12

Local Antecedents and Trigger Events: Policy Implications of Path Dependence for Cluster Formation

David A. Wolfe and Meric S. Gertler

A critical issue for any consideration of the genesis of clusters and the role for policy in their origin is the relative importance of chance events, or serendipity, as opposed to rational or intentional actions. As greater attention is focused on the promotion of clusters as an economic development policy tool, the question of whether, and how, they can be fostered assumes greater significance. Central to this debate is the role of path dependencies created by small, initial—often chance—events, as opposed to the role played by deliberate actions by both private actors and public sector agencies in contributing to the genesis of clusters. The concept of path dependency has been adopted by a wide range of disciplines to analyze and explain a broader range of social phenomena—sometimes in a rather deterministic fashion. The concepts of path dependency and lock-in as they have been developed in evolutionary economics are complex and somewhat counterintuitive in the sense that they set out to explain how structured patterns of development—across both space and time—can result from seemingly chance or contingent occurrences. It has proven effective in explaining why and how certain technologies prevail in the competitive setting of the marketplace, although they may not always be technologically superior. The challenge in applying the concept to other disciplines and problems—such as the genesis of clusters—lies in determining precisely what aspects of a developmental path or trajectory can be attributed to underlying causes or preconditions, and what aspects are the result of chance or contingent events.

The evolutionary approach to economics reminds us that economic systems change over time, but in ways that are shaped and constrained by past decisions, chance events, and accidents of history—implying a certain degree of serendipity. Current decisions and events are not determined by past ones, but they are conditioned by them. As a result of past choices and events, certain possibilities are easier to pursue in the present, others less so. This is key to appreciating the implications of path dependency for locational theory and accounts of cluster formation. However, extending the concept of path dependency from the narrowly technological to the social and political dimension raises a series of problems—both for academic researchers and for active policymakers. The challenge lies in reconciling the role that chance events play in launching a specific technology on the path to market dominance or a particular region to enjoying a concentration of firms in a cluster with the scope for subsequent change in broader institutional structures and settings.

As it has been applied to locational theories of cluster development, path dependency has downplayed the role of serendipity or chance occurrences in launching the initial genesis of individual clusters in specific locations and overemphasized the subsequent advantage enjoyed by these regions against potential competitors. It suggests that the trajectory of specific regions and localities is rooted in a series of economic, social, and cultural factors that influence their development over time. The presence, or absence, of key institutional elements of the local innovation system may affect both their innovative capacity and their potential to serve as nodes for cluster development. However, path dependency should also remind us that the confluence of these factors in a specific location may have initially resulted from a set of chance events or occurrences rather than the conscious designs of private or public agents. This poses a significant challenge for policymakers charged with the goal of promoting the emergence and development of clusters in their local or regional economy. The following chapter explores relation of path dependence to previous theorizing in the fields of economic geography and locational analysis and its contemporary value for both understanding the genesis of clusters and the practical constraints on policy designs for promoting their development.

Path Dependence and the Origins and Growth of Clusters

The concept of path dependence originates with the desire of evolutionary economists to account for the factors which determine the selection

mechanisms that exist within the process of technological choice and the natural trajectories that emerge from those patterns. Brian Arthur (1994) and Paul David (1997) used path dependence to explain how and why certain technologies emerged and prevailed over competing technologies in periods of rapid innovation when the marketplace was characterized by a number of alternative technological designs. Paul David defines a path-dependent sequence of economic changes as one in which important influences upon the eventual outcome can be exerted by temporally remote events, including those dominated by chance elements rather than systematic forces. He suggests that in a dynamic process, positive feedbacks are generated by strong technical complementarities on the supply side of markets, and/or the interdependence of customer preferences operating on the demand side. These may arise as well from learning effects and habituation associated with the sunk cost effects of new technologies—such as those involved in learning how to use a new program. But he also insists that the concept of path dependence does not mean that economic outcomes are predetermined. Instead he quotes approvingly from Douglas North to reinforce his point that 'contingent probabilistic events have a place throughout the dynamic process' (David 1997). North argues that 'At every step of the way, there were choices—political and economic—that provided real alternatives. Path dependence is a way to narrow conceptually the choice set and link decision-making through time. It is not a story of inevitability in which the past neatly predicts the future' (North 1990: 98–9). AQ1

There is a closely related idea within the evolutionary approach—that of increasing returns. It refers to a process in which, once a particular economic change occurs, it becomes self-reinforcing. Brian Arthur, who is equally credited with elaborating the concepts of path dependence and increasing returns, maintains that in many areas of economic activity, stabilizing forces do not seem to operate; rather, positive feedback amplifies the effects of small economic shifts. The presence of positive feedbacks and the phenomenon of increasing returns make possible many equilibrium points rather than the single equilibrium point posited by the neoclassical model based on the notion of diminishing marginal returns. Once a set of chance events or a series of small historic accidents push the technological trajectory of a new product or process onto a certain path, the prevailing technology may become locked-in regardless of the purely technical advantages of the competing alternatives. The initial advantage may be acquired through small, seemingly insignificant events and the triumphant variant is not necessarily the technically superior or more

efficient one. Its dominance may be based purely on the fact that it was the first to gain wider acceptance in the marketplace, which many supplying businesses, distribution networks, supporting technologies and users, and a large community of users and developers, all converged on its design. However, once it establishes a lead, further technological development is locked into the trajectory or path set by the dominant products. Competing technologies that were available at the outset quickly fade from view and become little more than historic footnotes (Arthur 1988*a*, 1988*b*).

Evolutionary economists, historic sociologists, and economic geographers have expanded on the original application of the concept. While the specifics of the application vary across this range of disciplines, social scientists suggest that path-dependent analysis shares several common features. In the first place, it involves the study of causal processes that are sensitive to a series of events which occurred in the early stages of the causal sequence. Events that occur early in the sequence tend to exert a disproportionate influence over the long-term development path of the sequence. Secondly, these early events involve a high degree of chance or contingency that cannot be explained purely on the basis of the starting conditions or initial factor endowments. Similar starting conditions may lead to a wide range of possible outcomes. This fact makes it particularly difficult to forecast patterns of development based on the initial conditions. Finally, once the chance events have occurred, the path-dependent sequence exhibits a more deterministic pattern, involving a large degree of irreversibility. In economic and geographic systems, the degree of irreversibility is strongly reinforced by the effects of increasing returns to scale (Mahoney 2000: 510–11).

The complementary concepts of path dependence, increasing returns and lock-in have obvious relevance for understanding the historic paths taken by regional clusters. Once a regional cluster establishes itself as an early success in a particular set of production activities, its chances for continued growth tend to be high. While this may be to some extent reducible to the success of dominant lead firms in the region, the more interesting aspect of this process has to do with the collective processes and forces at work: local social and economic institutions and culture. By the same token, ailing places may also face great challenges in improving their fortunes, for the same reason. Once a path-dependent trajectory of decline becomes established, institutional and cultural lock-in will make deviation from this path a serious challenge.

The rich geographic literature on path dependence, increasing returns and lock-in has its own distinctive evolutionary trajectory. Within this

literature three broad approaches can be distinguished that emphasize different aspects of the dynamics in regional development. The first approach focuses on the way in which the initial economic activity in a region triggered by accidents of history tends to become self-reinforcing. The second approach, following in the tradition of Marshall (1927), places AQ2 greater emphasis on the influence of agglomeration economies and supply side externalities. The concentration of critical factors of production in specific regions tends to reinforce the effects of increasing returns in the region. Finally, a more recent tradition focuses on the extent to which the benefits derived from externalities in the form of knowledge spillovers are frequently tied to ensembles of related capabilities. From this perspective, the economic advantages conferred by the institutional infrastructure of the region are a vital element in the supply architecture for learning and innovation. We build on these previous findings when addressing the issue of how to promote the growth of cluster-based development within the nexus of innovation, experimentation, and learning.

Highly influential have been the classic works by Myrdal (1957), Hirschman (1958), and Kaldor (1970) on disequilibrium models of regional economic development. These authors endeavored to show how initial economic activity, triggered by accidents of history and geography, become self-reinforcing over time and lead to growing geographic unevenness and inequality. In Kaldor's version of the story, early growth in the core region sets in motion private and social dynamics based on increasing returns to scale. Myrdal and Hirschman similarly outline a process of circular and cumulative causation, defining an evolutionary path in which backwash or polarization effects (such as selective out-migration of skilled labor and the net outflow of capital from peripheral to core regions) outweigh spread or trickle-down effects so that initial growth in the core region begets further growth, and initial disadvantage in peripheral regions becomes amplified over time. In this manner, initial events trigger long-term processes of interregional divergence which are extremely difficult to reverse. Indeed, the primary motivation for all three of these authors was to justify why public sector intervention at the national level was necessary in order to overcome these powerful, increasing-returns dynamics exhibited at the regional level.

One of the first to link the concept of increasing returns to the division of labor and, at least implicitly, the geography of production systems is Allyn Young (1928), who noted how the intricate set of interdependencies between firms in a well-developed social division of labor leads to increasing returns dynamics. Young's early insights have stimulated a

more recent reexamination of the dynamics of growth from an explicitly geographic perspective (Scott 1988; Storper 1999). In recent years, economists, such as Paul Krugman (1991*a*, 1991*b*, 1991*c*, 1991*d*) and Brian Arthur (1994), have drawn upon this rich tradition of earlier ideas within economics and geography to fashion more formalized models of territorial development. Krugman's intellectual debt (1991*a*, 1991*b*, 1991*c*, 1991*d*) to Kaldor and the other early adherents of increasing returns theories of economic development is especially clear. He builds on these ideas, as well as Alfred Marshall's original thinking (1920, 1927) on the nature of agglomeration economies, by specifying the types of supply side externalities that generate localized increasing returns. The first source is the large, deep pool of specialized labor created by the concentration of firms within the similar industries in the same location. The second arises from the fact that a local concentration of firms in the same industry can also support a larger number of specialized local providers of intermediate inputs and services, and thus reduce the cost to firms. Finally, the colocation of similar firms in a region can generate positive technological externalities or spillovers that can flow more easily among the similar firms than over longer distances (although Krugman is more skeptical about this externality). Overall, Krugman endeavors to show that the phenomenon of increasing returns is a key aspect of the process of industrial clustering that leads to a pathway of increasing sectoral specialization in particular regions over time (Baptista 1998: 27–9; Krugman 1991*a*, 1991*b*, 1991*c*, 1991*d*). While Krugman's work focuses on the way in which scale economies and positive externalities can feed the process of industrial clustering, Brian Arthur's work focuses more specifically on the way in which agglomeration externalities contribute to the concentration of firms in specific regions.

More recently, Maskell and Malmberg (1999) have argued that the competitive success of firms depends on their ability to develop sustainable, distinctive capabilities. These capabilities are most likely to arise from nonubiquitous and tacit forms of knowledge related to products, processes, and organizational routines within the individual firm. However, they will also arise from socially organized assets, such as localized, learning-based, interfirm relationships, that are not easily replicated by (groups of) firms elsewhere. Maskell and Malmberg (1999: 173) argue still further that a region's institutional architecture accumulates and changes incrementally over time, and 'thus represents the intricate contemporary interaction between elements of different ages ... from the very old (religion, beliefs, and values) to the recent/current (contemporary

industry standards, current regulations, etc.)'. Because of these properties, this institutional endowment can become a key part of a region's non-replicable asset base, thereby reinforcing durable local competitive advantages that are difficult for competitor regions to emulate. Gertler (2004) develops this idea more fully, showing how this regional institutional architecture consists of social structures that shape the attitudes, norms, expectations, conventions and—ultimately—the practices of individuals and firms in the region through informal and formal means of regulation.

These factors together determine the technological specializations of individual countries and regions; a pattern of specialization that may actually be increasing, despite the increasing reaches of globalization. They also provide an important clue for our understanding of how the trajectories of development for particular regions or local economies may be conditioned by the preexisting conditions—in turns of productive competencies in older technologies and products that firms located in the region enjoyed. In certain instances, this can help explain how a series of small chance events were able to take hold more successfully in the fertile soil of one region rather than another and launch it on a new path of development. However, one danger with these interlinked concepts is that they can serve as a double-edged sword—both to explain the social and technical bases of success for certain regions, but also to suggest the existence of constraints on the potential for others.

The Origin of Clusters: Theoretical Foundations and Empirical Findings

While policymakers seem intent on finding policy-relevant solutions to this problem, the academic literature has been less successful in formulating a clear and consistent set of answers to this question, leaving the field to a host of consulting firms that have emerged to guide municipal and regional governments through an increasing array of cluster initiatives (Sölvell, Lindqvist, and Kertels 2003). Still, Michael Porter, widely recognized as one of the leading authorities on cluster research and policy, is surprizingly clear on the factors that contribute to cluster formation and equally clear on the potential role that policy can play in their formation. He does not phrase in terms of path dependence and increasing returns; rather, he traces the roots of a cluster to his well-known diamond model of competitive dynamics. That is, cluster emergence depends on the local conditions for factor input, demand, firm strategy and rivalry, the presence of related

and supporting industries and, finally, historic circumstances (serendipity) and policy.

Although in principle he affords all four corners of the diamond equal weight as factors contributing to the seeding of clusters, he clearly privileges the role of factor input conditions such as specialized skills and talent, specific areas of expertise in the research infrastructure, an attractive physical location, and especially supportive infrastructure (Porter 1998: 237). The key assets that determine the viability of a cluster are firm based. Of particular importance is the emergence of a core or anchor firm for the cluster. Whole clusters can develop out of the formation of one or two critical firms that feed the growth of numerous smaller ones. Examples of the role played by this kind of anchor firm can be found in the case of Medtronic in Minneapolis, or MCI and AOL in Washington, DC, or in the Canadian case, by the role of Northern Electric (now Nortel) in the genesis of the Ottawa telecom cluster.

Once a cluster is launched by this combination of locational assets, chance events, and entrepreneurial dynamism, Porter affords a strong degree of importance to the role of increasing returns and feedback. The emergence of a major anchor firm in the cluster acts as a magnet for the local cluster, attracting both allies and rivals to locate in the region to monitor the activities of the dominant firm. This is the case with San Diego, where Nokia, Ericsson, and Motorola all located their CDMA wireless research efforts to benefit from Qualcomm's leadership in the field, or in Ottawa, where Cisco and Alcatel both acquired local firms to benefit from the high degree of optical and telecommunications expertise in the region, largely spun out of Nortel, the cluster's anchor firm. This raises the critical issue for policy analysts—what precisely is the relationship between the local antecedents that formed the basis for the genesis of the cluster and the specific events that triggered its emergence? And which of the two elements is most amenable to policy influence and which is the product of broader economic factors less likely to respond to policy stimuli?

The Knowledge–Entrepreneur Nexus

A more fully developed explanation of the way antecedent conditions are transformed by trigger events into the genesis of a cluster is precisely what is missing in prevailing cluster formation theories. While it is important to acknowledge that the concept of chance does not lend itself to formal theorizing, closer examination of numerous cases suggests that we

actually can isolate specific factors which constitute the trigger for cluster emergence. Feldman, Francis, and Bercovitz (2005) present a descriptive model which provides this link by placing entrepreneurship at the center of the process of cluster formation. Entrepreneurs act as the key agents who build on the existing base of institutional assets that provide the local antecedents for cluster formation. Entrepreneurial activity stimulates the development of industrial clusters over time in a series of three phases. The first stage constitutes the latency phase in which a strong base of labor skills or human capital, or a significant research infrastructure is created in a region. The presence of these underlying assets is not sufficient on its own to trigger the process of cluster formation. What is required is some external shock to the regional economy that dramatically alters the opportunity cost for entrepreneurship and new firm formation. It may come in the form of a major downsizing in government laboratories or the unwillingness of a large research-intensive firm or laboratory to pursue new technological opportunities. These shifts ultimately lead employees, whether they are laid off in the downsizing or frustrated by the inability to pursue new commercial possibilities, to reexamine the opportunity cost of starting their own firms (Feldman, Francis, and Bercovitz 2005). The likelihood of this occurring is further enhanced when a movement along the technological frontier in key industries opens up a range of new opportunities for these entrepreneurs to exploit. Such technological shifts are frequently associated with a realignment of leadership positions among national economies, but they can also have the same effect on regional and local economies, as entrepreneurs in new localities are the first to perceive and act upon the potential created by these shifts (Zysman 1996). One of the reasons why the uptake of these opportunities occurs more rapidly in these new regions is that there is no lock-in to the existing technologies or production paradigm that prevailed previously.

In this phase the cluster evolves further as entrepreneurs establish their own networks and build the deep institutional structures that constitute the industrial system or supply architecture of a region described earlier. Once a critical mass of new start-up firms has emerged, the entrepreneurial founders of the firms begin to form the support organizations needed to both sustain their own activities and encourage new entrepreneurs to take the plunge. These organizations engage in a range of activities, including peer-to-peer mentoring and the creation of angel networks that are essential to diffusing the knowledge and skills needed to grow and expand the cluster. Further, the establishment of these organizations raises the profile of the emerging cluster in both the local economy and more distant ones

and helps generate the kind of buzz that attracts new entrants and talent into the regional economy. In such a way, the institutional assets of the region are extended (Maskell and Malmberg 1999; Storper 1999). The final stage occurs when there is a fully functioning entrepreneurial environment where the success of the initial start-ups creates additional possibilities for new ones as well as spin-offs. This stage is also marked by the emergence of local VC to fund the activities of the second tier of start-ups (see also Chapter 9). This can come either from the success of a few first generation entrepreneurs cashing in and beginning to redeploy their assets or from venture capitalists from outside the region, drawn to it by the perceived explosion of investment opportunities.

This perspective is echoed in the work of Swann and Prevezer (1998) on high-technology clusters. The seeding of clusters at particular junctures is strongly influenced by relatively minor historic events. They see positive feedback as a key factor playing a central role in the formation of clusters. Firms are drawn initially to a specific location by strong demand for their products or services in the location, a large supply of highly skilled or scientific labor, and a network of complementary strengths in neighboring firms; once the cluster has begun to develop, this process is accelerated by the presence of a critical mass for firms due to the positive feedback (or increasing returns engendered by colocating with similar firms). The further development of clusters is affected by two key dynamics: entry factors that attract new entrants to a cluster, and growth promoters that support the growth of incumbent firms in the cluster. The feedback process is important in accelerating the growth of clusters by enabling more sharing and transmission of tacit knowledge. Such knowledge spillovers primarily occur through labor mobility and/or the informal sharing of knowledge among technical staff at different firms.

The case studies in this volume, as well as others, serve as examples of the way this process has evolved in a number of instances. Kenney and Patton (Chapter 3) underline the coevolution of technologies and institutions with respect to the origins of the Silicon Valley high-tech cluster. However, they also put a high priority on the role played by the underlying assets of the region, as do Braunerhjelm and Halverson in their analysis in Chapter 7 of the factors that led to the emergence of the Danish/Swedish biotechnology cluster Medicon Valley. Principally the universities (both publicly and privately funded) and corporate research laboratories provided the intellectual space for the growth of the Silicon Valley cluster—going back to the role of Frederick Term as in the prewar and early postwar period, in fostering a strong degree of entrepreneurialism among Stanford

engineering and computer science graduates and encouraging some of them to found their own firms. In the period from the 1970s onward, in which the contemporary contours of the cluster emerged, the computer science and electrical engineering departments at UC Berkeley contributed significantly to the process of new firm formation—a role that is often overlooked in the preeminence assigned to Stanford University. Two corporate research laboratories, IBM's San Jose laboratory and the Xerox Palo Alto Research Centre, provided additional sources of new technologies for commercial exploitation and the entrepreneurs to bring them to market. The essential contribution made by these four key components of the research infrastructure is that they developed new technologies for start-up firms to exploit and attracted both talented personnel and entrepreneurs to the region (Kenney and Patton, Chapter 3).

A comparable case in the Canadian context that illustrates the long-term impact of building a strong research infrastructure is the contemporary information technology cluster in Waterloo, Ontario. The University of Waterloo, inaugurated in 1957, was established due to a confluence of local and national demand for more sophisticated and technical educational institutions. The strong postwar expansion of local industries generated a rising demand for technically trained labor that was not being met. Many local business leaders felt that the future competitiveness of the region depended on the establishment of world-class educational facilities. These concerns led to the creation of the University of Waterloo—a school that would specialize in a scientific and technical curriculum. Acutely conscious of the financial limitations that would exist for a new university; the local business advocates developed a unique solution in the form of the Waterloo Plan. This plan called for a new type of education to be offered on a cooperative basis with local industry. In sharing the burden of technical training with industry, the university would be able to support double the number of students (as one class rotated out to cooperative placements, another would take its place in the classroom), provide a greater depth of education—both theoretical and practical—and build a closer relationship with industry in order to anticipate employment needs, secure additional funding and ensure that classroom education remained on the cutting edge. Over the next three decades, the University of Waterloo came to be widely recognized for the strength of its mathematics, computer science and engineer programs, as well as the unique aspects of its cooperative system. In the late 1970s, these long-term investments by the local community and two senior levels of government bore fruit as key spin-offs from the university began to seed

the growth of local high-tech industry. While cluster formation was not an essential part of its mandate, the University of Waterloo has served the regional economy in two important ways: by providing a pool of local talent and by transferring cutting-edge knowledge, either in the form of entrepreneurial spin-off companies or through patenting, licensing, consulting or joint research projects (Nelles, Bramwell, and Wolfe 2005).

The Multifaceted Dynamics of Cluster

While the Silicon Valley, Medicon Valley, and Waterloo cases provide strong evidence of the way in which antecedent conditions lay the groundwork for the conditions that trigger the entrepreneurial spark, another Canadian case provides a striking illustration of the way in which external shocks to the same can provide the trigger mechanism. The roots of the Ottawa telecom cluster can be traced back to the presence of federal government laboratories in the national capital region, many of which underwent substantial expansion during the research intensive period of World War II. This dense research infrastructure provided the fertile ground on which the telecom cluster took hold. However, the external shock which was delivered to the region took the form of the consent decree signed between the US Department of Justice and AT&T and its subsidiary, Western Electric in 1956, forcing them to make patent holdings available to other firms without charge and release technical information to outside suppliers. Up to that point, Western Electric had owned 44 percent of Northern Electric, the dominant equipment supplier to Bell Canada, but the consent decree forced the withdrawal of the US firm from the Canadian market. Western Electric gradually terminated its patent and information agreements with Northern Electric, out of fear that their liberal provisions would have to be extended to other North American firms. By 1962, AT&T and Western Electric had divested themselves of their holdings in Bell Canada and Northern Electric (Macdonald 2000).

Cut off from its easy access to US patents and technical information, the primary sources for its product designs and development, Northern Electric realized that it needed to develop its own in-house R&D capacity to replace the designs previously licensed from Western Electric. It began the search for a location for the new research facility and, despite the fact that most of its manufacturing was in Montreal and southern Ontario; it eventually bought a substantial tract of land on the outskirts of Ottawa to be the home for Bell Northern Research. The main attraction of the

national capital region was the large concentration of scientists and engineers employed at the National Research Council laboratories and the Defense Research Board in technical areas of interest to Northern Electric. The Defense Research Board also offered training courses in advanced technologies, such as transistors and was authorized to transfer its technical innovations to firms in the private sector. Bell Northern Research recruited leading research scientists and engineers to its laboratories from outside the region and even the country—many of whom ultimately became leading entrepreneurs in the Ottawa telecommunications and photonics cluster (Chamberlin and de la Mothe 2003; Macdonald 2000).

The central role played by inadvertent government policy in seeding the Ottawa cluster is paralleled to some extent in Maryann Feldman's account of the emergence of the telecommunications cluster in the Washington–Baltimore corridor. Feldman's analysis emphasizes the importance of entrepreneurship in stimulating the genesis of that cluster. She traces the roots of the entrepreneurial drive to the massive wave of downsizing and outsourcing that occurred in the US federal government in the late 1970s and 1980s. As a result of this trend, employment conditions in the federal public service became less secure and future prospects deteriorated. In the same period, public sector pay scales lagged behind those for executives in the private sector. An increased emphasis on outsourcing goods and services for the federal government provided a further inducement for prospective entrepreneurs to leave the government and start firms to supply goods and services back to their employer. Other policy initiatives launched in the early 1980s facilitated the licensing and transfer of technology from federal laboratories and provided further support for innovation in small businesses. 'Enterprising scientists licensed technology out of their own university or government research labs to start new companies and chose to locate the new companies near their existing homes' (Feldman 2001: 878). The strong concentration of federal research expertise in the nation's capital established the research infrastructure for the growth of a cluster, as in the Ottawa case. Although cluster creation was not the primary goal of the federal government's downsizing, the inadvertent role played by public policy in the formation of the cluster cannot be overlooked, together with an environment conducive to entrepreneurial initiatives.

Similar dynamics can be discerned from the evolutionary paths of biotech regions. In their analysis of Boston–Cambridge, Massachusetts, and the San Francisco Bay Area, Owen-Smith and Powell (Chapter 4) argue that

these two leading biotech regions evolved along very different trajectories that continue to generate distinctive outputs. In the case of Boston, publicly funded research organizations (PROs—i.e. universities and hospitals) provided the initial knowledge base that led to subsequent commercial application. They suggest that academic rivalry between Harvard and MIT was an especially important characteristic of this regional network in its early years. Many of the Boston area biotech firms were founded by senior professors from these two institutions, most of whom kept their academic affiliations. With the passage of time, as the local network matured, these biotech firms developed a larger number of relations with both venture capitalists and other biotech firms. By the end of the 1990s, Genzyme and Biogen had developed a large number of linkages to other local biotech firms. Despite this late proliferation of firm-to-firm linkages, PROs, such as MIT, Harvard, and Massachusetts General Hospital, were still very important elements of the Boston network at the end of the 1990s.

By contrast, biotech firms in the Bay Area exhibited strong ties to the local VC community from the earliest days of the industry's development (late 1980s)—the original pattern being set when UCSF biochemist Howard Boyer partnered with venture capitalist Bob Swanson to establish Genentech in 1976. Owen-Smith and Powell attribute this pattern to 'the prospecting and matchmaking efforts' of the venture capitalist community, from which many firm founders emerged. To the extent that academic researchers were involved in firm start-ups, they tended to be at much earlier stages in their careers and were considerably more likely to leave their home institutions (either temporarily or permanently). In subsequent years, local network connections developed to include some linkages to PROs, but these were dwarfed by the rapid growth in linkages to other biotech firms. One of the densest local subnetworks developed around two lead firms established early in the region's evolution: Genentech and Chiron. Given their less academic origins and closer links to the venture community, Bay Area firms tended to pursue more commercially focused, exploitative research along incremental trajectories.

The larger implications of the work by Owen-Smith and Powell for our analysis are considerable. First it is clear that, despite the knowledge-intensive nature of biotechnology, the direct role played by universities in stimulating initial local development through spin-offs and commercialization can vary dramatically, even between two admittedly successful regions. Second, this finding also reminds us of the perils of reading off causal relations from spatial associations in ex post analysis of successful clusters: in the case of biotech at least, the local presence of Stanford,

UCSF, and UC Berkeley may not have been adequate, on its own, to have seeded the development of a world-leading biotech cluster in the Bay Area. The catalytic accelerator appears to have been the density of local VCs—itself a legacy of earlier rounds of venture-based industrial development focused on ICT-based firms. Restating this finding in terms of the language adopted above, the antecedent conditions are likely to vary dramatically between different biotech regions. On the basis of this analysis, Owen-Smith and Powell strongly caution against the formulation of standard models of regional innovation-based success to guide policy intervention.

Notwithstanding the critical role of a strong knowledge base in science-driven clusters, Romanelli and Feldman (Chapter 5) emphasize the strong connection between the entrepreneurial factor and cluster emergence. Entrepreneurs bring several key capabilities and assets to these processes that position them as key agents of cluster genesis and evolution. Foremost, they embody a creative spark—that is, an ability to identify viable new business opportunities amidst considerable uncertainty concerning technologies and markets. Successful long-term cluster growth also depends on local entrepreneurial firms' ability to spin-off new second-generation firms at later stages in the cluster's development. These dynamics have been especially visible in San Francisco, Boston, and San Diego, while New York, Los Angeles, and Washington, DC have been relatively less successful in generating second waves of entrepreneurial spin-offs, or in attracting entrepreneurs migrating into the region from elsewhere. Romanelli and Feldman attribute the relatively poor performance of the latter regions to their failure to generate a strong community of biotherapeutics entrepreneurs. In contrast, entrepreneurs in San Francisco, Boston, and San Diego were strongly bound together by histories of cross-institutional collaboration as well as common educational and research backgrounds.

Policy Implications

Emphasizing the importance of chance events and the central role played by entrepreneurial initiative in the genesis of clusters does not eliminate the role for public policy. While it is virtually a commonplace to state that governments cannot create clusters by fiat or direct policy intervention, the preceding account of the evolutionary and path-dependent character of cluster genesis makes it clear that government policies play a critical role at many different stages of cluster formation and growth. It is important to be clear about the most valuable initiatives at the individual stages of

cluster development. The critical insight that the evolutionary perspective affords is that multiple locational outcomes are possible in the early stage of cluster formation, as shown by Scott (Chapter 2) and Kenney and Patton (Chapter 3). This potential makes it difficult, if not impossible for regional policymakers to target the development of specific clusters (Lambooy and Boschma 2001).

Conversely, the importance of local antecedents for cluster development means that policy, across multiple levels of governance, can contribute to the accumulation of key assets in a specific location. As evident from several of the case studies in this volume, as well as pervious studies, chance events that act as triggers for cluster formation or the entrepreneurial spark occur within a specific historic and geographic context. Frequently, it is public sector agencies that are critical in establishing the local antecedents that define this context. The public sector encompasses federal, state or provincial, and local governments; as well as public research institutes like Canada's National Research Council or US government laboratories and institutions of higher education (although this would include leading private universities in the USA). In some instances, private sector research laboratories or contract research organizations can also lay the groundwork for the emergence of clusters, with strong support from public sector funding. While the ultimate impact of these policy interventions cannot be fully anticipated at the outset, over the long term, those policy interventions that strengthen the research and institutional infrastructure of a region or locality have the greatest potential to act as attractors for a cluster of firms (Wolfe and Gertler 2004).

Public policies that create a strong knowledge base in the regional economy and contribute to the creation of a well-educated workforce establish the local antecedents that can support the emergence of clusters. While a strong research infrastructure and a thick labor market are distinctly local phenomena, in most industrial countries they are not exclusively the result of local, or even state and provincial government policies; the presence of the senior level of government lurks in the background. Several of the cases reviewed above underline the important roles played by different scales of political jurisdiction in the genesis of clusters. The literature on path dependency and divergent national trajectories, and the importance of culture, reinforces the point that national institutions shape the context for local development (Gertler 2002; Zysman 1994, 1996). Thus clusters can be seen as being nested within, and impacted by, other spatial scales of analysis, including regional and national innovation systems, each of which adds an important dimension

to the process of knowledge creation and diffusion that occurs within the cluster.

The case of Silicon Valley clearly illustrates the way in which these differing scales of governance impact on the development of local clusters. The cluster exists within the distinctive features of the US system of innovation—with its unique system of laws, regulations, and conventions governing the operation of capital markets, forms of corporate governance, R&D, and other relevant factors. A number of these features are central to the story of Silicon Valley's growth and development, including the highly decentralized nature of the postsecondary education system with complementary and interlocking roles for both the federal and state governments. The federal government played a central role as the initial customer for many of the early products of the cluster. For most of the 1960s, the US defense and space programs consumed the largest portion of the cluster's output of integrated circuits. The US government was also the primary funder for much of the critical R&D that underpinned the growth of key segments of the computer and related industries in the cluster (National Research Council 1999*a*). Even in the celebrated case of Xerox' Palo Alto Research Centre mentioned above, the initial staffing of key laboratories benefited immeasurably from the extensive research networks that had previously been developed through the Department of Defense's Advanced Research Projects Agency (Hiltzik 1999). Once the cluster began to emerge in the 1960s and 1970s, institutional change—such as subsequent changes in capital gains tax rates, the tax treatment of stock options, and the rules governing investments in VC by pension funds—coevolved to further strengthen the cluster by facilitating the growth of a VC industry. As shown in this volume (Chapters 3 and 9), this seems to be a decisive step in the emergence of high-technology clusters. Hence, understanding the multiple factors that influence the development trajectory of a cluster and ultimately its economic performance is necessary.

The other cases considered in this volume and in the previous literature provide equally clear evidence of the critical contribution made by policy interventions from all three levels to the genesis and growth of the clusters studied. In the case of the Capitol region in the USA, the dense concentration of federal laboratories constituted the breeding grounds for a whole new generation of entrepreneurs in the telecom and biotech sectors. However, a series of federal policy interventions in the early 1980s, in response to the perceived decline in the competitiveness of the US economy, reduced the barriers and increased the incentives for nascent entrepreneurs to exploit the commercial potential of intellectual property

generated by public research funding in these government laboratories (National Research Council 2003). The passage of the Federal Technology AQ3
Transfer Act in the USA in 1986, which stimulated the creation of Cooperative Research and Development Agreements (CRADA) allowed federal agencies to partner with small firms to develop new technologies (Guston 1998). A large number of the biotech firms that emerged in the Capitol region in this period were the product of CRADAs with government laboratories. Finally, the introduction of the Small Business Innovation Research Program in 1982 that set aside a certain portion of federal R&D funding for small business provided a critical source of funding for small business start-ups in the Capitol region, with local firms receiving a
significant proportion of funding under this program (Feldman et al. AQ4
2003; National Research Council 1999*b*). Collectively, this major shift in US policy at the federal level in the early and mid-1980s provided a powerful impetus for capitalizing on the crucial knowledge base in the research infrastructure of the Washington, DC region and stimulating the entrepreneurial impulse in the cluster.

In the Canadian cases considered above, the role of the federal and provincial governments in building the local research infrastructure and building up the resources of highly skilled labor was equally critical. In the Waterloo, Ontario case, the mobilization by local business leaders to secure a charter for a new university, financed with federal and provincial funding, and their foresightedness in structuring a curriculum around math, sciences, and engineering and creating a pioneering program of cooperative education, all laid the groundwork for the future emergence of a strong information technology cluster. In this case, it was the specific pattern of interaction of dynamic, visionary leaders at the community level, with the increase in combined federal and provincial funding for postsecondary education that strengthened the local antecedents essential for the emergence of the information technology cluster. In the case of Ottawa, the Canadian capital, the dense concentration of federal government laboratories in telecommunications served as the magnet that drew Northern Electric's primary research facility to the region.

This is corroborated by the case studies in Chapters 6–9 on the emergence of biotechnology clusters in China and Denmark/Sweden, and of the ICT clusters in Ireland and Israel, which all points to the crucial role played by policymakers. In some cases, policies have intervened quite strongly—particularly in China, Ireland, and Israel—in order to build institutions and markets, whereas more general policies have been pursued in other cases (e.g. Sweden and Denmark). Orsenigo emphasizes in

his survey in Chapter 10 that policy matters, while simulation exercises in the following Chapter 11 stress the type of policy required to foster cluster emergence.

When it comes to increasing entrepreneurial activity, the policy instruments are far less clear. As some of the cases discussed above illustrate, government policy played a critical role in stimulating the genesis of the cluster, but often in a completely inadvertent manner. In these instances, governments were pursuing policies designed to achieve other goals, but the consequences of the policy triggered the kind of chance occurrence that path dependency describes. It seems to be a key element of cluster development in the third and last stage of cluster emergence, particularly to provide an environment conducive to the entering of second and third generation start-up firms. At this stage government policies to sustain the entrepreneurial drive are perhaps the most important. These can include a broad range of government policies to support upgrading the innovative capacities of firms and promote the rapid diffusion of technologies, networks to foster greater interaction among the emerging SME's, as well as providing much needed mentoring programs for newly minted entrepreneurs. Often local high-technology industry associations emerge with support from local and regional government agencies to play this role. A key barrier that a rapidly growing cluster often runs up against is an adequate supply of the critical skills needed to feed the growing firms. This is a policy area where local universities and colleges have played a crucial role, often with the backing of state and provincial governments, in expanding training and research programs in the areas of most crucial need. The formation of angel networks and the attraction of VC into the locality can also be supported by appropriate government policies (Feldman, Francis, and Bercovitz 2005; Porter et al. 2001). In general though, government policy at the third stage of cluster formation and development is much more varied and is often tailored to meet the needs of the specific region and locality in which the cluster is located.

Furthermore, means to evaluate and strengthen the policy supports for cluster development at this stage is crucial, for example through strategic planning or innovation-based strategic planning at the regional level. The strategic planning process is valuable for helping regions develop a shared understanding of their local assets and identifying the area's unique local characteristics that support the development of regional industry clusters. These include knowledge economy assets (such as workforce skills, knowledge and research development, creativity, advanced telecommunications infrastructure, quality of place, and financial capital), collaborative

institutions and organizations (such as regional development organizations, professional networks, research consortia, and entrepreneurial support networks), and the regional mindset (values and attitudes that encourage innovation, entrepreneurship, and collaboration). Strategic planning exercises have also been used to identify key gaps in the region's mix of assets as well as common opportunities that may be exploited by its existing or emerging clusters. The common framework for understanding the region's potential and the shared vision generated through such a planning exercise can also help mobilize support at the local level for key activities needed to boost the cluster (Gertler and Wolfe 2004; Porter et al. 2001).

Conclusions

The overall lesson extracted from the case studies and the previous literature is that the path dependencies for cluster creation are highly variable and that the chance events which provide the trigger for cluster formation can come from many sources. There is a strong element of serendipity in virtually all of the cases described above and any policy analyst or cluster consultant that would try to design a formula for cluster growth on the basis of these lessons would be wildly optimistic, to say the least. However, virtually all of the cases strongly reinforce the point made concerning the intersection of historic context and chance occurrence in launching a regional or local economy along a certain trajectory of development. Public sector involvement can affect cluster trajectories in a variety of ways, though the impacts are often unpredictable and even, in some instances, unintended. Whether intentional or inadvertent, one of the most effective public policies for seeding cluster development is a sound investment in building the research infrastructure and educated labor base in a region. The establishment of a strong local talent pool of highly skilled and knowledgeable workers both feeds the growth of the local entrepreneurship in the cluster as increasing returns begin to take hold, and attracts outside firms and entrepreneurs to the cluster to gain access to the local buzz.

Similarly, the ability, or inability, of the local or regional economy to develop the underlying conditions of trust and social capital that contribute to the presence of a learning economy may create a condition of lock-in to a specific innovation trajectory. A related question that needs to be explored is whether the conditions that can provide a supportive culture and institutional framework for a specific regional or local economy can

also be influenced by direct intervention, and if so, how effectively. The presence, or absence, of key institutional elements of the local or regional innovation system may affect both their innovative capacity and their potential to serve as nodes for cluster development. Many clusters enjoy the knowledge assets and research infrastructure that are necessary for the development of an innovation-based development strategy, but they differ dramatically in their capacity to mobilize these assets in the pursuit of such a strategy. Experience demonstrates that local communities can formulate strategies to alter their economic trajectory and improve their chances of economic development. The successful initiation of this kind of process depends on the ability to collaborate across boundaries—both geographic and social.

AUTHOR QUERY

[AQ1]: North 1990 is not cited.
[AQ2]: Marshall (1927) is not listed.
[AQ3]: National Research Council 2003 is not listed.
[AQ4]: Feldman et al. 2003 is not listed.