A commentary on Grazia Ietto-Gillies’ paper: ‘The Theory of the Transnational Corporation at 50+’

John Cantwell, Rutgers University, Newark, New Jersey, USA
cantwell@business.rutgers.edu

Keywords: transnational corporations, Hymer, Vernon, internalisation theory, Scandinavian School, Cantwell, Kogut and Zander, New Trade theory and multinationals, nation-states and transnationals, history of economic ideas

This paper offers a splendid overview and a succinct summary of the theory of international business. It should be especially helpful for Ph.D. students in this field, and perhaps for other scholars that are coming into the area from other specialisms, or considering doing so. The article extends the author’s recently revamped book, Transnational Corporations and International Production: Trends, Theories, Effects (Ietto-Gillies, 2012), which book I commend and indeed which I use myself as a central text on my own doctoral course in the Theory of International Business. The author knows already of my views on many of the issues she discusses, both because she refers to some of my earlier work in the paper, and from some direct correspondence that we had in the past over her book, when she was writing it or re-writing it.

Ietto-Gillies is right to recognise that the issues addressed by the theory have altered over the 50+ years as the environment has changed, and with it the nature of the subject; and by the migration of the subject from departments of economics, and to a slightly lesser degree from departments of marketing and of finance, to becoming incorporated in its own right in business schools – most often within departments of management and sometimes in standalone departments of international business. I detect in the paper some sense of disappointment on the part of the author that the subject areas of economics and international business have diverged over the period described. As Ietto-Gillies has described, the theory of international business was spawned from economics a little over 50 years ago. The subsequent narrowing of the discipline of economics that has moved it away from the domain of international business studies will be well known to World Economics Association members. The philosophically deep, historically rich and complex thinking of scholars such as John Dunning or Dick Nelson, impoverished if overly formalised, which was so welcomed in the economics profession of the 1960s (or in any earlier era) would not be so welcome today. Nelson has termed what has been lost or relegated to a secondary place within the economics discipline as ‘appreciative theory’ (see e.g. Nelson, 1998). Appreciative theory is closer to the complexity of real-world social and economic systems – and to their changing historical context – than is the deductive logic, axioms and more restrictive assumptions of formal theory. Evolutionary economics in the tradition of Nelson and Winter, in common with international business studies, has instead returned to the traditions of classical political economy in relying on appreciative theory as the primary driver of analysis that explains real-world processes, and which, in modern terminology, we would call multidisciplinary and interdisciplinary in nature.

International business as a subject area has moved steadily further down this road of multidisciplinary theory-building grounded on empirical observation, despite adopting some
quantitative methods that simplify relationships and which come with a language taken from certain natural science contexts (hypothesis testing and controls) that is questionable when applied to complex social systems in which we cannot conduct controlled experiments. However, since formal representations of international business theory tend to be associated with quantitative applications, they are also designed to have a direct connection with some aspect of empirically observed phenomena (rather than model-building for its own sake, not directly linked to any observation, as in much neoclassical economic theory). The purpose of this formal structure of tests against the evidence is generally seen as a capacity to draw conclusions that have conceptual implications which feed into the enrichment of, and the new contribution to, some relevant appreciative theory. As Nelson argues, when a subject is progressing well, there is a largely positive and constructive relationship between appreciative theory and formal theory, or formal representations of relationships. Given this relative openness of the international business subject area, it is not surprising that inputs have been accommodated or absorbed from beyond economics, and especially from those working on the sociology and psychology of management.

Ietto-Gillies and I can agree that it would indeed be wonderful if the broader approach of classical political economy, driven by appreciative social and behavioural science theorising, were to be revived beyond the realm of international business studies, and to be embraced – as it once was – in economics as a whole. However, in at least some specialised fields of study, such as in international business, in innovation studies, in much evolutionary economics, and in some economic history, this traditional form of approach has been preserved and revitalised. It would be welcome for it to return to the economics discipline as a whole at some stage, but the discussion of this aspiration lies beyond the scope of our present exchange.

Ietto-Gillies does a good job of showing how the theory of international business has evolved in terms of the analysis of its dominant actor, the transnational corporation (TNC), and in evaluating the strengths and weaknesses of the various components of that theory that have emerged along the way. I would add the interpretative comment that the theory has become steadily more multi-level as it has moved from macro level appraisals to incorporate more micro elements. This of course is related to the points I have made already, about the move from economics and finance towards management and strategy, and interest in more micro and individual aspects of the subject from scholars with backgrounds in sociology or psychology. While Hymer and Dunning began by analysing patterns of foreign direct investment (FDI) at the country and industry level, and Vernon explained cross-country patterns of trade and FDI, those such as Buckley and Casson or Johanson and Vahlne shifted attention to the firm level, and the more recent scholarship of those like Birkinshaw has brought this down to the firm sub-unit or subsidiary level. More recently still, work has begun at the project level, which is appropriate where intra-firm networks become increasingly interconnected with inter-organisational networks, as hinted at in Ietto-Gillies's reference to the trend away from internalisation and towards externalisation (outsourcing, subcontracting and the like). Our theories of international business need increasingly to be adapted to handle a synthesis of these various levels. While it is quite common to discuss multi-level analysis simply in terms of the required statistical methodology, the point I would emphasise here is that it calls for theory which is complex and multi-disciplinary in nature. Therefore, it is unsurprising that the trends in the international business field towards the explicit incorporation of different levels of analysis on the one hand, and towards more multidisciplinary and interdisciplinary approaches on the other, have gone hand in hand with one another.
The recognition of a change in the international business environment from internalisation towards externalisation leads me to two other comments on Ietto-Gillies’s article. First, Ietto-Gillies is right to draw attention to the conflation of ownership and control in the literature on international business from Hymer onwards. As conveniently also reflected in the definition of FDI used by the agencies responsible, an ownership stake in a business abroad above some threshold share of equity is supposed to be both necessary and sufficient to ensure control over the management of that (thus) subsidiary company. In the earlier era of internalisation it seemed quite appropriate to associate the ownership of assets with the capacity to control the use of those assets. Of course, it was always understood that sometimes such control might not be actively exercised despite the existence of a majority holding, or equally sometimes control might be exercised over enterprises such as dependent suppliers despite a lack of ownership in their business, but these were often regarded as minor qualifications – the exceptions that merely proved the rule. Today, however, we must acknowledge that firms often exercise control over much wider international business networks in forms that are commonly known as ‘global production networks’ or ‘global value chains’, in which substantial parts of the network or chain are not owned, but are effectively controlled or orchestrated by the flagship firm. This led Dunning, for example, in his later work, to shift away from the traditional definition of the TNC in terms of the ownership of income-generating assets abroad, and towards defining the TNC instead as a firm that takes the lead responsibility for the orchestration of international business networks (see e.g. Dunning and Lundan, 2008).

I mention this not because of any novelty about Dunning’s definition of the firm or TNC in terms of its network of control rather than its ownership of assets – this idea and discussions around it have been around for a long time – but rather because the Dunning and Lundan account shows how this conceptualisation of the TNC has now become central to mainstream thinking in the international business field. However, while many of us found conceptually attractive the Cowling and Sugden (1987) definition of the TNC in terms of the strategic coordination of production facilities across national borders, it has to be admitted that it is very difficult to operationalise this definition empirically, and this problem remains unresolved. This is why in most empirical work the firm continues to be treated as a legal entity that owns assets and employs people, which is usually how it is also required to report its operations as a firm, and it is not generally defined as the coordinator of a network of business activity that ranges well beyond the facilities it owns and the people it directly employs, through a variety of formal (contractual) and informal (cooperative) relationships. It may well be that, to empirically operationalise the construct of an informal business network coordinated by a TNC, we need to work at the project level rather than at the firm or corporate level, especially since these networks depend often on decentralised structures of coordination (unlike in the earlier Cowling and Sugden definition of the firm in terms of coordination from a single centre).

As an aside, the new trade theories discussed by Ietto-Gillies are less able to explain the spread of global production networks than are the new firm-level approaches to international trade that have emerged from more empirically oriented economists working on trade and TNCs (e.g. Feinberg and Keane, 2006). In this perspective, which is echoed in recent work in international economic geography (see Iammarino and McCann, 2013), the expansion of both international trade and TNCs is to be explained – not by a change in transport costs or trade barriers (so long as these continue to remain relatively low by historical standards) – but by a change in inventory holding costs made possible through ICT-based innovation and the associated organisational innovations, most notably the just-in-time system. While the econometric demonstrations of this have been in terms of intra-firm or intra-
TNC trade (using the traditional definition of a firm in ownership terms), and this is still largely necessary for measurement reasons, the underlying explanation applies just as readily to the emergence and growth of global value chains that incorporate various partner or affiliated organisations into the wider international networks of the TNC. All this, of course, reinforces the point that Ietto-Gillies has rightly stressed from my own work, namely, that trade, FDI and contractual partnerships are largely complementary in processes of TNC growth, and indeed they have become ever more so in recent times.

A second observation here is that I doubt whether internalisation theorists would accept Ietto-Gillies's claim that they are at a loss to explain the trend towards externalisation in the last three decades. Casson, in particular, always saw the processes of internalisation or externalisation as entirely symmetrical, moving readily from one to the other as the nature of transaction costs shifted in either or both the market or non-market means of coordination of economic activity (see e.g. Casson, 1979). A large part of the explanation for changes in transaction costs within each mode of coordination would be the kinds of changes in the environment which Ietto-Gillies describes. While it is true that transaction cost economics has tended to focus on manager vs. manager or manager vs. shareholder (principal-agent) conflicts, it can also be applied to manager vs. worker conflicts in traditional class or industrial relations terms – Coase's main original point of reference was the employment contract with a firm. These conflicts can be examined in terms of the scope that exists for rent seeking behaviours within (or beyond) the firm. However, where I do think Ietto-Gillies's argument is well taken in her discussion of these issues is that the relevant transaction costs of alternative modes may be influenced by pro-active management strategies, and not just by an exogenous shift in the environment beyond the control of any individual decision taker. So strategies have co-evolved with the environment, and the move towards externalisation is, in part, deliberately designed to increase the capture of rents by strengthening bargaining positions. These aspects of active management and the pursuit of power (rents), rather than efficiency (profits), are indeed neglected in most transaction cost approaches in the international business field.

However, in considering the extent to which the established theories of the TNC can be equally well adapted to explain either internalisation or externalisation, I do see some greater difficulties with the evolutionary approach to internalisation originally set out by Kogut and Zander (1993), unless that is thoroughly reworked. Their theory depicts the firm as a social community characterised by certain shared values, which encourages and lowers the costs of internal knowledge transfer (relative to external transfer), and hence promotes the internalisation of knowledge development and exchange within the TNC. Yet we now appreciate that social groups or communities (business networks) can often be formed successfully externally as well as internally, as argued at least implicitly above with reference to the new definition of a TNC as a coordinator of international business networks with both internal and external elements. Indeed, contemporary social network analysis has more often adopted such a person-based rather than a strictly organisation-based notion of ties in assessing network relationships. So, while conventional internalisation theory can readily be inverted to become a theory of externalisation, this is not so evidently the case with Kogut and Zander's interpretation of the logic for internalisation in terms of organisational sociology. One can, of course, depict functioning business networks as social groups or communities, but generalising this approach to potentially apply equally to internal or external networks challenges the Kogut and Zander interpretation of the firm as a kind of privileged social community. If we try and avoid this difficulty by re-defining the TNC (like above) to consist of close social ties rather than the ownership of assets, then it might be objected that we would run into the same sorts of worries over tautology as have plagued the transaction cost version
of internalisation theory. The question would become the conditions under which close social ties and shared values come about externally as well as internally. Indeed, once we start down this route there is no reason on the other side of the story to suppose that social ties and shared values always exist or work well internally, especially in large and geographically disparate TNCs. Moreover, as noted earlier, social groups may still contain divergent interests, and so the original Kogut and Zander story tends to downplay the existence of potential intra-group conflicts and rent seeking (on the relevance of which, see Mudambi and Navarra, 2004). These aspects would follow more naturally from transaction cost reasoning, even if these accounts of the existence of the TNC have often been somewhat narrower in character than might have been appropriate (as just discussed above).

I think that Ietto-Gillies and I would agree that the long-term shift from a trend that was predominantly towards internalisation in the 1960s and 1970s to a trend that is predominantly towards externalisation today, has to be seen in terms of an historical evolution in the socio-economic paradigm that characterises the environment for international business. Ietto-Gillies seems to think of this change mainly in terms of the playing out of the conflict between capital and labour, and the interplay between TNCs and states (on which more below). Instead I see the new paradigm more broadly as the emergence of a new system of production, and consequently of the relationships between the actors within it, closer to the notion of paradigm change proposed by Freeman and Louçã (2001) and Perez (2002). The shift from a system of production driven by economies of scale in large plants towards one driven by economies of scope and flexibility in the information age, readily explains the change from internalisation to externalisation. (For a further discussion of these issues and their implications for the TNC, see Cantwell, 2013.) As part of the same process of the fragmentation of production and the fine slicing of the value chain, the nature of work has been transformed and the bargaining strength of trade unions has been weakened. These combined processes of change in the economy and society are reflected in the changing nature of organisational and social structures in the new system of production. However, I would not describe these changes as ‘external’ (to the firm) – since, as Ietto-Gillies also rightly emphasises, firms and the pursuit of their interests have been an integral or endogenous part of this economic and social change. So it would be more accurate to say that these changes reach well beyond the scope of the firm and its organisational form.

As Ietto-Gillies has correctly observed, as part of this same process of change, labour has become more mobile across firms, across sectors, and across countries. In the previous mass production system a worker might spend an entire lifetime in one type of job, beginning from an apprenticeship, while now the onus is on the individual’s capacity to become multi-skilled and to multi-task, which may lead to more varied opportunities for at least some people. The linkages between science and technology, and hence between firms and universities, have steadily risen as technology has become increasingly science- and information-based. As products have become more multi-technology in character, and technological knowledge itself within any field has become more complex, understanding (know-how and know-why) in any area of expertise increasingly requires access to a raft of supporting knowledge beyond the scope of specialists within a given field. This requirement to combine knowledge from more diverse domains necessitates a wider range of formal and informal ties to other firms and universities, and the need to develop a greater capacity to source knowledge internationally.

Turning to the eclectic paradigm or ‘ownership, location, and internalisation’ (OLI) framework, I think that Ietto-Gillies misses an important distinction between Dunning’s concept of ownership advantages and what later were called (by Rugman and others) firm-specific advantages. Ietto-Gillies follows a common belief that ownership advantages and
firm-specific advantages are equivalent, which belief I suspect came about around the time that attention shifted in international business scholarship from the country level to the firm level, as I described earlier. Yet both Dunning and Vernon had a notion of TNC capabilities that incorporated some collective elements in their home country of origin. The term ‘ownership advantages’ is the shortened version of what Dunning had called the ‘advantages of the nationality of ownership’. In other words, he intended to refer not to the ownership of assets (another common mistake) but to the advantages associated with having emanated from some specific home country. These would therefore include capabilities accessed through inter-organisational networks in the home country, and access to home country institutions, as well as capabilities held in-house in the TNC itself (firm-specific advantages). While in the internalisation era there seemed little need to emphasise this distinction, in the contemporary period of externalisation Dunning’s distinction becomes ever more vital and prescient, since capabilities are held in business networks and not just in in-house facilities. Likewise, location advantages are not just host country advantages, but refer to resources and capabilities associated with any unit of observation of a host location, as appropriate to the context examined – sometimes a sub-national region, or a cross-national region like the EU. I will come back to this issue in considering Ietto-Gillies’s discussion of the role of the nation state in international business below.

Another aspect of the eclectic paradigm that I think is worth emphasising is its flexibility and adaptability. This means that the way in which the eclectic paradigm has been interpreted and used over time has changed, and indeed has undergone more than one transformation as it has evolved (see e.g. Eden and Dai, 2010). Critics of the eclectic paradigm have often seen this versatility of the eclectic paradigm and its theoretical openness as a weakness, but in my view it has actually been its greatest strength and the reason for its continued centrality in the international business field. Originally, the chief objective of the eclectic paradigm was to provide a synthesis of the various economic theories of international business, and a framework within which they could be compared on some common ground where they offered genuinely competing explanations of a common phenomenon. Although it was probably unanticipated at the time, the eclectic paradigm is sufficiently general in nature that it has continued to fulfil a similar function but now in a broader analytical context as the domain of international business theory has expanded. Today, the eclectic paradigm offers a template for incorporating and relating a wider range of multidisciplinary perspectives and theories on the subject. It serves as the analytical means by which the field is still brought together and becomes more than just the sum of its various disciplinary parts. I am sure that this is a development that would have made Dunning very contented, since he placed great store on shifting the field of international business in a more interdisciplinary direction (Dunning, 1989).

Finally, I come to Ietto-Gillies’s remarks on the role of nation states, which rightly call our attention to the need to re-introduce the political science dimension into the multidisciplinary mix that constitutes our current thinking on TNCs, which has been relatively neglected since the time of Vernon (1971) – although those such as Dunning or Kobrin had been writing especially on governments and international business, and the role of public policies. Recent work by those like Henisz, Makhija or Cuervo-Cazurra has been linking TNC strategies to their interactions with governments and political structures, which has dovetailed quite nicely with more sociologically-grounded work on institutions and international business. So in one sense I suspect that Ietto-Gillies is here pushing at an open door into a branch of international business theory development in which a process of revival seems to be already under way and which is likely to draw in further research interest in the near future.
However, although it is certainly true that the borders between countries are political boundaries, I do think that it oversimplifies matters from an international business perspective to think of the crossing of borders by the TNC merely in terms of encountering a different governmental, regulatory and policy regime. Countries and regions of the world have been separated by the barriers of geographical distance for so long in history, that the constraints of distance have only gradually begun to diminish over the past few hundred years, and especially since the transport and communications revolutions of the mid-19th century which Ietto-Gillies mentions. From this longer-term perspective, the emergence and growth of nation states in this same historical epoch since the Middle Ages is a reflection of these – by now inherited – human and cultural boundaries, rather than the reason for them. What those of us also in the innovation studies field call national systems of innovation are differentiated not just because the system of government and regulatory structures are distinct, but because of a range of other associated formal institutions, and an even more complex set of informal institutions or ways of doing business. The connections between firms, and between firms and non-firm actors vary greatly across countries, and not just for the reasons of the specificities of government policy and regulation. It can be argued e.g. that the reason inward FDI penetration in Japan is so low for its level of economic development has little to do with formal barriers or regulatory constraints on foreign enterprise, but rather with distinctive ways of doing business and forming inter-organisational network relationships that are often misunderstood or misinterpreted by Western firms. Moreover, as alluded to in passing earlier, once we mention national systems of innovation we come naturally as well to the role of regional systems of innovation, both at the sub-national level and at the supra-national level. While each of these levels of regional entity also has its administrative authority structures, they too are not fully described as locational units by these political and regulatory features. So although Ietto-Gillies is certainly right to ask us to bring the role of states more into our discourse, in my view it overstates matters to say that this is the only aspect of locational variety encountered by TNCs which is not experienced also by purely domestic firms. As geographical distance rises, so does institutional variety and differences.

When Ietto-Gillies speaks of the advantages that TNCs have in their negotiations with nation states, owing to their greater degree of centralised power to move resources across global space, I believe she describes best the situation for the finance function, which is one aspect of the system of production that has also been caught up in the effects of socio-economic paradigm change mentioned earlier (see Perez, 2002). Thus, there are contemporary debates over whether governments can find a way of effectively taxing TNCs where they have most of their productive activities and their sales, and of curbing the use of offshore financial centres that offer low or zero tax rates. However, nation states are in a relatively stronger position when it comes to the regulation of markets and competition – at least in larger countries, and when it comes to access to knowledge and skills – states have more leverage everywhere. Indeed, TNCs may suffer from regulatory confusion or inter-country disagreements, where there are competing authorities with claims, e.g. in the case of mergers and acquisitions (M&As) that are approved in the US but not in the EU, or vice versa. As well as the benefits of globally dispersing their value chain activities, this has also led to an increased vulnerability to global risks for TNCs, in the case of a natural or environmental disaster or a major civil or political disturbance in any part of a global supply chain.

Rather than seeing matters simply in terms of the degree of bargaining power that nation states can exercise in extracting rents from TNCs (although I understand the point about all actors contributing fairly to the tax base of a country), we might ask more generally under what conditions countries can benefit from globalisation, just as TNCs do. I would suggest that a virtuous circle may be created where they cultivate some centres of
specialised skills and knowledge in a locally differentiated field of expertise; these centres then exchange knowledge globally with poles of suitably complementary capabilities elsewhere. This is how contemporary global cities have been revitalised or emerged, drawing upon a variety of economic and business ties with other such cities. The form of institutional environment that is most conducive to such internationally interconnected, but locally specialised, growth might be described as a relatively open society, rather than a relatively closed one. Once again this openness refers to a wide variety of formal and informal institutions that characterise a society, so what I mean by openness is very far from being reducible (say) to a country's trade and FDI regime. It is above all an openness to new knowledge and ideas, from outside as well as from within. In this context, we should not see cultural and institutional diversity across countries as a barrier to be overcome, but rather as a rich source of diversity, the combination of which can generate advantages both for TNCs and countries. It should be further noted that for TNCs to take advantage of such cross-border diversity, they require decentralised and distributed organisational structures, rather than the centralised direction of resources that may apply with respect to the finance function. This logic of organisational decentralisation is a critical part of the current trend towards externalisation emphasised earlier. I might add that, as quite rightly argued by Shenkar (2001), ‘cultural distance’ is a misleading concept. The commonly used measures of this construct are even more misleading, not least as the liabilities of foreignness are not symmetrical – it is often easier to move in one direction than the other, meaning a substantial disparity between inward and outward FDI. To take advantage of international variety, TNCs need to learn about the distinct ways of doing business in each location, and this requires them to become locally embedded and to build up the status of a local insider in business networks.

Therefore, I would argue that TNCs need to be understood, not just in the context of nation states, but more generally in the context of locational diversity of various kinds (including differences in political and regulatory environments), in which the degree of locational diversity across countries is of a qualitatively different order of magnitude to that experienced by domestic firms, even in a large country. For international business and, in particular, for the innovative TNC, the main reason why this matters is due to the far greater diversity in the settings encountered for interaction with local capabilities across locations, and in the distinct nature of those locally differentiated capabilities. Operating in such a diverse set of environments, in different national systems of innovation, provides TNCs with the opportunity to create more diverse kinds of knowledge, and to discover a much broader range of new combinations of knowledge. This brings us back to my earlier discussion of the change in the nature of the TNC itself as an actor, since TNCs are distinguished from other firms by what Kogut and Zander (1993) called their combinative capabilities. These have gradually enabled TNCs to serve increasingly as system integrators across international networks that connect a series of other actors, each with very different kinds of knowledge and capabilities. So, as Ietto-Gillies rightly remarks, our notion of TNC strategy should not be confined to rivalrous interactions with others, but must increasingly recognise and emphasise strategies with respect to cooperative relationships in local and international business networks (including those with non-firm actors). As she suggests, these may, in turn, be a major source of advantages relative to a TNC’s major competitors.
References


SUGGESTED CITATION:


http://www.worldeconomicsassociation.org/files/journals/economicthought/WEA-ET-3.2-Cantwell.pdf