

Coase and international business: the origin and development of internalisation theory

Article

Accepted Version

Casson, M. ORCID: <https://orcid.org/0000-0003-2907-6538>
(2014) Coase and international business: the origin and development of internalisation theory. *Managerial and Decision Economics*, 36 (1). pp. 55-66. ISSN 1099-1468 doi: <https://doi.org/10.1002/mde.2706> Available at <https://centaur.reading.ac.uk/38244/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1002/mde.2706>

Publisher: Wiley

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Coase and International Business:

The origin and development of internalisation theory

Mark Casson

Keywords: INTERNALISATION; COASE; MULTINATIONAL ENTERPRISE;
INTERNATIONAL BUSINESS; KNOWLEDGE

Address:

Mark Casson

Department of Economics

University of Reading

PO Box 218

Reading RG6 6AH, UK

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

Tel: (0) (44) 118 378 8227

Word count: 8267

Date: 10 August 2014

Abstract

The internalisation theory of the multinational enterprise is a significant intellectual legacy of Ronald Coase. US direct investment in Europe became highly political in the 1960s, and neoclassical trade theory had no explanation. A theory of the multi-plant enterprise was required, and internalisation theory filled this gap. Using Coasian economics to explain the ownership of production plants, and the geography of trade to explain their location, internalisation theory offered a comprehensive account of MNEs and their role in the international economy. This paper outlines the development of the theory, explains the Coasian contribution, and examines in detail the early work of Hymer, McManus and Buckley and Casson. It then reviews the current state of internalisation theory and suggests some future developments.

Introduction

This paper explores the intellectual legacy of Ronald Coase in the field of international business (IB) studies. His principal legacy is the internalisation theory of the multinational enterprise (MNE). Initially considered as a branch of applied economics, internalisation theory is now a core element of the modern field of IB studies. 'Internalisation' refers to the fact that MNEs replace external markets in proprietary knowledge and semi-processed products with internal managerial coordination.

Internalisation theory was developed in the 1970s to explain the growth of multinational enterprises and the spread of foreign direct investment (Buckley and Casson, 2009a). It provides an explanation of why multinational business activity is concentrated in innovative knowledge-intensive industries, and in industries where the quality of components and raw materials is difficult to measure and control. Before internalisation theory it was widely believed that multinational firms transferred capital to a foreign country (Kemp, 1961; MacDougall, 1960; Penrose, 1956), while afterwards it was recognised that it is mainly knowledge that they transfer; capital is transferred, if at all, mainly to protect the knowledge and to appropriate profit from its exploitation abroad (Casson, 1979).

Internalization theory focuses on imperfections in intermediate product markets (Rugman, 1981). Two main kinds of intermediate product are distinguished: knowledge flows linking research and development (R&D) to production, and flows of components and raw materials from an upstream production facility to a downstream one. Most applications of the theory focus on knowledge flow (Dunning and Lundan, 2008). Proprietary knowledge is easy to copy when intellectual property rights such as patents and trademarks are weak. Firms therefore protect their knowledge through secrecy. Instead of licensing their knowledge to independent local producers, firms exploit it themselves in their own production facilities. In effect, they internalise the market in knowledge within the firm. Internalisation leads to multinationality because knowledge is a public good (Buckley and Casson, 1976). Development of a new technology is concentrated in a single R&D facility, and the knowledge is transferred to subsidiaries abroad. The firm becomes the owner of production plants in different countries and therefore (by definition) a multinational.

Firms do not always internalise markets: internalisation occurs only when the benefits perceived by the firm exceed the costs. When internalisation leads to foreign investment the firm may incur political risks, and also commercial risks due to its unfamiliarity with the foreign environment. These are known as costs of doing business abroad (Hymer, 1976) arising from the 'liability of foreignness' (Zaheer, 1995). When the costs of doing business abroad are high a firm may license or subcontract production to an independent foreign firm; or it may produce at home and export to the country instead. Firms without special knowledge may become multinational if they need to internalise supplies of components or raw materials in order to guarantee quality or continuity of supply, or if there are tax advantages from transfer pricing.

Other variants of internalisation theory have emerged. Magee's (1977) 'appropriability theory' is similar to Buckley and Casson in some respects, and like Rugman (1981) emphasises applications to trade policy. Hennart (1982), meanwhile, emphasised the role of authority relations within the firm, and subsequently extended his approach to analyse headquarters-subsidiary relations.

Dunning (1977) employed internalisation theory as a component of his eclectic paradigm or OLI model. In the OLI model, knowledge was identified as an important source of 'ownership advantage'. Dunning claimed that possession of an ownership advantage was a necessary condition for a firm to become multinational. Internalisation theorists disagreed, however; they argued that if quality control and transfer pricing are sufficient for multinationality then ownership advantage cannot be necessary. Dunning replied that the ability to internalise could also be classified as an ownership advantage, but thereby exposed the concept of ownership advantage to the criticism that it was merely tautological (Williams, 1997).

Internalisation theory is related to transaction cost theory through common dependence on Coase's (1937) essay on the nature of the firm. The two theories are not the same however. Internalisation theory focuses on links between R&D and production whereas transaction cost theory focuses on links between one production facility and another. Transaction cost theory typically attributes market imperfections to bounded rationality and 'lock in', whilst internalisation theory emphasises asymmetric information and weaknesses in property rights (Williamson, 1975, 1985). Transaction cost theory is typically applied in a domestic context, whereas internalisation theory was developed specifically for an international context. Whilst some of these differences may be partly semantic (e.g. bounded rationality can subsume asymmetric information), most of them are substantive, and they explain why the two literatures have remained distinct.

Prior to internalisation theory, the study of IB was largely focused on the foreign business environment, and in particular the economic, financial, political and cultural dimensions of doing business abroad (Rugman and Collinson, 2012). Internalisation theory provided a theory of the MNE, and thus augmented the IB field by demonstrating the interaction between the external environment and the internal knowledge flows between MNE parent firms and their subsidiaries. This interaction between external country-specific factors and internal firm-specific factors provides a basic template for contemporary analysis of IB strategy (Rugman and Verbeke, 1992, 2003; Rugman, Verbeke and Nguyen, 2011).

The view that multinationals transfer technology and not capital provided a major boost to the process of globalisation. The United Nations Conference on Trade and Development (UNCTAD) was strongly influenced by internalisation theory and the eclectic paradigm (UNCTAD, 1991-2014). It persuaded political leaders to encourage inward investment as a source of the new technologies required for economic development, thereby reversing their previous attitudes. Multinational profits were increasingly viewed as payments for knowledge and technology rather than as interest paid on capital, and foreign ownership

became accepted, in certain cases, as a necessary safeguard for foreign investors' intellectual property. This change in political attitudes accelerated the pace of globalization.

'The nature of the firm' as a foundation for international business theory

Internalisation theory is based directly on Coase (1937); unlike transaction costs economics, it is not based on a synthesis of Coase, Simon, Commons, North and others (Williamson, 1975). It is therefore useful to preface a discussion of internalisation theory with an examination of ideas from Coase that have been particularly influential in IB theory.

Coase believed that he had discovered a gap in economic theory, 'between the assumptions (made for some purposes) that resources are allocated by means of the price mechanism and the assumption (made for other purposes) that this allocation is dependent on the entrepreneur-coordinator' (p. 389). It was possible, he claimed, to bridge this gap using a novel concept: 'The main reason why it is profitable to establish a firm would seem to be that there is a cost of using the price mechanism. The most obvious cost of 'organising' production through the price mechanism is that of discovering what the relevant prices are. This cost may be reduced but will not be eliminated by the emergence of specialists who sell this information.' (p. 390).

Coase argued that the distinguishing feature of the firm is the supersession of the price mechanism. Within the boundaries of the firm, planning rather than price negotiation prevails. This approach, he claimed, is realistic 'in that it corresponds to what is meant by a firm in the real world', and is 'tractable by two of the most powerful instruments of economic analysis developed by Marshall, the idea of the margin and that of substitution, together giving the idea of substitution at a margin' (pp.386-7). The margin comes in to play when discussing the size of the firm. 'A firm becomes larger as additional transactions (which could be exchange transactions co-ordinated through the price mechanism) are organised by the entrepreneur...' (p.393). Coase analysed the margin in terms of benefits and costs. Beyond a certain point, the marginal cost of organising an additional transaction increases because the entrepreneur is more likely to make mistakes, although this may be mitigated by economies of bulk-buying and the exercise of monopsony power.

Coase suggested that firms of similar managerial capability would be of similar size. If two firms, one large and one small, are both attempting to take over a firm with which they trade then, other things being equal, the marginal cost will be higher for the larger firm, and so the smaller firm will make the higher offer and secure the acquisition; as it result the smaller firm will grow relative to the larger firm, thereby tending to equalise their sizes (pp. 395-6).

Although the firm supersedes the price mechanism, prices still ultimately regulate the economy. Under socialism the state determines the scope of the market system, but under capitalism, the scope of the market system is determined by individual actions, and the cost of using markets determines whether the formation of firms is profitable. Thus 'in a competitive system, there is an 'optimum' amount of planning!' (p. 389, fn. 3).

Coase's basic unit of analysis is the national economy. Following Adam Smith (1776), he assumes that within the national economy there is a division of labour (p. 389). This creates a range of specialised activities, many of which are connected by flows of intermediate products. The coordination of a division of labour does not necessarily call for planning (p. 398). In a market economy the prices of intermediate products coordinate flows. The relationship between intermediate products and final products is crucial. If demand switches from one final product to another then changes in final product prices will induce changes in intermediate product prices. This will lead to substitution between intermediate products and thereby adjust supply.

The division of labour can take place both *within* plants and *between* them, although Coase does not always clearly distinguish between the two. The distinction is important for IB theory because it focuses almost exclusively on the inter-plant division of labour, and specifically on international flows of intermediate products between plants located in different countries.

Coase analyses the intra-plant division of labour by focusing on the open-ended nature of employment contracts, which allows an entrepreneur to reallocate a worker to different tasks as circumstances change. The entrepreneur substitutes a single long-term contract for a multiplicity of short-term contracts which would be costly to negotiate (pp. 391-2). For similar reasons entrepreneurs buy capital equipment outright or lease it on a long terms basis.

Coase analyses the inter-plant division of labour in terms of vertical integration. He argues that vertical integration is encouraged by a desire to maintain long-run continuity of supply, to obtain flexibility through informal coordination, to compensate for lack of product standardisation, and to avoid sales taxation, rationing or price control imposed by government on inter-firm sales (p. 393).

Critical appraisal of Coase's analysis

While internalisation theory is directly inspired by Coase, it draws only selectively on his work. There are some aspects of Coase's work, such as his repeated attacks on Knight (1921) (pp. 392, 394, 398-401) which seem mis-judged in the light of later research. In other respects, Coase fails to set out clearly the context in which his analysis is to be construed. In particular, he fails to analyse complementarities, which are such an important feature of multi-stage production processes.

Coase is so keen to emphasise the role of prices in coordinating inter-related substitutions that he ignores complementarities altogether. For example, when discussing intra-plant coordination he ignores the role of teamwork, as emphasised by Alchian and Demsetz (1972), and focuses exclusively on internal substitution, e.g. workmen moving between departments (p. 387) and the re-allocation of space within a department store (p. 388).

By ignoring complementarity, Coase ignores the possibility that a firm exists simply to combine productive inputs in appropriate proportions, as in neoclassical theory. This is not a

fatal flaw, but it may explain why his work did not achieve much impact until the limitations of neoclassical theory became so evident in the 1970s.

The principle of complementarity is important in analysing multi-stage production, which in turn is the basis for theories of vertical integration (Warren-Boulton, 1978). In multi-stage production, the amount of input from the previous stage is often fixed in proportion to the output supplied to the next stage. The principle extends to hierarchical systems involving different tiers of production. A conventional motor car, for example, is assembled from four wheels, two headlights, one engine, and so on. These components are in turn assembled from smaller components, and so on. Coordinating flows of intermediate products involves matching outputs at each stage or tier. It is sometimes possible to adjust proportions by managing wastage at each stage, but the scope is usually limited.

Given fixed proportions throughout the production process, long-run substitution between components is impossible so long as the technology remains unchanged. Substitution decisions arise mainly in response to short-run shocks (Wadeson, 2013). Suppose, for example, that the supply of wheels to motor manufacture is disrupted; then inventories of wheels may be drawn down (to be replaced later), or workers may be switched from the motor car production line to the manufacture of wheels. If the wheels are sourced from overseas then labour may be put on short time in the domestic assembly plant and on overtime in the overseas wheel plant until the shortage has been resolved. The key issue is inter-temporal substitution in response to uncertain conditions. The context is inventory control in a fixed proportions process. In this context imperfection in forward markets, and in insurance markets, encourage internalisation (Arrow, 1975).

Long-run substitution under fixed proportions occurs only through substitution between technologies. In this context, internalisation arises when the same firm controls two or more alternative technologies for producing the same product. In simple neoclassical theory, technological substitution is fully internalised because technology is not only a public good, that can be shared, but also a common good, that is freely accessible to all; the neoclassical firm exists simply to allow the entrepreneur to combine inputs in the optimal proportion. But Coase does not discuss choice of technology at all.

The discussion of technology leads on to another problem with Coase. This concerns the knowledge and know-how used by the firm. Knight postulates that entrepreneurs exploit superior knowledge that allows them to forecast future demand better than other people. Coase criticises Knight on the grounds that such knowledge will not be internalised but will be sold through external markets as advice. He appears to suggest that, although markets in ordinary intermediate products may be internalised, markets in knowledge will not.

‘... the fact that certain people have better judgement or better knowledge does not mean that they can only get an income from it by themselves actively taking part in production. They can sell advice or knowledge. Every business buys the services of a host of advisers. We can imagine a system where all advice or knowledge was bought as required.’ (pp. 400-1)

He then goes on to mention the use of contracts for the sale of knowledge. If Coase's criticism of Knight is interpreted literally, it seems that he did not consider that internal exploitation of superior knowledge was a plausible rationale for the firm.

If this interpretation of Coase is correct, then his stance appears quite ironic in view of later developments. Interest in Coase's work was re-kindled in the 1970s specifically by a desire to understand the logic of the multi-plant firm. An obvious rationale for the multi-plant firm is the sharing of proprietary knowledge between plants through internal markets. Coase's negative stance on the internalisation of knowledge may explain why he never personally applied his ideas to the MNE. But other scholars were on hand to make the connections between internalisation, knowledge and the MNE, as this paper will now show. These scholars drew on Coase selectively, choosing the bits they found useful and ignoring the bits they did not.

Early IB applications of Coase: Hymer and McManus

Before Coase's ideas had been adopted, economists explained foreign investment in terms of capital flow, as indicated above. In neoclassical theory, foreign investment was analysed by relaxing the assumption of immobile factors of production that underpinned conventional trade theory. With immobile capital, rates of interest and rates of capital can differ between countries. When capital becomes mobile, it will flow from low-profit countries to high-profit countries until the discrepancy in rates of return has been eliminated, at which point an equilibrium international distribution of capital will prevail. Capital only flows so long as adjustment is required, and it flows in only one direction at a time.

Neoclassical theory analysed capital flow as if it were a financial flow, but the analogy did not work well. Evidence on MNE investments indicated that capital tended to flow from high profit countries to low-profit countries rather than the other way round. Capital also flowed in both directions at once, and the balance of capital flow varied between industries.

The earliest known application of Coase's thinking to IB is by Stephen Hymer (1976), whose dissertation, awarded in 1960, was only published sixteen years later. His debt to Coase is only implicit in his thesis, but is explicit in his later work:

'The great merit of Coase's work is that it does not confuse scale economies at the level of the industry and scale economies at the level of the firm. The former have a technological origin and result from the division of labour. [...] As for scale economies achieved by the firm, they result from organisation. The firm is an organism that coordinates activities, and the question of scale economies can be reduced to a problem of efficiency of information. [...] When coordination is effected by the firm, each unit transmits the information through the internal network of communication to the other units and to a 'central brain'...' (Hymer, 1990, p. 10)

Hymer was the first writer to set out a coherent alternative to the neoclassical view of foreign investment (Dunning and Rugman, 1985; Dunning, 2003; Buckley, 2006; Dunning and Pitelis, 2008). He clearly distinguished between foreign direct investment (FDI), which gives

a firm headquartered in one country control of assets located in some other country, and foreign indirect (or 'portfolio') investment which typically gives an individual the right to interest on a bond, or income from a lease. He distinguished two main motives for FDI. One was concerned with cartelisation or collusion: bringing colluding plants under common ownership makes collusion easier to organise. The other was concerned with the exploitation of a firm-specific advantage, such as a unique product design. It may be difficult to rent or sell the advantage, and so the firm exploits the advantage entirely by itself:

'The control of the foreign enterprise is desired in order to remove competition between that foreign enterprise and enterprises in other countries. Or the control is desired in order to appropriate fully the returns on certain skills and abilities.'

'It frequently happens that enterprises compete with each other because they sell in the same market or because some of the firms sell to other firms. If the markets are imperfect, that is, if there is horizontal or bilateral monopoly or oligopoly, some form of collusion will be profitable. One form of collusion is to have the various enterprises owned and controlled by one firm. This is one motivation for firms to control enterprises in foreign countries.'

'The other main motivation is that firms are very unequal in their ability to operate in a particular industry. A firm with advantages over other firms in the production of a particular product may find it profitable to undertake the production of this product in a foreign country as well. The firm could also rent or sell its skill rather than undertake itself the foreign production. Which method it chooses depends largely on the degree of imperfection in the market for this skill. If the market is imperfect the owner may not be able to appropriate fully the returns to the ability unless he controls its use.' (Hymer, 1976, pp. 26-26)

These two motives are in fact connected, although Hymer does not emphasise the point. This is because the advantage discussed in connection with the second motive may be the source of the monopoly power in the first motive. When the same advantage is exploited in different countries, local producers share the advantage and gain by collusion.

This discussion prefaces a more detailed discussion of licensing later in the thesis. But in contrast to modern discussions of licensing, which emphasise property rights and asymmetric information, Hymer focuses exclusively on problems relating to market structure, and in particular to deviations from perfect competition:

'Consider a firm deciding whether to license its advantage in a particular industry. Suppose that in this particular industry considerations are such that there will be only one or, at most, a few firms in each market. If the firm attempts to license its product, it will encounter two problems which could be solved by owning the operating enterprises. In the first place, it will be selling its advantage to enterprises which possess monopoly power in their markets. A sequential monopoly problem is therefore involved. Integration could then increase joint profits. The second problem of licensing arises from the difficulty of controlling price and output. To achieve maximum profits, the firm must specify the precise use to each firm, and this is not always possible under anti-trust laws. Alternatively it could let the firms compete,

but this may result in a loss of profits. If the firm which possesses the advantage does not license but instead undertakes the operations itself, there is less difficulty in achieving maximum profits.’ (Hymer, 1976, pp. 48-49)

Hymer’s emphasis on the role of monopoly is reflected in his concluding remark:

‘It is important to stress that both of these reasons for not licensing disappear if there are many buyers of the advantage’ (Hymer, 1976, p.49).

This is incorrect, because if there were many buyers it would still pay the licensor to select a subset of the potential buyers and award each an exclusive share of the market.

Internalisation then could then be used to police market shares and prevent unintended competition between producers.

Hymer’s emphasis on monopoly power led him to later embrace a Marxist interpretation of the MNE (Hymer, 1972). Together with his premature death, this may have diminished the later influence of his early work; for some years the only published account of his thesis was a summary in Kindleberger (1969), which omitted his discussion of the licensing decision.

A fellow-Canadian, John McManus built on Hymer’s work using a more philosophical approach. He began by noting, quite correctly, that

‘...the term ‘international’ plays no substantive role in the analysis... We would equally well be discussing the interstate firm, the intercountry firm, or any organization that centralises control of productive activities in separate locations.’ (McManus, 1973, p. 33)

He then sets up the economic problem in very general terms:

‘If the actions of a set of producers in different countries were completely independent, there would be no gains to be made by centralising the control of their operations within an international firm. But if a set of autonomous producers in different countries were interdependent, there would be circumstances in which the reallocation of their resources would yield a higher total wealth for the group as a whole.’ (McManus, 1973, pp. 38-39)

These circumstances involve a ‘cost of transacting’, which tends to be high when quantities are difficult to measure. They include costs of enforcing contracts and adapting long-term contracts to evolving conditions.

‘Given transacting costs, there will exist ... dimensions of interdependence that will not be constrained efficiently, so that some actions will be taken by individual producers without their taking into account, at the margin, all the effects that their actions have on the values of the resources of other members of the set. [...] The only way in which the firm can obtain changes in the actions of the set of producers is by changing the constraints facing individual members. [...] The set of producers who choose to combine their resources within a firm will mutually agree to an elimination or reduction of non-discretionary pecuniary constraints as a determinant of their individual decisions so that management control of their actions can be exercised’ (McManus, 1973, p. 45)

McManus discusses three Canadian case studies, involving textile producers, Holiday Inns and automobile plants, the most interesting being his analysis of franchising by the Holiday Inn chain. He concludes by examining the inter-industry distribution of US investment in Canada, arguing that constraints on property rights rather than tariff barriers explain the desire of US MNEs to exercise control over their Canadian operations.

Internalisation theory: the systems view of Buckley and Casson

The most highly cited statement of internalisation theory in IB is Buckley and Casson (1976). They adopt a 'systems view' of the global economy (Buckley and Casson, 1998, 2009b). Following Babbage (1832), they portray, in stylised form, the division of labour in a knowledge-based economy. Coasian theory is then applied to analyse the coordination of this economy.

The focus is a particular industry, and a specific technology within that industry. In this context 'technology' can be construed generally to include any knowledge-based asset to which some firm has exclusive access. There are three main activities in the industry: production, distribution and R&D. These activities are linked to each other by flows of intermediate products. Distribution is linked to customers by a flow of final product. With three corporate activities and customers who consume the product there are four activities in the industrial system. Figure 1 illustrates the linkages between them. The linkages involve both intangible flows of knowledge, illustrated by grey lines, and tangible flows of goods and services, shown as black lines. Knowledge flow is two-way, and product flow one way, as indicated by the arrows (Casson, 1995, 2000).

R&D generates proprietary knowledge, which is embodied in a new product or production process. Knowledge flows from R&D to production, and experience feeds back from production to R&D, creating a learning loop. Marketing requires a link between R&D and distribution; the technical claims for the product need to be explained to sales representatives, and the representatives need to feed back information regarding customer satisfaction. Intermediate product flows from production to distribution; this represents final product flowing from factory to warehouse. The final product then flows on to the customer.

Altogether, there are four linkages that need to be coordinated. The linkage involving the final product is always external to the firm. The other three linkages can be either internal or external. There are five possible ownership structures, as indicated in Table 1.

A more sophisticated analysis would refine the division of labour by distinguishing between basic research and applied development, upstream and downstream production, and between marketing and the logistics of distribution.

The unit of analysis is the global economy rather than the national economy, with the global economy being sub-divided into national jurisdictions by boundaries whose positions may have little economic logic. Each country is a potential location for production, distribution and R&D. Each activity takes place in a particular type of facility; production in a plant,

distribution in a warehouse and R&D in a laboratory. Each facility has three identifiers: the function it performs, the firm that owns it, and the country where it is located.

Knowledge is a public good; once it has been produced it is easy to share. As a result there is just a single R&D facility for each technology. The cost of transferring knowledge to a foreign plant may be high when cultural differences make communication difficult, but the cost of transfer is lower than the cost of reinventing the knowledge at the local plant.

Production, by contrast, affords only limited economies of scale; it is therefore economic to replicate production in different locations when trade costs (transport costs, tariffs and non-tariff barriers) are high. It is assumed that distribution is normally located close the final consumer, especially when after-sales support is important. For a given technology in a global industry, therefore, there will be one R&D facility, several production plants and many distribution facilities. This may change over time, e.g. if transport costs fall and communications improve, the number of production and distribution facilities may shrink as operations are rationalised into fewer larger units.

R&D is privately financed and, in the absence of subsidies, its costs must be recovered from sales of the final product (Arrow, 1962). It is assumed that R&D has been financed by a firm that owns the knowledge that has been generated. This firm can either license the use of the knowledge, or exploit it itself, in which case it internalises the link between production and R&D.

Buckley and Casson (1976) formalise Hymer's analysis of the licensing decision. When intellectual property rights (patents, trademarks) are weak, licensees may imitate or covertly improve the technology. When licensing different firms to serve different market areas, the licensor may find it difficult to control competition between licensees serving adjacent areas; indeed, the threat of such competition may induce licensees to under-bid for licenses in the first place. They may also under-bid because they are uncertain about the quality of the technology they are being offered. If the licensor provides too much information in advance to reassure the bidders then they may inadvertently give away the very knowledge that they wish to charge for (the so-called buyer-uncertainty problem).

Internalisation of production and distribution may also be advantageous. Missing forward markets may encourage the centralisation of inventory policy. Tax differentials may encourage the 'transfer pricing' of wholesale product. Quality control may be improved, and inter-related investments in production and distribution better coordinated.

Buckley and Casson also clarified the role of market structure by distinguishing its role in intermediate product markets from its role in final product markets. Consider final product markets first. When firms exploit proprietary knowledge there is a natural tendency for monopoly to prevail. When R&D is privately financed, monopoly power provides an important mechanism by which an innovative firm recovers the sunk fixed costs of its R&D. To minimise the adverse effects of monopoly power, it is helpful if a monopolist can discriminate in pricing by segmenting the final product market and charging each segment

according to their willing to pay. Even where there are competing product varieties, each producer will have a degree of monopoly power because the different varieties will not be perfect substitutes for each other. Such monopoly rents promote variety and, indirectly, consumer choice. The regulation of monopoly will typically reflect government priorities. Governments that wish to stimulate innovation through private enterprise will be more tolerant of monopoly than governments that wish to maximise the social benefits of innovations that have already been carried out.

The situation is different in intermediate product markets. Unique or differentiated final products do not necessarily require unique differentiated intermediate products, e.g. a novel design of final product may be generated by recombining standard components in a novel way. Market structure problems in intermediate products markets are most likely to be generated by economies of scale or large sunk costs at certain stages of production, creating unilateral or bilateral market power. This may in turn distort prices, not only at the stages directly affected, but throughout the entire production chain. Internalisation can eliminate such distortions, permitting an internalising firm to appropriate the efficiency gains.

Internalisation theory achieved early success in explaining a number of empirical paradoxes. The costs of licensing explained why FDI was concentrated in high-technology and knowledge-intensive industries. The link between knowledge and R&D explained why FDI flowed mainly from research-intensive economies such as the US. It flowed to other fairly rich but less research-intensive countries, e.g. in Western Europe, because their consumers could afford the sophisticated products generated by the new technologies (Vernon, 1966). FDI flowed in both directions because in any given industry there were non-US rivals to US companies that found it profitable to invest in the US.

Internalisation theory also led to novel forecasts which turned out to be correct. Marxist political economists in the 1970s claimed that global capital was becoming concentrated, and that very soon almost every industry would be dominated by giant managerial corporations headquartered in the US (Rowthorn, 1972). Internalisation theorists, however, argued that innovative firms would increasingly rely on arm's length contractual arrangements. They would keep tight control of their knowledge but delegate production and distribution increasingly to firms with better local knowledge.

Internalisation theorists took this view because they recognised the costs as well as the benefits of internalisation. As Coase suggested, these costs were likely to increase with the size of firm. The logic of knowledge as a public good suggested that research-intensive firms would wish to expand their markets geographically. In order to manage the scope and complexity of their operations as they grew, therefore, licensing, franchising and subcontracting would all become increasingly attractive options. These 'arm's length' contractual arrangements involving external markets would allow innovative firms to serve a global market without requiring local knowledge; furthermore, the political risks of doing business in certain countries could be reduced by using local partners instead.

Subcontracting and franchising were more attractive than licensing where knowledge was very valuable, because the firm lost direct control of just one activity (production and distribution respectively) whereas under licensing it lost direct control of both. This analysis also turned out to be correct.

The years of consolidation

From the 1980s onwards the IB literature has proliferated, and the range of topics addressed has increased significantly. More attention has been focused on joint ventures and strategic alliances (Buckley and Casson, 1996). The dynamics of internationalisation have been considered more carefully, using a 'real option' approach to FDI in which firms sequence market entry to maximise learning economies (Buckley, Casson and Gulamhussen, 2002; Rugman and Li, 2004).

The generality of the theory has also been put to the test. Could it explain outward FDI from newly industrialising countries (NICs) as well as from mature Western economies like the US? This controversy began with Kojima's (1978) claims that Japanese motives for FDI were different from Western motives, and has continued with calls for new theories to explain FDI from 'emerging market economies' such as China (Ramamurti, 2012). While these debates have revealed some limitations of early analysis (playing down the role of natural resources, and the impact of government subsidies) they have never effectively challenged the Coasian principles on which internalisation theory is based.

The range of industries studied has also increased. Initially the focus was on high-technology industries such as pharmaceuticals and motor vehicles. The question arose as to whether internalisation was just a theory of manufacturing FDI, or whether it encompassed FDI in services as well (Boddewyn, Halbrich and Perry, 1986). It is certainly true that much of the early language in which the theory was expressed implicitly assumed a manufacturing context, but it was quickly established that it applied to the service sector too. The theory was refined to include the prominent role of partnerships in knowledge-intensive professional service industries (Enderwick, 1988).

There have been some negative developments too. Internalisation is essentially a theory of the boundaries of firms, and it needs to be combined with a theory of location in order to provide a comprehensive account of MNEs and FDI. The interface between internalisation and location economics has received relatively little attention, however. Links with trade theory have been developed by Markusen (1995, 2002), but IB scholars have taken little notice of this work. Economic geographers such as Dicken (2010) have also taken a strong interest in FDI but, as McCann and Iammarino (2013) note, much more remains to be done in this area.

Another negative development has been a controversy with resource-based theorists, which has generated considerable heat and very little light (Kogut and Zander, 1992; Tallman, 2003). The critics contend that the emphasis of internalisation theory on ownership, property rights and contracts distracts attention from the factors that really matter in explaining MNEs

and FDI, which they maintain are primarily social, cultural and organisational. In effect, they claim that the firm should be considered, not as a nexus of individual contracts, or even of relationships, but rather as a self-perpetuating organic entity. They do not directly attack the economic logic of internalisation, but rather the methodological individualism on which the theory is based. They claim that assuming knowledge to be a public good denies the existence of communication costs, and that assuming that knowledge can be licensed denies that sharing knowledge is a social activity. These criticisms ignore the fact that the social factors have always been recognised in internalisation theory. They are reflected in the structures of revenues and costs that influence the behaviour of a profit-maximising firm. These include marketing costs, technology transfer costs, and the costs of doing business abroad, all of which play a prominent role in the theory. Economic analysis of the IB system requires some simplifications, however, and it is quite appropriate to abstract from very detailed aspects of social interaction when constructing a rigorous account of the nature of the firm.

A continuing research agenda

Perhaps because of these distractions, some of the basic insights of Coasian economics, set out in the 1970s literature, have not been followed up. From a Coasian perspective, a global industry comprises a set of facilities – laboratories, production plants, warehouses, etc. – which transform factor services into final products through inter-dependent multi-stage processes. Because of barriers to trade, different national markets are often served using different sets of plants and distribution facilities. The number and location of each of these plants is, in principle, endogenous. When transport costs, communication costs and political relations change, the configuration of plants will change as well.

But plants are characterised, not only by their function and location, but by their ownership. Who will own which plants? This is where the logic of internalisation is crucial. The costs of doing business abroad tend to favour local ownership. If external markets were all perfect it is highly likely that all plants would be locally owned, as this would maximise security of ownership and simplify the monitoring of activity. But when intermediate product markets are imperfect it may pay to internalise them by bringing the relevant facilities under common ownership and control. This is the origin of the MNE. The direction of FDI depends on where the MNE is headquartered.

In a knowledge-intensive industry close monitoring of R&D may be a strategic imperative, and so the ownership of production may be effectively dictated by the location of R&D. This raises another issue, however. If R&D dictates the location of headquarters, then the choice of R&D location may be influenced by its role as headquarters location too. Thus R&D may be undertaken in a country that makes a good headquarters because choosing the right location for the headquarters is just as important as choosing the right location for R&D.

In general, the ownership and location of plants is interdependent. Ownership influences location and location influences ownership. But ownership and location may also influence the number of plants. A large market in a relatively poor country may ideally be served by

local distribution, but political risks may encourage the location of the facility in a neighbouring country instead. This facility may then serve the neighbouring country's domestic market too. Thus instead of two distribution centres there is just one.

A consistent application of Coasian principles therefore suggests that in any given global industry, contractual arrangements will lead to a configuration of production in which the number of plants of each type, their ownership and their location are simultaneously optimised through appropriate contractual arrangements. These contractual arrangements will determine the number of independent firms, the scope of each firm's operations and the country in which each is headquartered. The scope of a firm's operations is reflected in the types of facilities it owns, the number of facilities of each type, and the location of each facility.

From this perspective, the set of firms that emerges within the global economy is entirely endogenous. This view contrasts markedly with the predominant view of firms in the modern IB literature, in which the existence of firms is regarded as self-evident, and the analysis is predicated on the prior existence of a given type of firm.

From an internalisation perspective, the boundary of every firm is dictated in part by the boundaries of other firms, so that in equilibrium every firm is satisfied with the boundaries that prevail. The configuration of boundaries reflects a set of environmental factors that include the nature of technology, the geographical distribution of resource endowments, the structure of transport and communication costs, and the state of international relations (Casson and Wadeson, 2013). When these factors change, the boundaries of firms will change in response; new firms will emerge, existing firms will die, and survivors may merge, diversify, or divest. In this way internalisation theory captures the organic nature of the economy in a rigorous way. Internalisation theory is therefore well-equipped to explain systemic changes in the global economy in a way that theories which focus on a single firm can never do.

There is still some way to go before the modern IB literature fulfils the promise and potential of the early literature in the field. IB literature needs to return to its roots, and to the insights of internalisation theory. It needs to re-focus away from the individual firm and its internal social structures towards the place of the firm within its industry, and within the global economic system as a whole.

References

- Alchian, Armen A. and Harold Demsetz (1972) Production, information costs, and economic organization, *American Economic Review*, 62, 777-795
- Arrow, Kenneth J. (1962) Economic welfare and the allocation of resources for invention, in National Bureau of Economic Research, *The Rate and Direction of Inventive Activity*, Princeton, NJ: NBER, 609-626
- Arrow, Kenneth J. (1975) Vertical integration and communication, *Bell Journal of Economics*, 6 (1), 173-183
- Babbage, Charles (1832) *On the Economy of Machinery and Manufactures*, London: Charles Knight
- Boddewyn, Jean, M.B. Halbrich and A. Perry (1986) Service industry multinationals: Conceptualization, measurement, theory, *Journal of International Business Studies*, 17 (3), 41-57
- Buckley, Peter J. (2006) Stephen Hymer: Three phases, one approach? *International Business Review*, 5, 140-147
- Buckley, Peter J. and Mark C. Casson (1976) *The Future of the Multinational Enterprise*, London: Macmillan [Basingstoke, Hants: Palgrave Macmillan, 25th Anniversary ed. 2001], 112pp.
- Buckley, P.J. and Mark Casson (1996) An economic model of joint ventures, *Journal of International Business Studies*, 27 (5), 849-876, reprinted in Buckley, P.J. and Mark Casson *The Multinational Enterprise Revisited: The Essential Buckley and Casson*, Basingstoke: Palgrave Macmillan, 2010, 118-146
- Buckley, P.J. and Mark Casson (1998) Analysing foreign market entry strategies: Extending the Internalization approach, *Journal of International Business Studies*, 29 (3), 539-561, reprinted in Buckley, P.J. and Mark Casson *The Multinational Enterprise Revisited: The Essential Buckley and Casson*, Basingstoke: Palgrave Macmillan, 2010, 177-204
- Buckley, Peter J. and Mark Casson (2009a) 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
- Buckley, Peter J. and Mark C. Casson (2009b) *The Multinational Enterprise Revisited*, Basingstoke, Hants: Palgrave Macmillan
- Buckley, Peter J., Casson, Mark and M. Azzim Gulamhussen (2002) Internationalisation: Real options, knowledge management and the Uppsala approach, in V. Havila, M. Forsgen, & H. Hakansson (eds.), *Critical Perspectives on Internationalisation*, Oxford: Elsevier Science, 229-261
- Casson, Mark (1979) *Alternatives to the Multinational Enterprise*, London: Macmillan

Casson, Mark C. (1995) *The Organization of International Business*, Cheltenham: Edward Elgar

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

Casson, Mark and Nigel S. Wadeson (2013) The economic theory of international supply chains: A systems view, *International Journal of the Economics of Business*, 20 (2), 163-186

Coase, Ronald H. (1937) The nature of the firm, *Economica* (New series), 4, 387-405

Dicken, Peter (2010) *Global Shift: Mapping the Changing Contours of the World Economy*, New York Guildford, 624pp.

Dunning, John H. (1977) Trade, location of economic activity and the 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, 395-418

Dunning, J.H. and Alan M. Rugman (1985) The influence of Hymer's dissertation on the theory of foreign direct investment, *American Economic Review*, 75 (2) Papers and Proceedings, 228-232

Dunning, John H. (2003) Some antecedents of internalization theory, *Journal of International Business Studies*, 34, 108-115

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

Dunning, John H and Christos N. Pitelis (2008) Stephen Hymer's contribution to international business scholarship: An assessment and extension, *Journal of International Business Studies*, 39, 167-176

Enderwick, Peter (1988) *Multinational Service Firms*, London: Routledge

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

Hymer, Stephen H. (1972) The internationalization of capital, *Journal of Economic Issues*, 6 (1), 91-111

Hymer, Stephen H. (1976) *The International Operations of National Firms: A Study of Direct Foreign Investment*, [MIT PhD dissertation, 1960] Cambridge, MA: MIT Press

Hymer, Stephen H. (1990) The large multinational corporation: An analysis of some motives for the international integration of business, (trans. N. Vacherot, intro. M.C. Casson), in Mark Casson (ed.) *Multinational Corporations*, Aldershot: Edward Elgar, 1990, 3-31 [originally published in French in *Revue Economique*, 19 (6), 1968, 949-73].

- Kemp, Murray C. (1961) *The Pure Theory of International Trade and Investment*, Englewood Cliffs, NJ: Prentice-Hall
- Kindleberger, Charles P. (1969) *American Business Abroad*, New Haven, CT: Yale University Press
- Knight, Frank H. (1921) *Risk Uncertainty and Profit*, Boston: Houghton Mifflin
- Kogut, Bruce and Udo Zander (1992) Knowledge of the firm, combinative capabilities and the replication of technology. *Organization Science*, 3(3), 383-397
- Kojima, Kiyoshi (1978) *Direct Foreign Investment: A Japanese Model of Foreign Direct Investment*, London: Croom Helm
- MacDougall, G. Donald A. (1960) The benefits and costs of private investment from abroad: A theoretical approach, *Economic Record*, 36, 13-35
- Magee, Stephen P. (1977) Multinational corporation, industry technology cycle and development, *Journal of World Trade Law*, 11, 297-321.
- Markusen, James R. (1995) The boundaries of multinational enterprises and the theory of international trade, *Journal of Economic Perspectives*, 9 (2), 169-189
- Markusen, James R. (2002) *Multinational Firms and the Theory of International Trade*, Cambridge, MA: MIT Press
- McCann, Philip and Simona Iammarino (2013) *Multinationals and Economic Geography*, Cheltenham: Edward Elgar
- McManus, John (1973) The theory of the international firm, in G. Paquet (ed.), *The Multinational Firm and the Nation State*, Toronto: Collier Macmillan, 66-93, reprinted in Mark Casson (ed.) *Multinational Corporations*, Aldershot: Edward Elgar, 1990, 32-59
- Penrose, Edith T. (1956) Foreign investment and the growth of the firm, *Economic Journal*, 66, 220-235
- Ramamurti, Ravi (2012) What is really different about emerging market multinationals? *Global Strategy Journal*, 2 (1), 41-47
- Rowthorn, Robert (1972) *International Big Business*, Cambridge: Cambridge University Press
- Rugman, Alan M. (1981) *Inside the Multinationals*, London: Croom Helm [Basingstoke, Hants: Palgrave Macmillan, 25th Anniversary edition, 2006] xxxi + 164pp.
- Rugman, Alan M. and Alain Verbeke (1992) A note on the transnational solution and the transaction cost theory of multinational strategic management. *Journal of International Business Studies*, 23(4), 761-771

- Rugman, Alan M. and Alain Verbeke (2003) Extending the theory of the multinational enterprises: Internalization theory and strategic management perspectives. *Journal of International Business Studies*, 34(2), 125-137
- Rugman, Alan M. and Jing Li (eds.) (2004) *Real Options and International Investment*, Cheltenham: Edward Elgar
- Rugman, Alan M., Alain Verbeke and Quyen T.K. Nguyen (2011) Fifty years of international business theory and beyond. *Management International Review*, 51 (6), 755-786
- Rugman, Alan M. and Simon Collinson (2012) *International Business*, 6th ed., Harlow: Pearson
- Smith, Adam (1776) *An Inquiry into the Nature and Causes of the Wealth of Nations*, Glasgow edition, Oxford: Oxford University Press, 1975
- Tallman, Stephen (2003) The significance of Bruce Kogut and Udo Zander's article, 'Knowledge of the firm and the evolutionary theory of the multinational corporation', *Journal of International Business Studies*, 34, 495-497
- UNCTAD (1991-2014) *World Investment Report* Geneva: United Nations [annual, various issues]
- Vernon, Raymond (1966) International investment and international trade in the product cycle, *Quarterly Journal of Economics*, 80, 190-207
- Wadeson, Nigel S. (2013) The division of labour under uncertainty, *Journal of Institutional and Theoretical Economics*, 169 (2), 253-274
- Warren-Boulton, Frederick R. (1978) *Vertical Control of Markets: Business and Labor Practices*, Cambridge, MA: Ballinger
- Williams, Barry (1997), Positive theories of multinational banking: Eclectic theory versus Internalization theory, *Journal of Economic Surveys*, 11 (1), 71-100
- Williamson, Oliver E. (1975) *Markets and Hierarchies: Analysis and Anti-trust Implications*, New York: Free Press
- Williamson, Oliver E. (1985) *The Economic Institutions of Capitalism*, New York: Free Press
- Zaheer, Srilata (1995) Overcoming the liability of foreignness, *Academy of Management Journal*, 38 (2), 341-363

Table 1: Possible ownership strategies for serving a single representative customer

	Owner of R&D	Owner of production	Owner of distribution
Fully integrated	Firm A	Firm A	Firm A
Licensing to integrated licensee	Firm A	Firm B	Firm B
Franchising	Firm A	Firm A	Firm B
Subcontracting production	Firm A	Firm B	Firm A
Licensing to licensee who franchises	Firm A	Firm B	Firm C

Figure 1: Schematic view of the division of labour in a knowledge-intensive global industry

