The Theory of the Transnational Corporation at 50+.

Grazia Ietto-Gillies
London South Bank University
and
Birkbeck University of London

Abstract.
The paper briefly summarizes the historical evolution of TNCs and their activities. It then introduced the major theories developed to explain the TNC. There is an attempt to place the theories historically, within the context of the socio-economic conditions and of the relevant economic ideas in which they were developed. The following theories are discussed. Hymer, market power and control; Vernon’s international product life cycle; the internalization theory; Dunning’s eclectic framework based on: Ownership, Location, and Internalization (OLI) advantages; The Scandinavian School; the evolutionary approaches of Cantwell and of Kogut and Zander; the New Trade theory applied to the TNC; the role of nation-states in the strategic behaviour of TNCs. There are some critical comments at the end of each presentation. A brief analysis of key elements in the theories, their differences and commonalities follows in. It is pointed out that the pattern of development shows tensions between the following interconnected elements: (1) Contents and methods of interest to Business Schools and to Economics Departments; (2) Static versus dynamic approaches; (3) Emphasis on efficiency versus strategic elements; (4) strategies towards rivals as well as towards other players in the economic system such as labour, governments and suppliers; (5) and single- versus multi-disciplinary approaches.

Key Words: Transnational Corporations; Hymer; Vernon; internalization theory; Vernon; Scandinavian School; cantwell; Kogut and Zander; New Trade theory and multinationals; nation-states and transnationals; history of economic ideas.

1 iettogg@lsbu.ac.uk A version of this paper was presented at the 17th Annual Conference of the European Society for the History of Economic Thought (ESHET) on Economic Theory and Business Practice: Their Relations Through the Ages, Kingston University, London 16-18 May 2013.
1. Introduction

The theory of the transnational corporation (TNC)² and of its defining activity – foreign direct investment (FDI) – were born with the seminal doctoral dissertation of Stephen Hymer (1960 [1976]). Prior to it there have been theories of cross-border movements of capital and theories of imperialism³. The TNCs as such played no part in either. Theories about international capital movements were developed within the neoclassical tradition and following, mainly, the framework of neoclassical theories of trade, specifically Heckscher (1919) and Ohlin (1933)⁴. The theories of imperialism were developed within the Marxist tradition whether the relevant authors maintained their Marxists roots or not (Hobson, 1902; Luxemburg, 1913; Lenin, 1917; Bukharin, 1917).

What has been developed since Hymer’s work is a variety of theories dealing with different aspects of the TNC. Their focus ranges from: why firms become transnational; to the modalities of their activities; to FDI as their main activity; to why some countries become host or home (or both) for TNCs and FDI.

This paper considers the main theories, developed since Hymer’s, which have the TNC as a focus. The next section gives a brief excursion into the activities of the TNCs in the XX century and beyond. Section three summarises the main theories by presenting them – as far as possible – in historical sequence. Section four analyses the key elements in the theories presented in Section three. Section five summarises and concludes.

2. TNCs and their activities in history⁵

The antecedents of the modern TNC can be traced very far back into history. Transborder direct business activities go back many centuries indeed before the formation of nation-states. The Medici Bank can be considered a company with such direct business activities. In later centuries, the chartered companies such as the East India Company, The Royal African Company and the Hudson Bay Company had some elements in common with the present TNCs but the differences are too large for them to be seen as forerunners. Hymer (1971), following Chandler (1962), traces back the origin of the TNC in joint stock companies established from the mid-nineteenth century. But what is the distinguishing characteristic of modern TNCs compared to previous companies?

The distinguishing way of doing business abroad, the one that characterizes the transnationals compared with other companies, is direct production and generally direct business activities abroad. In order to engage in these direct activities, the TNCs establish affiliates abroad and acquire the ownership and control of their assets. This gives them a long-term interest in the strategies and management of the foreign enterprises which they control. But what do we mean by control?

Control can be seen as ownership control: what percentage of ownership secures majority in decision making. But it can also signify ability to organize and manage particularly at a distance. Such ability requires fairly speedy communication and transportation systems. It also requires appropriate internal organization of companies as they grow. Innovations of the technological and organizational types have made it possible to secure such control and therefore they have made it possible for the modern TNCs to develop and grow from the XXth century onward.

² A variety of adjectives and nouns are used to indicate this particular type of firm. Adjectives include ‘international’, and ‘multinational’; the nouns include: ‘firm’, ‘company’, ‘corporation’, ‘enterprise’. I prefer the adjective ‘transnational’ because it conveys the fact that these corporations can organize, manage and control activities across countries rather than just operate in several of them independently.

³ A summary of those pre-WWII theories is in Ietto-Gillies (2012: Part II).

⁴ Other main contributions are Nurkse (1933) and Iversen (1935).

⁵ For a more extensive treatment of the issue discussed in this section see Ietto-Gillies (2012: Part I).
The growth in the number of TNCs worldwide and in their operations has progressed steadily after WWII. The increase has been very considerable since the mid-1970s. In 1968–9 the number of TNCs originating from 14 developed countries was 7276 (Ietto-Gillies (2002a: 12, table 2.1). This figure is likely to be very close to the total number of world TNCs at the time. The latest World Investment Report (UNCTAD 2012; statistics table 34) estimates the total number of TNCs worldwide to be 103786.

Various elements have contributed to the growth of TNCs and specifically the following. (a) The developments in transportation and in the communications technologies and costs. (b) The organizational innovation within large companies and institutions. (c) The favourable political environment after the Second World War. (d) The liberalization and privatization programmes of many developed and developing countries in the last 30 years.

Elements (a) and (b) have made control at a distance possible. All four elements together have greatly facilitated and encouraged companies to invest abroad. There has been large growth in the value of foreign direct investment world wide as well as in the growth of other modalities of internationalization for which TNCs are responsible. For example some three quarters of world trade originates with TNCs. The growth in the number of transnationals and in their activities is also reflected in changes in the sectoral structure and connected changes in the geographical structure of TNCs’ activities. Between WWI and WWII most FDI was by resource-seeking TNCs and therefore most FDI was in developing countries. After WWII most FDI was in manufacturing. It was by developed countries’ TNCs and directed towards other developed countries. For example, a manufacturing US corporation investing in the UK or Canada.

The development of ICTs has allowed the vertical division of the production process and the location of various components into different type of countries; labour intensive components located in developing countries and those requiring high skills and latest technologies located in developed countries. From the late 1970s onwards we have seen a surge of FDI in services directed to both developed and developing countries. In terms of modalities the first few decades after WWII saw most of FDI taking the greenfield – i.e. real investment and accumulation – mode. From the 1980s onwards most FDI have taken the mergers and acquisition (M&As) modality. The different pattern has implication for competition as well as for the level of activity and employment in the host country.

Regarding the political environment (elements c and d above), the post WWII decades have seen considerable changes. The 1960s and 1970s were seen as decades of confrontation between TNCs and national governments, particularly those in developing countries. There were large numbers of nationalisations of foreign affiliates, particularly in developing countries. As neoliberalism took hold and spread we saw confrontation slowly turning into cooperation between national governments and TNCs (Dunning, 1993: ch. 13). Far from threatening nationalisations, many governments in developing countries followed in the footsteps of some developed countries in engaging in large-scale privatisations. The privatised assets were often bought by foreign companies. UNCTAD (1993: fig. 1, p. 17) shows that the number of nationalisations peaked in the mid-1970s and became non-existent after the mid-1980s. Privatisations started in the mid-1970s and increased very rapidly in the 1980s and 1990s. The 1990s saw a wave of protests by anti-globalization movements against international institutions at the heart of globalization and neoliberalism including TNCs. The protests faded away leaving few traces. Neoliberalism led to the big financial crisis of 2007-08. The austerity policies that followed have recently (2012-13) led to questions about the tax arrangements of TNCs via their transfer prices policies.

---

6 The new wave of protests (in Turkey and Brazil) currently taking place are indirectly focused on TNCs and globalization issues.
7 On transfer prices and their effects see Ietto-Gillies (2012: ch. 20) as well as Eden (2001).
3. The main theories

Hymer’s seminal work

Stephen Hymer was a Canadian economist doing doctoral research at the Massachusetts Institute of Technology in Cambridge, USA. He became intrigued by the motivations behind the large foreign investment by US corporations in a growing number of countries including his own. He died in a car accident in 1974, aged 39 and his dissertation was published posthumously in 1976. Hymer’s work constitutes a radical departure from the conventional neoclassical approach of the time. It opened a whole new research programme in the area of international production. Follow-ups, refinements and new twists to the theory are continuously coming out.

In order to understand the relevance of Hymer’s contribution as well as the novelty of his approach, we must remember that, when he was writing, there was no theory of foreign direct investment as such. There was no perceived need to consider direct investment as a special case; indeed the concept of foreign direct investment had not been developed before Hymer’s breakthrough. The then prevalent neoclassical theory explained movements of capital across borders via differentials in interest rates. However, as Hymer noted:

- FDI does not necessarily involve movement of funds from the home to the host country. In fact, direct investment is, at times, funded in other ways including borrowing in the host country or via retained profits.
- FDI often takes place both ways so that both countries involved are originators and host to FDI.
- FDI tends to be concentrated in particular industries across various countries, rather than in a particular country across various industries.

These three characteristics are incompatible with the neoclassical explanation for movements of capital based on differentials in interest rates. Hymer thus saw the need to differentiate between purely financial investment (i.e. from portfolio investment) and investment by large firms for production purposes. His demarcation criterion between foreign direct investment and portfolio investment is control. Direct investment gives the firm control over the business activities abroad; portfolio investment does not. By acquiring control of foreign assets the firms removes conflicts with local competitors. It does so by giving the controlling firm more market power and thus intensifying the imperfections in the market structure. The existence of structural market imperfections is, in fact, one of the key assumptions of Hymer’s theory: market imperfections and the search for market power are a key determinant of FDI.

Moreover, market power is affected by companies’ strategies including the ones leading to control of foreign assets and production. The type of imperfections he considers are structural ones, that is, those imperfections arising from the market structure, for example from an oligopolistic structure in which a few large firms dominate the market.9

Later works by Hymer are more in the Marxist tradition. They emphasize the contradictory and conflictual nature of capitalist production and deal with the following issues. (a) Effects of MNCs’ activities on: labour; politics; the nation-state and its government. (b) The effectiveness of economic policies (Hymer, 1966; 1975; Cohen et al., 1979: chs 9 and 11). (c) The division of labour (Hymer, 1970, 1971; 1972; Cohen et al., 1980).

---

8 The theories presented in this section as well as other theories are discussed at greater length in Ietto-Gillies (2012: Part III). See also Cantwell (2000).

9 Transactional imperfections are considered by Hymer (1968), a paper which seems to have little relationship with his main work (1960 [1976]) and with his later research.
Vernon’s international product life cycle

Raymond Vernon was working on what became a well known theory at the same time as Hymer and indeed up the road from where Hymer was working: at the Harvard Business School. The economic context of Vernon’s theory is one of expanding technologies and markets for new mass consumption products such as washing machines. It was also one of increased internationalization as barriers to movements of products and capital gradually came down after WWII. The theoretical background to his approach must be sought in the technological gap theories of trade (Posner 1961) and in the theories of the product’s life cycle (Kutznets 1953). In fact, while Hymer’s point of departure is the firm, Vernon’s is the product. How new products emerge; how they impact on the innovating firm and to the industry structure in which the firm operates; how the firm is affected by the progress of the product through its life; how the product progresses through its life in national and international markets and production locations.

Vernon begins with the assumption that enterprises in any one of the advanced countries of the world have equal access to knowledge. However, this does not mean an equal probability of application of such knowledge to the development of new products. It is the consciousness of opportunities and the responsiveness to such opportunities that vary from one entrepreneur to another. Such consciousness and responsiveness are associated with the market conditions in which entrepreneurs operate; this makes knowledge inseparable from the decision-making process about its use. Therefore knowledge is not an exogenous variable.

In the 1960s and 1970s the US market offered unique opportunities for the exploitation of knowledge and its embodiment in new products because:

- It was a market in which consumers had high average income per capita.
- It was a very large market; hence even minority tastes were likely to provide a fairly large market.
- It was characterized by high unit labour costs and a large supply of capital; it was, in other words, a market abundant in capital and scarce in labour.

For these reasons the new product would be located in the US. Such location would secure flexibility of adaptation to possible problems and to requirements of consumers. Adaptation is more easily achieved if production takes place near its initial development location. Moreover, when first launched into the market, the product enjoys a large amount of differentiation and thus a semi monopolistic position. It will have low price elasticity of demand and high income elasticity.

However as the product matures and the market expands there will be the threat of imitators. Expanding foreign demand – usually in other developed countries – will first be met by exports. At a later stage direct production in Europe may replace exports in response to the following: the emergence of competitors in European countries; possible import controls; and possible lower production costs in Europe. As the product becomes standardized, competition increases and the search for lower production costs starts. This last phase in the life of the product is likely to lead to the location of production in developing countries and to the sourcing of developed countries’ markets – including the US itself – from this production.

The key elements in Vernon’s highly dynamic theory are: innovation in products which gives the firm a temporary monopolistic position; interaction between the life of the 1979: ch. 6) within the firm, the industry and the international arena (in particular between developed and developing countries).
product, the degree of competition in the industry and the geography of trade and of FDI/production.

Many criticisms can be levelled at the theory\textsuperscript{10}. It was developed in the 1960s and reflected the economic environment of the times as Vernon himself recognized in a courageous article which dissects critically his own theory (1976). Moreover, as time went by, not only the economic environment and the differences between the US and European economies changed, we also saw significant technological changes. The development and wider effects of ICTs brought shorter product lives as well as changes in the sequences of location of international production. Moreover, it should be noted that the concentration of the theory on the product more than on the firm does not allow a full analysis of the competitive position of the firm and how it can be affected by product diversification strategies.

The internalization theory

The post war expansion in Western economies saw concentration of production and increase in firms’ size. Concomitantly with – as a consequence of - these developments companies adapted their organizational systems to cope with new functions, new products or new geographies in the more complex resultant structures.

The internalization theory of the TNC reflects these changes in the economic environment. It was developed on the back of Coase’s analysis of the firm (1937) and it also benefitted by Williamson’s later developments (1975; 1981). It started with a paper by McManus (1972); a fuller development was achieved with Buckley and Casson (1976). Further contributions include Teece (1977); Rugman (1981); Caves (1982) and Hennart (1982).

Buckley and Casson concentrate on a particular type of market imperfections; transaction imperfections as in Coase’s analysis. When markets present transactional imperfections there is an incentive to internalize. Why do firms internalize? What are the limits to internalization? There are benefits of internalization and there are also costs; the balance between the two will determine the limit to internalization\textsuperscript{11}. The benefits of internalization stem from \textit{transactional market imperfections} and relate to one or more of the following situations.

- When there are long time lags between initiation and completion of the production process and, at the same time, futures markets are non-existent or unsatisfactory.
- When the efficient exploitation of market power over an intermediate product requires discriminatory pricing of a kind difficult or impossible to implement in an external market, though possible to implement internally.
- When imperfections would lead to bilateral concentration of market power and thus to an unstable situation under external markets.
- When there is inequality in the position of the buyer and seller regarding knowledge on the value, nature and quality of the product; the resultant buyer uncertainty may encourage forward integration.

\textsuperscript{10} See for example Cantwell (1995).

\textsuperscript{11} This equilibrium approach to the firm is taken up by the New Trade theories applied to the TNC as discussed below.
• When there are imperfections deriving from government intervention in international markets such as the existence of ad valorem tariffs, restrictions on capital movements, discrepancies in rates of taxation. The two most important areas of internalization relevant to TNCs are markets for intermediate products and markets for knowledge. Before the Second World War the major factor that contributed to the emergence of MNEs was demand for primary products, leading to vertical integration across frontiers and to internalization of intermediate markets. Since WWII the major factor has been the growth in demand for knowledge-based products coupled with the difficulties of organizing efficient external markets for intangibles and knowledge. A TNC implies internalization across national boundaries. Buckley and Casson (1976: 45) write on this issue: ‘There is a special reason for believing that internalization of the knowledge market will generate a high degree of multinationality among firms. Because knowledge is a public good which is easily transmitted across national boundaries, its exploitation is logically an international operation.’ So the conclusions seem to be that imperfect markets generate incentives to internalize; the market for knowledge is highly imperfect, so there are strong benefits in internalizing it.

The internalization theory of the TNC is still a very successful and widely used theory. However, there are some doubts about it. There is the question of whether the theory is tautological as the authors themselves recognize. Casson (1982: 24) writes: ‘Internalization is in fact a general theory of why firms exist, and without additional assumptions it is almost tautological.’ Buckley (1983: 42) expresses similar doubts when he writes: ‘At its most general, the concept of internalisation is tautological; firms internalise imperfect markets until the cost of further internalism outweighs the benefits.’

In terms of relationship with the economic context to which the theory is supposed to apply the following should be noted. When the theory was first developed, there had been decades of firm’s growth via internal expansion leading to the concentration and large firms in many industries. However, the last 25 years have seen a great increase in outsourcing and generally in firms’ activities being contracted out and bought on the market. Yet, these are the decades when the internalization theory has been most successful within the international business community. The two patterns – decades of internalization followed by decades of externalization – cannot be explained by the same theory of internalization.

Moreover, the following should be noted. The internalization theory tries to explain why firms prefer the FDI rather than licensing route to growth, thus why they prefer internalization to market-based relationships. However, even accepting that internalization is to be favoured because it cuts transactional costs, it is not clear why firms should prefer the FDI rather than the exporting route: the first implies internalization across borders; the latter modality implies internalization within the nation-state.

**Dunning’s OLI advantages**

John Dunning worked on international production issues from the 1950s onwards and until his death in 2009. His early research was on the factors leading to the high productivity of American investments in British manufacturing. In his (1977) he developed a ‘systemic’ theory – whose origin he traces to his earlier work (Dunning, 2000b) – designed to explain internationalization modes and processes. He developed a framework for considering: (a) all main modalities of internationalization and specifically FDI, exports and licensing thus attempting to address the criticism of the internalization theory mentioned in the last paragraph of the previous subsection; (b) issues of why and when firms invest in foreign countries; and (c) issues of why certain countries become attractive for inward FDI.
Dunning’s wide framework was built around the analysis of three sets of advantages: Ownership, Locational and Internalization (OLI) advantages.

1. **Ownership advantages** are those that are specific to a particular enterprise. They constitute competitive advantages towards rivals and enable the company to take advantage of investment opportunities wherever they arise. This set of advantages links Dunning’s theory to Hymer’s.

2. **Locational advantages** are those advantages specific to a country which are likely to make it attractive for foreign investors.

3. **Internalization advantages** are all those benefits that derive from producing internally to the firm; they allow it to bypass external markets and the transaction costs associated with them. They are, essentially, benefits of operating within hierarchies rather than markets. This set of advantages links Dunning’s theory to the internalization theory and, of course, to Coase’s theory of the firm.

Foreign direct investment takes place whenever:

- The enterprise concerned possesses “…net ownership advantages vis-à-vis firms of other nationalities in serving particular markets” (1980: 275).
- The enterprise derives benefits from internalizing the use of resources in which it has an advantage rather than selling them on external markets, e.g. via licensing.
- The country where the FDI takes place must offer special locational advantages to be used in conjunction with those deriving from ownership and internalization.

Dunning’s theory has been for many years the main reference framework for many pieces of international business research. It gives a clear, well defined framework and its multi variables structure makes it easy to apply to almost any country, firm and time. Each of the above three sets of advantages (OLI) can include a long list of variables from which researchers can choose in the adaptation of their research to the specific context they are interested in. Thus the theory seemed to be always applicable independently of specific circumstances. This wide applicability made the theory irrefutable and rather than strengthen it, it weakened it. A theory that is always applicable may be tautological and loses its usefulness and scientificity. Moreover, most of the criticisms that were levelled at the internalization theory apply also do Dunning’s because it also relies on internalization.

Dunning was well aware of the weaknesses of his framework and, in later years, further develop it. He also worked on many other aspects of international business. Specifically he developed interesting work in the following areas:

- Operationalization via contextualization of the three sets of variables (Dunning 1993a and 2000a).
- Dynamization (Dunning 1993b) of his eclectic theory.
- The relationship between international production and countries’ development patterns (Dunning 1981: ch 5; Dunning and Narula 1996).
- Incorporations into the framework of new and growing organizational forms such as mergers and acquisitions and inter-firm collaborative agreements (Dunning 1997).

**The Scandinavian School**

The international business researchers we have discussed so far concentrated on countries which have been traditionally involved in FDI such as the US and UK. However, the 1960s and 1970s saw also many other countries involved in international direct production. A group of Swedish economists and management/marketing/strategy academics (Johanson and Wiedersheim-Paul 1975; Johanson and Vahlne 1977 and 1990) became interested in studying the position of smaller countries and their companies’ strategies towards international activities. Their interest focuses on strategies in relation to the stages and modalities of internationalization that companies go through. The authors link the stages and
modalities to the timing of internationalization activities. The timing determines the modality of establishment of operations abroad; it also affects the amount invested and the type of country in which the operations are established, starting with the nearest countries in terms of both spatial and psychic distance.

The authors analyse two internationalization patterns. The first one is designed to explain the increasing involvement in a single foreign country. The second pattern explains involvement in a variety of countries. The theory is very dynamic in that it considers time sequences and also because the resources already committed in a country impact on further decision. Thus decisions about the future modalities, countries and the amount of resources to be committed abroad depend on the path already followed in internationalization in terms of resources committed, modalities followed and countries of involvement.

The conclusion is that involvement in any single foreign country will proceed cautiously and in accordance with the following stages in the establishment chain:

- exports via agents;
- setting up of sales subsidiaries;
- setting up of production subsidiaries.

The above sequence is the result of state and change aspects in which knowledge and uncertainty play a large role. The dynamic sequence is linear in two ways: because each stage leads to the next one and because each new stage involves a larger commitment of resources than the previous stage.

The second internationalization pattern refers to the spread of internationalization from one foreign country to others. Here the sequence is also dynamic and linear proceeding by stages from the foreign country(ies) psychically closer to those more distant. Psychic and spatial distances tend to be strongly related.

The dynamism in the theory links it to Vernon’s. Both theories consider stages and time sequences. In Vernon’s case the stages relate to the life of the product and they affect its production location and its markets. They also impact on the competitive environment in which the firm operates at the various stages in the product life. In the case of the Scandinavian School the stages refer to the modalities and locations of internationalization. This is a theory not about the product but about the firm and its internationalization strategies. The relationship between the product – in terms of innovativeness, technology or potential demand – and the firm is not considered; neither is the market structure in which the firm operates.

**Evolutionary approaches to the theory of the TNC**

Coase (1937) questioned the relationship firm-market and the reasons for the very existence of the firm. However, questions about the internal functioning of the firm and its objectives were left unanswered or, indeed, were not asked at all: the firm continued to be a black box. Its opening started with Penrose (1956) – a work that had very little impact when first published though it is now, deservedly, recognized – and with Nelson and Winter (1982) a work that had a considerable impact since publication.

None of these two works dealt directly with the multinational company. However the competence-based theory of the firm which they expounded had a big impact on other authors working on international business and specifically on John Cantwell as well as on Kogut and Zander.

Cantwell (1989) takes on the competitive advantages view of Hymer and of Dunning but goes a step further. He considers such advantages not as exogenous but as created by the
firm itself. Specifically they can be created in the field of innovation and technology within which the firm becomes the generator of its own advantages. The theory is thus injected with: (a) realism because it attempts to answer the question: where do advantages come from? and (b) dynamism because it links created advantages to changes within the firm and its environment.

According to Cantwell the TNCs are in a particularly strong position to develop their ownership advantages in innovation. By operating in many countries - often characterized by diverse knowledge and innovation contexts – they can acquire knowledge from the localities and use it to further their innovative activities. In this process the TNC is aided by its involvement in two types of networks: (1) its own internal network between the various units of the firm spread in a variety of geographies; and (2) external networks between units of the firm and suppliers/distributors, consumers and partners in collaborative ventures.

The latter networks enable units of the TNC to acquire knowledge from their external environment. This knowledge is incorporated within the unit and also transferred to other units of the TNC via its internal network. The TNC with its geographically diversified structure, its variety of organizational interactions with the external environments and its internal network is in the best position to accumulate innovation and technology across countries and through time. The internal networks raise issues of control of the subsidiaries by the headquarters of the company. The external networks raise issues of the degree of embeddedness of the subsidiary into the local economy. The acquired knowledge gives the TNC advantages in all its modalities of operation from FDI to export to licensing. Here is one of the several criticisms of the internalization theory on the part of Cantwell: in the real world FDI and exports are complementary not an either or situation. There is an empirical basis to this criticism since the TNCs are responsible not only for all FDI but also for over three quarters of world trade (UNCTAD 1996). Moreover, over one third of world trade is intra-firm, i.e. between different units of the same company though across different countries.

Thus knowledge spills over from the locality to the TNCs. It also spills over from the TNC to the local economy. The absorptive capacity of the locality becomes crucial for the innovation benefits of both the local economy and the unit of the TNC operating within it. There is a dynamic interaction and a cumulative causation mechanism between ownership advantages and locational advantages and both can be seen as endogenous and created. Thus the separation of ownership from location advantages in Dunning’s OLI framework may be misleading.

Kogut and Zander’s analysis focuses on the role played by knowledge in the boundaries of the firm, i.e. the extent to which the firm decides to expand via internalization or through external, contractual relationships. Kogut and Zander start by criticizing the standard view on the boundaries of the firm: the internalization theory. In the latter the boundaries are set by the failure of the market to protect knowledge and by market transaction costs. Moreover, in the internalization view the boundaries of the firm are independent of its ownership advantages.

Kogut and Zander (1993) key insights are the views of the firm as a social community and the development of knowledge as a product of the social group. They write: ‘…firms are social communities that serve the efficient mechanisms for the creation and transformation of knowledge into economically rewarded products and services’. In Kogut and Zander (2003: 511) we read on this point: ‘…knowledge exists in networks and in institutionalized contexts.’

---

12 The links between internal and external networks and knowledge diffusion have been explored in the management/organizational analysis literature (Forsgren et al. 2005; Hedlund, 1986; Hedlund and Rolander 1990; Bartlett and Ghoshal 1988 and 1991; Ghoshal and Nohria 1997).
Whenever knowledge is embedded in – and dependent on – social structures, it is more context-specific and, largely, tacit. This makes it less likely to be codifiable, teachable and transferable to other social settings. The social community setting of knowledge development means that: (a) knowledge is more likely to be tacit because emerging from shared experiences and procedures; and (b) further knowledge development is likely to emerge from the shared experiences. Here we have clear pointers towards the fact that the social nature of the firm and of groups within it lead to a specific type of knowledge and to ownership advantages and value creation. The authors write: ‘Cooperation within an organization leads to a set of capabilities that are easier to transfer within the firm than across organizations and constitute the ownership advantages of the firm.’ (p. 627). The social community setting of the firm apply also to the TNC because its subsidiaries tend to share identities and values or, at least, they share them to a higher degree than each subsidiary would share with independent external firms.

For Kogut and Zander, the limits to the firm are, therefore, set not by market failure but by the firm’s efficiency in acquiring knowledge. They write: ‘In our view, firms are efficient means by which knowledge is created and transferred […]. Through repeated interactions, individuals and groups in a firm develop a common understanding by which to transfer knowledge from ideas into production and markets. In this very critical sense, what determines what a firm does is not the failure of the market, but the firm’s efficiency in this process of transformation relative to other firms.’ (p. 631).

Moreover, the authors see knowledge as the main source of ownership advantages and there is, therefore, interaction between ownership advantages and internalization. The ownership advantage characteristic of knowledge is enhanced by the fact that tacit, uncodifiable knowledge is also more difficult to imitate: knowledge is therefore an advantage on which the firm can further build up without fears from rivals’ imitations.

Knowledge is cumulative. Older knowledge is more easily codifiable and therefore more easily transferable outside the boundaries of the firm. The costs of technology transfer vary with the degree of tacitness of the related knowledge. Thus established technology is not a public good; it is transferable at a cost and the cost varies with the accumulation of experience and learning about codification procedures.

As knowledge becomes more codifiable with the passage of time, the company is likely to move from internalization to externalization, from FDI to licensing in international operations. The sequence and its timing depend on the degree of tacitness and codifiability of the knowledge specific to the firm.

These two evolutionary theories have much in common as well as many differences. Cantwell’s approach is very critical of the internalization theory while Kogut and Zander’s is, ultimately, a theory of why firms internalize. Their view is that the limits to internalization and thus the boundaries of the firm are not set by transactional market failures – as in Coase and in the internalization theory – but by the efficiency of the firm in developing, spreading and utilising knowledge. Both theories can be seen to have elements of created ownership advantages and thus of market power. However, the impact on – and interaction with – the local economy so prominent in Cantwell is absent in Kogut and Zander.

New Trade Theories and the TNC

The evolutionary theory of the firm was an attempt to move the theory of the TNCs more towards the real world and away from the neoclassical theory of the firm. It was also an attempt to inject into the theory elements from disciplines other than economics and more in line with organizational sociology (Kogut and Zander). Meanwhile other forces were pushing more towards the directions of an ‘economics-only’ theory of the TNCs and one more
strongly embedded into partial equilibrium analysis and the neo-classical framework. The 1980s saw the development of New Trade theories (Krugman, 1985; 1991a; 1998) in which the trade pattern could be linked to increasing economies of scale and its advantages for countries on the basis of their factor endowments. These developments gave way to a considerable amount of research in which the New Trade theories and its models could be used to explain regional development and agglomeration as well as developed v developing countries’ trade. They were also used to draw policy implication from those analyses.

This framework, however, cannot explain direct production in other countries by TNCs. Essentially, if there are external economies of agglomeration and the internal economies are plant economies, then it can only make sense to produce in one location/country and supply other markets through exports. There is a basic conflict and tension between a theory that predicts clustering of production activities and a reality of companies that spread their activities in space sometimes horizontally, sometimes vertically, sometimes both ways.

At the theoretical level it is possible to solve the conundrum by adjusting some of the assumptions, and this is what some economists have done. The assumption of capital immobility – underlying much trade theory – has obviously to be removed when dealing with theories of direct foreign production and FDI, which by their nature imply capital mobility. Moreover, constraints to the movements of products are sometimes introduced such as barriers to trade.

However, the main adjustment is in the treatment of internal economies of scale. They are split into two types:

- internal economies at the level of plants;
- internal economies at the level of the firm.

These economies separately or together are of the Chamberlinian type i.e. they are internal to the firm. They are therefore analysed in the context of imperfect competition. The first type of economies – those at the plant level – are linked to more traditional fixed inputs, those deriving from traditional physical assets such as machinery; they give rise to fixed costs. The second type of economies derives from such inputs and assets as organizational, technological, managerial/marketing; the services deriving from them are of benefit to - and can be used by - the company as a whole and therefore by its head office as well as by its affiliates. These are joint inputs within the firm because they can be used by different parts of the firm for the same product and/or for different products. No matter how many plants (and affiliates) are going to use these inputs, the marginal cost of using the inputs in additional plants – at home or abroad – is low or negligible. In addition to this, the industry as a whole may also achieve scale economies of the external, Marshallian type.

The authors dealing with this specific theory in the context of TNCs have developed models for FDI from developed countries directed towards (a) other developed countries (Markusen 1984 and 1995); and (b) developing countries (Helpman 1984 and 1985; Helpman and Krugman 1985). This is done by changing the assumptions.

These highly theoretical models do not seem to fit the facts very well. Krugman (1998: 15) writes: ‘preliminary efforts . . . have found that such models are not at all easy to calibrate to actual data; in general, the tendency toward agglomeration is stronger in the models than in the real economy!’

There are several problems and contradictions within the overall framework. They go from using a Chamberlinian monopolistic competition framework for companies that operate

---

13 The modelling of equilibrium under economies of scale became possible after the development of the mathematics behind imperfect competition and increasing returns (Dixit and Stiglitz 1977).
under oligopolistic condition; to the lack of consideration for exports as an alternative to FDI.
This is a point in common with the internalization theory which is the starting framework for
the New trade theories. It is a problematic point because – as we saw above – TNC are
responsible for most world trade as well as for all FDI.

In my view the most problematic element is the fact that the analysis is largely a
spatial analysis: transport costs play a big role in the outcome of where different plants will
be located. It is essentially a theory of spatial location of production – a field that geographers
have been interested in for decades. As such it does not distinguish between plant location
within a single nation-state or across nation-states. In other words, the nation-state is hardly
relevant in the New Trade theories.

The role of the nation-states

I have argued elsewhere (Ietto-Gillies 2012: ch.14) that the existence of nation-states is the
very reason why we need theories of the TNC as opposed to theories of the firm in general.
Yet the nation-state - as determinant of TNCs’ activities if not in terms of effects on it - has,
largely, been ignored by most theories. Hymer did write on the State and the nation-state and
their interaction with the international firm. This was part of his post-dissertation research
which had a Marxist orientation. He was particularly interested in the effects of TNCs’
activities on the abilities of national governments to develop and implement policies.
However, Hymer did not consider the relevance of the nation-state for the explanation of
TNCs and FDI.

Nation-states are relevant because they are characterized by different regulatory
regimes regarding (a) labour and social security systems; (b) fiscal systems; (c) currencies;
(d) industrial policies including incentives to businesses; and (d) environmental and safety
standards. The differences in these regulation regimes allow companies that can truly
organize, manage and control their operations transnationally to arrange their activities as to
benefit from these differences. Their transnationality puts them in a position of advantage
towards actors - with whom they interact – who are not able to operate transnationally, or not
to the same extent. Such actors include: labour; governments of nation and regional states;
suppliers. Moreover, their international operations allow them to build up also advantages in
terms of risk spreading and of acquisition of knowledge from the various localities in which
they operate.

Transnationality therefore gives the TNC the ability to develop strategies that
maximize their bargaining power towards other actors such as labour or governments. In
particular labour working for the same company in countries with differing labour law, trade
unions, social security systems cannot organize effectively or not as effectively as if it were
all working for the very same company within the same country. These strategies of
transnationality thus result in a strategy of fragmentation of labour by national geographies.
Moreover, in the last 30 years there have also been strategies of organizational labour
fragmentation. Outsourcing strategies lead to labour working for a variety of companies and
have thus weaken its bargaining power. Outsourcing can take the international route via
offshoring strategies. The organizational fragmentation of labour can, in this case, combine
with geographical (by nation-state) fragmentation.

It was pointed out above how the internalization theory is at a loss to explain the trend
towards externalization in the last three decades. This trend can only be explained if we bring
in socio-political-economic elements linked to reactions to the power of labour vis-à-vis large
corporations. Outsourcing and externalization in general were developed from the 1980s
onwards as organizational strategies leading to the fragmentation of labour and thus to its
lower bargaining power. Such power had increased in earlier decades characterized by internalization.

Thus a full study and understanding of TNCs require them to be placed in the context of nation-states. Moreover, it requires their analysis to be made in terms of strategic rather than efficiency/equilibrium behaviour. Strategic behaviour has, at times, been considered in the literature on firms and TNCs. But it is usually in the context of strategies towards rival firms. Hymer’s, Dunning’s and Cantwell’s analyses of firm’s advantages can be seen in the context of advantages and strategic behaviour towards rival companies. These are certainly very important. However, it is worth stressing also advantages towards other players in the economic system from labour to governments to suppliers. Advantages towards any of those will result in higher profits and thus in advantages also towards rivals.

4. Key elements in the theories and their context

As with theories in any other field, the ones we discussed above must be seen in the economic and social context in which they were developed. They must also be seen in the context of the ideas and theories prevalent at the time. These elements have been pointed out in the subsections above. Nonetheless there are also elements specific to the TNCs theories and to some of these elements we now turn.

The theories here presented are only a subset of all those developed to explain the TNC and its activities. However, they are the ones which have been most successful in terms of acceptance by the research and pedagogy community. Such community does not coincide with the economics academic community. In fact theories of the TNCs have been most successful with the wider international business community. This is a very large and active community mostly clustered around Business Schools or a variety of Business Departments from Marketing and Strategy to Organizational Analysis to Economic Geography. Economics Departments have – largely – ignored the theory of the TNC following the very first rejection of Hymer’s work. Why should that have been so? Why the neglect of an institution that has been so relevant to economies and societies particularly after WWII? I can only attempt possible reasons: from the feeling that there is nothing special about the TNC over and above the large corporation; to the difficulty of analysing messy institutions characterized by various locations and types of activities and slotting them into neat categories and analyses particularly of the mathematical type. It is interesting, therefore, to note that the strictly economics community has taken an interest in the TNC and developed the New Trade theory applied to TNCs when they were able to develop neat, equilibrium models of it. These are not very realistic but they can be taught as part of the general economic curriculum.

Against economic-only theories there have been theories developed with reference to other disciplines from marketing – Vernon as well as the Scandinavian School - to sociology of organizations – Kogut and Zander and, to some extent, Cantwell. One effect of the multidisciplinary contexts is the fact that the methodologies used tend to be more wide ranging and diversified compared to what one normally sees in economics-only type of theories. They range from traditional econometrics to qualitative methods based on large scale or on elected, in-depth interviews.

Moreover, the emphasis of theories developed in multidisciplinary contexts tends to be wider than the purely efficiency/equilibrium analysis of economics. Strategic elements are

---

14 This statement does not apply to the theory presented in the last sub-section.
15 A very clear textbook (Barba Navaretti and Venables 2004) has been developed explaining TNCs and their activities entirely in terms of the New Trade theory applied to the MNCs. No mention of the historical development of TNCs nor of other explanations for their emergency, development and activities.
brought in usually with reference of company’s strategies towards rivals. However, strategies towards other players in the economic systems – labour, governments and suppliers – can also be brought in to reflect their relevance for the pattern of internationalization.

Whether developed in the context of Economics or Business studies, the theories can differ in terms of their dynamic versus static approaches. Those developed by Vernon, the Scandinavian School and Cantwell have definitely more dynamic elements; in the case of Vernon and Cantwell the dynamic elements are endogenous to their theories.

Most theories emphasize – directly or indirectly – market imperfections and market power. However, these can be of two types: structural imperfections in which large TNCs operate in imperfect markets and have varying degrees of market power sometimes endogenously built by their own strategies (as in Cantwell’s theory). Imperfect markets can be – directly or indirectly – traced down to oligopolistic structures (Hymer; Dunning; Cantwell) or to monopolistic competition (Vernon; New Trade theories). Imperfections may also be of the transactional type, a’ la Coase. The internalization theory – Buckley and Casson; Dunning - falls into the latter category.

It should also be noted that the theories vary in terms of what it is that they are trying to explain: from FDI only; to a variety of modalities of international business; to the TNC as a firm; to the position of different countries.

5. Summary and conclusions

The paper briefly summarizes the historical evolution of TNCs and their activities. It then introduced the major theories developed to explain the TNC. The presentation is in historical sequence. There is also an attempt to place each theory in its socio-economic and history of ideas contexts. The following theories are discussed. Hymer, market power and control; Vernon’s international product life cycle; the internalization theory; Dunning’s eclectic framework based on: Ownership, Location, and Internalization (OLI) advantages; The Scandinavian School; the evolutionary approaches of Cantwell and of Kogut and Zander; the New Trade theory applied to the TNC; the role of nation-states in the strategic behaviour of TNCs. There are some critical comments at the end of each presentation. A brief analysis of key elements in the theories, their differences and commonalities follows in Section four. It is pointed out that the pattern of development shows tensions between the following interconnected elements: (1) contents and methods of interest to Business Schools and to Economics Departments; (2) static versus dynamic approaches; (3) emphasis on efficiency versus strategic elements; (4) strategies towards rivals as well as towards other players in the economic system such as labour, governments and suppliers; (5) and single- versus multi-disciplinary approaches.

The transnational companies dominate our economies and more research should be devoted to them by the economics community. I firmly believe that such community would benefit from multi- and interdisciplinarity links with relevant fields as well as from a historical and history of ideas approach to the field. I also believe that equilibrium analyses may be inappropriate to such institutions and that reality should be at the forefront of analysis. It may also be time for economists working in paradigmatic approached other than the neo-classical one to take a more active interest in this key actor of contemporary economies.

References


