The exaggerated death of geography: learning, proximity and territorial innovation systems

Kevin Morgan*

Abstract

Globalization and digitalization have been presented as ineluctable forces which signal the ‘death of geography’. The paper takes issue with this fashionable narrative. The argument that ‘geography matters’ is pursued in three ways: first, by questioning the ‘distance-destroying’ capacity of information and communication technologies where social depth is conflated with spatial reach; second, by arguing that physical proximity may be essential for some forms of knowledge exchange; and third, by charting the growth of territorial innovation systems.

Keywords: learning, physical proximity, organizational proximity, territorial innovation systems

JEL classification: O31, O32, P16, R11

Date submitted: 17 March 2002   Date accepted: 30 January 2003

1. Introduction

Rarely more than a minority sport in conventional economics, the spatial dimension is now being written off completely in some quarters on account of the twin processes of digitalization and globalization, processes which supposedly signal the ‘death of geography’. Even within economic geography itself, some theorists are beginning to question the prominent role which physical proximity is assumed to play in shaping the spatial distribution of economic activity. It is now being suggested, for example, that distance per se is not necessarily an impediment to the acquisition and diffusion of knowledge, even of tacit knowledge, because organizational or relational proximity can act as a surrogate for physical or geographical proximity.

Paradoxically, at the same time as geography is being laid to rest in some conceptions of the ‘knowledge economy’, its significance is being affirmed by others, especially by evolutionary theorists of innovation, where it is deemed to be an important influence on learning, innovation, and development. To explore these competing narratives in more detail the paper aims to address the following three themes.

First, it aims to examine the roots of the ‘geography is dead’ thesis and to argue, among other things, that this thesis grossly over-estimates the distance-destroying capacity of information and communication technologies (ICT) by conflating spatial reach with social depth. Because information diffuses rapidly across organizational and territorial borders, it wrongly assumes that understanding does too.

* School of City and Regional Planning, Cardiff University, Glamorgan Building, King Edward VII Avenue, Cardiff, CF10 3WA, Wales, UK.

email <morgankj@cardiff.ac.uk>
Second, it examines the rival claims of organizational versus physical proximity and suggests that, while these are not mutually exclusive, something gets lost when we communicate at a distance, even when the participants know each other well. Drawing on evolutionary theories of innovation, and recent corporate practice, it argues that firms continue to set a high premium on physical proximity in the organizational design of complex activities.

Finally, it explores the growth of territorial innovation systems and suggests that, for all their limitations, these constitute another reason as to why geography matters, and that is because the spatial economy is a politically structured system in which nations, regions, and cities seek to influence the trajectory of economic development in their respective jurisdictions. One of the key questions here of course is whether learning and innovation are organic, self-activating processes, as they tend to be in core regions, or whether they can be consciously induced through collective action in the context of less favoured regions.

2. Digitalization and globalization spell the death of geography?

The ‘geography is dead’ thesis has much in common with economistic theories which proclaim the death of the nation-state because of the growth of multinationals or because of the growth of global markets (Ohmae, 1990). With the advent of ICT, especially the seemingly instantaneous communications capacity of internet and intranet technologies, it is often assumed that space–time relations have been so radically compressed that it is possible to completely annihilate space with time. The rapid diffusion of ICTs (in OECD countries at least) certainly offers firms new and hitherto unavailable opportunities to restructure their activities: for example, to reorganize work practices, to strike a new balance between centralization and decentralization of command and control functions, and to engage in telemediated products and services.

At the macro level the effects of ICT are potentially even more transformative and, from the standpoint of this paper, two kinds of transformation are worth noting. First, the tradability of output is already changing in new and unpredictable ways, particularly in the service sector, where a new generation of tradable services is emerging as a result of the ICT-driven separation of production and consumption. This means that many service sector jobs which were once considered to be place-specific, and ‘sheltered’ from international competition, are becoming less dependent on the places where the service is actually consumed—even electronic surveillance, so it is claimed, can be conducted on the other side of the globe (Cairncross, 1997).

The other major transformation concerns the effects of ICT on the ways in which information and knowledge are produced, stored and diffused. Perhaps the most important part of this transformation is the fact that ICT accelerates the codification of knowledge and modifies the balance between codified and tacit knowledge. Codified knowledge, being explicit and standardized, can be transferred over long distances and across organizational boundaries at low cost and ICT enables such knowledge to be made available more quickly and more cheaply than ever before. In contrast, tacit knowledge, being personal and context-dependent, is difficult to communicate other than through personal interaction in a context of shared experiences. The process of codification is sometimes likened to a spiral movement in which tacit knowledge is transformed into codified knowledge, followed by a movement back to practice where new forms of tacit knowledge are developed and this spiral movement lies at the core of individual and organizational learning (Nonaka and Takeuchi, 1995; Foray and Lundvall, 1996).
The ‘death of geography’ thesis draws heavily on these two particular transformations: tradability, because it allegedly frees the provision of services from their point of consumption; codification, because it allegedly reduces knowledge to a universally accessible form of information—and information, for some ‘digital beings’, is reducible to bits. One such ‘digital being’ is Nicholas Negroponte, the director of the MIT Media Laboratory, who champions an extreme version of the thesis when he says:

The digital planet will look and feel like the head of a pin. As we interconnect ourselves, many of the values of a nation-state will give way to those of both larger and smaller electronic communities. We will socialize in digital neighbourhoods in which physical space will be irrelevant and time will play a different role’ (Negroponte, 1995, p. 6).

These sentiments are echoed in the economic and business school literature in particular (Martin, 1996; Cairncross, 1997). Indeed, as ICTs become more powerful, and as virtual reality becomes more mimetic, this kind of ‘spaceless’ thinking could gain more credence despite the fact that it is profoundly misplaced. The notion that cyberspace will ever evolve into a genuine surrogate for geographic space is at best doubtful and this is fundamentally because it is difficult to imagine the rich diversity of physical proximity, where the nuances of body language and face-to-face communication convey as much as (if not more than) verbal communication, being matched by virtual proximity. This brings us to the most serious shortcoming of the ‘geography is dead’ thesis, namely that it conflates spatial reach with social depth and hence fails to recognize that it is the latter, with its wider scope for social reciprocity, which is the essential prerequisite for deep learning. Instructively, while academic theorists continue to debate the relative merits of physical versus virtual proximity, corporate managers seem to have resolved the issue, finding ‘the quality of face-to-face interaction higher than the electronic variety, even between people who know each other well’ (Lorenz, 1995).

To avoid sterile polarizations between physical and virtual proximity, between geographic space and cyberspace, the most defensible position might be to acknowledge that these intersect with one another in a complex fashion: that is to say, cyberspace is not a paraspace, a separate realm to geographic space, but forms part of ‘an experiential continuum in people’s lives’ (Dodge and Kitchin, 2001). Virtual proximity may well be a surrogate for physical proximity in the context of standardized transactions, but not in the context of transactions which are high in complexity, ambiguity and tacitness. Far from being mutually exclusive, then, ICTs and face-to-face communication will co-evolve as complementary mechanisms (‘emails and hallways’, ‘wheels and wires’, in other words), with the precise combination depending on the nature of the transaction and the degree of familiarity of the participants. But the fundamental point is this: digital technologies may be adept at maintaining communities that are already formed; but they are not so good at creating them in the first place (Brown and Duguid, 2000).

3. Organizational versus physical proximity: learning, knowledge, and distance

Although it remains under-explored, the role of geography is beginning to be more widely appreciated in evolutionary theories of innovation and technological change. In contrast to the traditional neo-classical approach, which takes as resolved some of the biggest questions in economic development—like what firms know and how they learn, for
example—the chief merits of the evolutionary approach are twofold: the realism of its core propositions about economic behaviour and its focus on dynamic rather than static analysis. By placing learning, knowledge, and innovation at the centre of its analytical agenda, evolutionary political economy seeks to understand how this trinity contributes to the uneven processes of capitalist development (Nelson and Winter, 1982; Dosi et al., 1988).

In traditional neo-classical theory all agents are assumed to be equally capable of ‘optimizing’ because economic competence (broadly understood as problem-solving skills) is thought to be relatively abundant, when in actual fact it is scarce, idiosyncratic and unevenly distributed as between individuals and firms (Pelikan, 1988; Foss and Knudsen, 1996). There are significant variations, in other words, in firms’ knowledge bases and major differences in their capacity for creating knowledge from within and absorbing it from without. The uneven distribution of economic competence, which is firm-specific and partly tacit, helps to explain the wide variations in corporate performance and why apparently superior organizational forms diffuse slowly, if at all, within and between sectors, regions, and countries (Nelson, 1991; Dosi and Coriat, 1994). In short, while all capitalist firms nominally share the same profit-seeking goals, what differentiates them—in terms of competence, organization, culture, and cognitive frameworks, for example—seems so much more striking than what unites them (Cooke and Morgan, 2000).

One of the paradoxes of the ‘knowledge economy’ is that it has spawned greater uncertainty, especially for the firm, the key repository of productive knowledge. The most palpable sign of this heightened uncertainty is the burgeoning debate about how to measure and report ‘intangible assets’ (R&D, proprietary know-how, intellectual property, brands, workforce skills, organizational competence, networks of customers and suppliers, goodwill, and the like). A growing chorus of critics maintains that conventional ‘balance sheet’ accounting is based on a fiction—namely that the valuations which auditors produce reflect the real value of the firms they audit. The over-emphasis on physical assets (land, plant, capital, etc.) and the under-emphasis on intangible assets transmits totally inappropriate signals to managers, employees, shareholders, and investors. At one level the information deficiencies can be read as the result of accounting shortcomings (e.g. the fact that spending on intangibles is treated as a current expense, while spending on physical and financial assets is capitalized). More fundamentally, however, the ‘information failures’ concerning intangibles are better understood as being rooted in the unique attributes of these assets—like high risk and the absence of markets for example (Lev, 2001).

These preliminary points need to be made because glib references to the ‘knowledge economy’ tend to obscure the problems facing the firm as it struggles to manage its intangible, relational and knowledge-based assets, particularly how it measures the returns to non-physical investment. In its more apocalyptic forms, the rhetoric of the ‘knowledge economy’ elides the fact that firms have to cost-justify their outlays on knowledge-creating assets, a discipline that is more of an art than a science—but an art which is easier to practice in some countries than in others, as we shall see in the following section.

No less of an art is the task of putting existing, untapped knowledge to better commercial effect, a frustratingly difficult task because it involves the vexed question of tacit knowledge. Although this is a recalcitrant asset from a managerial standpoint, the incentives to harness tacit knowledge, through better ‘knowledge management’ routines for example, are growing and the main incentive was expressed by Lew Platt, the former
chief executive of Hewlett-Packard, in the celebrated statement: ‘If HP knew what HP knows, we would be three times as profitable’ (Caulkin, 1998). Here Platt was referring to the untapped knowledge in a company which could otherwise claim to be one of the most successful knowledge-creating companies ever.

The renewed interest in tacit knowledge is largely due to its perceived social and spatial significance when learning and innovation are at a premium: socially, because tacit capabilities like team skills and organizational routines constitute the core competence of firms; spatially, because tacit knowledge, being person-embodied and context dependent, is locationally ‘sticky’, a characteristic which helps to explain the clustering of knowledge-intensive activities (Storper, 1997; Maskell et al., 1998; Gertler, 2001b).

Tacit knowledge was the name given to knowledge that cannot be articulated by Michael Polanyi, who famously captured its essence by saying: ‘We can know more than we can tell’ (Polanyi, 1966). Tacit knowledge was contrasted to explicit or codified knowledge, a formalized knowledge which could be transferred in a de-personalized manner through technical blueprints and operating manuals, etc. Being personal and context-dependent, tacit knowledge represents disembodied knowhow that is acquired directly through interactive learning (Howells, 1996). Originally designed to contest the notion of a depersonalized exact science which produced a wholly ‘objective knowledge’, Polanyi’s insights had a wider application, and they were successfully applied to the field of organizational capability in Nelson and Winter’s seminal text on evolutionary economic theory (Nelson and Winter, 1982).

Like Polanyi, Nelson, and Winter do not draw a hard and fast line between tacit and codified knowledge, stressing instead the need to identify the ‘degree of tacitness’ involved in a skill or a process. Given current debates about the feasibility and the desirability of codification (Cowan et al., 2000; Johnson and Lundvall, 2001), it is worth recalling Nelson and Winter’s judicious observation that costs matter here. That is to say, the relevant question is not whether some knowledge is in principle articulable or necessarily tacit, but whether the costs of codification are sufficiently high so that the knowledge remains in fact tacit. The relative significance of the tacit dimension will depend, therefore, on a combination of costs and context:

The knowledge contained in the how-to-do-it book and its various supplements and analogues tends to be more adequate when the pace of the required performance is slow and pace variations are tolerable, where a standardized, controlled context for the performance is somehow assured, and where the performance as a whole is truly reducible to a set of simple parts that relate to one another only in very simple ways. To the extent that these conditions do not hold, the role of tacit knowledge in the performance may be expected to be large (Nelson and Winter, 1982, p.82).

The problem of codifying tacit knowledge is further compounded by the metrics used to assess skill, creativity, and intelligence: the dilemma here is that the most valuable problem-solving skills (i.e. ‘practical intelligence’), much of which is acquired through everyday activities, often unconsciously, tend to elude conventional tests for academic and emotional intelligence.

Perhaps the most systematic treatment of tacit knowledge to date is the work of Nonaka and Takeuchi, who draw on the Japanese corporate experience to develop a theory of ‘knowledge conversion’ in which tacit knowledge is progressively converted into more widely accessible organizational knowledge through an intensely iterative, spiral-like process of collective learning (Nonaka and Takeuchi, 1995). Although these authors attach a great deal of significance to tacit knowledge, they never suggest that the tacit
realm is unknowable or untappable. But tapping it is not easy because ‘knowledge
conversion’ is a hugely demanding organizational exercise: indeed, far from being
a technical fix for a select few at the top, this process makes enormous demands on
the entire workforce. The key point to emphasize about their theory is that ‘the most
powerful learning’ comes from direct experience, from face-to-face communication and
from the use of the body not just the mind.

They also highlight the role of trust in expediting organizational learning. Building trust
requires ‘the use of mutually understandable, explicit language and often prolonged
socialization or two-way, face-to-face dialogue that provides reassurance about points
of doubt and leads to willingness to respect the other party’s sincerity’ (Nonaka and
Takeuchi, 1995). Although there is a lively debate about trust underway in the social
sciences—especially about how it is secured and how it is sustained—the evidence suggests
that this relational asset carries costs (like lock-in) as well as benefits. The main benefits
of trust would seem to be first, that it saves time and effort to be able to rely on others;
second, that it reduces risk and uncertainty; and third, that it expedites learning because
the parties are privy to thicker and richer information flows on account of the fact that
people divulge more to those they trust (Storper, 1997; Cooke and Morgan, 2000).

The literature on trust and cooperation also suggests that these relational assets are
more likely to develop where the participants expect to meet again, in other words where
‘the shadow of the future’ looms large over the present (Axelrod, 1984). This provides
a context for reciprocity: a good deal of informal know how trading takes place, even
among rival firms, precisely because of the expectation that the information which A
provides B today will be reciprocated in kind tomorrow (von Hippel, 1987). Although
these exchanges can of course take place at a distance, providing the climate of reciprocity
exists, they are easier to organize in the context of physical proximity (Malmberg, 1997).3

Crucially, however, the significance of physical proximity will ultimately depend on the
complexity of the project (e.g. the degree of tacitness involved) and the socio-spatial
context (e.g. the degree of physical and cultural distance involved). For example, the
literature on technology transfer is littered with examples of projects which were
compromised by a failure to appreciate that users need a good deal more than
hardware from suppliers: what is needed above all is mutual understanding, and this
requires a common code through which information can be understood (Lundvall, 1988;
Gertler, 1995). Most technology transfer research concurs with Teece when he argues that
projects with a high tacit component require nothing less than ‘intimate personal contact’
to succeed (Teece, 1981).

In their different ways these stylized accounts signal a simple, but fundamentally
important truth: namely that something gets lost, or degraded, when individuals and
organizations communicate at a distance, even when they know each other well. Although
this was well understood in traditional economic geography, in more recent variants the
costs of a spatially distantiated division of labour have received far less attention than the
benefits which are said to accrue to the firm from this form of organization. For example,
Massey’s pioneering work on spatial divisions of labour remains a robust statement about

1 The notion that trust requires long-term relationships is challenged by Gernot Grabher in his excellent
analysis of communities of practice in the advertising industry, where project teams work on a short-
term, task and finish basis. But because these teams can expect to recombine at some point (creating a
‘shadow of the future’ effect), there is a requirement for trust, but it tends to take a diffuse rather than a
personal form (Grabher, 2001).
the social processes and spatial patterns of uneven economic development, particularly how large firms allocate different corporate functions to different regions, with the result that spatial hierarchies come to mirror corporate hierarchies (Massey, 1984). The corporate benefits may be clear here, but what of the costs? Clearly some of the costs stem from the separation of R&D and production, a division that reflects a deeper, and more debilitating separation in some firms between conception and execution.

Before addressing the key question—which is whether organizational proximity can substitute for physical or geographical proximity—it is worth probing further into this problem of distantiation because it involves far more than spatial distance. With the professionalization of the R&D function the 'lab' became more socially exclusive and more spatially separate from other corporate functions, and the shortcomings of the linear model of innovation owe a lot to this separate identity. The barriers to learning and innovation in the linear model were actually exposed 40 years ago, when spatial distantiation was far more modest (Burns and Stalker, 1961). Here we encounter some familiar problems:

1. the 'one-way traffic of designs from laboratories to production shops'
2. the divisions between engineers, production and sales staff had features 'which can best be called linguistic'
3. development engineers freely conceded that, for them at least, the production workshop was 'a terra incognita'.

Anticipating a whole series of contemporary themes Burns and Stalker observed that 'as laboratories grow larger, and specialist groups multiply, there is a danger of some essential channels of communication becoming attenuated or severed merely because of the presence of so many channels of communication around the individual'. This led them to conclude that 'the fewer the links in the chain from development to production, the more, that is, development and production were forced to learn each other's language, the more effective, speedy and trouble-free was the passage through of designs'. The fact that this prescient observation was first made over four decades ago illustrates the point that recognizing a problem does not dispose of it—that is to say, organizational innovations do not diffuse as quickly as we think.

Perhaps what resonates most deeply today about the Burns and Stalker study is their emphasis on innovation as a kind of 'linguistic' project, in which language, meaning, identity and direct communication were the most essential elements of successful 'knowledge management'. In other words they were arguing for a 'shared language' through which the different functions of the firm could talk and understand each other and through which the firm could secure organizational coherence and some commonality of purpose.

Subsequently, this critically important insight would be re-stated in many different ways: for example, Arrow spoke of 'the need for codes which are mutually understandable', codes which imposed 'a uniformity requirement' (Arrow, 1974); Nelson and Winter thought of 'prevailing routines' as a 'truce' which helped to regulate potentially destructive conflict within the firm (Nelson and Winter, 1982); and more recently Dosi and Marengo have underlined the significance of a 'common language' through which members of the firm can develop a shared cognitive framework for the purposes of communication and coordination (Dosi and Marengo, 1994).

If this is what constitutes organizational proximity then we need to remember that it is a moving target, a process not an event, an aspiration which is never wholly attained because, in practice, the large firm is too heterogeneous to meet Arrow's 'uniformity
requirement’. Large firms face enormous problems when they seek to create and sustain a ‘shared language’ throughout the organization, especially between R&D and so-called ‘downstream’ functions. To illustrate the problem of organizational proximity, and what firms are doing to address it, let us very briefly consider three corporate cases: Xerox serves to highlight the basic problem and GE and BMW highlight the organizational innovations being used to overcome it.

Creating a ‘shared language’ across the firm is not easy at the best of times, but it is especially difficult on the cusp of a new technological era, a problem which is perfectly illustrated in the story of how Xerox supposedly ‘fumbled the future’. As we now know, it was a group of pioneering scientists at Xerox’s Palo Alto Research Center (PARC) that first developed the elements of the personal computer in the 1970s. Despite this major technical achievement...

... most of the extraordinary knowledge generated at PARC never crossed the boundary between the scientists in Palo Alto, and the development engineers in Dallas or the management in Stamford. The engineers found the scientists arrogant, and almost unintelligible. Management found them naive and unrealistic. The scientists, for their part, regarded almost everyone in the corporation outside their own community as ‘toner heads’—unable to think of the world beyond photocopiers (Brown and Duguid, 2000).

That it was Apple, not Xerox, which developed the PC was largely due to the fact that, during a visit to PARC, one of Apple’s founders ‘was able to see what Xerox management could not, the potential of what PARC had generated. So Apple licensed what it could and replicated what it could not. The knowledge that stuck within Xerox leaked readily through its front door’ (Brown and Duguid, 2000). One of the many implications of this classic case study is that, instead of having a ‘shared language’, Xerox had multiple languages, each being the preserve of a particular community of practice, and cognitive distance was in this case compounded by the physical distance between the sites. Erica Schoenberger has convincingly demonstrated that Xerox was organizationally unable to learn from its ‘peripheral’ R&D teams, whether these were based in California or Japan (Schoenberger, 1999).

GE is instructive because, over the past decade, it has sought to avoid the problems at Xerox by embedding R&D in a wider strategy designed to create a ‘boundaryless organization’. In fact the key aim is to create new incentives for sharing ideas so that information and knowledge circulate more freely rather than being hoarded for personal gain. While each of GE’s operating divisions has its own R&D facility, the corporate R&D centre at Schenectady, in New York State, is the intellectual hub of the company. Before the reforms of the 1990s Schenectady culture was akin to a campus-style lab, in which scientists had little or no incentive to commercialize their technical projects: indeed, it was not uncommon for them to ‘throw an idea over the wall to the business division, sit back and say my job is done’ (Dickson, 1992). When exhortation failed to change the culture, GE introduced two structural changes to the way central R&D operated. First, the operating divisions have to directly fund more R&D projects at the centre, giving them a stronger vested interest in what happens at Schenectady. Second, the centre’s staff has to spend more time in the divisions, where they get to know each other in face-to-face situations. The combined effect of these two reforms was to create more of a shared destiny, and therefore a ‘shared language’, between the R&D centre and the operating divisions.

The BMW case highlights an even more extreme way to achieve a ‘shared language’, namely co-location. To ensure that the product development process is as integrated as
possible BMW embarked upon a radical experiment in which some 6000 professional staff are co-located at its Research and Engineering Centre, to the north of Munich, in what is believed to be the largest single concentration of vehicle engineering expertise in Europe. In the belief that R&D staff are most productive when they can interact on a *face-to-face* basis, the architecture of the Centre has been designed in such a way that no one has to walk more than 50 metres to meet a colleague. Despite its name the Centre is much more than a conventional R&D facility because it represents an unprecedented co-mingling of skills, including research, design, development, manufacturing, personnel, procurement, and patents. Such extreme co-location is designed to achieve one fundamental goal—to reduce the development cycle of new models. This is the process where iteration between different disciplines is most important and where tacit knowledge is most pronounced, and co-location is deemed to be the key mechanism for tapping these intangible assets and for developing a ‘shared language’ through which to do so (Cooke and Morgan, 2000).

Since Xerox is not an isolated case, it is clear that informal divisions within the large firm are not a sign of organizational abnormality; on the contrary they are part of the normal state of affairs, an issue which is fruitfully explained by reference to the literature on ‘communities of practice’. The latter are close-knit, practice-based groups in which ideas and tacit knowledge diffuse rapidly because members are bound by a shared understanding and a common identity (Brown and Duguid, 1991; Wenger, 1998). Although these practice-based communities are in the vanguard of learning and knowledge diffusion in the firm, their codes and practices are often at variance with official codes and practices, and the PARC vignette is an extreme example of how debilitating these internal divisions can become. Potentially powerful conduits of innovation, communities of practice are also a potential source of instability, hence they are a mixed blessing in governance terms.

The ‘communities of practice’ concept now lies at the heart of a new debate in economic geography—the terms of which have been admirably laid out by Meric Gertler—and it concerns the question as to whether organizational proximity can be a surrogate for geographical proximity as a means of producing and diffusing tacit knowledge (Gertler, 2001a,b). In a recent series of papers Ash Amin and others have questioned the ‘taken for granted’ propositions which appear in economic geography and evolutionary economic theory. Although these arguments do not seek to pronounce the ‘death of geography’, they do in fact challenge the significance of physical proximity. In a concise formulation Amin asks:

> Is it not ‘relational proximity’—more specifically, ongoing organisational routines and the social practices of collectives implicated in a common venture—rather than geographical proximity, that constitutes the ‘soft’ architecture of learning? Such relational proximity might, of course, draw on face-to-face contact, but it can also be achieved at a distance (Amin, 2000; see also Amin and Cohendet, 1999; Oinas, 2000).

Stimulating as it is, this argument is problematical in at least three ways. First, it tends to juxtapose relational and organizational proximity on the one hand with geographical proximity on the other, a form of spatial fetishism which the authors endeavour to contest in every other respect. The spatial fetishism lies in the assumption that there is something called ‘geographical proximity’ which does not involve *relational* proximity, implying that the social interactions which constitute ‘local’ action are somehow natural, primordial, or automatic, when in fact they have to be actively constructed like any other relational asset, whatever the spatial scale (Cooke and Morgan, 2000).
The second problem is an over-exaggerated sense of what can be accomplished at a distance, whether it is through the virtual proximity of digital technology or the occasional proximity associated with business travel. Although there is a mechanism for transferring tacit knowledge across organizational boundaries and national borders (namely internationally mobile communities of practice), the latter do not offer the same scope for reciprocity, serendipity, and trust that is afforded by sustained face-to-face contact, a point freely conceded by some of the originators of the communities of practice concept (Brown and Duguid, 2000; Gertler, 2001b).

Thirdly, to the extent that communities of practice are confined within the firm, their learning opportunities would seem to be narrower than the opportunities on offer in the ecologies of knowledge which characterize advanced regional clusters. Writing of Silicon Valley, one of the most celebrated ecologies, Brown and Duguid (2000, pp.168–169) make a more general point about the intersection of locality and organization:

For the ecology to flourish, however, it evidently needs not just a range of capabilities, but a close range. The informal links … develop directly and in close quarters. In the Valley, people live in and out of each other’s pockets, and this helps them see what’s doing, what’s doable, and what’s not being done. This close proximity not only shows how to attack a particular niche, it provides the ability to see a niche before it is visible to most eyes … Density of firms, practices, and practitioners also promotes reliable risk- and trust-assessment … So distance is far from dead, even where distance technology is at its most advanced (emphasis added).

The spatial core of these ecologies of knowledge may be a regional cluster, but the outer boundaries might straddle multiple spatial scales, from the local to the global, because some of the firms which constitute the ecology will be multi-locational organizations. This point certainly merits more attention because there is a tendency to juxtapose, as alternative models of learning and innovation, the localized business networks of the industrial district model with the more formal and distantiated networks of the large firm.

It is also important to be clear about what this defence of geography does not entail. It does not mean that tacit and codified knowledge are being treated as separable entities, nor does it portray the local as ‘a unique source of tacit knowledge for competitive advantage’ (Amin, 2000). Still less does it mean that the tacit-codified distinction corresponds, spatially, to the local-global dichotomy (Allen, 2000). The most defensible view of tacit knowledge is not that it is immobile and confined to the ‘local’, but that it is person-embodied, context-dependent, spatially sticky, and socially accessible only through direct physical interaction.

These are the special features of tacit knowledge which help us to explain what otherwise looks like a remarkable aberration in a supposedly ‘hyper-mobile’ global economy, that is the phenomenal spatial concentration of R&D activities in the home base of the innovating firm—memorably referred to as ‘an important case of non-globalisation’ (Pavitt and Patel, 1991). Equally instructive, this SPRU study also found that the proportion of innovative firms’ activities which were performed at home tended to increase with the technological intensity of the industry and the firm, a sign of the premium which firms attach to having highly tacit activities co-located at the early stages of a major innovation (Feldman, 1994; Malmberg, 1997; Audretsch, 1998).

Admittedly, the spatial pattern of corporate R&D is becoming less starkly concentrated, as firms perform more of their innovative activities abroad, but this should not be construed to mean that these knowledge-intensive functions are becoming
impervious to geography. Rather than being binary opposities, globalization and localization in this respect tend to be complementary processes because overseas affiliates seek to tap into local clusters of expertise, a process which tends to enhance rather than erode national and sub national patterns of specialization (Cantwell, 1995; Archibugi and Michie, 1997).

Aided and abetted by ever more sophisticated digital technology, travel, and a modicum of cross-cultural ‘fusion’, organizational proximity may be a partial substitute for geographical proximity, especially for people who are already part of a community of practice, but partial is the operative word (Blanc and Sierra, 1999). A key question for future research therefore is not which form of proximity is better, since both are necessary, but rather how will they co-evolve in practice at a time when ‘localized’ learning and knowledge networks are evolving into complex ecologies composed of different organizations that straddle multiple spatial scales?

4. The scope and limits of territorial innovation systems

If geography is being buried in some quarters it seems to be undergoing a remarkable re-birth in others. That two radically different narratives can co-exist seems untenable until one realizes that they tend to be addressing different aspects of the same picture. The ‘death of geography’ school is fixated by the pace and scale of globalization, with its standardizing imprimatur; the ‘geography matters’ school, on the other hand, is impressed by the tenacity of spatial differentiation, with its national, regional, and local nuances. These two tendencies—standardization and differentiation—constitute a permanent dialectic in the spatial economy, making ‘geographical outcomes a two-way street between localization and diffusion, not a one-way highway to dispersion’ (Storper, 1997).

If the ‘forces of globalization’ were as ineluctable as they are said to be, then national patterns of development might be expected to converge around some world norm. But historical reality tells a different story. Far from converging around some bloodless norm, the advanced OECD countries actually exhibit very different sectoral patterns of technological and trade specialization, and these patterns show considerable stability over time, with little or no sign of convergence, implying that ‘geographical proximity continues to play a very significant role for knowledge flows’ (Guerrieri, 1999; see also Verspagen, 1993; Pavitt and Patel, 1994). Specialization itself seems to confer certain advantages, so that ‘being specialised appears to be even more important than choosing the “right” field’ (Archibugi and Pianta, 1992). Such high levels of specialization help countries to secure leading or even dominant positions in sectors where they have developed a finely honed expertise, like pharmaceuticals in the UK, machinery in Germany, fine chemicals in Switzerland, mechanical engineering in Italy, and electronics in the US for example.

These national forms of specialization and comparative advantage are very often based on distinctive subnational formations, be they localized clusters or core regions. In the European Union for example, just 12 core regions account for nearly half of all research and technological development, and these ‘islands of innovation’ represent the core regions of the advanced member states. This pronounced regional pattern of specialization appears to be even more durable than the national patterns referred to earlier (Cooke and Morgan, 2000; Breschi, 2000).
Whatever the spatial scale, these deeply-embedded patterns of specialization reflect the fact that the growth of know-how (managerial, technological, and organizational) is a profoundly cumulative, path-dependent process—a process shaped less and less by natural endowments and more and more by competencies and capabilities built over time and ‘channelled into specific trajectories by increasing returns’ (Maskell et al., 1998).

Evolutionary political economy rightly allots an important role to the institutions which shape, and which are in turn shaped by, these deep developmental processes. Like all structures, these institutions are both the medium for, and the result of social action: in other words they enable and constrain what firms and other agents wish to accomplish. Although the evolutionary account remains the most convincing of all the stylized accounts of learning and innovation, it nevertheless leaves much to be desired as regards the interplay between its macro- and micro-level narratives. Exactly how, for example, does a national system of innovation influence the behaviour of its firms?

Contrary to fashionable notions of ‘techno-globalism’ and ‘borderless worlds’ the national environment remains a highly significant operating milieu for firms, even for so-called multinational firms. Simply consider the following for example: in the main OECD countries some 90% of production is for the home market; domestic investment by domestic capital far exceeds direct investment overseas plus foreign investment at home; national stock exchanges tend to trade in domestic stock; multinational firms are more accurately referred to as national firms with international operations; labour markets and industrial relations are largely governed by nationally specific regulatory regimes; and national borders are proxies for cultural, political, linguistic, and cognitive affinity (Wade, 1996; Berger and Dore, 1996). It is against this background that the concept of a national system of innovation (NSI) was developed, a concept which has been defined in narrow technological terms (Freeman, 1987) and more broadly as a nationally structured social system of interactive learning (Lundvall, 1992). Whatever the nuances the key elements of such a system tend to include some or all of the following:

1. the R&D system, particularly its sectoral composition and the division of labour between publicly funded and business funded R&D spending
2. the education and training system, particularly the division between academic and vocational skills
3. the financial system, particularly the interface with industry and its capacity to provide ‘patient capital’
4. the network of user-producer relationships and the norms of interaction (e.g. exit versus voice-based relations)
5. the associational capacity of the system, that is the extent to which firms forge dynamic linkages with their institutional milieu, be it local, regional, national or international (Cooke and Morgan, 2000).

The NSI helps to explain why firms, even multinational firms, tend to have a ‘national character’, a cognitive framework which influences the way they look at the world, the way they do things, how they discount time and therefore how they calculate opportunity and risk. Although the NSI does not in any sense determine corporate behaviour, it certainly renders some courses of action easier than others. A classic example here would be the differences between the national systems in Germany and the UK: the financial system in the UK, being more ‘short-termist’ and less ‘patient’ than in Germany, helps to privilege short-term divided payments over long-term R&D outlays. Another contrast would be
skills: a woefully inadequate vocational training system makes it more difficult for UK manufacturing firms to emulate German productivity levels or German quality product strategies for example. This is not to say that some UK firms cannot equal or surpass their German counterparts, rather that certain courses of action are encouraged, and thus easier to adopt, in some national systems than in others.

Although it demonstrates why national patterns of development remain important, the NSI literature leaves a number of questions unanswered, four in particular. First, this literature tends to focus on the formal science and technology system, as though learning was synonymous with and confined to R&D activities. This bias makes it difficult to pick up the very important processes of informal learning and organizational innovation which take place in traditional sectors in large countries like Italy or in small countries like the Nordic countries (Maskell et al., 1998).

Second, the relationship between national and sectoral patterns of innovation is still under-developed despite some promising work on ‘systems of innovation’ (Edquist, 1997). In particular we need a much better understanding as to why strong sectors manage to develop in weak national systems, like pharmaceuticals in the UK for example, and how national systems interact with sector-specific ‘technological regimes’ (Malerba and Orsenigo, 1994).

Third, the dichotomy between macro- and micro-level narratives cries out for more attention because we simply don’t know enough about the different ways in which firms actually use their national systems. The uneven distribution of economic competence means that wide variations in firm behaviour will co-exist in each national system, but what does the system do for ‘laggard’ firms? Conversely, do ‘leading’ firms compensate for deficiencies in their national system by doing more in-house or by using alternative national systems for certain activities? Multinationals, for example, are trying to ‘graft’ some features from their domestic system on to their new system, and we need to know if this ‘mix and match’ strategy is leading to more hybrid national systems.

Finally, just as firms behave differently within each national system, so do the localities and regions which compose the ‘national’ economy. One of the most serious gaps in the classical NSI literature was its silence on subnational institutions, mechanisms which can play an important role as bridging institutions in diffusing knowledge and keeping local firms abreast of new practices (Cooke and Morgan, 1994, 1998).

Over the past decade this subnational level has attracted considerable attention, even from some mainstream economists who, having discovered ‘geography’, have proclaimed it to be alive, well and an important factor in understanding a country’s growth dynamics (Porter, 1990; Krugman, 1991). The recent literature on subnational territorial development has spawned a bewildering array of terms to cover the different permutations that are claimed to exist, from informally arranged local clusters to formally constituted regional innovation systems. Originally triggered by the discovery of ‘industrial districts’ in Italy, mono-industrial areas where dense local networks seemed to confer scale and scope advantages to small firms acting in concert, the subnational realm has opened up new perspectives on learning and innovation as interactive and reflexive processes (Camagni, 1991; Maskell et al., 1998).

Indeed, the most sophisticated attempt to explicate the territorial dimension of these twin processes suggests that the guiding metaphor of economic development needs to be revised. Instead of it being cast in exclusively mechanical terms, with hard inputs and outputs, it should also be viewed as a process of ‘conversation and coordination’, where economies are understood as ‘stocks of relational assets’ in which ‘untraded
interdependencies’ (the conventions and informal rules that coordinate economic life) need to be introduced to help explain the phenomenon of localization (Storper, 1997).

Building on Marshall’s notion of ‘localization economies’ (pools of skilled labour, specialized intermediate inputs, knowledge spillovers, and a supportive industrial atmosphere) Porter has done most to popularize these ideas under the rubric of spatial clusters, which he defines as:

... geographic concentrations of interconnected companies and institutions in a particular field. Clusters encompass an array of linked industries and other entities important to competition. They include, for example, suppliers of specialised inputs such as components, machinery, and services, and providers of specialised infrastructure. Finally, many clusters include governmental and other institutions—such as universities, standards-setting agencies, think-tanks, vocational training providers, and trade associations—that provide specialised training, education, information, research, and technical support (Porter, 1998, p.78).

Whereas economic geographers tended to treat clusters as a special case of economic development, Porter claims that clusters are ‘a striking feature of virtually every national, regional, state, and even metropolitan economy’ (Porter, 1998). In contrast to Porter’s somewhat simplistic notion of clusters, more rigorous researchers have argued that the cluster literature has failed to substantiate its claims about extensive locally traded transactions, with the result that the concept of localization economies remains ‘elusive’ (Malmberg and Maskell, 2001).

Malmberg and Maskell bring a more robust analysis to the cluster debate by distinguishing between horizontal and vertical relationships and by demonstrating that localization economies can be independent of the degree of internal interaction. In other words clusters can exist even if there are no locally traded transactions (the vertical dimension) because a more important dimension may be the knowledge-creating effect of similar firms being able to monitor each other at little or no cost (the horizontal dimension). Far from challenging the existence of spatial clustering, this important argument highlights the need for a more rigorous analysis as to why the cluster exists and what benefits it confers on firms—and they suggest that clusters exist less because of cost reduction or input–output reasons, but primarily because of the scope for enhanced knowledge creation:

When firms co-locate, a spatially defined community is usually formed that makes it easier for them to bridge communication gaps resulting from heterogeneous knowledge endowments. The innovative capabilities of firms are enhanced because co-location can provide them with an arsenal of instruments to obtain and understand even the most subtle, elusive and complex information of possible relevance. Hence the process of clustering tilts the balance between advantages of specialization and costs of coordination so that a higher level of knowledge creation can be obtained. The ability to de-code and utilise knowledge residing elsewhere is not a phenomenon to be captured by input/output analyses of trade flows or accounts of business contact patterns (Malmberg and Maskell, 2001, pp.17, 18).

If the nature of inter-firm relations needs to be better understood in the cluster debate, so too does the role of territorial innovation systems. To the extent that we can speak of local or regional innovation systems—which essentially consist of the firm and its subnational network of institutional support—we need to remember that these are not national systems writ small, though they might involve elements of a national system which have been regionalized, like research laboratories for example. Generally speaking the smaller the spatial scale of the ‘system’ the more open and porous it will be, with the result that local firms will have many non-local interactions (Howells, 1999).
Despite Porter’s deceptively simple definition of clusters, there is no such thing as a standard cluster, hence the hazards of generalizing from a small number of celebrated cases. For example, at one end of the cluster spectrum we might locate ‘technology districts’, rapidly evolving production systems engaged in product-based technological learning, like Silicon Valley or Greater Boston, which are said to be ‘the most important form of territorial economy that exists today’ (Storper, 1997). At the other end of the spectrum we might locate a mono-industrial district which is based on a relatively stable technology, like the artisanal production system associated with Parmigiano-Reggiano cheese in Emilia-Romagna for example (de Roest, 2000).

This brings us to a disturbing and intriguing paradox: the growing interest in clusters, among theorists and policy-makers alike, is paralleled by an increasingly ambiguous evidence base. That is to say we actually know much less than we think we know about how firms actually learn, particularly as regards the interplay between learning and proximity, be it physical or organizational proximity (Glasmeier and Fuellhart, 1996; Oinas, 2000). Far from evoking caution from the architects of policy, however, this knowledge deficit has been sidelined as policy-makers throughout the OECD prescribe cluster-building regional policies for all regions whatever their circumstances.

Like clusters, subnational territorial innovation systems may also be more problematical than we think, at least if we distinguish between genuine innovation processes that have assumed a territorial form and the more common situation whereby localities and regions have created an enterprise support system for the express purpose of promoting innovation. Many ‘regional innovation systems’ seem to fall into the latter category and, once again, there is no such thing as a standard system (Braczyk et al., 1998; Cooke et al., 1998). The specification of a territorial innovation system needs to be more than an inventory of the institutions and the interactions considered necessary for success. Studies which have examined ‘micro-innovation systems’ from the bottom-up standpoint of the firm suggest that some or all of the following conditions need to be present to sustain the claims: localized patterns of communication, search, learning, knowledge-sharing, and innovation. Having examined the evidence one sobering conclusion was that these systems were not commonplace because ‘there are many sub-regions (and indeed regions) which lack these concentration and localization benefits because of low density, peripherality, lack of dynamic, innovative firms and institutions and being simply knowledge and information poor’ (Howells, 1999).

Just as Hayek considered it a ‘fatal conceit’ to think the state could be a surrogate for the market, since the latter was by nature a de-centralized discovery mechanism, so it may be a planner’s conceit to think that ‘institutional thickness’ is always necessary for successful innovation. Some technology districts are not thickly constituted with supporting institutions, whether public or private. In some of these technology districts the burden of innovation is largely carried by competent, leading-edge firms, in association with other like-minded firms. This helps to explain the ‘mystery’ as to why some highly successful technology districts, like Silicon Valley for example, seem to be so ‘under-populated’ with supporting institutions (Saxenian, 1994).

This point needs highlighting because the recent ‘institutional turn’ in economic geography is wont to give the impression that supportive institutions matter as much, if not more than, the firms at the heart of the innovation process. This point also has implications for the new generation of regional innovation strategies in less favoured regions. It is not that these strategies are wrong to emphasize the role of supportive
institutional networks, just that the latter cannot be a substitute for a local corporate sector, which is by definition weak in peripheral regions (Morgan and Nauwelaers, 1999; Landabaso, 2001; Morgan and Henderson, 2002; Oughton et al., 2002).

The rapid diffusion of regional innovation strategies in OECD countries has been partly kindled by evolutionary theories of learning and innovation, theories which admit that the subnational institutional environment can play a role, albeit a modest role, in stimulating learning, innovation and development. Geography matters more than globalists and digitalists may think, but uneven development is a sobering reminder that it matters for different reasons in different regions. At the dynamic end of the spatial development spectrum clustering may be one of the forms through which localized learning and innovation take place, a process that tends to be organic and self-activating. It is at the less dynamic end of the spectrum, in the context of less favoured regions, where we encounter one of the biggest questions in political economy today, that is whether localized learning and innovation can be consciously induced through judicious public intervention and new forms of collective action. The evidence from Europe’s less favoured regions may be less than inspiring, but it at least suggests that, even here, the development process is not set in aspic.

It would be a tragedy for these poor regions if their embryonic efforts to promote localized learning and innovation were to be over-burdened with the unrealistic expectations of cluster-building regional policies since clusters, contrary to what Porter seems to think, are not uniformly relevant or appropriate. If less favoured regions are to become something other than they are today, especially if they are to develop a more robust endogenous capacity for innovation and development, they will need to adopt a twin-track approach. They’ll need to recognize that local circumstances are the only meaningful point of departure for a genuinely attuned regional strategy and they’ll also need to recognize that local resources are a necessary but not a sufficient condition for progress.

References


