976). Oligopolistic reaction in European and Canadian direct investment in the United States. *Journal of International Business Studies*, **7**(2): 43–55.
- Fudenberg, D. and J. Tirole (1991). *Game Theory*. Cambridge, MA: MIT Press.
- Graham, E.M. (1978). Transnational investment by multinational firms: A rivalistic phenomenon. *Journal of Post-Keynesian Economics*, **1**: 82–99.
- Hennart, J.F. (1982). *A Theory of Multinational Enterprise*. Ann Arbor, MI: University of Michigan Press.

- Hymer, S.H. (1976). *The National Operations of International Firms*. Cambridge, MA: MIT Press.
- Iammarino, S. and P. McCann (2013). *Multinationals and Economic Geography*. Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.
- Kemp, M.C. (1970). *The Pure Theory of International Trade and Investment*. Englewood Cliffs, NJ: Prentice-Hall.
- Knickerbocker, F.T. (1973). *Oligopolistic Reaction and Multinational Enterprise*. Cambridge, MA: Harvard University Press.
- MacDougall, G.D.A. (1960). The benefits and costs of private investment from abroad: A theoretical approach. *Economic Record*, **36**: 13–35.
- Markusen, J.R. (2002). *Multinational Firms and the Theory of International Trade*. Cambridge, MA: MIT Press.
- McManus, J.C. (1973). The theory of the international firm. In G. Paquet (ed.), *The Multinational Firm and the Nation State*. Toronto: Collier Macmillan, pp. 66–93, reprinted in M. Casson (ed.) (1990). *Multinational Corporations*. Aldershot, UK and Brookfield, VT, USA: Edward Elgar Publishing, pp. 32–59.
- McManus, J.C. (1975). The costs of alternative economic organizations. *Canadian Journal of Economics*, **8**: 334–50.
- Milgrom, P. and J. Roberts (1982). Limit pricing and entry under incomplete information: An equilibrium analysis. *Econometrica*, **50**(2): 443–59.
- Petit, M.L., F. Sanna-Randaccio and R. Sestini (2012). R&D and foreign direct investment with asymmetric spillovers. *Economics of Innovation and New Technology*, **21**(2): 125–50.
- Ricardo, D. (1817 [1951]). *Principles of Political Economy* (ed. P. Sraffa). Cambridge, UK: Cambridge University Press.
- Richardson, G.B. (1964). *Economic Theory*. London: Hutchison (reprinted by Routledge, 2003).
- Richardson, G.B. (1998). *The Economics of Imperfect Knowledge: Collected Papers*. Cheltenham, UK and Lyme, NH, USA: Edward Elgar Publishing.
- Robbins, L. (1932). *An Essay on the Nature and Significance of Economic Science*. London: Macmillan.
- Robinson, J.V. (1933). *The Economics of Imperfect Competition*. London: Macmillan.
- Rowthorn, R. and S.H. Hymer (1971). *International Big Business*. Cambridge, UK: Cambridge University Press.
- Rugman, A.M. (1981). *Inside the Multinationals: The Economics of Internal Markets*. London: Croom Helm [25th Anniversary Edition, 2007].
- Rugman, A.M. (2005). *The Regional Multinationals*. Cambridge, UK: Cambridge University Press.
- Rugman, A.M. (2014). Letter from the editor. *Multinational Business Review*, **22**(3): 200–204.
- Rugman, A.M. and J.R. D’Cruz (2000). *Multinationals as Flagship Firms: Regional Business Networks*. Oxford: Oxford University Press.
- Smith, A. (1776 [1976]). *Inquiry into the Nature and Causes of the Wealth of Nations* (Glasgow Edition). Oxford: Oxford University Press.
- Spender, J.C. (1989). *Industry Recipes*. Oxford: Blackwell.
- UNCTAD (2011). *World Investment Report 2011: Non-equity Modes of International Production and Development*. Geneva: United Nations.
- Vaitsos, C.V. (1974). *Inter-country Income Distribution and Transnational Enterprises*. Oxford: Oxford University Press.

- Verbeke, A. (2013). *International Business Strategy*. Cambridge, UK: Cambridge University Press.
- Vernon, R. (1971). *Sovereignty at Bay*. New York: Basic Books.
- Weber, A. (1929). *Theory of the Location of Industries* (ed. C.J. Friedrich). Chicago, IL: University of Chicago Press.
- Zaheer, S. (1995). Overcoming the liability of foreignness. *Academy of Management Journal*, **38**(2): 341–63.

APPENDIX: THE FUNDAMENTAL FRAMEWORK OF INTERNALIZATION THEORY

This Appendix sets out the basic economic principles that were implicitly assumed in *The Future*. The content of the Appendix is fairly conventional but the presentation is unusually philosophical, and is written specifically for the benefit of IB scholars. Key sources include Robbins (1932), Richardson (1964, 1998) and Casson (1997).

Free enterprise: Economists generally focus on some variant of a free enterprise market economy. In this scenario, individual action is entirely voluntary. Certain individuals step forward and offer to produce and sell particular products. Each producer specializes on a single product, or a narrow range of products. They act as market-makers, setting up production and selling product to their customers. Harmony is achieved in a general equilibrium where demand for each product equals its supply, everyone spends exactly what they earn, and producers break even or make a profit.

Methodological individualism: It is individuals that take decisions, not institutions like governments or firms. In a free society, individuals are free to associate, and to establish institutions such as firms. Institutions have an identity but they do not have ‘a will of their own’, despite the fact that their members may have common interests and share certain values and beliefs. Decisions made on behalf of institutions are taken by individual members who occupy specific roles that authorize them to act on behalf of others. Some decisions may be taken by committees, but even then a committee simply aggregates the opinions of its individual members using principles such as a chairperson’s prerogative or majority vote.

Coordination: Institutions have an important role in reconciling conflicts and coordinating economic activity. Coordination harmonizes the plans of individuals. They make it possible for individuals to agree on what they do before they act. Individual actions become more predictable and so expectations of other people’s actions are more likely to be correct. This

eliminates confusion and reduces waste. In economic terms, coordination promotes 'Pareto efficiency'.

Property rights: Property rights play an important role in coordination. They determine who has what rights over each unit of resource. This avoids conflict over how each resource is to be used.

Ownership: Ownership of a resource confers the right to control its use, to consume it or to sell it. It also confers the right to exclude other people from it. Ownership is a general right that comprises a bundle of specific rights. Ownership is rarely absolute: owners exercise control subject to powers reserved to other parties (e.g., government).

Specific right: A specific right is the right to use a specific resource for a specific purpose at a specific time. Specific rights are too numerous to itemize. The exercise of ownership is always constrained by any specific rights that the owner may have granted away; hence ownership is often referred to as a 'residual right'. Most property rights are *alienable*, but some are not. A person cannot sell themselves into slavery, but they can sell their labour to a firm. They cannot alienate self-ownership, therefore, but they can alienate specific rights. Alienable rights are normally *tradeable* (they can be bought and sold and not just given away).

Contracts: Contracts can be entered into for the sale of property rights. Contracts for the exchange of ownership rights do not necessarily involve the physical exchange of resources at the time the contract is made.

Contracts as a coordinating mechanism: In a free enterprise economy contracts play an important role in the coordination process. Exchange of ownership places resources under the control of the most appropriate parties, and these parties then decide how their resources are to be used.

Economic opportunities: An economic opportunity exists when scarce resources are being wasted. It is an opportunity to improve coordination. Recognition of an opportunity takes place inside a person's head. A typical opportunity involves responding to unsatisfied customer needs by developing a product or service that is superior to existing offerings.

Projects: The exploitation of an opportunity involves a project. A typical project involves an initial input of resources followed (if successful) by the output of a product. Projects involve commitment, as set-up costs may be sunk (i.e., non-recoverable). Projects take time and may be risky. Radical projects may involve product or process innovation, or the development of new markets or new sources of supply. People who specialize in recognizing and exploiting opportunities are often described as

entrepreneurs; people who undertake radical projects are 'Schumpeterian entrepreneurs'.

The nature of the firm: A firm is an institution that promotes coordination by facilitating the exploitation of economic opportunities. The firm manages projects relating to production and trade. In a centrally planned socialist economy all organization of production and trade is carried out by state-owned enterprises. In a free enterprise economy, by contrast, private individuals can establish firms and become shareholders in them. They are also free to work for firms for a wage or salary, and to purchase products from them.

Opportunity recognition: Opportunity recognition is carried out by private entrepreneurs rather than politicians or state employees. The firm acts as a *nexus of contracts* between shareholders, workers, suppliers and customers and, as an institution, can outlive all of them. Opportunities may be exploited by new or existing firms. New firms are often established by entrepreneurs, while existing firms may be managed by 'intrapreneurs', such as salaried chief executives.

Market: A market is a mechanism for allocating goods and services that have alternative sources and/or alternative uses. It matches up suppliers with users through negotiations over price. Firms are buyers of labour, raw materials and finance and sellers of product. Where firms sell to households they often quote non-negotiable prices, and households search across these prices to get the best deal. Where supplies vary in quality or reputation, prices may vary according to quality premia. Firms also quote wages to the workers they recruit. Firms therefore play an intermediating role, quoting prices to both their buyers and their sellers, and appropriating a profit from the margin between the two sets of prices.

Market efficiency: A market works efficiently when buyers have good knowledge of suppliers (and vice versa), there are many independent buyers and sellers, and buyers (sellers) perceive different sellers (buyers) as very close substitutes for each other. They work less well when buyers and sellers have little knowledge of each other, there are few of them, and no buyer or seller is a close substitute for any of the others.

Internalization of a market: There is an important distinction between final product markets, where firms sell to households, factor markets (e.g., the labour market), where firms buy from households, and intermediate product markets, where firms buy from each other. Firms cannot buy out households, but they can buy out other firms. This means that firms can internalize markets in intermediate products by bringing related activities

under common ownership and control. They cannot internalize final product or factor markets in the same way.

Plans: A project that exploits an opportunity requires a plan. A plan may be a written document or just an idea inside the entrepreneur's head. Some plans are incremental, small and simple, while others are radical, large and complex. In the latter case, the entrepreneur may employ managers to formulate and implement plans. Complex plans typically have a modular structure where interdependent activities are connected by flows of intermediate products. Such plans are characteristic of large MNEs; in the case of MNEs complexity is further increased by the scale and diversity of the global market.

Decentralization: Planning requires information. The entrepreneur may not have all the information required. Collecting all the information may be difficult. Managers can be recruited to supply it. With modular activities that are spatially dispersed, it may be better to devolve local planning to local managers. This avoids the cost of transmitting information to and from headquarters, and reduces the information burden on headquarters staff. Headquarters coordinates the linkages between the modular units but leaves internal coordination within each module to local managers. This can work well so long as 'internal' and 'external' linkages can be clearly separated.

Appropriation of profit: Profit is generated by the appropriation of surplus from a project. The firm provides a legal framework for appropriation. Shareholders invest capital and workers accept employment. The firm buys raw materials and sells finished products to its customers. Profit is equal to the excess of sales revenues over input costs. When wages are sticky, and outputs lag inputs, profits fluctuate more than wage income and so shareholders provide a degree of insurance to workers.

Cheating: The aim of a plan is to harmonize the actions of individuals, but individuals may benefit personally from deviating from the plan, even if they damage the interests of others. Shareholders appropriate profit by writing contracts with workers, managers, suppliers and customers in such a way that they appropriate as much of the surplus for themselves as possible. But their partners can cheat on these contracts. If contractual compliance is difficult to monitor then cheating may be covert, while if contracts themselves are difficult to enforce in law, it may become overt. (1) *Covert cheating:* When monitoring is difficult there is an incentive for employees to disobey orders, for example, by shirking; for suppliers to economize by reducing quality; and for customers to make false warranty claims. The motive for covert cheating is usually to reduce the costs of the cheat. (2) *Overt cheating*

is usually based on threats. When enforcement is weak workers may threaten strikes unless wages are raised, major suppliers may threaten to withhold supplies unless prices are raised, while major customers may boycott the product unless its price is reduced. The motive for overt cheating is usually to obtain a better price. In either case, cheating undermines shareholders' interests: the surplus available is reduced through inefficiency and a proportion of what remains accrues to the other parties instead.

Teamwork: If the shareholders cannot monitor their employees then they may consider making the employees shareholders themselves. A self-employed worker would sell their product to the firm and be paid only for what they supply. A shirker would receive less than a hard worker and shirking would be deterred as a result. The problem is that in a team any worker can disrupt the work of others and so there is no direct link between effort and output; effort is a necessary condition for output, but not sufficient, as it can be nullified by others who do not 'pull their weight'. Team members therefore normally work as employees of the same firm.

Interplant versus intraplant coordination: Teamwork is more common within a plant than between plants. In a multiplant production system distributed ownership is much more viable. In general, different groups of plants can be owned by different firms. The boundaries between the firms would lie along the linkages between individual plants owned by different firms.

Planning versus prices: In a modular multiplant system involving different firms, the linkages between the plants can be coordinated by contracts between the firms. Decentralized ownership of plants supports the decentralization of planning discussed above. Where intraplant coordination requires local information, the owner of a local plant has a strong incentive to collect and utilize such information efficiently because it directly increases their own profits. This encourages decentralization of ownership. But decentralization has disadvantages too. In particular, the enforcement of interfirm contracts may be weak. Covert cheating, for example, makes the quality of supplies difficult to assure when inspection is superficial and access to the suppliers' premises is denied. Overt cheating can also occur; for example, major investments may be compromised if suppliers decline to invest themselves, but instead exploit the situation to negotiate more favourable prices for their output. The decentralization of ownership therefore involves a trade-off. It can provide better local information, but also lower product quality and greater disruption too.

Long-term contracts: These allow an investor to buy supplies forward and to sell output forward, thereby assuring their profit at the time that their invest-

ment is made. To provide flexibility in a volatile environment, long-term contracts may be made contingent on volatile factors that impinge on prices. Such contracts are particularly useful in coordinating interdependent investments made by independent firms. However, long-term contracts become very complex unless the contingencies are specified in terms of simple rules or procedures. Furthermore, overt cheating can still be a problem.

Competition: The threat to disrupt supplies or demands in order to get a better price only has credibility if the owner of the plant has no alternative sources of supply or demand. Competition in intermediate product markets provides an owner with the option of switching customers or suppliers. It thereby eliminates the dependence of any individual plant on any single supplier (upstream monopoly power) or any single buyer (downstream monopsony power). Competition has two main requirements: (1) economies of scale in the supply of inputs and the use of outputs must be limited, for otherwise supply or demand will be concentrated on a single plant, which will normally be owned by a single firm; (2) the costs of switching are low.

Trust: Appropriation of profit is often identified as selfish. Selfishness is sometimes described as a private vice that can be a public virtue, but the virtuous outcomes only occur through competition under the rule of law. Cheating is selfishness that violates the rule of law and is not virtuous at all. The social engineering of trust is an economical way of controlling cheating. Instead of using the law to punish cheats, morality can be used to make cheats punish themselves, either privately through feelings of guilt or publicly through shame. Moral obligations may, in principle, be universal, but in practice they are strongest within social groups. In international business, they may be stronger within groups that share a language, culture or religion. Cheating on international contracts may be lower within such groups, and this may encourage trade to be concentrated inside groups rather than between them.

Determination of the boundaries of the firm: The principles presented above identify six key factors that favour decentralized ownership:

- *Independence of local operations:* Intra-plant coordination can be effected independently of operations at other plants.
- *Importance of local information:* Information on local conditions needs to be acted on quickly; it is context-dependent and therefore difficult to construe at a remote headquarters.
- *Product quality is easy to monitor and assess,* so that covert cheating is easy to detect.

- *Costs of switching are low*, so that competition in intermediate product markets discourages overt cheating.
- *A strong rule of law* discourages cheating in general, and overt cheating in particular.
- *Strong socially engineered trust* discourages cheating and provides an effective solution to weaknesses in the rule of law.

If a linkage in a production system satisfies these conditions then it is easy to coordinate that linkage through intermediate product trade between two independent firms. If some or all of these conditions are not satisfied, however, then it may be more efficient to internalize the intermediate product market instead. The linkage is then coordinated, not by arm's length trade, but through internal management instead.

It is advantageous for the owner of a firm to shift its boundaries to points where the conditions for arm's length trade apply. In other words, certain parts of its production system are hived off to independent firms using arm's length contracts, while other activities are retained within the ownership of the firm. The internal activities form a core of activities linked to each other internally and to other activities externally.