Maximilian Benner

What do we know about clusters?  
In search of effective cluster policies
Abstract

Geographical clusters are in the focus of a large breadth of literature and have been discovered as a preferred policy tool by policymakers around the world. The vast literature suggests differing explanations for the advantages clusters offer to constituent firms. Yet not all of these explanations have been empirically confirmed. Before cluster policy can act as a reliable tool for economic policy, there must be a certain degree of clarity about what kind of advantages a cluster can bring to constituent firms. Therefore, the following article reviews different explanations of cluster advantages from the literature. Its aim is to identify concrete mechanisms through which these advantages materialize. On the basis of these mechanisms the article identifies questions in cluster theory that are still open and defines steps for further research to develop a comprehensive theory of cluster policy.
What do we know about clusters?
In search of effective cluster policies

Introduction

Geographical clusters of firms that operate in the same or in related industries have captured the attention of academics and policymakers alike. Interest in agglomeration has resurfaced in the 1980s. This was due to the analysis of industrial districts and clustered industries in Italy, Baden-Württemberg and California and the debate about a new industrial paradigm of “flexible specialization” as a successor to the hitherto prevailing Fordist mass production paradigm. In 1990 Michael Porter added a business-studies perspective on clustering based on his notion of competitiveness. His contribution received widespread attention by policymakers in many industrial countries who looked for new ways to cope with structural change and resulting threats of high unemployment.

It was roughly a hundred years ago when Alfred Marshall tried to find reasons for the existence and viability of industrial districts. More recent accounts focus on the question of why clustering still, and maybe even more, occurs in an era in which improved information and communication as well as transportation technologies are supposed to render spatial distance less relevant. The success of Silicon Valley, however unique it may be, indicates that even industries such as software concentrate in just a few locations worldwide, although in the Internet age they supposedly should be able to locate everywhere. What is it that makes the presence in a cluster so attractive to companies in a particular industry? After almost three decades of research, what do we know about clusters? What is it that we can still not explain?

This article's purpose is to summarize the major findings of cluster theory. In addition, it aims at preparing a solid basis for developing a comprehensive theory of cluster policy. If economic geography wants to provide a foundation of political strategies grounded in theory, cluster theory needs policy relevance. This paper tries to identify mechanisms that may be at work within clusters. If further research assesses ways to politically influence these mechanisms, they could be used as a starting point for developing sound cluster policies.

First, the article gives a short overview of cluster definitions and associated problems. This is necessary to clarify the cluster notion and to emphasize fields of agreement and disagreement in the literature. Then it addresses the emergence of clusters in order to illustrate what role cluster theory can play in explaining the cluster phenomenon. This role is examined in detail in the next sections. They turn to the main strands of reasoning explaining why clusters develop. Thus the article asks how the presence of companies in clusters affects efficiency, uncertainty, and creativity. In order to develop a precise image clustering processes, it lists the mechanisms that are derived from these strands of reasoning. Building on these mechanisms, it draws conclusions for building a sound theory of cluster policy.

What is a cluster?

As Martin and Sunley (2003, 12) demonstrate, there is a wide array of differing definitions of the term that are not necessarily in harmony with each other. Porter (1998a, 197-198) defines clusters as “geographic concentrations of interconnected companies, specialized suppliers, service providers, firms in related industries, and associated institutions (for example, universities, standards agencies, and trade associations) in particular
fields that compete but also cooperate.” Roelandt and den Hertog (1999, 12) use a different definition: “Clusters are often cross-sectoral (vertical and/or lateral) networks, made up of dissimilar and complementary firms specialising around a specific link or knowledge base in the value chain.” Here the notion of spatial proximity is completely absent. In contrast to Porter, their definition does not view clusters as agglomeration plus interrelations between actors but focuses solely on their interrelations in the form of networking. Still other definitions focus solely on spatial proximity (cf. Enright 1998, 337). Further complicating the task of defining the cluster notion are different terms such as industrial districts, innovative milieus, regional innovation systems, and the distinction between regional and industrial clusters (cf. Enright 1996). Closely related to the difficulty of defining the cluster notion is the question of its spatial scale. Porter applies his cluster definition to all spatial scales: “The geographic scope of a cluster can range from a single city or state to a country or even a network of neighboring countries” (Porter 1998a, 199) This vagueness invites a host of criticism (cf. Martin and Sunley 2003).

In sum, cluster theory has not reached a generally accepted definition of (1) what of a cluster is, (2) the role of spatial proximity, (3) the appropriate spatial scale on which clusters occur, and (4) the role of interrelations between the cluster’s actors. It is necessary to come up with a working definition of clusters that accommodates the wide variety of reasons given in the literature to explain possible advantages (and disadvantages) of clusters but still offers a certain degree of objectivity in separating between what can be seen as a cluster and what cannot. Here every spatial agglomeration of enterprises within the same or within related industries primarily on the local and the regional, but also on the interregional, national and supranational scale are regarded as clusters, provided they are not primarily based on physical factors. Enterprises are considered to be within the same or within related industries if they are part of the same supply chain or if they are close to it. Note that explicit networking is not a precondition for the existence of a cluster in this definition. This wide definition makes it possible to develop a broad range of mechanisms. It is then necessary to ask on which spatial scale(s) each single mechanism can work.

How do clusters emerge?

Before elucidating the advantages an established cluster may offer to its constituent actors, we must address the emergence of clusters. The evolution of clusters is a path-dependent process. It is often triggered, at least in part, by chance events, or by a combination of planning and chance. An example for the role of chance events in cluster evolution is discernible in the origins of the Swiss watch industry. Its emergence was heavily influenced by Calvinist edicts against luxury and “useless” jewelry, the flight of Protestant huguenots from prosecution in other European countries to Switzerland and the specialization of jewelers among them on watches (Bumbacher 1995, 127). Krugman (1991, 35, 60) demonstrates the role of chance events in the case of the carpet manufacturing cluster of Dalton, Georgia, whose origins can be traced back to a wedding gift that reintroduced a technique uncommon at that time. Feldman (2001, 867-878) regards cuts in public employment in the region around Washington, D.C. as an important trigger for the establishment of new enterprises in biotechnology and information and communication technologies. Even the genesis of Silicon Valley, although heavily influenced by the vision of Stanford University’s manufacturing dean and provost Frederick Terman, was shaped by chance events that were not subject to foresighted planning. Nobel laureate William Shockley’s decision to found the semiconductor manufacturer Shockley Transistor in Palo Alto after returning from New Jersey may at least in part have been due to his familial ties within the region. These ties may have complemented Terman’s encouragement for Shockley to settle there (cf. Saxenian 1994, 22-24; Castilla, Hwang et al. 2000, 230; Lécuyer 2000. 161).
However, chance events could occur in any region roughly in proportion to its population. Thus they cannot sufficiently explain why some regions develop into clusters while others do not. Chance events, and sometimes also foresighted planning, can act as triggers for a path-dependent process that leads to the development of a cluster. Glückler (2007, 620) defines path-dependence as “a concept of cumulative causation in which a certain sequence of events creates unequal propensities for future events. Though path-dependent change allows for inferences from a present on future states of development, it is subject to contingency.” This means that the specific sequence of events leads to a certain regional trajectory that can result in a cluster.

Storper and Walker (1989) conceptualize cluster evolution in their theory of geographical industrialization. They suggest that industries actively shape their locational environments instead of taking them as given. In their words, “the basic patterns of industry location and regional growth can be produced by processes endogenous to capitalist industrialization, rather than by the exogenous placement of resources and consumers. Industrial location patterns are created through the process of growth rather than through a process of efficient allocation of plants across a static economic landscape. That is, industries produce economic space rather than being hostage to the pre-existing spatial distribution of supplies and buyers” (Storper and Walker 1989, 70). This model seeks to explain why clusters of new industries emerge in locations distant from established industrial core regions, as was the case with the semiconductor industry in Southern California’s Silicon Valley. However, while increasing returns lead to a path-dependent growth process of those locations that develop into clusters in the clustering phase, the model does not directly explain processes that lead to increasing returns. Thus it cannot predict which locations grow into clusters (cf. Markusen 1996, 293).

The concept of regional branching (cf. Frenken and Boschma 2007; Boschma and Frenken 2009) introduces product diversification as an explanation for localization. Diversification into new activities leads to regional branching when new industries grow out of old ones or emerge through a recombination of existing industries’ competences. Boschma and Frenken (2009, 7) argue that “when firms diversify (but not many will do so because of the risks involved), they will show a higher propensity to diversify into technologically related instead of unrelated industries, because of the firm-specific routines they have built over the years (e.g. reducing switching costs), and because of the opportunities the regional environment provides.” While companies tend to prefer diversification through incremental innovation, radical innovations are more likely to be pursued through labor mobility and spin-off creation. As their commercialization does not necessarily require the use of existing routines, it may be started in a different location. Thus, the more radical an innovation is, the more open is the window of locational opportunity (Frenken and Boschma 2007, 644). Yet, this model does not predict either which locality will take use of an open window of locational opportunity. Still, it provides an explanation of agglomeration dynamics after the initial localization impulse. If a new location emerges after a radical innovation, following incremental innovations can lead to its further growth. However, the model cannot explain why some locations develop stronger agglomeration dynamics as others after localization.

How exactly do locations develop into clusters after the initial triggering events? How do they manage to grow as clusters subsequently? Cluster theory must identify mechanisms leading to increasing returns to scale but shaping clusters selectively. The main groups of approaches in cluster theory can be, although rather roughly, grouped into three categories. They address the questions of how companies’ presence in clusters may enhance efficiency, reduce uncertainty, and induce creativity.

**How can clusters enhance efficiency?**

Roughly, there are three strands of reasoning focusing on efficiency in clusters. The first one emphasizes the availability of specialized suppliers and qualified labor within clusters. This rationale can be traced back to
Marshall (1920). He stated these two reasons together with technological spillover effects as the main advantages of clustering. A specialized labor pool constitutes an incentive both for employers and for workers to locate in the cluster, since colocation can alleviate the consequences of business cycles and thus reduce the risk of not finding adequate labor or employment. If companies are not completely and simultaneously affected by business cycles, workers who lost their jobs during a cyclical downturn have a higher chance of finding employment elsewhere in the cluster. In contrast, during a cyclical upturn employers have a higher chance to find workers within the localized labor pool (cf. Krugman 1991, 38-45; Bathelt and Glückler 2003, 81). Specialized suppliers have an incentive to locate in a cluster if they encounter internal economies of scale in their own production process. Then they can capture the benefits of being close to the market and concentrating their production capacities at the same time by locating their entire production in the cluster (cf. Krugman 1991, 49-52; Bathelt and Glückler 2003, 82).

The latter reasoning leads directly to the second broad strand of approaches. This is a central part of what Martin and Sunley (1996) call “geographical economics”. In his core-periphery model Krugman (1991, 14-25) emphasizes the interplay of transport costs and economies of scale, albeit generally on a higher spatial scale. Under the assumption of mobility of both labor and capital, the model predicts that producers will concentrate all of their production activities in a location that offers the highest demand. This way they can capture economies of scale if transport costs are low in comparison to the achievable economies of scale. If transport costs are comparatively high, serving more markets with several manufacturing sites may become more advantageous. If transport costs are zero, concentrating all production activities at any site enables capturing economies of scale without necessarily leading to agglomeration at the location with the highest demand (cf. Bathelt and Glückler 2003, 79-81). Although Krugman applies his model to larger agglomerations like the “manufacturing belt” in the U.S., his reasoning fits within the rationale of local supplier availability mentioned above that originally dates back to Marshall.

The third strand of approaches emphasizing clusters’ advantages regarding efficiency centers on the relevance of transaction costs. These costs are “ex ante costs of drafting, negotiating, and safeguarding an agreement and, more especially, the ex post costs of maladaptation and adjustment that arise when contract execution is misaligned as a result of gaps, errors, omissions, and unanticipated disturbances” (Williamson 1994, 103, emphasis in original). Instead of assuming rational behavior of the neoclassical homo oeconomicus, transaction cost economics highlights imperfect information and an incomplete capability of market actors to process information. Consequently, contracts governing complex transactions can often be expected to be incomplete. Parties acting opportunistically can take advantage of this incompleteness (cf. Williamson 1981, 552-554). In order to protect itself from the danger of the counterparty's opportunism, each party engages in search, information gathering and monitoring activities that makes them incur additional costs. According to the characteristics of a particular type of transaction such as frequency, uncertainty and transaction specificity of investments, Williamson (1979; 1981) describes organizational arrangements that are able to reduce those transaction costs. Between the extremes of markets and hierarchies, there are a host of possible hybrid or network forms of organization.

Clusters can be seen as one of those hybrid forms. Authors of the so-called “Californian school” (e.g. Scott 1983; 1988a; 1988b; 1988c; Storper 1989) argue that proximity within clusters enhances efficiency by enabling interactions between constituent companies at lower transactions costs than would be possible over larger spatial distances. Similar to the argument of “flexible specialization” brought forward by Piore and Sabel (1984), they assume a shift away from Fordism towards a new industrial paradigm of “flexible production” whose flexibility is basically achieved by a deeper social division of labor. According to Scott (1988a, 176), “when changes in economic conditions bring about intensified uncertainty and instability in production and increased competitiveness in final markets, then internal economies of scale and scope within the firm begin to break down so that the entire production system is liable to display strong symptoms of horizontal and ver-
tical disintegration. Such disintegration enormously enhances flexibility in the deployment of capital and labour for it permits producers to combine and recombine together in loose, rapidly shifting coalitions held together by external transactional linkages.” In order to counter fastly changing circumstances on the market, vertical disintegration leading to a deeper social division of labor can stabilize flows of work, minimize the risk of overcapacities and maximize specialization advantages. This is because vertical disintegration can act as a means to harmonize different production structures with diverging optimal scales and thus to react more flexibly to changing end user demand. In addition, distributing subcontracting among various suppliers can alleviate the risk of major disruptions in supply (cf. Scott 1983, 245; 1988c, 26).

However, vertical disintegration increases the need for cooperation between the parties. Thus it causes potentially higher transaction costs. This problem can be countered by spatial agglomeration: “The greater the spatial dispersion of producers, the more onerous these costs will be. The immediate consequence is that selected sets of producers with particularly elevated intragroup interaction costs will tend to converge around their own geographical centre of gravity and thus to engender definite nodes of economic activity on the landscape” (Scott 1988a, 176-177). Spatial proximity can cut the cost of gathering information and of searching suppliers and workers (Scott 1988c, S. 27). In addition, it reduces the need for coordination of production processes by enabling intensified personal contact and thereby lowers communication costs.

However, the role ascribed to spatial proximity by these approaches depends on the prevalence of buyer-supplier relations. McCann (1995, 563-564) emphasizes that often “a large proportion of firms have few or no trading links with other local firms in the same industry, even when there is a strong spatial clustering of a particular industrial sector.” In addition, why, does vertical disintegration plus spatial agglomeration occur instead of vertical integration in order to reduce transaction costs? This goal could be better achieved within a vertically integrated enterprise (Maskell 2001, 926-927; Malmberg and Maskell 2002, 438).

It may be true that, following Scott’s rationale, vertical disintegration offers flexibility advantages that may justify incurring higher transaction costs, and that these costs can be reduced somewhat by spatial agglomeration. Yet, it is questionable whether the peculiar market conditions Scott assumes are prevalent in all industries that are prone to clustering. Thus it is not sure whether vertical disintegration does indeed offer substantial advantages over vertical integration in other industries than those studied by the Californian school.

Consequently, we must analyze other reasons for spatial clustering. We must move beyond static efficiency that is examined by the above-mentioned approaches towards dynamic efficiency. Harrison, Kelley and Gant (1996, 234) highlight the difference between both perspectives: “Static agglomeration economies are said to occur when the unit costs of production of a business enterprise or establishment are lower in the context of relatively dense clusters of other firms or specialized resources, such as skilled labor and infrastructure, than would be the case if the typical business were located elsewhere. Dynamic agglomeration economies refer to the heightened prospect for technological learning to occur (not simply reductions in unit costs of production with a given technology) in relatively dense urban places, districts, or clusters compared with less dense locations.” Porter (1994, 36-37) considers static efficiency with price competition to have become less important: “The basis of competitive advantage has shifted from static efficiencies to the rate of dynamic improvement. It is not the inputs or scale the firm possesses today, but its ability to relentlessly innovate and upgrade its skill and technology (largely intangible assets) in competing.” If this is true, approaches focusing on static efficiency gains of clustering cannot fully explain spatial agglomeration, although they can complement explanations based on dynamic advantages. Other reasons like the reduction of uncertainty and the stimulation of creativity can fill this void.


**How can clusters reduce uncertainty?**

Clusters can reduce the uncertainty economic agents are facing. The above-mentioned approaches ascribe a rather limited, if any, role to socio-institutional aspects of economic action. In contrast, approaches focusing on uncertainty reduction view economic action as embedded in social relations. Central of this argument is the concept of embeddedness following Granovetter (1985). He regards it as a new perspective between what he calls “oversocialized” and “undersocialized” views. In this context, embeddedness means that economic action is part of a social structure that is similar to what political scientists call social capital that is embodied in the relations between agents (Coleman 1988; Putnam 1993).

Granovetter (1990) distinguishes between relational and structural embeddedness. Relational embeddedness determines the quality of a dyadic relation between two agents. In contrast to the neoclassical concept of the *homo oeconomicus*, under relational embeddedness parties do not necessarily behave opportunistically once they have the opportunity to do so. Rather, by repeated interaction agents may build trust which can reduce uncertainty (Harrison 1992, 477-478; Glückler 2001, 214-215). This can facilitate complex transactions and lead to time savings (cf. Uzzi 1996; 1997). Structural embeddedness determines the quality of a wider structure of social relations within a network. Trust in dyadic relations can be transmitted to agents’ reputation within a network (Glückler 2001, 215; 2005, 1732-1733).

Uzzi (1996; 1997) examines the role of embeddedness in what he calls embedded ties in buyer-supplier relationships. These ties are intensive relationships that, in contrast to arm’s-length relationships, are not governed merely by price and quantity. Embedded ties have three main functions. First, they enable the emergence of trust that leads to each party’s willingness to provide extra efforts in excess of its contractual obligations without being able to enforce direct reciprocation. Second, these ties enable fine-grained information transfer that exceeds mere price and quantity information that are exchanged in arm’s-length ties. In embedded ties, parties can learn about each other’s preferences over time. Third, the parties can develop joint problem-solving arrangements. Instead of threatening to exit the relationship, which is the ultimate sanctioning measure in arm’s-length ties, these arrangements offer more flexible solutions. Over time the parties can develop problem-solving routines. This reduces uncertainty since each party can better predict how the other one will react if a problem arises (Uzzi 1996, 677-679; 1997, 43-47).

However, those embedded ties do not necessarily have to occur between spatially proximate parties. Following Glückler (2001, 222-224), embeddedness is not *per se* a spatially delimited phenomenon. Rather, it is a socio-economic phenomenon that can be examined in a spatial perspective. Building embedded relations requires continued interaction between the parties. Although modern transport infrastructure enables face-to-face contact over large distances, continuous personal contact can be maintained easier in spatial proximity. Thus, embedded ties can lead to the reduction of uncertainty in interactions among cluster participants.

Additionally, uncertainty can be reduced through social capital embodied in norms, rules and conventions that may exist within a cluster. For example, Piore and Sabel (1984, 32) view such institutions as an important factor in Italian clusters: “The constant reorganization of production was possible only on the condition that everyone knew, and was known to abide by, a long list of rules of fair behavior: had these rules required formal application, they would have prohibitively delayed shifts from one grouping of firms to another.” The similar concept of institutional thickness emphasizes the role of institutions including organizations. Amin and Thrift (1994, 15) define institutional thickness “as the combination of factors including inter-institutional interaction and synergy, collective representation by many bodies, a common industrial purpose, and shared cultural norms and values.” Storper (1995; 1997a; 1997b) regards a cluster as a nexus of untraded interdependencies. They “take the form of conventions, informal rules, and habits that coordinate economic actors under conditions of uncertainty; these relations constitute region-specific assets in production” (Storper 1997b, 5).
They include informal institutions like rules, conventions and a common language (cf. Storper 1995, 206). According to Storper (1997b, 38), conventions “include taken-for-granted mutually coherent expectations, routines, and practices, which are sometimes manifested as formal institutions and rules, but often not.” Untraded interdependencies like conventions can generate mutual trust. They can constitute specific regional advantages if they exist in a region.

Notwithstanding their differing definitions, the above-mentioned concepts have in common their emphasis on the relevance of non-market relations in economic transactions. Their major role is their ability to reduce uncertainty. While they don't necessarily require close spatial proximity, they may be facilitated by the relative ease of face-to-face contact within clusters.

On the other hand, a too high degree of embeddedness within clusters can have long-term negative consequences. Uzzi (1996, 683-685; 1997, 57-60) suggests a paradox of embeddedness, arguing that a too low as well as a too high degree of embeddedness can weaken a network. Grabher (1993, 260-264) distinguishes three kinds of lock-in situations that may obstruct the development of a cluster with a considerable degree of embeddedness. First, functional lock-in refers to structural aspects of companies and industries that weaken their ability to look for information originating from sources external to the network. Second, cognitive lock-in leads to a specific view of the world that determines which phenomena are perceived and how they are interpreted. Third, political lock-in due to close relationships between politics and business may enhance the other two forms. Lock-in situations can have negative consequences for the development of a cluster, especially when radical ruptures in technology or market conditions occur. This can lead to what Storper and Walker (1989) describe as shifting centers.

While approaches focusing on the reduction of uncertainty can certainly contribute to explaining clustering, they are incomplete. Even if building trust requires frequent face-to-face contact, in principal establishing embedded relations can equally be achieved over large spatial distances. The main disadvantage of maintaining embedded relationships over large distances in contrast to the spatial proximity of clusters is transport and transaction costs. Insofar, the benefits of clustering are still based on static efficiency. Consequently, the criticisms concerning approaches focusing on efficiency apply here too. Particularly, clustering of actors without relevant localized buyer-supplier relations can hardly be explained by this reasoning either.

On the other hand, embeddedness can be a prerequisite to the improvement of dynamic efficiency. If this is the case, embeddedness is relevant as it may enable learning processes that might not be equally possible without a certain degree of trust. In this perspective, embeddedness can be seen as the foundation the stimulation of creativity in clusters.

**How can clusters induce creativity?**

Approaches focusing on creativity center on dynamic instead of static efficiency. Participation in clusters is beneficial to constituent companies because it enhances opportunities for continuous improvement and innovation through technological progress. Non-market relations like untraded interdependencies can offer a foundation for creativity. Marshall’s (1927, 287) “industrial atmosphere” metaphor suggests there is something to clusters that fosters entrepreneurship, including learning and innovation. In the concept of the innovative milieu (e.g. Aydalot 1986; Camagni 1991a; 1991b; Maillat 1995; 1998), learning processes are rooted in the social fabric of the region. Similar to the embeddedness perspective, but centering on learning processes, this concept regards enterprises not as isolated actors. Rather, the emergence of innovation is embedded in the structure of social relations within the region.
How exactly is creativity induced in clusters? For analytical purposes, three main lines of reasoning can be distinguished. The first line centers on cooperative learning processes. They are a direct consequence of the role of trust in embedded relations as a facilitator for knowledge transfer. Central to these approaches is the notion of interactive learning as a collective social process. Edquist (1997, 1-2) suggests that “firms almost never innovate in isolation. In the pursuit of innovation they interact with other organizations to gain, develop, and exchange various kinds of knowledge, information, and other resources. These organizations might be other firms (suppliers, customers, competitors) but also universities, research institutes, investment banks, schools, government ministries, etc. Through their innovative activities firms often establish relations with each other and other kinds of organizations; therefore it does not make sense to regard innovating firms as isolated, individual decision-making units.” In this perspective, innovation is not a linear process that proceeds from basic to applied science to product development to production (cf. Smith 1994, 1-3). Rather, innovation is based on a continuous interactive process and can as well result from customer feedback or from experiences in production. Thus, all actors within the local milieu can possibly be sources of innovation (cf. Camagni 1991a; Lundvall 1992; Lundvall and Johnson 1994; Asheim and Isaksen 1997).

What Uzzi (1996; 1997) describes as fine-grained information transfer in embedded relationships needs not be limited to specifications of transactions between the parties. It can even lead to the long-term joint development of new solutions to technological problems that can enhance the partners’ competitiveness beyond their dyadic relationship. Lundvall (1988, 352-353) describes how in an interactive process a producer can gradually adjust the product to the user’s needs and offer him specific training. This process is characterized by a high degree of uncertainty and the danger of opportunistic behavior. Consequently, for such an interactive learning process to occur trust is needed and reputation can help.

Similar to Storper’s (1995; 1997a; 1997b) concept of untraded interdependencies, Maskell and Malmberg (1999a, 173-174) stress the role of what they call localised capabilities in collective learning. They consist of a region’s infrastructure, its natural resources, knowledge and skills available there, and the region’s specific institutional endowment. A region’s institutional endowment contains institutions like routines, habits, traditions, conventions and entrepreneurial spirit. It can enable the creation of specific kinds of knowledge while hampering the creation of others. This can be a source of competitive advantage: “In a knowledge-based economy, valuable localised capabilities will primarily be those which increase the ability of firms to create; acquire; accumulate; and utilise knowledge a little faster than their cost-wise more favourably located competitors” (Maskell, Eskelinen et al. 1998, 51). As competitive advantage must rest on heterogeneity, enterprises can build it on the basis of the localised capabilities available to them in the region where they are located. However, the internationalization of economic activity and the codification of knowledge leads to a transformation of former localised capabilities into ubiquities (cf. Weber 1909, 51-52) that other regions can imitate and use as well. Maskell and Malmberg call this process ubiquitification. It leads to the destruction of localised capabilities (Maskell 1999; Maskell and Malmberg 1999b).

Enterprises can react by focusing either on static price competition or on dynamic improvement through innovation. Dynamic competition through innovation is based on the ability of enterprises to create and use tacit, that is, non-codified knowledge faster than competitors before it can be codified. Tacit knowledge cannot easily be written down and transferred over large distances. It consists primarily of routines and experience and is embodied in persons and their relations with each other. The more easily knowledge can be codified and standardized and thus becomes available for possibly every user globally at low transaction costs, the more relevant new tacit knowledge becomes in dynamic competition through innovation. Local cultures and trust can help regions build competences that enable the creation of tacit knowledge through interactive learning. In so doing, regions can create new localised capabilities at least as fast as old ones are destroyed by ubiquitification. In this way they can maintain their regional specializations over time (Maskell 1996; 1999; Maskell and Malmberg 1999b).
In order to create and use new combinations of tacit knowledge, it must be transferred, e.g. in interactive learning processes. Von Hippel (1994) calls it sticky knowledge due to the difficulty to transfer it. This transfer depends on a shared social context (cf. Lundvall and Johnson 1994, 30; Gertler 2003, 78-79). It often needs face-to-face contact that is more difficult to achieve over larger distances than in close spatial proximity. Maskell and Malmberg (1999a, 180) stress the meaning of embeddedness in this process: “To communicate tacit knowledge will normally require a high degree of mutual trust and understanding, which in turn is related not only to language but also to shared values and ‘culture’.” However, spatial proximity is neither a necessary nor a sufficient condition for the transfer of tacit knowledge. Boschma (2005) describes other dimensions of proximity that can as well be relevant for the transfer of tacit knowledge. Spatial proximity can support the establishment of other kinds of proximity by facilitating face-to-face contact, possibly contributing to other dimensions of proximity in a cluster.

Whether interactive learning in cooperative relationships like buyer-supplier relations contributes to the advantages of clustering depends on how prevalent these ties are. If they exist, it remains questionable whether the advantages they offer to localized actors derive from their static or their dynamic efficiency. Even if a cluster shows a considerable degree of close buyer-supplier relationships, a causality between these embedded ties and higher innovative activity is hard to confirm. In sum, interactive learning in cooperative relationships cannot fully explain the dynamic advantages of clusters.

In contrast to approaches focusing on cooperative learning behavior, the second line of thinking centers on dynamic effects of competition. While interactive learning is relevant primarily within the vertical, institutional and external cluster dimensions, competitive approaches focus on the horizontal cluster dimension. Companies can learn from each other by mechanisms of knowledge transfer that are not necessarily dependent on personal contact. Brown and Duguid (2000, 22), referring to Silicon Valley, describe the potential role of clusters as a hotbed for knowledge exchange: “All this intermingling makes it almost impossible for people not to know what others are up to. And it gives participants the extra insight to interpret a product announcement, read a patent, understand the significance of a product, or use a new tool. Critically, then, the level of shared informal knowledge ‘in the air’ within a locality provides an unrivaled key for interpreting the formal knowledge produced there.”

In his theory of competitiveness Porter (1990) ascribes a central role to these mechanisms. Firm strategy, structure and rivalry are a part of his diamond model. They are relevant within the horizontal cluster dimension. Spatial proximity in clusters eases comparing competitors’ actions: “Rivalry with locally based competitors has particularly strong incentive effects because of the ease of constant comparison and because local rivals have similar general circumstances (e.g. labor costs and local market access) so that competition must take place on other things” (Porter 2000, 261). As clusters offer similar general conditions to all localized competitors, companies whose performance is falling behind need to examine their own actions in order to find the reasons. Thus the range of possibilities for finding solutions becomes smaller, as Maskell (2001, 928-929) explains: “Co-localized firms undertaking similar activities find themselves in a situation where every difference in the solutions chosen, however small, can be observed and compared. While it might be easy for firms to blame the inadequate local factor market when confronted with the superior performance of competitors located far away, it is less so when the premium producer lies down the street. The sharing of common conditions, opportunities and threats make the strengths and weaknesses of each individual firm apparent to the management, the owners, the employees and everyone else in the cluster who cares to take an interest.” Additionally, high competitive pressure forces companies to innovate: “Reinforcing these advantages for innovation is the sheer pressure–competitive pressure, peer pressure, and constant comparison–occurring in geographically concentrated clusters. The similarity of basic circumstances (e.g. labor and utility costs), combined with the presence of multiple rivals, forces firms to distinguish themselves creatively” (Porter 2000: 262). Per-
sonal ambition by entrepreneurs and managers to ascend in the local social hierarchy can reinforce this pressure (Porter 1998a, 219; 1998b, 83).

Spatial proximity allows companies to monitor each other more closely: “Just as people in a residential area simply cannot help noticing what their next-door neighbours do (regrettably, many would say), business firms often have remarkably good knowledge of the undertakings of nearby firms even if they do not make any dedicated efforts at systematic monitoring. If those neighbouring firms are in a similar business, it is more likely that the observing firm will understand, and learn from, what it sees” (Malmberg and Maskell 2002, 439). Monitoring can induce variation in turn, as the mere existence of an innovation signifies that there is at least one way to produce it (Maskell 1999, 42). In this way the whole process may even start anew. In contrast to approaches focusing on interactive learning in cooperative relationships, these mechanisms do not require trust or input-output relations (Malmberg and Maskell 2002: 439). Thus, they offer an explanation to the advantages of clusters where these relations do not exist to a considerable extent or concerning critical inputs.

The third line of thinking that tries to explain dynamic advantages of clusters centers on knowledge spillovers independent both of cooperative and of competitive behavior. These spillovers can occur through knowledge transfer in informal face-to-face contacts that are not necessarily aimed at acquiring knowledge. Marshall hints at these effects as one reason for clustering: “The mysteries of the trade become no mysteries; but are as it were in the air, and children learn many of them unconsciously” (Marshall 1920, 271). This is part of Marshall’s (1927, 287) “industrial atmosphere” with its many opportunities for intended and coincidental interactions that are facilitated by spatial proximity. An example for this are “cafeteria effects” (Camagni 1991b, 130) and knowledge transfer as a by-product of other social interactions. Brown and Duguid (2000, 22) describe this for Silicon Valley: “Competitors shop in the same stores, eat in the same restaurants, send their kids to the same schools, travel in the same car pools, work with the same suppliers, even smoke in the same groups outside the smoke-free office buildings. And, of course, many shuffle between firms, changing jobs with relative ease.” Labor mobility and personal business and private contacts create a climate that makes knowledge spillovers almost impossible not to occur – not only between companies, but also between them and research and education institutes (cf. Wolfe and Gertler 2004, 1076). Spin-off enterprises that locate within the cluster are another source of knowledge spillovers. Bathelt (1990, 170) and Porter (1990, 157) argue that spin-off enterprises often seek to locate close to the former employer of the founder in order to benefit from existing relationships and information channels. Venture capitalists can also contribute to knowledge transfer if they engage in intensive coaching of the start-up enterprises they finance (cf. Banato and Fong 2000, 304-306). While knowledge transfer does not necessarily require spatial proximity (cf. Bathelt, Malmberg and Maskell 2004; Boschma 2005), coincidental knowledge spillovers through face-to-face contacts can be expected to occur with higher frequency and probability in spatial proximity. They cannot fully be substituted by communication technology. Still, knowledge transfer from external sources is important in order to avoid lock-in. Bathelt, Malmberg and Maskell (2004) regard so-called pipelines (cf. Owen-Smith and Powell 2004) as an essential source to acquire knowledge. Once it has been be inserted into the cluster, it can circulate there through various mechanisms for knowledge transfer including knowledge spillovers.

Which mechanisms shape clustering?

The possible advantages of clustering can be achieved through a number of mechanisms that may work within clusters. While not every mechanism may be observable in each cluster, cluster theory offers a toolbox for case-study research. Table 1 lists possible mechanisms and their underlying reasoning, together with a selection of references of their empirical discussion. Some mechanisms can be a source both of static and dynamic
efficiency, making it difficult to distinguish between their effects in practice and to establish causality. As is obvious from Table 1, the empirical discussion of the various mechanisms is rather uneven. While some mechanisms like labor mobility, spin-off formation and vertical cooperation have been frequently discussed, there seems to be a need for further research on others like competition in the local social hierarchy and cafeteria effects. Interestingly, these mechanisms are often considered as relevant for the development of high-profile Silicon Valley. Given their unclear empirical relevance, one might caution against generalizing such effects. They may be discernible in highly visible singular clusters but could possibly be much less relevant in many other cases.

Table 1: *Mechanisms identified by cluster theory*

<table>
<thead>
<tr>
<th>Mechanism</th>
<th>Reasoning</th>
<th>Empirical discussion (selection)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recruitment of qualified new staff among alumni of higher education</td>
<td><strong>Efficiency:</strong> the availability of alumni leads to a localized pool of qualified labor (Marshall 1920; Krugman 1991)</td>
<td>• Bathelt (1990; 2001)</td>
</tr>
<tr>
<td>institutes</td>
<td><strong>Creativity:</strong> alumni transfer new knowledge into companies</td>
<td>• Lawson (1997)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Lawton Smith (1997)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Lindholm Dahlstrand (1999)</td>
</tr>
<tr>
<td>Labor mobility among companies or between higher education or research</td>
<td><strong>Efficiency:</strong> through the localized pool of qualified labor the effects of business cycles can be alleviated (Marshall 1920; Krugman 1991)</td>
<td>• Oakey (1985)</td>
</tr>
<tr>
<td>institutes and companies</td>
<td><strong>Creativity:</strong> Workers transfer experience and knowledge collected with the former employer (company, education or research institute) to the new one (Breschi and Lissoni 2009)</td>
<td>• Storper and Christopher (1987)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Scott (1988c)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Bathelt (1990)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Saxenian (1990; 1994)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Angel (1991)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Lawson (1997)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Lawton Smith (1997)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Almeida and Kogut (1999)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Lindholm Dahlstrand (1999)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Sternberg and Támasy (1999)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Lissoni (2001)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Nachum and Keeble (2003)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Power and Lundmark (2004)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Breschi and Lissoni (2009)</td>
</tr>
<tr>
<td>Student work in companies (e.g. as interns or student trainees or through</td>
<td><strong>Creativity:</strong> students transfer new knowledge from higher education institutes to companies</td>
<td>• Bathelt (1990; 2004)</td>
</tr>
<tr>
<td>writing theses)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spin-off formation</td>
<td><strong>Creativity:</strong> Workers transfer experience and knowledge collected with the former employer (company, education or research institute) to the spin-off (Audretsch 1998)</td>
<td>• Bathelt (1990)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Saxenian (1994)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Campoccia (1997)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Keeble, Lawson <em>et al.</em> (1998)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Lindholm Dahlstrand (1999)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Castilla, Hwang <em>et al.</em> (2000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Gibbons (2000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Lécuyer (2000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Feldman (2001)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Moore and Davis (2001)</td>
</tr>
<tr>
<td>Availability of venture capital (including financing through angel</td>
<td><strong>Efficiency:</strong> the relationship between a start-up (including spin-offs) and a localized venture capitalist can save transaction costs</td>
<td>• Oakey (1985)</td>
</tr>
<tr>
<td>investors)</td>
<td></td>
<td>• Bathelt (1990)</td>
</tr>
<tr>
<td>Reduction of uncertainty: in repeated interactions between a start-up and a localized venture capitalist trust evolves; the investment of successful venture capitalists reduces uncertainty for other agents through its signalling function (Ferrary and Granovetter 2009)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creativity: specialized venture capitalists can transfer knowledge into the start-up in interactive learning processes, exchange information among each other in joint financing projects and broker contacts that can be used for information exchange in turn (Saxenian 1990; 1994; Castilla, Hwang et al. 2000)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Cooperation between higher education or research institutes and companies | Efficiency: e.g. use of university laboratories by companies |
|---|
| Creativity: knowledge transfer through interactive learning processes (Edquist 1997), e.g. in contract research |

| Horizontal cooperation among companies (including cooperation in trade associations) | Efficiency: joint provision of infrastructure, training programs and joint representation of interests (Porter 1998a; Enright 1998; 2003) |
| Creativity: a local monopoly can provide incentives for innovation due to better possibilities for appropriation of innovation rents (Marshall-Arrow-Romer school; Marshall 1920; Arrow 1962; Romer 1987, cf. Glaeser, Kallal et al. 1992) |

| Vertical cooperation among companies | Efficiency: localized input-output relations can reduce transaction costs (Scott 1983; 1988a; 1988c) |
| Reduction of uncertainty: through repeated interactions in close localized input-output relations trust emerges (Uzzi 1996; 1997) |
| Creativity: in close localized input-output relations interactive learning processes can take place (Lundvall 1988; Lundvall and Johnson 1994; Edquist 1997) |

- Saxenian (1990; 1994) |
- Banatao and Fong (2000) |
- Castilla, Hwang et al. (2000) |
- Hellmann (2000) |
- Ferrary and Granovetter (2009) |
- Oakey (1985) |
- Bellini, Giordani, Pasquini (1990) |
- Saxenian (1994) |
- Lindholm Dahlstrand (1999) |
- Castilla, Hwang et al. (2000) |
- Bathelt (1990) |
- Bellini, Giordani, Pasquini (1990) |
- Saxenian (1992; 1994) |
- Oakey (1985) |
- Storper and Christopherson (1987) |
- Scott (1988c) |
- Bathelt (1990; 1991; 2001) |
- Saxenian (1991; 1994) |
- Amin and Thrift (1992) |
- Henry (1992) |
- Angel (1994) |
- Bonaccorsi and Lipparini (1994) |
- Angel and Engstrom (1995) |
- Park and Markusen (1995) |
- Uzzi (1996; 1997) |
- Campoccia (1997) |
- Lawson (1997) |
- Lawton Smith (1997) |
- Wever (1997) |
- Lindholm Dahlstrand (1999) |
- Sternberg (1999) |
- Sternberg and Tamásy (1999) |
- Arndt and Sternberg (2000) |
- Camagni and Capello (2000) |
- Enright (2000) |
- Lissoni (2001) |
- Bengtsson and Sölvell (2004) |
- Borrás and Tsagdis (2008) |
- Weterings and Boschma (2009) |
Intensive local competition | **Creativity:** spatial proximity in clusters offers similar conditions and thus makes corporate strategies comparable; the existence of an innovation can lead to an incentive for competitor to innovate and variate (Porter 1990; Porter 2000; Maskell 2001; Malmberg and Maskell 2002) | • Dorfman (1983)  
• Glaeser, Kallal et al. (1992)  
• Saxenian (1994)  
• Feldman and Audretsch (1999)  
• van der Linde (2003)  
• Weterings and Boschma (2009)

Competition in the local social hierarchy | **Creativity:** the perspective for entrepreneurs and workers of advancing in the local social hierarchy can intensify local competition (Porter 1998b; Enright 1998) |  

Cafeteria effects | **Creativity:** in coincidental social contacts, e.g. in cafeterias and in leisure activities workers can exchange knowledge (Brown and Duguid 2000) | • Saxenian (1994)  
• Nachum and Keeble (2003)

Social networks | **Creativity:** labor mobility and education in local education institutions enable entrepreneurs and workers to build networks in which knowledge can be transferred | • Saxenian (1990; 1994)  
• Lawson (1997)  
• Lawton Smith (1997)  
• Lissoni (2001)  
• Østergaard (2009)

Considering the vagueness of most cluster definitions concerning the spatial scales of clusters, these mechanisms must be examined in their spatial extent. Assuming a high spatial mobility of qualified workers, labor market mechanisms can be expected not just on the local and regional level, but also on the national and to a certain degree even on the supranational scale. Knowledge spillovers through coincidental interactions like cafeteria effects, in contrast, should by far be strongest on the local level. Between those extremes, most mechanisms might be strongest on the regional and the local levels, while on the national level macroeconomic conditions may play the major role in influencing cluster development.

**Conclusion: Do we know enough to build sound cluster policies?**

There is a considerable gap between cluster theory and cluster policy. With the notable exception of Porter’s (1990) broader theory of economic policy aimed at promoting competitiveness, most cluster theories do not draw direct and concrete conclusions for policy. In order to build sound cluster policies, we must be aware of what we know about clusters and what we still don’t know. While there is much we already know about how clusters work, many questions are still open. Gaps in the empirical examination of some mechanisms leave room for further study in order to determine whether those respective mechanisms are less relevant per se or whether they just have not yet been proved satisfactorily.

An important insight is the high relevance of the local labor market to which various mechanisms are connected in the context both of efficiency and creativity, as Table 1 demonstrates. In practice many cluster initiatives focus mainly on promoting cooperation either among companies or between companies and research institutes. Considering the high relevance of mechanisms based on the local labor market, concentrating more on these mechanisms might prove justified. A further caveat relates to the difficulty of distinguishing between advantages of static and of dynamic efficiency and thus observing causality between mechanisms that lead to clustering advantages according to both perspectives. For example, recruitment of qualified new staff may lead to both efficiency and creativity advantages. These different lines of reasoning that apply to the same mechanisms must be kept apart. This requires further research that analytically distinguishes between different lines of reasoning and examines their effects on the concerning mechanisms separately.
Generally, what we know is that proximity within clusters may enhance efficiency and induce creativity, with the reduction of uncertainty potentially supporting both. We know about certain mechanisms that can be the drivers of these effects, although further research might identify more of them. To a certain degree, we can establish causality between mechanisms and their effects, although mixed degrees of empirical confirmation on several mechanisms may pose a caveat.

Further research is needed to bridge the gap between cluster theory and cluster policy. Although we know a lot about these mechanisms, research on how cluster policy can shape them is still insufficient. In order to build solid cluster policies led by theory, several questions must be addressed:

1. **How susceptible are effective cluster mechanisms to policy influence?**

   Probably not all mechanisms can be governed effectively through policy incentives. While policymakers might find it comparatively easy to induce cooperation between research institutions and companies, influencing labor mobility might prove more difficult. A closer examination of each mechanism could provide a solid basis for the consecutive questions.

2. **Which instruments can influence those mechanisms?**

   Precise instruments that can shape each of the mechanisms that are susceptible to political influence need to be identified. The causality of their impact on the respective mechanisms needs to be established.

3. **On what spatial scale (supranational, national, regional, or local) can these instruments be used?**

   Generally, most mechanisms can be expected to be strongest on the local and regional level. Thus, instruments used on these levels might be regarded as highly relevant for designing and carrying out cluster policies. While instruments used on the national or supranational scales could set conditions for the development of clusters, they must be filled by action on the regional and local scales.

4. **Which actors on each level can influence those mechanisms by using those instruments?**

   Potential actors on various spatial scales need to be identified. Their ability to use the instruments mentioned above and their interrelations in doing so must be examined. Implications for their respective approaches and the structure of organizations might ensue. Answering these questions could contribute to the development and implementation of sound cluster policies based on cogent theory. A possible outcome is that the high hopes connected with cluster policies in a lot of regions worldwide might prove to be exaggerated. Yet there is a chance of clarifying the potential of the cluster notion as a tool of economic policy. In the end, cluster policy might develop from a chaotic concept regarded as a policy panacea (cf. Martin and Sunley 2003) into a specific device in the economic policymaker’s toolbox – one that does not cure all economic woes but instead can be used to solve specific problems and achieve well-defined goals in regional economic development.
References


