Entrepreneurial Ecosystem Co-creation, Appropriability and Clusters: A Conceptual Framework

Christos N. Pitelis
University of Cambridge
Judge Business School
Trumpington Street
CB2 1AG
e-mail: c.pitelis@jbs.cam.ac.uk

Submission for the 52nd European Congress of the Regional Science Association
International

* We are grateful to Giovanni Dosi, the late John Dunning, and to Mark de Rond, Neil Kay, Anita McGahan, Bart Nooteboom, Roger Sugden, David Teece, Paul Tracey, and Konstantinos Vrettos for useful comments and discussion on earlier drafts. Errors are ours.
Entrepreneurial Ecosystem Co-creation, Appropriability and Clusters: A Conceptual Framework

ABSTRACT
Extant literature on clusters underplays the role of entrepreneurs and entrepreneurial management in creating and co-creating organizations, markets and supporting ecosystems. We employ transaction costs, resource-knowledge-capabilities and power-control-based theories to first provide a comparative static, governance-based perspective on clusters. Towards this purpose we employ the notion of dynamic transaction costs and propose the term ‘joint productive opportunity’. We then build on this to propose an entrepreneurial theory of the emergence, evolution and co-evolution of markets, ecosystems, and clusters. We explain the emergence of clusters in terms of entrepreneurial activity motivated by value appropriation. By increasing the overall value co-created, and given each firm’s value appropriation architecture, participant firms in a cluster benefit more from operating within a cluster than without. We suggest that clusters can involve advantages that help engender superior appropriation of value as compared to alternatives. We claim that clusters like other forms of economic organization are the outcome of purposive entrepreneurial action that involves the creation and co-creation of organizations, markets and supporting ecosystems. In this wider context, the emergence and resilience of clustering is explicated in terms of value appropriation, not just value creation potential. We suggest that in cases where a degree of strategic choice is involved, the choice of location by entrepreneurs can help trigger a market development and ecosystem and cluster co-creation process. In this context, the process of cluster co-creation is partly endogenized. Entrepreneurial managers, faced with a degree of choice, will help co-create clusters and be part of them, for as long as they can appropriate more value in this way than through alternatives.

KEYWORDS: Clusters, comparative governance, entrepreneurship, market and ecosystem co-creation

JEL Classification Codes: L26, L14, L22, N20
1. Introduction

Over the past twenty-five years or so, the literature on “clusters” has grown rapidly. The diagnosis and upgrading of clusters is now arguably one of the most favored forms of supply-side competitiveness policy worldwide (Malmberg and Maskell, 2002; Martin and Sunley, 2003; Porter, 2003; Matthews, 2010). However, the conceptual foundations of “clusters” have remained weak. This is in part because the theory of clusters has been based on their absolute advantages and not on a comprehensive comparative governance-based framework that compares clusters to alternative forms of economic organization, such as markets, hierarchies and other modes that can involve co-operation, particularly inter-firm cooperation (IFC). Perhaps more importantly, the literature on clusters has focused on the issues of co-location and social embeddedness. In so doing, it has failed to leverage ideas from strategy and entrepreneurship scholarship, notably the role of markets and ecosystem co-creation, in order to explain the emergence, evolution and co-evolution of clusters.

We draw on the aforementioned ideas to first develop a comparative static, governance-based theory of clusters. We then build on this to propose an entrepreneurial, co-evolutionary theory, based on the concept of entrepreneurial “market and ecosystem co-creation”. In this theoretical framework, an important reason why clusters emerge is the capabilities of entrepreneurial management to orchestrate market and ecosystem creation and co-creation. We also offer propositions and discuss case examples of cluster development, in line with our proposed theories. In section (2), we critically assess extant theory on clusters. Section 3 provides a comparative static, governance-based conceptual framework, that builds on ideas from transaction costs, the resource-knowledge-capabilities-based views, the industrial organization (IO) market power versus efficiency perspective, and the economics of agglomeration and co-location. Following on from this, we propose an appropriability-informed entrepreneurial theory of clusters, based on the concept of market and ecosystem co-creation. The concluding section discusses limitations and scope for further research.

2. Extant theory on economic organization and “clusters”

2.1 Introductory and Definitional Issues

Interest in, and literature on, agglomerations, “clusters”, “networks”, “industrial districts”, “webs”, local production systems, regional systems of innovation, innovative milieux, neo-Marshallian nodes, etc, was reinvigorated by Michael Porter (1990, 1998a, 1998b). Cluster creation policies have risen from relative obscurity to being seen by some as the only, or best, supply-side strategy for the international competitiveness of regions and
nations (Porter, 2000; Mesquita, 2007; Matthews, 2010; Engel and del-Palacio, 2011; Delgado et al., 2011)

Work on cluster-type forms of economic organization go back to Alfred Marshall (1920), who stressed certain benefits from co-location such as the availability of skilled labor and knowledge. More recent contributions include work by, among others, Enright (1996), Sabel (1989), Saxenian (1994), Markusen (1996), Gordon and McCann (2000), Breschi and Malerba (2001), Maskell (2001), Martin and Sunley (2003), Mesquita (2007), Bell et al. (2009), Arikan (2009), Matthews (2010). On the whole, such literature focuses on advantages of clusters in terms of co-location, social embeddedness and value creation.

There are many attempted definitions of “clusters”; Porter's (1998a) views clusters as geographical concentrations of inter-connected companies, specialised suppliers, service providers, firms in related industries, associated institutions (for example, universities, standard agencies, trade associations) in a particular economic field, that compete but also cooperate (see e.g., Porter, 1998a: 77-90; 2000). Four major elements of clustering have emerged from the theory and the evidence discussed in the literature. These are geographical agglomeration (co-location), linkages, and competition with co-operation (co-opetition). A fifth element refers to a perceived shared objective or vision by cluster members. I offer the following definition of clusters to cover all these characteristics:

“Clusters” are geographical agglomerations of firms in particular, related, and/or complementary, activities, with perceived shared objectives, and exhibiting horizontal, vertical intra- and/or inter-sectoral linkages, operating within a supportive socio-institutional setting, and co-operating and competing in national and international markets.

The above is the definition of a developed or mature cluster, such as the Cambridge high-tech cluster, Silicon Valley or the Hsinchu science-based cluster in Taiwan. Real-life clusters are usually the result of a path-dependent, evolutionary, socio-political, historical and entrepreneurial process that unfold in real time (Locke, 1995; Feldman, 2001; Feldman et al., 2005; Simmie and Martin, 2010). As such, they will tend to differ as to the extent to which they are characterized by all the above features (Gordon and McCann, 2000; Arikan, 2009). Gordon and McCann for example identify three ideal-typical models of ‘clustering’ – the pure agglomeration, industrial complex, and social network models. They suggest that each model, or ideal type, exhibits different logic and characteristics (including the nature and degree of IFC). In addition, elements from the three ideal types can co-exist in some real life clusters.
A common explanation of clusters involves the listing of a number of absolute advantages of clustering. These include agglomeration and external economies from co-location, the concentration of skilled human resources in the region, social embeddedness and capital that reduce transaction costs due to trust, the flexibility and entrepreneurship of small firms involved in clusters, economies of diversity, as well as the existence of untraded interdependencies. The latter are said to emerge from dense inter-firm linkages and networks of associations and interest groups. The last mentioned, along with the associated creation of “organizational fields” (Scott, 1994), are said to be capable of reducing “cognitive distance” (Aoki, 2010). Other explanations include benefits from aggregating diverse interests and mediating conflicts and diffusing information (Locke, 1995; Krugman, 1991; Gordon and McCann, 2000; Nootenboom, 2008). In addition to such “supply-side externalities”, further benefits of clustering include the reduction of consumer search costs and associated demand side effects (McCann and Folta, 2009). On the negative side, there can be agglomeration dis-economies (due, for example, to “congestion effects”) and also the possibility of institutional and organizational sclerosis-inertia and lock-ins. For example, “embeddedness” can help create such advantages that eventually become disadvantages (Robertson et al., 2008). In some cases success can eventually engender failure (Martin and Sunley, 2003, 2006; Matthews, 2010). More recently, knowledge and associated cognition-based theories of clusters aim to explain clusters in terms of their knowledge-value-creation advantages through enhanced associational cognition, learning, increased variation, and deepened division of labor. In this framework the additional value created through clustering is said to be able to offset any costs (Lawson, 1999; Maskell, 2001; Aoki, 2010).

The early literature on clusters lacked a solid conceptual foundation (Martin and Sunley, 2003). The classic works on “industrial districts” (e.g., Becattini, 1978; Brusco 1982, 1986; Pyke et al., 1990), were rich in the historical, descriptive, social, cultural and institutional detail of particular industrial districts, such as Prato (in textiles), Bologna (in machine tools), and Sassuolo (in ceramic tiles). However, this literature lacked well developed theoretical foundations, especially of the comparative governance type. The focus has been on the absolute advantages of “industrial districts”. Despite its superior analytical sophistication, the more recent literature on increasing returns that drew on Young (1921), the new economic geography (e.g., Krugman 1991, 1998a, 1998b; Audretsch, 1998), urbanization and localization economics (Dicken, 2007), and “clusters” (notably Porter 1990, 1998a, 1998b) has focused on the absolute advantages of clusters, not their comparative ones. Porter (1998a, b), for example, explored in some detail the positive impact of such advantages on
innovation, productivity and competitiveness of the cluster, region and nation (Delgado et al., 2011).

2.2 Perspectives on Inter-firm Cooperation (IFC) and (their Implications on) Clusters

IFC can be defined as quasi-stable and durable, formal or informal arrangements between two or more independent firms, aiming to further the perceived interests of the parties involved. In this generic definition, IFC involves independent firms (hierarchies) that pursue their interests, without resorting to full integration, but in a way that involves more stable relationships than a spot-market contract; IFC is thus “between market and hierarchy” (Williamson, 1996; Ménard, 2004). IFC can take various forms, such as subcontracting/outsourcing equity joint ventures (EJVs), strategic alliances, strategic networks, “clusters”, etc. (Child and Yan, 1999; Dussauge and Garrette, 1999; Gulati et al., 2000; Ménard, 2004; Nooteboom, 2008). An important characteristic of such arrangements is that to varying degrees co-operation can co-exist with competition between firms, especially when co-operating firms are involved in similar activities (Teece, 1992). In cases like strategic alliances or alliance portfolios, for example, cooperation can be in one activity, while firms compete in other areas (de Rond, 2003; Child et al., 2005; Ozcan and Eisenhardt 2009; de Figueiredo and Teece, 1996).


We can distinguish three major analytical perspectives on IFC: the “industrial organization” (IO), the transaction costs and the resource-knowledge-capabilities-based. There also exist variants within each perspective. Given extensive coverage, e.g., in Faulkner (1995), Child and Faulkner (1998), Mockler (1999), Gomes-Casseres (2001), de Rond (2003), Williamson (2008), Bell et al. (2009), Arikan (2009), here we provide a bird eye’s view.

From the IO perspective, IFC is seen partly in terms of anticompetitive activity. Firms have an incentive to cooperate so as to raise price-cost margins (Cowling and Waterson,
Game theoretic work on IFC such as Axelrod’s (1984, 1997), provides additional important reasons and evidence as to why firms may have an incentive to co-operate - in terms of retaliatory (“tit-for-tat”-type) strategies of economic agents (Child and Faulkner, 1998; Parkhe, 1993). Other, efficiency-based, arguments within IO, such as Demsetz’s (1973), can also be used to explain IFC in terms, for example, of production-side, synergy-related efficiency gains. However, within standard industrial organization, there is no comparative governance-based analysis between of modes of cooperation and competitiveness.6

The origins of the transaction costs explanation of IFC go back to Coase’s (1937) classic article. Coase focused on the nature, or existence, of the firm, vis-à-vis the market, and attributed integration by firms to high market transactions cost. Following the emergence of debates on co-operation, Oliver Williamson (1985, 1996) originally viewed such co-operation (and what he called hybrids) as a transient phenomenon, a progression from market to hierarchy.7 In this perspective, there can be transaction costs benefits of co-operation vis-à-vis markets, but generally not vis-à-vis hierarchy. Some explanations of “clusters” rely on transaction costs arguments (see Bell et al., 2009). Important contributions in the transaction costs tradition involve Hennart’s (1991) analysis of joint ventures, Oxley’s (1997) analysis of alliances, and Williamson’s (2008) analysis of outsourcing. An implication of the transaction costs perspective is that ceteris paribus, in transactional terms (and subject to increases in organizational costs not fully offsetting any transaction costs saving), IFC is an inferior form of organizing economic activity in terms of economizing behavior than integration by firms - hierarchy.8

The transaction costs analysis is predicated on the assumption, that the “production side” can be assumed to be constant for different modes of the organization of economic activity (Williamson, 1985). Remarkable but rarely commented upon assumption is that production costs don’t change with different organizational arrangements, only governance costs change. This need not, and realistically cannot be the case (Langlois and Robertson, 1995; Aoki, 2004). Different governance arrangements, are likely to possess different capabilities, advantages and disadvantages. The resource-knowledge-capabilities-based perspective is based on the explicit observation that institutions and organizations differ in terms of their value-creation and capture capabilities, (Teece, 1986; Gulati and Singh, 1998). For Penrose (1959), firms are superior to markets in terms of their endogenous creation of knowledge, innovation and value. For other “resource-based” and capabilities scholars, for example Peteraf and Barney (2003); Mahoney and Pandian (1992); Kor and Mahoney (2000),
Dosi et al. (2000); Teece et al. (1997); Teece (2007), Loasby (1999, 2010), each firm, too, is unique in its ability to create and appropriate value (Eisenhardt and Schoonhoven, 1996).

If one recognizes that the production-value creation side varies with governance mode, then one needs to analyze the differential advantages and disadvantages of each governance mode both in terms of its costs and in terms of value-creation and capture benefits. For example, even if IFC is inferior to integration in terms of transactions cost, it may be superior in terms of value creation benefits (Kale et al., 2001). This indeed is a basis of the knowledge-based theory of clusters (Lawson and Lorenz, 1999; Lawson, 1999; Kim and Park, 2002; Lorenzen and Maskell, 2004; Malmberg and Maskell, 2002; Maskell, 2001; Tallman et al. 2004; Storper and Venables, 2004; Arikan, 2009). The pertinent cost benefit calculus is more complex than that based on the transactions cost alone. (Colombo, 2003). Penrose did not originally deal with intra-firm co-operation. Later, she recognized its importance (Penrose, 1995) and her last published paper focused on networking and IFC (Penrose, 2008). She felt that networking blurred the boundaries of the firm and attributed networking to its production-side benefits. She did not explicitly consider the transactions cost side. In addition, despite reference to Richardson’s seminal piece, Penrose did not take up Richardson’s resource-capabilities-based analysis of the division of labour between market, integration and co-operation.

2.3 A Comparative Governance Perspective?

As Coase (1937), Richardson (1972) and Williamson (1975) have convincingly argued, the absolute advantages (or costs) are less relevant than the comparative ones for purposes of understanding economic organization. The argument that clusters may reduce transactions cost, because of “trust”, social capital, and lower cognitive distance, (vis-à-vis arm’s length transactions), need not imply that clusters can do so more than “integration” (“hierarchy”). Similarly, Porter’s discussion of the benefits of “related and supporting industries” does not address the question whether and why an integrated larger firm does not possess the same, or even larger benefits, than a cluster, either through full integration or by creating or co-creating a production system, like for example Toyota, or Research in Motion in Southern Ontario. As noted, even the impact of “embeddedness” in industrial districts on innovation, can be positive but also negative (Robertson et al., 2008). In all cases, unless one explains the comparative advantages and disadvantages of clusters, one only has a partial and incomplete explanation of the phenomena in hand.
Despite progress, it is arguable that a remaining problem with much of the extant literature is that it tends to focus on the absolute advantages of clusters, and fails to adequately address the question “why clusters?” relative to alternatives, notably “integration” by firms and as compared to markets and other types of economic organization, involving cooperation between firms. In addition, and importantly, extant theory fails to adequately address the role of entrepreneurs and entrepreneurial management in creating and co-creating clusters, as well as organizations, markets and ecosystems, with an eye to appropriability. Indeed, it is arguable that appropriability and value capture are all but ignored in cluster theory – the focus instead being on efficiency and value creation. We suggest this to be a major limitation. Below we aim to fill these important gaps in the literature.

Building on Penrose (1959), Richardson (1972) produced a template, or “good practice”, for the choice of mode of organizing economic activity, in terms of his proposed concepts of similarity and complementarity of activities. He defined similar activities as those that require the same capabilities, while complementary activities as those that require complementary capabilities. In Richardson’s schema, similar and complementary activities are best integrated in a single firm. Dissimilar yet complementary activities are best undertaken through co-operative arrangements. Markets are best, when activities are both dissimilar and non-complementary.

Richardson’s contribution was arguably the first conceptual framework aimed at explicitly dealing with IFC in a comparative institutional manner. It preceded Williamson’s (1975) classic book; it considered three modes of economic organization; and importantly it employed the now popular capabilities lens (Hamel, 1991; Parkhe, 1991; Teece et al., 1997; Inkpen, 1998; Loasby, 1999; Dosi et al., 2000, 2008; Dosi and Marengo, 2007; Teece, 2007), and the concept of activities (Porter, 2008).

Seminal as it is for its timing and originality, Richardson’s contribution has weaknesses. His focus was exclusively on production costs efficiency in terms of requisite production capabilities. He did not deal explicitly with other potential advantages of alternative modes, for example those arising from the “revenue” side. However, the revenue-value creation side is important per se. Different modes may have differential advantages, for example in knowledge, learning and value creation. The currently extensive knowledge-based view of clusters (Maskell, 2001; Arikan, 2009) draws on, and supports, this view. Richardson did not deal explicitly with this aspect. While he developed a Penrosean, resource-capabilities-based view, he went both beyond Penrose (by focusing on production cost
efficiency) and at the same time ignored Penrose’s important contribution; her focus on the revenue side of firms and their ability to generate knowledge and innovation better and/or faster than markets. In this context, Richardson’s analysis complements the transactions cost approach, by looking at the production efficiency side. On the other hand, it did not consider explicitly transaction costs, knowledge-creation, as well as power-control-related factors.\textsuperscript{10}

As noted above, much of extant work on clusters focuses on the perceived absolute advantages of clustering. The work of Michael Porter (1998a, 1998b), demonstrates no conceptual framework as such, and makes no reference to transactions cost, resources, capabilities or the work of Richardson. Accordingly, Porter’s work is \textit{sui generis} and detached from important contributions on IFC and other contributions on clusters. Arguably, the useful elements of Porter’s work need to be placed in context, and combined with other insights, for further progress to be made. Recent insights from the transactions cost lens (Bell et al., 2009) and/or the knowledge-based one (Maskell, 2001; Arikan, 2009), go beyond Porter’s early analysis. However, they do not attempt a comparison of the advantages and disadvantages of IFC and clusters vis-à-vis markets, integration/hierarchies, and IFC, in terms of a comprehensive set of available conceptual perspectives, both in a static comparative governance and in a co-evolutionary, appropriability-informed entrepreneurial framework.

We may conclude from the above that to-date there exists no comprehensive comparative governance-based framework on clusters, that leverages extant literature on economic organization and IFC. More importantly, there is no theory of clusters based on appropriability-informed entrepreneurial agency, in engendering organization and/or market and ecosystem co-creation. We attempt to address these limitations in the next section.

3. Towards a Comparative Static, Governance-Based and an Entrepreneurial Theory of Clusters

3.1 A Comparative Static, Governance-Based Approach

The type of IFC present in the cluster may simply be “the way things happen around here”, history-specific and path-dependent. In cases, however, where a higher degree of choice is present, (such as in EJVs between firms, especially of different nationalities, such as Western and Chinese), non-integration can be due to government restrictions or other external impediments, for example unavailability and/or unwillingness of targets to be taken over (Williamson, 1975; Child and Yan, 1999). Such “negative” reasons or constraints are possible and can be due, for example, to policies by governments. However, not all or even most,
instances of IFC are the outcome of constrained choices. In many cases of EJVs, strategic alliances, or the choice of licensing over foreign direct investment (FDI), IFC is chosen by firms as preferable to full integration (Hennart, 1991; Faulkner, 1995; Oxley, 1997; Inkpen, 1998; Gomes-Casseres, 2001; Colombo, 2003; Pitelis and Teece, 2010). In addition, the choice of spot contracting is always available, at least theoretically, under Williamson’s (1975) assumption that markets are seen as the baseline default case\textsuperscript{11}. The question of why IFC can therefore be due to affirmative reasons.

Below we first aim to explicate the choice of mode, for cases where a degree of choice is available, by employing three important concepts from extant literature. The first is the concept of “productive opportunity”, introduced by Penrose (1959); the second is “dynamic transaction costs”, employed by Langlois and Robertson (1995); and the third is power and control (including market power that involves the reduction of competition), employed by Hymer (1979), Bain, (1956), and Porter (1980, 1987). In all, these three concepts encapsulate the essence of the three major perspectives discussed so far (power-efficiency, transaction costs and resource-knowledge). Leveraging these concepts allows us to develop a more comprehensive comparative governance-based framework. This is because the extant literature employs at most a comparison of two perspectives (e.g. Hamel, 1991: Parkhe 1993; Eisenhardt and Schoovenhoven, 1996; Colombo, 2003). Having done this, we then employ Richardson’s concept of similarity and complementarity of activities, to consider whether and how this modifies our previous baseline analysis.

Penrose (1959, 1960) defined “productive opportunity” as the managerially- perceived dynamic interaction between the internal environment of the firm and its external environment. As such “productive opportunity” incorporates issues of external competition/power \textit{à la} IO, but also cognition, internal resources and capabilities. Penrose claimed that “productive opportunity” determines a firm’s growth and the limits to such growth. While she did not deal with the “nature” of firms per se, it can be claimed that the Penroseean advantages of firms vis-à-vis markets, in terms of knowledge creation, innovation, endogenous growth and productivity can serve as a reason for the existence of the firm, over and above transactions cost considerations. IFC allows firms to leverage their respective “productive opportunities”. The sum of these productive opportunities, or the combined or joint productive opportunity, can exceed the productive opportunities of each separate party, and may exceed the sum of the two productive opportunities of the firms, if these do not co-operate.\textsuperscript{12} This can be due to complementary resources, other synergies and importantly to the combined knowledge-base and vision of the co-operating firms. Market transactions are
inferior to co-operation in terms of this argument. Two independent firms cannot normally, through spot contracts, gain access to the resources, knowledge and vision of the other firm to the same extent (Parkhe, 1991; Inkpen, 1998; Kale et al., 2001). In addition, ceteris paribus, coordination through the market involves higher transactions cost. It follows that in terms of both “combined productive opportunity” and transaction costs, IFC will tend to be superior to spot market contracts.

When comparing IFC to integration, however, our previous observation is modified. Ceteris paribus, IFC involves higher market transaction costs than integration. If these exceed any increases in intra-firm transactions cost (organizational costs) due to integration, hierarchy will tend to be superior to co-operation in transactional terms. The case of the “joint productive opportunity” is more complex. At one level, the integrated firm will have a more coherent joint productive opportunity: for example, consider the case of a multinational firm taking over an independent firm in another country, but letting it operate with its existing resources as a quasi-independent business unit, or subsidiary, so as to leverage its subsidiary skills (Mockler, 1999; Madhok, 2006; Papanastassiou and Pearce, 2009). This can give rise to an enhanced “joint productive opportunity” which is more “coherent” than that of two independent co-operating firms (Teece et al., 1994). Clearly, in some cases, an integrated entity may lose out in terms of “productive opportunity” potential if, for example, useful resources of the acquired firm are being lost to the new entity. However, on the whole, integration will tend to dominate IFC, both in terms of joint productive opportunity and transactions cost. The above analysis raises the question why does IFC exist at all. Richardson’s concepts of similar and complementary activities help moderate the baseline observations and address this question. If firms are involved in dissimilar activities, the joint productive opportunity of the integrated body may not be as well aligned as that of two firms involved in similar activities. This is because the dis-similarity of activities and their underlying capabilities can blur a firm’s vision, by engendering confusion as regard the firm’s focus. This argument is akin to problems related to unrelated diversification (Porter, 1987). In addition, the internal organizational costs of the integrated unit will tend to be higher. Even the control advantages of hierarchy are put into question if activities are dissimilar, not least because of knowledge co-ordination-related problems. On the other hand, when activities are dissimilar but complementary, IFC acquires logic (in that there is mutual potential advantage in acquiring complementary knowledge). In the case of dissimilar activities, IFC may be superior to integration. In the event of non-complementary and dissimilar activities, markets may be superior, as the integration of dissimilar and non-complementary activities is likely to
lead to high organizational costs (if integration makes any sense in this context, to start with) and IFC makes limited sense, as there is little of mutual benefit to be gained.

From the above it follows that from the point of view of a choice-theoretic perspective, it is only in the case where firms are involved in dissimilar activities, which are also complementary, that IFC may be superior to both market and integration. This results in Proposition 1.

Proposition 1. The net advantages of IFC in terms of transaction costs-joint productive opportunity can exceed those of integration in cases involving complementary but dissimilar activities.

The value of Proposition 1 is that it helps re-establish the Richardsonian contribution, under more general conditions than examined by Richardson himself. Moreover, it highlights some limitations of this perspective, notably its choice-theoretic pedigree and underlying assumptions of sufficiency of knowledge-information. Such a degree of strategic rational choice, and near perfect and full knowledge, are normally absent from reality (Loasby, 1996), or even from bounded rationality-based transactions cost analyses (Williamson, 1985). For example, there is little in the analysis of Richardson and Porter by way of an explanation of how resources are being recognized as complementary and/or similar and how control-power is exercised.

The work of Penrose (1959) can in part help address these lacunae. In Penrose production-specific and relevant knowledge is engendered intra-firm. This affords entrepreneurial managers the appreciation, and even imagination, of ways in which resources and activities can be similar and/or complementary, as well as ways in which control can be exercised through knowledge (Foucault, 1980). It is also instrumental in explaining how firms can develop and leverage know-how and supporting implementation skills and capabilities, not least through learning by doing. Such implementation issues are not adequately addressed in clusters literature (Klepper, 2010; Matthews, 2010).

As already noted, clusters can be seen as a special form of economic organization that can involve cooperation between firms and with other organizations and institutions (Saxenian, 1994; Porter, 2000; Cooke and Huggins, 2003; Aoki, 2004). Differences between “clusters” and forms of IFC such as EJVs, strategic alliances or subcontracting, including the existence in clusters of geographical co-location, a usually lower degree of formality of linkages, as well as a lower direction-control and strategic choice-intentionality. In addition, firms in some clusters, such as Silicon Valley and the Cambridge high tech cluster, need not
be involved only in complementary and dis-similar activities, but also in the whole gamut of similar, dissimilar, complementary and non-complementary activities.

In the context of our earlier analysis, the existence of the cluster can be attributed to the differential advantages of such special attributes, notably those arising from co-location and embeddedness, which are in addition to factors favoring IFC. Such advantages of clustering can limit integration, even when there exist similar and complementary activities and scope for rational choice. Indeed, the benefits of clustering for each firm may make them unavailable as a target of a take-over, thus reducing the population of potential targets for potential acquirers. Clearly integration will be taking place in a cluster, too, but full integration of all firms with similar and complementary activities may fail to realize for the reasons detailed above, even assuming willingness on the part of potential acquirers. This is the case even when a cluster is dominated and has even been co-created by a firm, such as in the case of the Toyota production system. There follows Proposition 2 below.

**Proposition 2. Clusters are a form of economic organization that can involve IFC, with net advantages that can render it superior to integration, even when cluster firms are involved in similar and complementary activities.**

To summarize, we proposed that clusters can be seen to as a form of economic organization that can involve IFC, that exhibit special characteristics and advantages as compared to other forms of economic organization, including integration (Aoki, 2001). In addition such characteristics-advantages can also lead to regional joint-productive opportunity effects, and indeed a ‘regional industrial identity’, (Romanelli and Khessina, 2005), that may suffice to outweigh the advantages of integration (and/or limit its availability), even when firms are involved in similar and complementary activities and there exists scope for rational choice. This runs counter to Richardson’s prediction, which failed to consider the wider benefits of clustering.

### 3.2 Appropriability, Entrepreneurship, and Market and Eco-system Co-creation

A limitation of the comparative static, governance-based perspective, but also of much of the cluster literature, is that it focuses on efficiency and value creation, with little explicit regard to appropriability. This is not for example the case in some IFC literature, which recognizes the importance of appropriability (Gulati and Singh, 1998). However, the constituent firms in clusters are usually involved in business for much the same reason as are other firms, namely to capture as much value as possible out of their perceived value creating
advantages, actions and action potential (Pitelis and Teece, 2010). A theory of clusters that
fails to consider value appropriation is incomplete, especially given that value creation need
not always lead to value appropriation (Teece, 1986). It is therefore important for value
capture to be addressed and for an explanation to be provided as to when and how cluster
firms can avoid what Augier and Teece (2008) call the ‘zero profit condition’.18 This also
brings centre-stage the issue of appropriability-informed agency, in the form of
entrepreneurship and entrepreneurial management.

This issue has not, to our knowledge, been adequately addressed in cluster literature.
An exception is Feldman et al. (2005), who note the need of entrepreneurs and cluster firms to
act so as to satisfy their business interests. Sacchetti and Sugden (2003) note that networks
can also acquire positions of power that serve the interests of their members. Aoki (2004)
oberves that knowledge exchanged in clusters such as Silicon Valley, is non-proprietary, yet
intensive interactions make clusters a more coherent institutional arrangement, than predicted
in terms of transactions cost analysis. While we share this view, we feel that the value capture
potential of clusters is underconceptualized. We submit that the viability and even the
emergence of clusters is predicated on two major factors that facilitate value capture. The first
is value co-creation, given the existing value capture architecture of each cluster firm. The co-
location and embeddedness of firms and other institutions and organizations in a cluster helps
create a supporting eco-system (Nelson, 1993), and co-create value, thereby increasing the
overall pie of created appropriable value. Given each cluster firm’s extant value capture
strategy/architecture, firms will benefit from operating in a cluster if and when the part of co-
created value they can appropriate is higher than the value they could have appropriated, had
they been able to appropriate more value but from a smaller pool. In turn, value capture
strategies are being developed by entrepreneurs who can set-up organizations (firms) and help
coopurate markets, eco-systems and clusters, so as to appropriate co-created value (Pitelis and
Teece, 2010). Differently put, when firms can capture more value through integration, or
indeed spot market transactions, and given the choice (which as we noted need not always be
available), they may opt out of a cluster. This results in Proposition 3.

Proposition 3.Given the choice, entrepreneurial managers will help co-create a cluster and
operate within it if/when the value they can appropriate through requisite
value capture strategies by standing alone, is lower than the co-created value
they can capture by being part of a cluster.
Besides addressing the appropriability issue, Proposition 3 also deals with the implicit assumption that cluster firms can either be part of the cluster, or just not exist. This need not be the case in all cases, especially for larger firms (but even smaller ones), that could decide to relocate, and they sometimes do (Faulkner, 1995; Gomes-Casseres, 2001). An example is the ceramics firm Royal Doulton of the North Staffordshire cluster, which had fully relocated – outsourced its production (Sacchetti and Tomlinson, 2006). This renders important an analysis of factors that induce outsourcing yet keep certain management and design activities in the original location.19

Our focus on appropriability-driven entrepreneurship also brings centre stage the issue of dynamics. In real life firms, IFC, clusters and indeed markets emerge, evolve and co-evolve in real time (Nelson and Winter, 1982, 2002). This invites a co-evolutionary perspective that starts from first principles. These involve the recognition of human, here mostly entrepreneurial, activity under conditions of uncertainty, change, limited, bounded and/or procedural rationality, thin or absent markets and human action, which can be adaptive, reactive, proactive, and indeed anticipatory (Mises, 1949). Uncertainty can be radical, that is not subject to probabilistic calculus (Knight, 1921), and entrepreneurial perception and indeed imagination can drive human action (Loasby, 1996, 2010). In such a world, human agency, environmental structure, and the nature, objectives and evolution of organizations and institutions are co-determined and co-evolving (Nelson and Winter, 2002; Dosi et al., 2000; Dosi et al., 2003). Having said this, we submit that purposive human agency, here in the form of entrepreneurial action, is of the essence, and it should be our starting point (Penrose, 1959; Augier and Teece, 2008). This is because institutions and organizations cannot be created without humans. If human agency is critical, so are the objectives of economic agents that inform and motivate their actions of economic agents. We suggest that economic agents mostly aim to capture value out of their perceived appropriable value creating advantages, actions and action potential.20 The above is applicable in most contexts, albeit it can take different forms.21

In market economies an important way by which economic agents achieve their objectives is through the use of market transactions. These can involve the buying and selling of commodities, including the selling of labor. In the context of thick and relatively efficiently functioning markets, a way to profit is by identifying and/or creating and exploiting market opportunities (Klein et al., 2010).
The process of opportunity identification, creation and exploitation can be performed by individual agents-entrepreneurs. On many occasions such a process can be facilitated by the setting-up of an organization, such as a firm. Organizations can offer advantages which are not often available in markets. They can help co-create value through human co-operation within the cohesive shell of an organization more than through the market, for reasons relating to knowledge creation (Penrose, 1959), transaction costs and property rights (Coase, 1937; Williamson, 1975; Hart and Moore, 1990), enhanced market power, and power over labor (Marglin, 1974; Dosi, 1995), co-specialization of assets (Teece, 2007), and many other reasons. Such advantages can be sanctioned by the government (such as for example the bestowment of limited liability in certain business concerns). By leveraging the advantages of organization, economic agents (or teams thereof) that set them up (identified in the literature as principals or entrepreneurs), can take advantage of market opportunities, create such opportunities, and exploit market imperfections often more effectively than when they operate as stand alone individuals (Klein et al., 2010).

The advantages of organization become even more pronounced, in conditions where markets are thin and even non-existent. In such cases organization can help create and co-create markets, something that markets can simply not do. The Toyota production system is a case in point, but it is by no means unique. In some way or another, most firms create and co-create markets and ecosystems. This attribute of firms is arguably an important reason for their very emergence (Pitelis and Teece, 2010). It allows the economic agents that set them up to capture value. In this sense, organization is superior to the market from the point of view of facilitating the effectuation of entrepreneurial objectives.

The decision to set up an organization normally involves also a locational dimension. Despite advances in globalization and the information and communication technologies (ICT), the vast majority of organizations and firms involve a territorial dimension. This necessitates the choice of a location and an audit of its advantages and disadvantages. The concept of locational advantages has much currency in the International Business (IB) literature, notably in the work of John Dunning (1980, 1998). In the OLI (Ownership, Location, Internalization) framework, Dunning discusses extensively how locational advantages can help influence the choice of a (foreign) location by firms. Such locational advantages are also applicable within national boundaries and are employed here to further our argument.
In cases where a degree of choice is available, an aspiring entrepreneur (or team of entrepreneurs) will employ their perceived and even imagined, ‘productive opportunity’ (Penrose, 1959) to choose where to locate. This involves the best possible fit between the perceived advantages of the entrepreneurs and their chosen activity, and the location. The comparative advantages of the last mentioned can be in terms of resources, transport, communications and more generally factor-endowment advantages within the remit of the perceived-imagined value chain (and eventually value system) of the entrepreneur-to-be. Once a location has been selected on the above basis, it is arguable that we already have an ecosystem-cluster in the making. In the absence of mistakes or incompetence, the very selection of a location suggests that other entrepreneurs in similar and related activities may choose the same location for much the same reasons, but now also because of the added advantage of the region in terms of the existence of the pioneering entrepreneur-firm. In this context, the firm, the market, and the cluster are being co-created through entrepreneurial actions. Case histories of clusters such as Detroit, Silicon Valley, the Cambridge high tech cluster, and the Taiwan Hsinchu cluster, are fully in line with these processes (Klepper, 2007; 2010; Evans and Garnsey, 2009; Simmie and Martin, 2009; Matthews, 2010).

In the case of Akron, Detroit, for example, Buenstorf and Klepper (2009) note the “chance location of Goodrich in Akron in 1871” (p. 705). Klepper (2007) also emphasizes the importance of early locational choices by four early successful entrants, and goes on to suggest that successful spinoffs arising from disagreements, as well as organization reproduction and heredity (Klepper, 2010), can help explain the emergence of clusters better than agglomeration economies. Aoki (2010) also points to the role of spinoffs, resulting from voluntary exit of scientists.

In cases where markets are absent, a pioneering firm can help create the market. In the case of McDonald’s expansion to Russia for example, the company had to effectively create the market and eco-system for its around 300 ingredients required for its products. Often this involved setting up companies with an eye to eventually turn them into independent suppliers. In other cases, such as Toyota, the production system is being co-created with a higher reliance on pre-existing markets (Dyer and Nobeoka, 2000). In both cases, however, a cluster eventually emerges. The firm, the cluster, and even the market and ecosystem are co-created through entrepreneurial action. While this is more obvious in the case of top-down production systems, similar considerations apply for the case of organically emergent clusters. The perceived ‘productive opportunity’, defined to include the locational advantages of the
selected region, act as the catalyst for the emergence of the firm, the cluster, and in the cases of thin or inexistent markets – the market and ecosystem too.

Extant government policies and/or the presence of related and supporting organizations, such as a university or a supporting organization, can be part and parcel of the perceived locational advantages. Subsequent (or even prior) government intervention can help create and/or improve such locational advantages, thus facilitating the emergence of the cluster (Matthews, 2010). Both can be at least partly endogenous to entrepreneurial actions (Klein et al, 2010).

Once a cluster emerges, it is possible to explain integration and cooperation within and without clusters, in terms of the concepts employed in our comparative static framework, applied “in real time” (Langlois, 1992). This requires an appreciation of the fact that entrepreneurial action can be on the basis of anticipatory change, which itself can be partly endogenous. For example, McDonald’s internalizes first, with the intended purpose to eventually externalize, thus co-creating and anticipating the conditions that will allow it to apply its ideally preferred mode of franchising. However, in some cases, in order to franchise, it first needs to create the franchisees-to-be. Such behavior can also help explain why in such cases integration is not the preferred option and it does not take place even when firms operate in related and supporting activities. Integration is prevalent partly because co-location was pre-selected by entrepreneurs for its perceived intended and unintended advantages, including market co-creation. This reverses the usual chain of causality adopted in comparative static frameworks.

In all, productive opportunity by entrepreneurs-to-be that includes perceived locational advantages, alongside the advantages of control-power, dynamic transactions cost and capabilities of organization in effecting value appropriation, often through market and ecosystem creation and co-creation, help explain why firms and clusters emerge in real time (as well as why integration is not more common in cases of co-located firms operating in related and supporting activities).

The above leads to our fourth proposition,

Proposition 4. In real life and time, clusters result from entrepreneurial efforts to capture value from perceived ‘productive opportunities’, which often require the setting-up of organizations and the creation and co-creation of markets and ecosystems in particular locations.
In the above context, firms, clusters, markets and ecosystems are co-determined and co-evolving, albeit not contemporaneously, as a result of entrepreneurial pursuit of value appropriation. They involve purposive co-creation.

Our propositions are in line with experience in real life clusters. If one looks at the classic case of Silicon Valley, the Cambridge and Hsinchu clusters, extensive analyses suggest the existence of locational advantages (such as a Science Park or government contracts), which motivate some pioneers to locate in the region. Emergent grown clusters, often involve value co-creation advantages, and distinct value capture models and strategies. These presumably are perceived as preferable to cluster-based firms than alternatives, e.g. relocation, or integration. Matthews’s (2010) analysis of the Hsinchu cluster in Taiwan highlights the role of appropriability-informed agency and entrepreneurship, as well as public-private partnership for market extension and cluster creation. Even in cases such as the Toyota production system, the existence of a dominant firm, is not followed by integration and engenders value co-creation benefits. Acquisitions (in this case by Toyota) do not take place, as the co-creation benefits emerge exactly from the partial independence of the parties involved.22

In another case, that of the North Staffordshire ceramics cluster, while cluster firms were involved in similar and complementary activities, little merger activity had taken place for as long as international competition did not threaten the export markets (Sacchetti and Tomlinson, 2006). When merger activity did take place, it was mostly of a defensive nature, in response to the exogenous pressures discussed, and with questionable benefits (Padley and Pugh, 2000). Value capture strategies involved a focus on quality differentiation with a focus on the wider (not just local) domestic and global markets. As soon as that was undermined by the aforementioned factors, the cluster entered a phase of decline.

Similar considerations apply for other clusters. Yet, it would be dangerous to try to generalize from case studies. Instead, our focus here was to examine the analytical issues involving clusters and to address the problems flowing from the absence of a comprehensive comparative governance conceptual framework that is not just absolute advantage-based. The framework considers market, ecosystem and value co-creation and value capture.

4. Conclusions, limitations and policy limitations

The extant literature on clusters emphasizes their absolute advantages in enhancing efficiency and creating knowledge and value. Our contribution consists of the following: (1)
The adoption of a comprehensive comparative governance perspective (first static and then dynamic), that looks at the relative (not just absolute) advantages of clusters, in terms of all major conceptual lenses – transaction costs, power-control and resource-knowledge creation. Towards this purpose we employed the notion of dynamic transaction costs and proposed the term ‘joint productive opportunity’.

This analysis helped re-establish one of Richardson’s predictions, under more general conditions. It also helped highlight the limitations of this prediction and his overall analysis. Richardson’s work is all but ignored in cluster analysis. When combined with Penrosean insights, one can address issues of knowledge generation and implementation. We claimed that the differential advantages of clustering help explain the limited occurrence of integration in clusters, even when Richardsonian conditions favouring integration are satisfied.

We explained the emergence of clusters in terms of entrepreneurial activity motivated by value appropriation. By increasing the overall value co-created, and given each firm’s value appropriation architecture, participant firms in a cluster benefit more from operating within a cluster than without. By often ignoring value appropriation, extant theory on clusters makes no prediction as to whether and when firms will choose to co-create and keep operating within a cluster.

(2) We claimed that clusters like other forms of economic organization are the outcome of purposive entrepreneurial action that involves the creation and co-creation of organizations, markets and supporting ecosystems. In this wider context, the emergence and resilience of clustering is explicated in terms of value appropriation, not just value creation potential. This differs from extant literature that bases its predictions on clusters on value creation potential alone.

(3) We suggested that in cases where a degree of strategic choice is involved, the choice of location by entrepreneurs, can help trigger a market development and ecosystem and cluster co-creation process. In this context, the process of cluster co-creation is partly endogenized.

Our analysis and predictions are testable and in line with case study examples of real clusters. We recognize that it is hard to generalize on the basis of individual cases. Another limitation of our analysis pertains to the issue of cluster boundaries, life cycles and the case of global cluster linkages. We have not addressed these important issues, which are the subject matter of further research23.
Our analysis can help inform managerial practice and public policy. For example, it can help national and multi-national firms to decide whether and how to locate in a particular location. This involves trying to compare own ‘productive opportunity’ (or that of a subsidiary) to that of a location and/or a cluster’s and explore whether and how it can create synergies and co-create markets and eco-systems in a way that allows it to capture more value than through integration. Public policy can also be brought to bear by supporting practices that may enhance a cluster’s productive opportunity, help increase the cluster’s social capital, eco-system and innovation system and/or encourage or discourage intra-cluster integration. The Hsinchu cluster is a case in point. We aim to pursue further these opportunities and hope to motivate other scholars to conceptualize further and undertake empirical work on these very important issues.

REFERENCES


ENDNOTES

1 The concept of shared developmental vision was proposed by Cooke and Huggins (2003), who also emphasize competition with cooperation between cluster firms.

2 For a different classification of five types of firm agglomerations (local production systems, Porter clusters, monopsonistic clusters, Marshalian industrial districts, and innovative millieux) in terms of three co-ordinates (structure of governance, innovation/learning, and thickness of inter-firm relationships), see De Propris and Driffield (2006).

3 For recent accounts of industrial districts and new economic geography, see respectively Belussi and Caldari (2009) and Jovanovic (2009).

4 Compare with a definition of strategic alliances as “inter-firm co-operative arrangement, aimed at achieving the strategic objectives of the partners” (Das and Teng, 1998, p. 491).
For wider considerations on power, markets and hierarchies, see Dosi (1995) and Hartman et al (1993).

This is not surprising, as the IO perspective is in effect mono-institutional. The only institution examined is the market and its (industry) structure (Coase, 1937, 1972).

Williamson (1996, p. 22), for example, states that “Joint ventures and alliances should sometimes be thought of as T-forms of organisation that permit the parties to remain players in a fast-moving environment. Each party being unable, by itself, to assemble and deploy the requisite resources in a timely way, the requisite resources are instead assembled by pooling. Thus constructed, both successful and unsuccessful joint ventures will commonly be terminated when contracts expire. Successful joint ventures will be terminated because the combined effort has permitted each to remain viable, to learn enough, and/or, buy time to go it alone. Unsuccessful joint ventures will be terminated because the opportunity to participate will have passed them by.” Williamson also acknowledges that “Our understanding of T-forms of organization is not good but is steadily improving.” – for which he cites works outside the transactions cost tradition, notably Nelson and Winter (1982), Dosi (1988), Teece (1992), Teece et al. (1994). T stands for temporary or transitional.

For a comparison of transaction costs and game theoretic view of IFC, see Parkhe (1993).

Despite this common focus, the Coasean and Williamson’s versions of transaction costs differ. Coase questioned strongly the concept of rationality and Williamson’s emphasis on asset specificity. Moreover, he did not pay as much attention to the case of opportunistic behaviour (Coase, 1993). See also Posner (2010).

Transaction costs enter Richardson’s story indirectly through the concepts of similarity and complementarity of activities. However, Richardson did not employ the terms of transaction costs theorizing explicitly.

As noted below, this assumption is both heroic and can sometimes hinder research. The reason is that when markets do not exist, they need to be explained from first principles, for example by looking at human action, notably entrepreneurial behaviour (Casson, 1982, 2005; Pitelis and Teece, 2010).

In this context we can speak in terms of the “productive opportunity” of the cluster or the region.

By focusing on existing firms, Penrose’s approach cannot help explain the existence of entrepreneurial capabilities in (cognitive and/or charismatic) leadership and authority, motivation, coordination, and cognitive and socio-political legitimacy, which is needed for organization creation to start with (as in Loasby, 1976; Fransman, 1994; Aldrich and Fiol, 1994; Langlois, 1998; Witt, 2007)

Note, however, that this can vary depending on the clusters. In some clusters, the presence of dominant firms can engender a higher degree of control (Markusen, 1996). However, this will normally still not be as high in an integrated firm.
Penrose’s (1959) idea of “interstices” (market segments with large firms cannot cover, thus leaving space for smaller ones), and Marshall’s (1920) “tendency to variation” within each activity, also support the existence of this argument. Similarly limitations to knowledge, such as bounded rationality, limited absorptive capacity and/or variety that can be handled by a large organization, can also constitute limits to integration (Kay, 1997). Aoki (2010) notes for the case of Silicon Valley, that takeover of successful smaller firms by the large established ones does take place. In this context, our claim is simply that integration does not take place in all cases in which Richardson’s conditions are satisfied.

As noted by Langlois (2004), modularisation driven by external economies and standardisation can also help benefit clusters compete with larger firms (such as Applied Materials), which also create their own ecosystem. Such competition between clusters and large firms can be an extra reason for non-integration. This, however, is likely to be dependent on large firms’ capabilities to increase their control over the expanded ecosystem through system integration capabilities.

Similar considerations apply in the case of work on alliances. For Adegbesan and Higgins (2010), this focuses on value creation, not value appropriation by specific firms.

In fact, some clusters, such as Italian industrial districts, can be one of the nearest types of organizational market structures to the IO ideal of ‘perfect competition’ (large number of usually small firms with similar products and technology). In this context, the possibility of ‘zero profit’ is very real indeed (‘zero profit’ being the major characteristic of ‘perfect competition’ and contestable markets, Baumol, 1982).

We submit that this lies in a cluster’s value co-creation advantages, given each cluster participant’s value capture strategy. While these advantages are similar in nature to those discussed in the literature, our emphasis on co-creation helps partly address the issue of capture too.

As noted by Coase (1972), this need not be interpreted as the pursuit of monopoly. Indeed, as noted by Penrose (1959), monopolistic restrictions, while not uncommon, usually fail to provide sustainable advantages to firms.

For example, in a centrally-planned police state, one way to appropriate value is by informing the police on dissident activity and/or become very innovative in avoiding police and state scrutiny.

The case of “alliance portfolios” is analogous (Ozcan and Eisenhardt, 2009).