The economic theory of the firm as a foundation for international business theory

Article

Accepted Version

The economic theory of the firm as a foundation for international business theory

Mark Casson

Keywords: MULTINATIONAL, INTERNALISATION, FIRM, HETEROGENEITY, COMPLEXITY, ORGANISATION, COORDINATION, SIZE, FAILURE

Address for correspondence:

Mark Casson
Department of Economics
University of Reading
PO Box 218
Reading RG6 6AA, UK

Tel: (44) (0) 10118 378 8227

E-mail: m.c.casson@reading.ac.uk

Second draft: 26 May 2014

Total word count 11,437

Abstract

The economic theory of the firm is central to the theory of the multinational enterprise. Recent literature on multinationals, however, makes only limited reference to the economic theory of the firm. Multinationals play an important role in coordinating the international division of labour through internal markets. The paper reviews the economic principles that underlie this view. Optimal internalisation equates marginal benefits and costs. The benefits of internalisation stem mainly from the difficulties of licensing proprietary knowledge, reflecting the view that MNEs possess an 'ownership' or 'firm-specific' advantage. The costs of internalisation, it is argued, reflect managerial capability, and in particular the capability to manage a large firm. The paper argues that management capability is a complement to ownership advantage. Ownership advantage determines the potential of the firm, and management capability governs the fulfilment of this potential through overcoming barriers to growth. The analysis is applied to a variety of issues, including out-sourcing, geographical dispersion of production, and regional specialisation in marketing.

1. Introduction

The importance of economic theory

The economic theory of the firm is central to the theory of the multinational enterprise (MNE). The switch from the neoclassical theory of the firm, based on the production function, to an institutional theory of the firm, based on market imperfections, was a crucial step in the development of internalisation theory. It was only when economists appreciated the importance of the distinction between the plant (the unit of production) and the firm (the unit of ownership and control) that the economics of multi-plant firms such as MNEs could be fully understood (Buckley and Casson, 1976; Rugman, 1981; Hennart, 1982).

Recent literature on the MNE, however, makes only limited reference to the economic theory of the firm. International business (IB) research increasingly focuses on the environment of the firm, and in particular on institutions in home and host countries (Peng, Wang and Jiang, 2008). Where the firm is discussed, it is often the resource-based theory of the firm, based on management and strategy, rather than the economic theory of the firm, that is used (Cantwell, 2014).

Internalisation theory was developed in order to explain why foreign direct investment (FDI) was concentrated in knowledge-intensive industries. It showed that the role of MNEs was to coordinate the transfer technology (and intellectual property in general) by bringing the source of technology (R&D) and the use of technology (production and marketing) under common control. This proved particularly useful in explaining market-seeking investment. In addition, MNEs can internalise flows of raw material and components, explaining resource-seeking and efficiency-seeking investments as well (Dunning and Lundan, 2008).

The principal application of internalisation theory has been to 'mode of entry' decisions, involving a choice between exporting, licensing and FDI. It was subsequently extended to strategic alliances. When it comes to other questions, however, IB literature generally turns to other theories with less economic content. To analyse which firms serve particular national markets, models of 'internationalisation' based on sociological concepts are often used (Johanson and Vahlne, 1977). To analyse the growth of the firm, the resource-based theory is usually employed (Pitelis, 2007).

This paper argues that such theoretical pluralism is unnecessary. Indeed, it can be harmful, because it prevents the development of an integrated body of IB theory. Rigorous theory must be based on common fundamental principles, and one obvious source of such principles is the economic theory of the firm (Dietrich and Krafft, 2012). This paper reviews these principles, focussing selectively on the principles that seem most relevant to IB theory. It shows that these principles can be used to integrate the analysis of mode of entry, internationalisation and growth. They can also address other important issues such as the survival of the firm.

Structure of the paper

Sections 2 and 3 set out the context of the subsequent discussion. The focus is on the problems of coordinating the division of labour in the global economy. An inter-plant

division of labour typically involves two types of intermediate product flow: tangible materials and intangible knowledge. Coordination is effected by institutions, and three types of institution are examined: government, firm and market. The discussion is grounded in the economic research of the 1930s into the relative merits of different institutional arrangements.

According to Coase (1937), intermediate product markets are internalised up to the margin where the benefit equals the cost. The benefits internalisation are extensively discussed in the IB literature, but the costs are not. Furthermore, discussion of benefits is sometimes misleading. The benefits are critically reviewed in sections 4 – 6. The aim is to derive new insights from original sources, and also to 'put the record straight' on certain points. Section 7 discusses the costs of internalisation. It is argued that the costs of internalisation increase significantly with size of firm. Many practical issues in the management of large organisations are connected with the costs of internalisation.

Section 8 analyses the trade-off between the costs and benefits of internalisation. The benefits of internalisation stem mainly from the difficulties of licensing proprietary knowledge, reflecting the view that MNEs possess an 'ownership' or 'firm-specific' advantage. The costs of internalisation, it is argued, are a reflection of managerial capability, and in particular the capability to manage a large firm. Managerial capability is an important resource, but it is not a firm-specific advantage. It is available to any firm that recruits managers with appropriate cultural and educational backgrounds, and life experiences (including experience of working in a large organisation). Without an ownership advantage managerial capability is of limited value, because there may be little to manage, but equally, ownership advantage without managerial capability will lead to under-performance. It is therefore the combination of the two that is characteristic of a successful large firm. A formal model is presented, with details in the appendix. This model shows how the internationalisation of an MNE is constrained not only by the 'tyranny of distance' and the severity of international competition, but by internal limits to size connected with the costs of internalisation. The analysis provides insight into why large firms may fail unexpectedly, and why firms in trouble 'run for home'. The results also feed into mainstream IB literature. The provide an analytical rationale for regional MNEs, grounded in diseconomies of size, and a possible explanation of why large firms resort to more extensive out-sourcing than small ones. The conclusions are summarised in section 9.

2. The spatial division of labour in a global economy

Division of labour: a typology

The analysis that follows focuses on the global economy and the role of MNEs within it. Within the global system there is a division of labour (Buckley, 2009). There is a *functional* division of labour between different types of facilities, e.g. production plants, distribution centres, retail outlets and R&D laboratories. Within a given function in a given industry, individual activities may be sub-divided; the overall process is modularised so that different

activities can be carried out using different resources; where the process is sequential, modularisation leads to a *vertical* division of labour.

A *spatial* division of labour allows operations that make intensive use of certain types of resource to concentrate in areas where that resource is abundant. It also allows production to be concentrated in a small number of locations in order to exploit economics of scale. Where knowledge is concerned, there is a tension between firms that wish to co-locate in order to share knowledge and those that wish to isolate themselves in order to protect it. With different countries involved the division of labour becomes *international*. With internalisation the coordination of an international division of labour involves an MNE.

An *industrial* division of labour involves specialisation between plants and laboratories producing different types of product. It is often described as *horizontal*, on the grounds that different industries operate in parallel, and each has its own distinctive sequence of operations. This is inaccurate however. Not all industries produce final products: the products of some industries are inputs to others; e.g. the output of the coal industry is an input to the steel industry and the output of the steel industry is an input into the engineering industry. In principle the same firm could be involved in all three industries, although this would be most unusual, for the reasons explained below. The division of labour is rarely purely horizontal, or purely vertical, for that matter; in practice, horizontal and vertical are inter-twined.

Intermediate products

In a functional division of labour, different activities are connected by flows of intermediate product. It is crucial to distinguish between the intra-plant division of labour, which occurs at a single location, and the inter-plant division of labour, which occurs between plants at different locations. It is only when different functions are carried out in different countries that the functional division of labour leads to MNEs. Within the theory of the firm many discussions of the division of labour focus on the intra-plant division of labour, e.g. teamwork, but in the study of MNEs it is the inter-plant division of labour that is crucial (Achian and Demsetz, 1972).

In a global industry, production plants feed product into foreign distribution centres, and distribution centres may pass it on to foreign retail outlets. This is essentially a vertical structure; however, if each production plant serves several distribution centres and each distribution centre serves several retail outlets then the system as a whole has a pyramid shape.

Knowledge flows have a different pattern. Knowledge flows directly into production, and possibly into retailing too; retailers need to understand how customers can use the product and they may benefit from knowledge derived from R&D. Because knowledge can be shared it naturally diffuses from a single location at which it is developed to multiple locations where it is exploited.

3. Alternative methods of coordination

Markets versus governments

The division of labour needs to be coordinated. Economic theory identifies three main methods of coordination: by government, firms or markets.

In the 1930s the ideological debate between capitalism and socialism was at its height (Lange and Taylor, 1938). In economic theory the focus was on government (representing socialism) *versus* markets (representing capitalism). The focus was on final product markets. Under socialism consumer goods would be supplied through rationing, or at regulated prices, whereas under capitalism they would be bought and sold at market prices, impersonally regulated by supply and demand.

Government coordination was identified with state planning. The argument in favour was quite straightforward: by centralising all information with an elite growth of planners the division of labour across the economy could be optimised using mathematical methods. With the benefit of hindsight it can be seen that it is difficult to implement central planning within a global economy composed of different nation states because international trade cannot be planned except through supra-national organisation. This creates a bias towards autarky and self-sufficiency, which leads to international competition for strategic resources such as industrial minerals, which in turn leads to war. Furthermore, within the nation state, a planning bureaucracy develops a monopoly of economic power and begins to appropriate rewards for itself. This exploitation dis-incentivises ordinary workers, who become uncooperative, so that the plans are not fulfilled. Bottlenecks emerge, together with black markets, and the system disintegrates from within. Historically the best example is Soviet planning, but modern examples can still be found in some developing economies.

Markets versus firms

Focusing mainly on government versus markets appeared anomalous even in the 1930s. By this time capitalism had evolved large managerial firms that administered the prices of their products (Berle and Means, 1932). If capitalism relied on markets, why was so much power vested in mangers, and why were firms so large? A popular explanation relied on economies of scale, linked to advanced technology and capital-intensive production. But economies of scale in production implied that the firm was large because it owned a single large plant, whereas many firms were large because they owned many small plants. Furthermore many large firms produced several products, rather than just a single product, and often undertook several consecutive stages of production.

Business writers of the time asserted that managers coordinated the economy, whilst economists asserted that, on the contrary, markets did so instead. Coase argued that both were correct. By focusing on intermediate product markets rather than final product markets, he showed that firms could internalise markets by bringing related activities under common ownership and control. Managers coordinated flows of intermediate product internal to the firm, whilst markets coordinated flows of final product external to the firm.

By focusing on intermediate products, Coase was also able to link managerial coordination to an internal division of labour. Instead of emphasising technological economies of scale at the plant level, his approach emphasised the division of labour at the firm level instead.

Firms and governments compared

According to Coase, a large firm may be viewed as a miniature version of the state. It can plan its internal division of labour because it owns all the resources involved. The main difference is that the firm is privately owned, its labour is free rather than directed, and competition may develop with other firms.

Firms can exist in a socialist economy as devolved decision-making units, although they operate within the framework of the state. In a market-based economy, firms operate subject to the discipline of the market rather than the discipline of the state. Although an internalising firm may be deemed to suppress an external market, it is ultimately market forces that govern internalisation, rather than the other way round. Internalisation is successful only if customers buy the final product, which means that internalisation must contribute to improving product quality or reducing costs. If internalisation reduces profitability then the firm may be unable to invest and in extreme cases it may fail. In a market economy, therefore, the degree of internalisation is ultimately governed by market forces rather than by managerial discretion. Although managers may favour internalisation because it creates more managerial jobs, it is ultimately profitability that governs internalisation, and this is determined by market forces.

There is a margin of substitution where either firms or markets can coordinate intermediate product flow. Efficiency requires that at this margin the benefits of internalisation are just equal to the costs. This leaves open the question of the nature of the benefits and costs and how they are measured.

Neoclassical economics shows that it is impossible to improve on the efficiency of a perfect market. The implication is that internalisation must be a product of market imperfections: the benefit of internalisation is equal to the cost of the market imperfections that it avoids. The key to assessing benefit is therefore to understand which particular types of market imperfection impede intermediate product markets.

The following sections examine three imperfections:

- Markets are uncompetitive and price discrimination is impractical
- Products are heterogeneous
- Traders are untrustworthy

There are many other imperfections that could be discussed (Casson, 1986); these have been selected because they are treated inadequately in the modern IB literature. The first two issues are largely ignored, whilst the third, though strongly emphasised, is often analysed superficially. The costs of internalisation are discussed later.

4. Imperfect competition and price discrimination

Competition emerges when different people recognise similar opportunities and set up firms to exploit them. The classic forum for competition is the final product market, where producers confront consumers. Competition based on freedom of entry into industry discourages the exploitation of consumers because any attempt by a firm to raise price will

attract entry, increase supply, reduce price, and restore profits to their normal level. Likewise, competition for free labour will ensure that labour is not exploited either.

It is widely held that monopoly is not only inequitable but also inefficient. It is argued that monopolised industries produce too little output because the price is so high that it restricts consumer demand. Strictly speaking, however, it is only differences in the degree of monopoly between industries that reduce efficiency, because if all prices were raised in the same proportion then relative prices would be unchanged and consumer purchasing decisions would not be distorted (although other decisions might be distorted instead) (Lerner, 1944). The argument against monopoly also assumes that the monopolist must charge the same price to all customers. This ignores the possibility of discriminatory pricing (Phillips, 2005). If the monopolist knows the maximum amount that each customer (or type of customer) is willing to pay then they can charge different prices to different customers depending upon how much they value the product. The main requirement is that they can prevent the consumers reselling to each other, or joining forces to form a buyer's club. If these conditions are satisfied, the marginal consumer pays no more than marginal cost and so the scale of output in each industry is efficient.

The efficiency of monopolistic price discrimination is widely used to support intellectual property rights (IPRs) that confer monopolies for the creation or discovery of knowledge. IPRs promote private enterprise in the creation of knowledge, but the argument against them is that they discourage dissemination by charging for access. However, if the owners of IPRs implement discriminatory pricing then no one is asked to pay more than they are willing to pay and so dissemination is not impaired (Casson, 1979). Indeed, private ownership encourages the active marketing of knowledge, so that more people may use the knowledge than before. On the other hand, the administrative costs of collecting payment may means that people with low valuations are denied effective access.

These arguments apply not only to final product markets but to intermediate product markets too. They suggest that efficient markets are either competitive, or involve discriminating monopoly. There are two main mechanisms by which competition is sustained. One involves a large number of suppliers confronting a large number of sellers, and the other involves a small numbers of buyers and sellers, but with potential entrants on either side waiting for an opportunity to join in (Baumol, Panzar and Willig, 1982). Intermediate product markets for agricultural products, linking farms to food processors, are a good example of competitive markets with large numbers of traders. Markets for mineral ores exemplify competition from potential entry; at any one time only a small number of large mines may be in operation, but there are usually other mines ready to be opened (or more likely re-opened) if price increases. Competitive entry and re-entry is easiest when the sunk costs of entry are small.

Under monopoly, market failure reflects the inability to discriminate. Consider, for example, the licensing decision. A technology owner serving the global market may prefer to license different firms in different countries because of their local knowledge. But it may be difficult to partition local markets in this way. If licensees can export then they can invade each others' territories; this threat will reduce the value of the licenses, and ultimately reduce the

technology owners' rents. The technology owner may therefore be obliged to use a single licensee for all markets, who will be less effective in each market and generate fewer rents for the licensor.

Inability to discriminate can also be an issue for ordinary intermediate product markets where production at certain stages exhibits economies of scale. Within a multi-stage production system (a 'value chain') one stage (say the upstream stage) may exhibit substantial economies of scale, so that industry production is in the hands of a single firm, whilst the downstream stage may exhibit constant return to scale, so that many small firms are involved. If the upstream firm sets a uniform monopoly price then downstream decisions will be distorted by the artificial scarcity of the intermediate input (e.g. excessive costs will be incurred in avoiding wastage) (Warren-Boulton, 1978). On the other hand, if the upstream firm charges all the downstream firms a two-part tariff, comprising a lump sum payment for the right to purchase and a unit price equal to upstream marginal cost then distortion will be eliminated. The efficiency gain will accrue to the monopolist, whose profits will increase as a result. But if the downstream firms can re-sell then the system will be undermined, as they can form a buyers' co-operative and pay the lump sum only once. Furthermore, with a downstream buyer's co-operative confronting an upstream monopolist, a bilateral monopoly may develop; competition breaks down, and exchanges of threats may ensue.

5. Heterogeneity

A careful reading of Coase (1937) suggests that the market imperfections that he had in mind were search costs (Casson, 2000): the costs of seeking out a suitable supplier or customer, and comparing the prices involved in alternative trades. Search implies that the product concerned is specific. A customer may be looking for a product that is novel, antique, or in some way unique, while a supplier may be looking in a specific locality for particular type of customer. In large markets there may be many suppliers and many customers. With many varieties of product it is important to match the supplier to the customer; a random pairing of customer and supplier may not give the customer exactly what they want, and cause the supplier to make a lower profit as a result (Roth and Sotomayor, 1990). To find an appropriate match it is normally necessary for each party to investigate a number of options. Search costs therefore include the costs, not only of searching out the partner with whom trade takes place, but making contact with other potential partners, to ensure that the chosen partner is an appropriate match.

Heterogeneity may occur naturally, as in different varieties of foodstuffs, or by design, as in different types of furniture or other household durables. Heterogeneity in time of delivery is also important for services and perishable goods, and is sometimes addressed through forward markets (Arrow, 1975).

With heterogeneity, different variants will have different values to a customer and they may have different production costs too. As a result there may be a spread of prices – the law of one price, which applies to homogeneous products – will not prevail. Customers therefore need to search across prices as well as across product characteristics.

Search costs can be reduced by *intermediation*. A market intermediator establishes contact with both buyer and seller, and thereby coordinates their decisions (Casson, 1982). In final product markets, retailers specialise in intermediation. A retailer may hold a range of different varieties (e.g. different brands) of related goods (e.g. household durables). Retailers typically hold stock: they take a speculative position by buying and re-selling the products in which they deal. However, intermediators can also act as brokers, e.g. by introducing buyer and seller to each other, and charging a fee for a successful match (e.g. auctioneers, estate agents).

Intermediation is also used in intermediate product markets. Agricultural products such as corn and coffee, and fuels such as coal and oil, are traded by specialist dealers on international exchanges. Variety is reduced by grading and standardisation.

Search costs can also be reduced by *internalisation*. Internalisation avoids the cost of a customer finding a supplier and, conversely, the cost of a supplier finding a buyer. In retail markets, for example, consumers may decide to 'do it themselves': repairing their own motor car, decorating their own home, and so on; this avoids having to find a suitable supplier, e.g. a local motor mechanic or decorator. Markets for consumer goods and services can also be internalised within the household, e.g. child care.

Internalisation is most common in intermediate product markets, however. While consumers can integrate backwards into production through do-it-yourself production, producers cannot integrate forward into consumption; thus forward integration occurs only in intermediate product markets and not in final product markets. In intermediate product markets, firms can integrate either backwards, e.g. by making rather than buying their components, or forwards, e.g. by retailing products they have produced themselves.

In practice the alternative to internalisation is often an intermediated external market. In fact, internal markets are usually intermediated too. Within an MNE with several subsidiaries, one subsidiary does not necessarily negotiate directly with another subsidiary – relations between subsidiaries are coordinated by headquarters instead. For example, upstream subsidiaries may not negotiate directly over intermediate product prices with downstream subsidiaries, but instead both groups of subsidiaries may submit tenders to headquarters, which then determines trade by matching up the successful bids.

Intermediation can have multiple levels. In final product markets bulk trading is often coordinated by wholesalers, who then sell to retailers who sell on to consumers. The same principle applies to intermediate products, and to internal markets as well. Consider an intermediate product in a global supply chain coordinated by an MNE. The MNE may establish regional headquarters to intermediate between global headquarters and national subsidiaries. The regional headquarters may negotiate with each other over regional intermediate product trade, and each region may then negotiate with its subsidiaries over the allocation of output. This point is developed further in sections 7 and 8.

6. Deception

Coase's emphasis on costs of search may seem unfamiliar to modern readers – with good reason, because modern institutional theories of the firm emphasise deception instead (Williamson, 1975, 1985). Deception may involve providing false information, or simply withholding information. A customer may have no intention of paying for a product, and a seller may substitute a poor quality product for a good quality one. The dishonest customer does not reveal their intentions to the seller, and the dishonest seller does no reveal the deception to the customer.

Deception may be deliberate or accidental. Deliberate deception is practiced by a selfish and dishonest person, but unintended deception may be practised by an incompetent person, e.g. a supplier does not realise that their product is faulty (Casson, 1997). The literature normally assumes the deception is deliberate, and for simplicity the same convention is followed here.

Deception creates a special type of matching problem. A successful trade is achieved by pairing a honest buyer and with an honest seller. If a dishonest buyer is paired with an honest seller then the honest seller may not get paid and, conversely, if an honest buyer is paired with a dishonest seller then the product will be faulty. If a dishonest buyer is paired with a dishonest seller then, in effect, no trade will take place; the buyer will receive a worthless product and pay nothing for it.

Deception therefore aggravates the search problem. Not only are there different varieties of product, but there are good and faulty variants of each (Akerlof, 1981). Once again, both intermediation and internalisation can help. An intermediator with a reputation for honesty can enter the market as a re-seller. This is most obvious with branded goods in consumer markets: buyers fear poor quality and sellers fear that they will not be paid. The intermediator can insist that the sellers from whom they buy accept payment in arrears and that the buyers to whom they sell pay in advance. In effect, those without reputation – the buyer and seller - insure those with reputation – the intermediator – against the risk of default. Since the intermediator trades upon their reputation they have no incentive to default themselves, unless there is the prospect of a major one-off 'sting'.

Internalisation removes the incentive for dishonesty at a stroke. Buyer and seller become the same person, and there is nothing to be gained by cheating on yourself. This is particularly important when there are problems of quality control. Suppliers often know more about the quality of their product than the buyer (asymmetric information), especially when the buyer cannot easily detect a fault (the product is an 'experience good' rather than an 'inspection good') (Nelson, 1970).

A key case concerns the marketing of proprietary knowledge. Internalisation theory highlights two distinct but related issues: a producer who is offered a technology under license may not know whether it is sound, or whether it is as exclusive as the licensor claims. Secondly the licensor may be unsure what the producer intends to do with the knowledge that they get, and whether they will respect the terms of the licence. The licensor cannot easily re-assure the buyer about the quality of the knowledge because if they divulge too much the producer may have no need to acquire the licence (Buckley and Casson, 1976). These issues

are distinct from, although related to, the problem of price discrimination between multiple licensees. They are also distinct from the question of whether the producer could understand and apply the knowledge embodied in the licence if they acquired it – an issue addressed in the resource-based literature (Barney, 1986).

Search costs arising from heterogeneity and costs arising from deception are not always properly distinguished in the IB literature. They are sometimes referred to collectively as 'transaction costs'. Search costs are certainly incurred in making transactions, but they are not included in Williamson's concept of transaction cost. According to Williamson (1975, 1985), transaction costs arise from 'opportunism', which is Williamson's word for calculated dishonesty, in which people may lie or strategically withhold information. But when heterogeneity is the issue, rather than deception, it generally pays to divulge information rather than withhold it. A customer who wishes to have a product delivered to their home would be foolish to withhold their address or give a false address, since delivery would fail. Thus the costs of transacting, considered in their totality, exceed the 'transaction costs' identified by Williamson, which exclude those costs that are purely due to heterogeneity.

Internalisation theory is more rigorous than transaction cost theory because it avoids conflating search costs and 'transaction costs' incurred by deception. A rigorous treatment of the costs of using the market, as described by Coase, would identify at least five constituents of cost (Casson and Lopes, 2013):

- Search costs incurred by heterogeneity of the product;
- Search costs incurred in seeking out honest and competent trading partners;
- Costs incurred in discouraging deception by dishonest trading partners;
- Costs incurred in mitigating the effects of dishonesty; and
- Costs incurred where prevention and mitigation fail.

While three of the five costs relate mainly to purely deception, two involve search, including one that involves an element of deception too. In the context of the subcontracting the production of a component, these costs would be:

- *Search only*: Cost of specifying the component and identifying a set of competing firms that can produce it
- *Search and deception*: Cost of selecting a subset of reputable firms using industry contacts and local knowledge
- *Deception only*: Cost of quality control of output, spot checks on production, etc.;
- *Deception only*; Cost of building safety features into the final product to mitigate the effects of failure in the component;
- *Deception only*: Loss of reputation when the final products fails because of a faulty component.

Internalisation can reduce or eliminate all of these costs, although its impact in any specific case will depend on the nature of the intermediate product involved. Novelty and complexity of the intermediate product will tend to encourage internalisation; in general, any

intermediate product whose quality is crucial to the performance of the final product is a candidate for internalisation.

7. Costs of internalisation

The costs of internalisation are treated in a cursory manner in the IB literature. This section examines the costs of internalisation, with special reference to the problems of coordinating a complex division of labour within a firm. The costs of internalisation are often addressed in terms of the agency problem, whereby shareholders cannot trust salaried managers to do a good job (Milgrom and Roberts, 1992; Jensen, 2000).

The agency problem shows that while internalisation may avoid deception in an external market, the deception may reappear in a different guise in the internal market. The problem of a dishonest buyer or seller is replaced by the problem of a dishonest manager – but at least the manager can be supervised, whereas the buyer and seller usually cannot.

Internalisation theory indicates that managers are employed to coordinate flows of intermediate product within the firm. Agency problems arise because, unlike intermediators in external markets, managerial intermediators do not buy and sell for their own profit. Any profits or losses they make accrue to their employers instead; the worst that can happen to a manager that makes a mistake is that they lose their job. Although well-motivated managers may do a better job than external intermediators, poorly motivated managers may do a worse job instead.

It is often alleged that as the scale of its operation expands, an organisation becomes increasingly bureaucratic, and that average costs of administration therefore increase. Such allegations are particularly common with respect to government, but they have also been directed at firms. Economists have argued that, while there may be technological economies of scale with respect to size of plant, there are often diseconomies of scale with respect to size of firm. These diseconomies are associated with the coordination of plants, and are managerial in nature. Increasing managerial diseconomies of scale may set an overall limit to the size of the firm.

In neoclassical economics managerial diseconomies are derived from an assumption that management is a fixed factor, specific to each firm (Marshall, 1890; Kaldor, 1934). Because of this fixed factor, there are diminishing marginal returns to variable inputs such as labour and capital, and so there is an optimal size beyond which the firm cannot profitably expand. In a competitive industry, any firm that attempts to expand beyond optimal size will be unable to compete with firms that operate only at optimal size, as its competitors will only just break even at the competitive price and so the larger firm will make a loss. On this view, equilibrium differences in the sizes of firms are due to different endowments of the fixed factor.

If the fixed factor is identified with the capability of a single owner-manager, such as the founder, then the size of the firm is dictated by their ability. This implicitly assumes that each firm has a single owner manager, however. An alternative view is that the owner of the firm

is the fixed factor, and that they may delegate authority to subordinate managers. On this view the fixed factor is the ability of the owner or entrepreneur to delegate authority in an efficient manner. The larger the span of control that the owner can exercise, the larger their firm can become.

The neoclassical analysis of firm size is somewhat abstract. Internalisation theory offers a more grounded explanation. As the firm grows, it structure evolves, and as it becomes more complex, it becomes increasingly difficult to coordinate. A firm does not necessarily expand simply by scaling up each activity in the same proportion. As scale increases, complexity increases too. The IB literature highlights the spatial dimensions of complexity: the unfamiliarity of foreign markets, problems of recruiting overseas labour, political threats, and so on. The product dimension is also important too: if the firm diversifies into different products then it may have to master new technologies, and recruit additional specialists with relevant knowledge (Wolf, 1977). Even if the firm simply expands its sales of existing products, the increased scale on which each product is produced may provide new opportunities to advance the division of labour. When producing a small amount of output, one worker may suffice at each stage of production, but when scale increases, and two or three workers are employed, productivity may be raised by specialising jobs at each stage. Coordinating people performing different tasks is more complicated than coordinating people performing identical tasks, and disruption (e.g. due to absence or illness) may be more serious because people cannot cover for each other so easily.

Management experience is an important factor in addressing complexity. As the size of the firm expands, managers who joined the firm at an early stage are confronted with problems they have not met before. As the complexity of the division of labour increases, record-keeping becomes more onerous. As the workforce expands, relations become impersonal, trust becomes weaker, and morale may suffer, and as specialisms proliferate rivalries may develop between different specialist groups. Some managers may have the ability to learn from experience, adapt, and ‘grow with the job’, but other may not. Recruiting experienced managers from established large firms into senior positions can avoid this problem, but it threatens the promotion prospects of long-serving staff; there are also costs of training and induction, as emphasised by Penrose (1959).

The threat of growing complexity can be confronted in two main ways, both of which have important implications for the strategy of the firm. One is to introduce new methods of coordination, and the other is to deliberate reduce complexity by foregoing opportunities. An efficient approach may involve combining the two approaches.

When it is difficult to observe directly how people behave, and to trust them, it is useful to measure the results they achieve. Transparency helps to hold managers to account. Divisionalisation offers one approach (Rumelt, Schendel and Teece, 1984). Different activities within the firm may be established as profit centres, trading with each other at internal transfer prices, e.g. upstream production and downstream production. Transfer prices can distinguish between under-performance in different centres.

Success, however, depends on transfer prices being set correctly. If transfer prices are too high then downstream production will be penalised, and conversely if they are too low then upstream production will be penalised instead. In the special case of internal bilateral monopoly, for example, where a single downstream facility is locked in to buying from a single external facility, price may be set through internal politics. So how is it possible to determine whether transfer prices are correct? One answer is to compare them with external prices. Where novel products are concerned, however, external markets may not have developed, and so external price information is not available. Furthermore, where the benefits of internalisation are high, all other firms in the industry may internalise as well, and so again no external prices are available.

A potential solution is to open up the internal market by allowing individual profit centres to trade externally if they wish (Casson, 1997). Thus upstream production can sell intermediate product to independent downstream producers, while downstream production can procure from independent upstream producers. Even if all trade is internalised, prices will be generated whenever a downstream producer in one firm can be persuaded to quote to an upstream producer in another firm. Indeed, where novel products are concerned, the opportunity to bid in an intermediate product market may encourage the emergence of independent start-up firms. The opening up of the internal market can give the firm the best of both worlds; search is facilitated because alternative sources of supply and demand can be explored; the quotes obtained can then be used to set internal prices, and individual profits centres can then decide whether to internalise or not.

Divisionalisation can be implemented by making each foreign subsidiary a profit centre. Authority for local decisions is then devolved to local management. Subsidiary autonomy reduces information flow because local decisions can be taken using local information, without seeking authority from the centre. Subsidiary autonomy therefore contributes to the growth of the firm whether or not the subsidiary undertakes R&D and contributes to product innovation (Papanastassiou and Pearce, 2009).

The second approach suggests that the firm will avoid strategies that increase complexity and embrace those that reduce it. When a firm is exploiting proprietary knowledge, its natural ambition is to supply the entire global market, but it may be thwarted by the complexity of coordinating global production (Langlois and Robertson, 1993). To reduce this complexity, the firm may choose to out-source more than a smaller firm, and to restrict the number of different locations at which its facilities are based, and possibly the distances between them too. These restrictions prevent it from exploiting the international division of labour to the full, but the loss of operating profit may well be offset by the savings of managerial costs.

Both approaches to reducing costs have their limitations, and when their potential has been exhausted the only strategy left is to limit the size of the firm. The firm may decide not to exploit the entire global market but to become a ‘regional’ multinational instead (Rugman, 2005). In the context of the theory, however, the ‘region’ does not have to be a connected space; in the nineteenth century for example, it could be the British Empire, and in the 21st century it may correspond to the diaspora of some particular ethic group.

8. Business failure and MNEs

It seems that the costs of internalisation are not only poorly understood by scholars, but are poorly understood by practitioners too. Managers who appreciate the benefits of internalisation but not the costs are liable to over-expand their firms; they do not appreciate the limit to size set out above. In an expanding firm the problems of managing size are a continuing source of surprise unless managers have worked in a large firm before. If inexperienced managers fail to learn as the firm grows then at some point internal coordination may break down, leading to a catastrophic failure of the firm. Large firm failure, in other words, may be understood as a failure to optimise internalisation in a growing firm.

Large firm failure, and the failure of established MNEs in particular, has received little study. This is surprising. It is widely appreciated that firms have a life-cycle: they are born, they grow, mature and die. However, different branches of the theory of the firm focus on different stages of a firm's life, and none provides a balanced account of all four stages. The same is true of IB studies. Start-ups are discussed in the born global literature, growth is discussed mainly from a resource-based perspective, whilst maturity is the focus of internalisation theory and the OLI paradigm. In both economics and IB there is little discussion of the death of the firm. This section considers the implications of internalisation theory for business failure, drawing on the previous analysis.

Deaths of firms are very common. Many small firms have only short lives, and many successful business people establish several firms before they finally establish the one that survives and grows (Westhead and Wright, 1998). The risk of failure in small business is so high that deaths of small firms attract little attention (Cressy, 2006). Failures of large firms are less common, and therefore attract more public attention; the more successful a firm has been in its early career, the more paradoxical its failure can seem. Such failures are also more serious – not only do owners lose money but employees lose their jobs and customers their sources of supply. But despite public attention, academic study remains limited.

Some firms are 'too big to fail'. Major retail banks and insurance companies, 'national champion' firms in high-technology industries, and defence contractors holding sensitive information, may all be protected from failure by bail-outs. Bail-outs usually occur under extreme conditions where no firm is willing or able to take over the failing enterprise. Its debts are too great, or its assets are totally valueless. More modest failures may be addressed by take-overs. Sometimes a merger may be used to 'window dress' a take-over; after the take-over brands are sold off, land is disposed of, and manufacturing plant is scrapped.

A crucial feature of failures is that they are often not foreseen, even by those closely involved. If the owners of firms had perfect foresight they would recognise that brands can go out of fashion, technologies obsolesce, and so on. If a rational owner recognised that nothing would arrest the decline of their firm then they would plan for their firm's demise, and run it down as part of an optimal exit strategy. They would stop investing, and 'sweat' the existing assets; they would also stop recruiting and encourage early retirement. But in fact owners often seem either to be unaware of their problem, or to believe that the answer lies in faster

growth; they stick to their existing 'business model' but apply it more intensively than before. This explains why take-overs are often hostile, and why bail-outs are sometimes government-imposed.

It is useful to distinguish between internal and external threats to survival. External threats are exemplified by imitation and new sources of competition, whilst internal threats are exemplified by failures of coordination.

Many external threats can be interpreted as the erosion of ownership advantage. As time passes, competition intensifies and decline sets in. Continuous innovation may be answer, but there may be diminishing returns to innovation within a given paradigm. The greatest threats may come from radical innovations unrelated to the initial advantage of the firm. This suggests that internal threats may constrain the firm's ability to respond to external ones. Managers may not possess the breath of vision required to understand potential sources of competition (Porter, 1980).

The ability to manage internal threats is a distinctive asset. It is complementary to conventional ownership advantage based on technology and product design. It represents management capability rather than the possession of intellectual property.

As the small business literature emphasises, the growth of the firm involves surmounting a succession of barriers (Storey and Greene, 2010). As it expands, a firm with management capability can adapt both by limiting complexity and by increasing the sophistication with which complex operations are managed. Management capability allows the firm to enter new markets and to produce in new location without losing control of it operations.

Management capability is independent of the advantage on which the growth of the firm is based. In the early stages of growth the firm does not require the full management capability because it is not sufficiently large to encounter the most serious size-related problems. But if the firm does not have management capability then it may fail to recognise the challenges of large-scale management until it is too late. It will not put in measures to slow down growth and counter bureaucratic pressures, and so will be overtaken by problems it fails to recognise and do not really understand. By contrast a firm with management capability can continue growing to a larger size because adapts its organisational structure. Growth may slow, however, in order to stay within the safety zone.

Managerial capability is a scarce factor, but unlike proprietary knowledge it is not a firm-specific asset. The fact that one firm possesses a managerial capability does not preclude another firm from possessing it too. The only firm-specific asset is the owner's ability to recruit capable managers more cheaply than others or, conversely, to a recruit manager of greater capability for the same salary.

Managers with experience of large scale organisation will command salaries that reflect a scarcity premium. The more relevant their experience, the greater their salary; the difference between their salary as a manager and their best alternative earnings measures the personal rent they derive from their job. The owners of the firm will extract a rent only if they can

search out capable managers more effectively than their rivals and then induce them to remain with the firm. A similar point is made in resource-based theory, but it originates with the theory of the firm (Knight, 1921)

9. Conclusions

This paper has re-affirmed the importance of the economic theory of the firm as a foundation for IB theory. The theory of the MNE developed by taking the Coasian model of the firm and adding two key ingredients: proprietary knowledge generated by the firm, and the spatial dimension of production. The concept of the knowledge-based firm was subsequently taken up by resource-based theory and the strategy literature, whilst the spatial dimension of the firm was developed further in economic geography (McCann and Iammarino, 2013). IB theory has broadened out to analyse the institutional environment of the environment but has, ironically, lost sight of the significance of the firm as an institution itself. Internalisation theory is now the only branch of IB theory that offers an logical integrated approach to knowledge and space based on the economic theory of the firm.

The spatial dimension of the firm is clearly crucial to IB theory. The spatial dimension can only be analysed by distinguishing sharply between the plant and the firm. Plants can exploit technological economies of scale. For the firm, economies of scale originate mainly from the exploitation of the knowledge it possess. Knowledge is a public good, and the marginal cost of exploiting it is usually below its average cost, because the average cost includes the cost of R&D and the marginal cost does not. So why do all knowledgeable firms not become global? One reason is that not all knowledge is technological; marketing knowledge, in particular, may be localised. Thus firms possessing only local marketing knowledge may be unable to expand abroad. Lack of foreign marketing knowledge can also impair the expansion of a technology-driven firm unless it partners with a foreign firm that possesses local knowledge or hires experienced local personnel.

There is another factor that constrains the expansion of a firm, however, and that is managerial capability. There are different forms of capability. The capability that constrains firm size concerns the ability to coordinate a complex division of labour. In this context ‘large’ refers principally to the number of employees, and only secondarily to the value of sales. Complexity refers to the geographical diversity of production and sales, the range of specialist skills required, and the number of different intermediate product flows.

This paper has argued that management capability is a complement to ownership advantage. Ownership advantage determines the potential of the firm, and management capability governs the fulfilment of this potential through overcoming barriers to growth. The paper shows how complexity can be limited by out-sourcing, licensing, and other arm’s length contractual arrangements. It also explains how internal complexity can be reduced by making internal markets more competitive. Internal complexity can also be addressed by simplifying management procedures to reduce information costs, and by making subsidiaries more autonomous. Failure to manage the increasing demands of internal coordination can result in

dramatic failures, resulting in hostile acquisitions or government bail-outs of the kind witnessed during and after the recent Banking Crisis.

References

Alchian, Armen A. and Harold Demsetz (1972) Production, information costs and economic organization, *American Economic Review*, 62 (5), 777-795

Akerlof, George A. (1981) The market for 'lemons': Quality uncertainty and the market mechanism, *Quarterly Journal of Economics*, 84 (3), 488-500

Arrow, Kenneth J. (1975) Vertical integration and communication, *Bell Journal of Economics*, 6 (1), 173-183

Barney, Jay B. (1986) Strategic factor markets: Expectations, luck and business strategy, *Management Science*, 32, 1231-1241

Baumol, William John C. Panzar and Robert D. Willig (1982) *Contestable Markets and the Theory of Industry Structure*, New York: Harcourt Brace Jovanovich

Berle, Adolph A., Jr. and Gardiner C. Means (1932) *The Modern Corporation and Private Property*, New York: Macmillan

Buckley, Peter J. (2009) The impact of the global factory on economic development, *Journal of World Business*, 44 (2), 131-143

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

Buckley, Peter J. and Mark Casson (2007) Edith Penrose's Theory of the Growth of the Firm and the strategic management of multinational enterprises, *Management International Review*, 47 (2), 151-173

Buckley, Peter J. and Mark Casson (2010) Entrepreneurship and the growth of the firm: An extension of Penrose's theory, in Mark Casson, *Entrepreneurship: Theory, Networks, History*, Cheltenham: Edward Elgar, 88-111

Cantwell, John A. (2014) Re-visiting international business theory: A capabilities-based theory of the MNE, *Journal of International Business Studies*, 45, 1-7

Casson Mark (1979) *Alternatives to the Multinational Enterprise*, London: Macmillan

Casson, Mark (1982) *The Entrepreneur: An Economic Theory*, Oxford: Martin Robertson

Casson, Mark (1986) The theory of vertical integration: A survey and synthesis, *Journal of Economic Studies*, 11 (2), 3-43

Casson, Mark (1997) *Information and Organization*, Oxford: Oxford University Press

Casson, Mark (2000) *Economics of International Business: A New Research Agenda*, Cheltenham: Edward Elgar

Casson, Mark and Teresa da Silva Lopes (2013) Foreign direct investment in high-risk environments: An historical approach, *Business History*, 55 (3), 375-404

Coase, Ronald H. (1937) The nature of the firm, *Economica* (New series), 4, 386-405

Cressy, Robert (2006) Determinants of small firm survival and growth, in Mark Casson, Bernard Yeung, Anuradha Basu and Nigel Wadeson (eds. *) Oxford Handbook of Entrepreneurship*, Oxford: Oxford University Press, 161-193

Dietrich, Michael and Jackie Krafft (eds.) (2012) *Handbook of the Economics and Theory of the Firm*, Cheltenham: Edward Elgar

Dunning, John H. and Sarianna M. Lundan (2008) *Multinational Enterprises and the Global Economy*, 2nd ed., Cheltenham: Edward Elgar

Hennart, Jean Francois (1982) *A Theory of Multinational Enterprise*, Ann Arbor: University of Michigan Press

Jensen, Michael C. (2000) *Theory of the Firm: Governance, Residual Claims and Organizational Forms*, Cambridge, MA: Harvard University Press

Johanson Jan and J.-E.Vahlne (1977) The internationalisation process of the firm: A model of knowledge development and increasing foreign market commitment, *Journal of International Business Studies*, 8 (1), 23-32

Kaldor, Nicholas (1934) The equilibrium of the firm, *Economic Journal*, 44, 60-76

Knight, Frank H. (1921) *Risk Uncertainty and Profit*, Boston: Houghton Mifflin,.

Lange, Oskar and Fred M. Taylor (1938) *On the Economic Theory of Socialism*, (ed. B.E. Lippincott), Minneapolis, MN: University of Minnesota Press

Langlois, Richard N. and Paul L Robertson (1993) Business organisation as a coordination problem: Towards a dynamic theory of the boundaries of the firm, *Business and Economic History*, 22(1), 31-41

Lerner, Abba P. (1944) *Economics of Control*, New York: Macmillan

Marshall, Alfred (1890) *Principles of Economics*, London: Macmillan

McCann, Philip and Simona Iammarino (2013) *Multinationals and Economic Geography*, Cheltenham: Edward Elgar

Milgrom, Paul R. and John Roberts (1992) *Economics, Organisation and Management*, Englewood Cliffs, NJ: Prentice Hall

Nelson, Philip (1970) Information and consumer behaviour, *Journal of Political Economy*, 78 (2), 311-329

Papanastassiou, Marina and Robert D. Pearce (2009) *The Strategic Development of Multinationals: Subsidiaries and Innovation*, Basingstoke: Macmillan

Peng, Mike W., Denis Y.L. Wang and Yi Jiang (2008) An institution-based view of international business strategy: A focus on emerging economies, *Journal of International Business Studies*, 39, 920-936

Penrose, Edith T. (1959) *Theory of the Growth of the Firm*, Oxford: Blackwell

Phillips, Robert (2005) *Pricing and Revenue Optimization*, Stanford, CA: Stanford University Press

Pitelis, Christos (2007) A behavioural resource-based view of the firm: The synergy of Cyert and March (196) and Penrose (1959), *Organizational Science*, 18 (3), 478-490

Porter, Michael E. (1980) *Competitive Strategy*, New York: Free Press

Roth, Alvin E. and M.A.O. Sotomayor (1990) *Two-sided Matching*, Cambridge: Cambridge University Press

Rugman, Alan M. (1981) *Inside the Multinationals*, London: Croom Helm

Rugman, Alan M. (2005) *The Regional Multinationals*, Cambridge: Cambridge University Press

Rumelt, Richard P., Dan E. Schendel and David J. Teece (eds.) (1984) *Fundamental Issues in Strategy: A Research Agenda*, Boston, MA: Harvard Business School Press

Storey, David J. and Francis J. Greene (2010) *Small Business and Entrepreneurship*, Harlow: Financial Time Prentice Hall

Warren-Boulton, Frederick R. (1978) *Vertical Control of Markets: Business and Labor Practices*, Cambridge, MA: Ballinger

Westhead, Paul and Mike Wright (1998) Novice, portfolio and serial founders: Are they different? *Journal of Business Venturing*, 13, 173-204

Williamson, Oliver E. (1975) *Markets and Hierarchies*, New York: Free Press

Williamson, Oliver E. (1985) *The Economic Institutions of Capitalism*, New York: Free Press

Wolf, Bernard M. (1977) Industrial diversification and internationalization: Some empirical evidence, *Journal of Industrial Economics*, 26 (2), 177-191

Appendix

This appendix specifies and solves a formal model of the MNE, and summarises the economic significance of the solution. It explains how the size and strategy of the firm are influenced by the costs and benefits of internalisation. The model determines the optimal size of an MNE, together with its optimal investment in R&D, degree of internalisation, and geographical concentration of production in the home country.

The modelling of internalisation costs is somewhat similar to the modelling of Penrose's constraints on growth (Penrose, 1959; Buckley and Casson, 2007, 2010). There is a fundamental difference, however, because the model focuses on optimal size instead of optimal growth. It is therefore in the tradition of Marshall (1890) rather than Penrose. Because the model focuses on size it is not so dynamic as a theory of growth but, unlike Penrose's model, it does not imply that the firm will grow without limit, or until it totally dominates its market.

Consider a firm that owns a proprietary technology used to produce a particular product. Suppose that there is a potential global market for the product. This global market is partitioned by political boundaries into a large number of relatively small national markets. In each market there is a fixed number of customers and each customer purchases only one unit. The firm can choose how to serve each market. When serving a foreign market the firm chooses between exporting, import-substituting FDI, and licensing. Other options, such as off-shore production and joint ventures, may be available. When serving its domestic market the normal choice is wholly-owned local production.

In each market the firm faces competitors who use alternative technologies. Competitive conditions vary between markets. In each market there is a strongest competitor that must be driven out of the market if the firm is to supply it. In each market the firm's technology is superior to the strongest competing technology by a proportional factor q. The firm's superiority is reflected in the quality of its product and not in its cost of production. The strongest alternative product is priced at unity in each market (in terms of the home-country currency), and the firm's product is priced at q. If it were priced at more than q then customers would not buy it, and if it were priced at less than q then the firm would fail to maximise revenue because it charged less than customers were willing to pay.

Output produced by the firm, whether as exporter or foreign direct investor, is coordinated internally by the firm. By contrast, output produced under license, or other arm's-length arrangements, is coordinated by others. Exporting and FDI therefore impose demands on the firm's management that licensing does not.

Two types of costs are distinguished: market-sourcing costs and overall coordination costs.

Market-sourcing costs are related to entry into specific markets. Serving a foreign market by exports incurs trade-related costs, whilst FDI incurs costs of international technology transfer, and also political risks. There are also costs of licensing. The unit cost of supply to any market is the sum of production costs and market-sourcing costs. Let all the national markets

be ranked in ascending order of their unit cost of supply. Typically the home market will be ranked low and distant markets with alien cultures and hostile governments will be ranked high. This ranking is specific to the firm; competitors based in other countries will normally have different rankings, because their geographical, cultural and political relations with each market will be different.

The unit cost of supply for each market is calculated on the basis that market sourcing strategy is optimised conditional on trade costs, international technology transfer costs, political risks and licensing costs. Optimisation takes account of overall coordination costs too. Because of market-specific factors, some markets will be served by exporting, some by FDI, some by licensing, and so on. Because of firm-specific factors, large firms may bias individual market supply decisions towards licensing in order to reduce the burden of coordination, and if they produce internally they may also bias production location decisions towards exporting rather than FDI. The cost of supply to any given market is the simply the cost of the chosen method, and is expressed as a unit cost, c.

When markets are ranked in this way, each successive unit of the firm's output costs no less than the previous one, and possibly more; on average, therefore, the cost of supply increases with the total number of units sold, x. It is assumed that this cost relationship is approximately linear; $c(x) = a + dx$. In this relationship, $c(x)$ is the marginal cost of supplying the xth unit of sales; the parameter a is a constant firm-specific component of marginal cost, reflecting production costs.

The parameter d measures the impact of 'distance', construed in general terms, on the marginal cost of supply. If the firm is headquartered in an idiosyncratic country then as the firm expands to serve lower-ranked markets the 'foreignness' of the markets that it encounters increases sharply and so b is high; conversely, if the firm is headquartered in a cosmopolitan country (e.g. a country with a large overseas empire) then foreignness increases only slowly as sales expand and so b is low.

It is assumed that the proportion of supply produced by the firm itself, either by exporting or FDI, is a fixed proportion, h, of the total amount supplied; conversely the amount of output supplied through licensing and other arm's length contractual relationships is $1 - h$. The assumption that h is fixed implies that the propensity to internalise is independent of distance; this is a strong assumption, because in practice the propensity may increase with distance, but relaxing the assumption would complicate the model.

It is also assumed that under internalisation, a proportion j of all output produced by the firm is produced in the home country, either for its home market or for export. Thus a proportion $1 - j$ of output is produced overseas (typically through FDI).

Overall coordination costs are related to the scale and scope of the firm's operations, considered as a whole. They depend on the output of the firm, which depends upon both sales, x, and internalisation, h. The larger the amount of production controlled by the firm, hx, the larger is the unit cost of supply to each market. Cost also depends on whether operations are concentrated in the home country; the larger the share of home country production, j, the

lower is the unit cost of coordinating supply. Let the marginal cost of overall coordination be e; then e is an increasing function of h and x and a decreasing function of j. Let the importance of overall coordination costs be measured by the parameter g; then $e = ghx/j$.

Adjusting h and j to minimise overall coordination cost may bias individual market supply decisions, as noted above. In particular, there may be pressure to license where internalisation would otherwise be preferred, and to produce at home where overseas production would be preferred. These biases will increase the cost of supplying individual markets; thus the marginal cost of supply to any market will increase by a factor $b = b(h, j)$.

Finally, the model includes costs of R&D. R&D incurs both set up costs and recurrent costs. The focus is on recurrent costs, which are fixed costs independent of the level of sales. It is assumed that rival technologies improve over time, and that to maintain its lead the firm must invest continuously in R&D. The quality of rival products improves at a proportional rate s, whilst their costs of production remain unchanged. To remain ahead of the competition and defend its premium price q, the firm must incur a recurrent cost of R&D, which increases with respect to both q and s.

In a steady state, the revenue of the firm is

$$R = qx \qquad (1.1)$$

Marginal costs comprise a fixed component, $a + b(h, j)$, and a variable component, $(d + (gh/j))x$. Total variable costs are the integral of marginal cost from zero to x:

$$V = (a + b(h, j))x + ((d + (gh/j))/2)x^2 \qquad (1.2)$$

Fixed costs are

$$F = f(q, s) \qquad (1.3)$$

Profit is

$$\Pi = R - V - F \qquad (2)$$

Suppose that the total size of the global market, X, is not binding, i.e. $x < X$. The first-order condition for a maximum of profit (2) with respect to x, conditional on h, j and q, is

$$q = (a + b(h, j)) + (d + (gh/j))x$$

whence

$$x = (q - a - b(h, j))/(d + (gh/j)) \qquad (3)$$

The sales equation (3) shows that the total volume of sales is greater, the greater the profit margin on production, the smaller the distance factor, d, the smaller the overall costs of coordination, g, the smaller the degree of internalisation, h, and the greater the degree of home production, j. The profit margin is higher, the higher the quality premium, q, the higher labour productivity (the lower is a), and the lower the potential for the distortion of market

supply decisions through excessive externalisation or excessive domestication of production (lower *b*).

The first order condition for internalisation, *h*, conditional on *j*, *q*, *x*, implies:

$$\partial b/\partial h = -gx/2j \qquad (4)$$

Equation (4) asserts that externalisation bias should be accepted in individual market supply decisions to a greater degree when overall coordination costs are significant (high *g*); sales are large because many countries are served (high *x*) and production is strongly internationalised (low *j*); high internationalisation of production increases complexity and warrants a compensating reduction in internalisation.

The first order condition for home-orientation of production, *j*, conditional on *h*, *q*, *x*, implies:

$$\partial b/\partial j = -ghx/2j^2 \qquad (5)$$

Equation (5) asserts that home production bias is more acceptable when overall coordination costs are significant (high *g*) and sales are large (high *x*); also when internalisation is high (so that complexity needs to be reduced) and when home production would otherwise be very small.

The first order condition for optimal quality, *q*, conditional on *h*, *j* and *x*, implies that R&D expenditure should be matched to the level of sales:

$$\partial f/\partial q = x \qquad (6)$$

The left-hand side of equation (6) measures the marginal cost of productivity improvement through R&D, while the left-hand side shows that the marginal benefit of productivity improvement is equal to the volume of sales. This results reflects the fact that knowledge is a public good within the firm and that improvements in quality (unmatched by rivals) are directly reflected in the price.

Under suitable conditions the four first-order conditions (3) – (6) determine the four endogenous variables *x*, *h* , *j*, *q*. This gives a full solution of the general form:

$$x = (a, d, g, s; b, f) \qquad (7.1)$$

$$h = (a, d, g, s; b, f) \qquad (7.2)$$

$$j = (a, d, g, s; b) \qquad (7.3)$$

$$q = (a, d, g, s; b) \qquad (7.4)$$

The independent variables are productivity *a*, which governs the basic profitability of production; the distance factor, *d*, which measures the difficulty of extending the range of markets served without substantially increasing the marginal cost of supply; *g*, which measures the importance of overall coordination costs; and *s*, that measures the quality

improvements achieved by rival firms. In addition, the size of the *b*-effect needs to be taken into account

Predicted impacts are summarised in Table A.1.

Optimal size, as measured by the volume of sales, *x*, is greater the higher the level of productivity (the lower is *a*), the lower the 'tyranny of distance', *d*, the lower the overall costs of coordination, *g*, the slower the innovation by rivals, *s*, and the greater the costs of biasing individual market sourcing decisions towards licensing and home production, *b*.

Some impacts are direct and others indirect. Productivity, tyranny of distance, overall costs of coordination, and the costs of bias all impact directly on sales. Innovation by rivals, however, has an indirect effect; it increases the cost of maintaining the quality advantage, thereby reducing the optimal price premium, which in turn reduces the marginal revenue from a sale.

Indirect impacts are even more important for other aspects of strategy. Many of the impacts on internalisation, geographical concentration and the quality premium are mediated by impacts on sales.

Some parameters have more indirect effects than others. The effects of innovation by rivals, *s*, are particularly subtle, as indicated in row 4 of the table. An increase in *s* makes it more expensive to maintain a substantial quality lead and so directly reduces *q*. A reduction in *q* reduces global sales, *x*, which in turn reduces the pressure to contain overall coordination costs. This reduces the pressure to minimise internalisation, *h*, and so indirectly increases *h*; it also reduces the pressure to maximise *j*, and so indirectly reduces it. The effects on *x*, *h* and *j* are therefore all indirect.

Table A.1 Impacts of exogenous factors on the equilibrium values of sales, internationalisation, home production and quality improvement

	x	***h***	***j***	***q***
a	-	+	-	-
d	-	+	-	-
g	-	+	-	-
s	-	+	-	-
b	-	+	-	-