Why do regions develop and change?: the challenge for geography and economics

Journal of Economic Geography **11** (2011) pp. 333–346 doi:10.1093/jeg/lbq033
Advance Access Published on 7 December 2010

Why do regions develop and change? The challenge for geography and economics

*Michael Storper**,**,***,†

*Department of Geography and Environment, London School of Economics, Houghton Street, London WC2A 2AE, UK
**Master of Public Affairs, Institut d'Etudes Politiques de Paris (''Sciences Po''), 13, rue de l'université, 75007 Paris, France
***School of Public Affairs, 3250 Public Affairs Building, University of California (UCLA), Los Angeles, CA 90095-1656, USA
†Corresponding author: Michael Storper. *email* <m.storper@lse.ac.uk>

Abstract

Explaining the growth and change of regions and cities is one of the great challenges for social science. The field of economic geography and associated economics has developed frameworks in recent years that, while tackling major questions in spatial economic development, are deficient in their ability to explain geographical develop in a causal way, and to incorporate principal forces for change.

Keywords: regional growth, urban growth, metropolitan growth, regional development, urban systems
JEL classifications: R11, R12, R23
Date submitted: 23 June 2010 **Date accepted:** 1 November 2010

1. Why focus on growth and change?

Explaining the growth and change of regions and cities is one of the great challenges for social science. Cities or regions, like any other geographical scale of the economic system, have complex economic development processes that are shaped by an almost infinite range of forces. There is a thorny question as to what social science should aim to do in the face of such complexity.

In developing countries, regional and urban development is—for obvious reasons—seen as an integral part of the broad process of industrialization, with research focused on the shaping of a city system, and the spatial dynamics of population and economic activity that underpin the urban-industrial tradition (Black and Henderson, 1999). The frames of reference employed to think about such processes are mostly borrowed from the 'long view backward', that is, to how the now-developed economies generated their urban systems at the time they developed (Fujita et al., 1999; Gabaix and Ioannides, 2004). Such borrowing has the advantage of looking at the long-term, at dynamics and change. But it necessarily must 'average' over many development experiences and hence may err in its attribution of causes or in the relationship between universally-present causes and nationally- or regionally-specific contexts. Along these lines, the currently-developed countries have widely-differing urban systems, in spite of their broad convergence in many areas of economic structure and performance. Moreover, even in the developed countries that are used as templates for theorizing causes, urban

and regional development patterns are changing through various combinations of deindustrialization and population loss, growth, sunbelts and frostbelts, metropolitan expansion, and stimulated by changing patterns of specialization and employment, while generating new patterns of incomes. The past may therefore be only the most basic of prologues for thinking about future dynamics, whether in developed or developing countries.

It would be unrealistic to ask any field of theory and research— especially in an area of such complex human-technical interaction as the spatial economy—to meet all these challenges fully. But a focus on change and causality, by which I mean on studying cities and regions as forward-moving development processes—should determine what is most relevant in defining the ambitions of the field. Concretely, then, the field should be able to respond to such questions as: why do regions grow? Why do some decline? What differentiates regions that are able to sustain growth from those that are not? What are the forces that cause per capita income to converge or diverge, and under what conditions do they operate? What are the principal regularities in urban and regional growth, and what are the events and processes that are not temporally- or geographically-regular but that affect pathways of development in durable ways?

2. The progress of the New Economic Geography and the New Neoclassical Urban Economics[1]

Two branches of spatial economics have developed in recent years. The New Economic Geography (NEG) emanates from the founding paper of Krugman (1991), whereas the New Neoclassical Urban Economics (NNUE) evolves from the older traditions in urban economics, offering an ambitious and inclusive set of inter-regional spatial equilibrium models. The two are substantially different from one another, but both have significant gaps in their ability to explain certain kinds of spatial-economic dynamics.

The NEG is principally concerned with production. The NEG sees spatial concentration of economic activity as an endogenous part of the economic process, and hence not dependent on 'first nature geography' such as the uneven distribution of natural resources, climate or proximity to coasts and rivers. Scale economies induce the concentration of workers and firms, each enjoying closer contact with its markets and access to a greater variety of inputs and products, if trade costs are even mildly positive (Krugman,1991; Fujita and Thisse, 2002). Moreover, trade costs are fully integrated into this way of thinking, something which was not previously possible. Trade costs can also have certain endogenous characteristics, because local interactions can generate further scale economies and increase the gap in trade costs between local and far away economic agents (Martin and Ottaviano, 1999).

These are significant achievements, though most of the underlying insights have been around quite a long time. The generation of development economists typified by Albert Hirschman, Gunnar Myrdal, Celso Furtado and Raúl Prebisch, themselves drawing on seminal thinkers such as Wassily Leontief and François Perroux, identified the circular and uneven developmental characteristics of the economy, through cumulative and self-reinforcing processes of agglomeration (Myrdal, 1957; Hirschman, 1958). Meanwhile, scholars of cities had long described Marshallian processes of economic

1 I make the arguments in this section in greater detail in Storper (2010).

agglomeration as the basis of city economies (Pred, 1977; Hall, 1998). These earlier thinkers emphasized intermediate linkages, home market effects, and local learning, just as does the NEG today; but they did not did not place their insights in a framework with a consistent view of trade costs, open economy relationships, labour mobility, economies of scale and secure microfoundations. The NEG thus represents a fundamental leap forward.

The New Neoclassical Urban Economics (NNUE) takes a different approach, which is no less ambitious. It attempts to bring into a single framework the behaviours of firms in choosing locations, of individuals and households in choosing residence, and of developers in shaping the built environment. At its heart are how firms seek to raise productivity and individuals to satisfy their preferences for income, paid amenities and non-monetized amenities. Significantly, the NNUE introduces local politics over land and housing as a key sphere in which the possibilities of firms and individuals are defined, unevenly from one place to another. Migration is made endogenous in this complex geographical force field. The NNUE also represents a sweeping conception of urban and regional dynamics.

The two are substantial rivals with one another. The NEG would explain the rise of Silicon Valley as the effect of a self-reinforcing process of agglomeration stemming from an initial accident; and it would explain high incomes there as the result of massive economies of scale and variety, on one hand, and the skill composition of labour demand in that agglomeration, on the other. The NNUE would say that workers seek interaction and hence go to Silicon Valley; that firms cluster there to accede to such workers and benefit from their productivity; and that households seek the amenities of the San Francisco Bay Area, but that real incomes are not higher than in other regions because of the resulting high cost of housing, which is itself in part due to restrictions imposed by amenity-seeking skilled worker/citizens, who limit land-housing development. Thus, the NEG centres on economies of scale in production and consumption, and spatial development as a vast checkerboard of monopolistic competition; the NNUE makes very conventional assumptions about the structure of production and consumption (that it is divisible), and focuses on optimal preference-satisfying behaviour on the part of firms, workers, households and builders, in a perfect-competition world. In other words, the spatial sorting mechanisms for people and activity are radically different in the two theories.

3. Limitations of the NEG and NNUE

Both the NEG and the NNUE have potentially important insights about certain aspects of the process of spatial-economic development; but both have substantial gaps in explaining causality and change, and many assumptions of both are unrealistic because they are driven by requirements of theoretical consistency rather than from what occurs in the real world.

The first major substantive limitation of the NNUE is that some of the assumptions employed to explain change are highly improbable. Principal among them is the notion of 'spatial indifference'. In this view, firms can substitute capital and labour to a significant degree as they choose among different possible locations. A firm can locate in a place with higher relative labour costs by substituting more capital, in the proportions dictated by its production function, but at the same time, it can opt for

alternative locations with lower relative labour prices by employing a higher proportion of labour. In other words, the firm (via its position along the production function) is considered to be a type of 'putty' that can be shaped by alternative locations (Glaeser, 2008). Linked closely to this indifference of the firm is that of the individual-as-household or individual-as-worker: it is assumed that there are significant elasticities of money wages for such things as housing size or amenities such as climate and culture, and that these affect the choice of location between regions (and not just the choice of neighbourhood within a region). People are considered to be 'putty' in the same way that firms are.

Firms certainly do choose locations according to the relative prices of factors at different locations; but there is scant evidence that they can be truly indifferent among regions in the way required by this assumption. At any given moment, the demand for capital and labour—as it is distributed among different locations—is driven first and foremost by the semi-fixed factor coefficients of the demanding firms. When there are strongly-positive establishment-level scale economies (the standard NEG case) and relatively low levels of inter-regional trade costs, firms 'go to' labour supplies, as in the cases of greenfield location or even offshoring at a global scale. When intermediate trade costs to suppliers are high, or the need for high skills or for high levels of turnover of labour dominate the firm's choice, firms 'go to' other firms and to pools of labour. Labour-capital substitutions on the margin are minor, and firms are not 'indifferent'. In the short-run, then, the economy does not work like a simultaneous substitution system across locations.

The NEG takes a different approach. Its big contribution is to place the phenomenon of localization or agglomeration at the centre of the geographical development process. It does this by introducing scale economies, and allowing factors of production (especially labour) to be mobile, and then specifies the degree of unevenness of development that will emerge as determined by the interaction of these scale economies with trade costs between firms and between firms and their markets. The NEG allows for multiple agglomerations within the same industry, even with very low trade costs, by allowing for product variety—a key realistic dimension of the modern economy. The NEG explains cross-hauling (intra-industry) trade among similarly-developed areas. Gains to trade in NEG come not only from Ricardian specialization, but from the geographically-differentiated productivity gains from agglomeration. The NEG has a different view of factor proportions and prices from the NNUE. Nominal wages can vary due to establishment-level scale, or to the existence of clusters, both of which raise productivity. So can the cost of living, where higher living costs are offset by access to consumer variety, which becomes cheaper in bigger home markets for the NEG, and represents a pure 'taste' amenity in the NNUE. The economic equilibria that emerge from this kind of thinking usually involve high levels of unevenness of output and, in some versions, of per capita income.

In the long-run, most NEG theorists continue to assume that interregional income convergence will emerge as the result of declining trade costs, allowing greater sorting of production between regions that compete, and due to increasing labour mobility. But in the NEG this is a much more temporally- and geographically-uneven process than in NNUE, because economies of scale—within and between firms—are a fundamental characteristic of the economy. Moreover, innovation and learning are admitted by some NEG theorists as potential sources of long-run productivity and factor price differences between places, even though they are very difficult to model. Because of this latter

feature, they are never integrated fully into the 'big picture' NEG models (Duranton, 2007; Fujita and Thisse, 2009). This leaves a strong difference between the NEG's sector-specific perspectives on spatial development, which push towards a vision of change and differentiation, and its general models (Duranton and Puga, 2001; Duranton, 2007), which often propose mechanisms of long-term smoothing out of development, such as bell curves in the distribution of population and economic activity (Thisse, 2010; Rossi-Hansberg and Wright, 2007). The NEG also eschews attempting explanation of the 'where' of agglomeration or the spatial income hierarchy (which places are on top or lower down), preferring to relegate these initial sources of difference to 'accidents of history,' instead of asking whether there are any regularities or structural determinants of such origins.

The NNUE and NEG have additional differences in thinking about change and development. In the NEG, scale economies in production lead to higher productivity and wages in certain places; which, in turn lead to concentration of a variety of goods at lower prices (consumer amenities), which in turn lead to a win-win choice for individuals of locating in places with high wages, low prices and high consumer variety. The 'amenities' are thus endogenous to the process of spatial concentration, driven by scale and transport costs. This canonical core-periphery model is appealing because it describes the reality of high-income city-regions. But it does not explain why people and firms would go to places with both lower wages and less variety of consumer goods or unpriced amenities. This is where the NNUE counters the core-periphery thinking of NEG by arguing that in some cases people will reduce their nominal wages in the periphery in order to get its better unpriced amenities and cheaper housing, such that sorting over many different types of locations with different characteristics is achieved. The NEG sees wages in the core regions as possibly incorporating a real wage boost from lower consumer prices (scale-supported variety and productivity), not a smooth, arbitrage-and-substitution-driven version of spatial equilibrium.

The big picture is completed in NNUE approaches by adding the notion that other sectors of the economy respond to these dynamics, but also—and this is crucial—shape them. The most important such sector is housing. Housing stocks respond to demand, which in turn depends on whatever is said to determine population growth (workers to jobs? Or jobs to workers?). This, however, can be shaped by political forces, such as land use regulation—involving sometimes an insider/outsider conflict of interest between those already in a place and those who would like to migrate there. However, in NNUE, it is then frequently assumed that the supply of housing will limit potential in-migration, in turn partially determining whether firms can come into a region and create jobs, and in the case where housing supply would limit population growth, the resulting labour force would be limited (hence placing limits on expansion of firms), driving up housing prices, nominal wages up and real wages down. In other words, in order to fully integrate this sector, it must be assumed that it has long-term power to enter into a variety of factor substitution and price adjustments, not just on the 'unimportant' margin, but in magnitudes that fundamentally affect the prices (wages) and quantities (population) of regional economies and bring about an equilibrium consisting of strong real wage/income convergence in the short-run and nominal wage/income convergence in the long-run.

All in all, the fundamental architecture of these theories is based on the notion that there are many forces that enter into parameters for locating jobs, people and housing, and that their elasticities of substitution are the main interest of researchers attempting

to find out what structures a given equilibrium arrangement of the spatial economy, and how such equilibria change over time. To the core question of what drives regional population change, both NEG and new neoclassical urban economics responds that 'people go to jobs, and jobs go to people, more or less simultaneously' (cf. Muth, 1971).

Both NEG and new neoclassical urban economics share the vision of this economy as a set of tradeoffs where there are no identifiable starting points that are causally more important or sequentially earlier than others in determining the subsequent pathways of change. They differ on the details, in the specific structural determinants, elasticities, time frames of adjustments, and some of the outcomes. This agnosticism about causes and sequences of change has the advantage of allowing many possible forces to be considered in an open manner. But it has the parallel disadvantage of eliminating, by assumption, the possibility of hierarchies of causes or directions of causality in generating major changes in the space economy. For example, if theory is neutral as to whether people go to jobs or jobs go to people, how can it explain what sets off a sequence of interactions that leads to development of some previously-underdeveloped places or de-development of others?

4. Firms come first, and people go to jobs

Are there some forces that are 'leaders' and others 'followers' in generating changes in spatial patterns? A realistic view of the history of spatial development suggests that what moves the spatial economy forward is the creation of new production possibility frontiers. At the leading edge of technology, this involves the innovation of processes and products; at the 'cost-price' frontier of the economy, it consists of the forces that open up new factor supplies or conditions of production (such as through liberalizing trade or improving transport infrastructure).

In the case of innovations at the technological frontier of the economy— such as major new types of products (including the birth of new sectors), applications of innovations to create new processes in other sectors, or development of new organizational capacities for delivering packages of services not previously possible— there are two consequences that have been observed time and again. The first is that such activities have a propensity to agglomeration, because they involve forms of transactional, output and informational uncertainty, and require scarce knowledge. These are costly attributes that can be managed through combinations of organizational fragmentation and the high levels of external transacting it engenders, as well as the pooling of resources (Duranton and Puga, 2004). This is one of the biggest overlaps between the NEG and EG; in fact, the latter had a flourishing literature on agglomeration in the 1980s, before the seminal NEG papers were published.

However, geographers cannot generally agree with how both the NEG and the NNUE interpret the broader *economic* effects of innovation and agglomeration. For economic geographers, the determination of factor and output prices in innovation-driven agglomerations is not a standard competitive equilibrium, nor in many cases does it even represent market-clearing (with no excess) in any realistic sense of that term. It matters little whether we call the spatial pattern of firms and people an equilibrium or not (a hot-button issue which has little practical importance). What is important is that in innovation-driven sectors, firms and places, factor and product markets have auctioning processes that raise prices above the standard zero-excess rule.

Thus, there tend to be high rates of capital accumulation by firms and high real wages (wealth accumulation) for people, and high real regional per capita incomes that diverge substantially from other regions for long periods of time. The endowment effects of these processes then have strong impacts on future states of the economy in general and certain regions in particular.

When such developments involve radical innovations—such as computing or biotechnology—they seem to create 'windows of locational opportunity', in the sense that they are not held down by pre-existing external economies or factor supplies (Scott and Storper, 1987). Their 'newness' obviates the advantages of older environments and may even make them into obstacles. In other cases, where innovation builds on existing industries, but is nonetheless substantial, the new activity may regenerate an existing production agglomeration, but substantially transform its growth and local development dynamics. In somewhat later stages, usually the early expansion of commercialization, only a very restricted number of regions can meet the basic requirements of firms, whatever the price of land, capital and labour in these regions may be.

If this Schumpeterian thinking is incorporated into theory, then it has enormous consequences for how regions are seen to 'capture' or attract these economy-driving innovative activities. In the perfectly competitive parts of the economy, some activities might have flexible factor coefficients and firms might be like 'putty'. This represents a small basket of activities, mostly in low-technology sectors. In other sectors, where the factor coefficients are largely fixed by technology, but products are standardized and trade costs are low, then the standard sorting of places/activities occurs through factor costs of places. But in the leading edges of the economy, both such mechanisms of spatial sorting are largely irrelevant. Put bluntly, firms have semi-fixed factor coefficients; but they have few locational choices in these sectors because inputs (people, skills and organizational competences) are rare and specialized and outputs are not standardized. Moreover, change is so rapid as to make competitive market-clearing inexistent over relevant locational decision-making and investment periods. This means that people go to jobs, both in a temporal sequence and as a representation of the hierarchy of causes that should go into the architecture of a theory.

A second case of major change in the spatial organization of the economy is more subtle than the first. What occurs when there are major 'shocks' to transport costs, big changes in openness to trade (e.g. changing of tariff or trade barriers), or other such 'events'? The sequence of adjustments is that firms respond to such conditions by re-organizing and re-locating, insofar as the shocks are big enough to overcome existing sunk costs or agglomeration economies. This in turn changes the location of labour pools that they can now access. Firms may, in this sense, 'go to people', and this is observable in contemporary delocalization to national peripheries and off-shoring to less-developed countries. From these initial starting points, secondary processes unfold, including the migration of people to jobs, and this in turn unleashes additional local development processes such as growth of home markets and local economic feedbacks. The causal sequence—what starts things off—and the hierarchy of causes, is clear.

The locally negative version of this process may involve a regional economy that is undiversified and which suffers a major 'relocation shock'. Once this occurs, there is the possibility of cumulative downward spirals—through unemployment and out-migration, but these are different from those predicted by standard theory. When jobs go away, the decline of non-tradable sectors leads to degradation of the local built

environment (absent significant public investment), because quantity adjustments of housing and building are slow. Prices decline, leading to accumulation of negative externalities and reputation effects or 'images' of the place. At the same time, migration may not be fluid enough to adjust local labour supply to demand (social relationships, psychological costs and such). But certain labour prices may be rigid due to institutions, conventions and contracts. This pathway of adjustment can be sufficiently complex on both price and quantity sides that it preserves outcomes that are far from any standard notion inter-regional nominal or real income convergence. If the resulting image degradation occurs at a larger regional scale, it can bring durable effects on subsequent investment and the pathway of economic adjustment, long after labour price rigidities have disappeared and even when housing and land become very cheap. This is the negative type of circular and cumulative causation that mirrors the positive type that is found in 'hot' growth centres. The process of spatial economic development is riddled with these dynamics, not the smooth substitutions and mean reversions around which most models of NEG and NNUE are built. This is the spatial version of the critiques of the equilibrium framework that have been made in a long line of thinkers from Keynes through Schumpeter and Douglass North.

Can such forces of discontinuous and complex change be brought 'within' economic geography and regional development theories? No social science theory can completely endogenize all causes of change. The problem with most spatial economics today (NEG or NNUE) is not that they totally ignore these forces, as that they bracket these causes and the pathways of their effects too strongly. They do not attempt to integrate them with systematically relevant time horizons. They are mostly backward looking. Because of their requirements of theoretical consistency, but in defiance of observable reality, they exaggerate the importance of a hypothetical arrival point of no-excess equilibrium. Hence, they obscure the out-of-equilibrium prices and quantities that are commonplace, and how they affect future endowments. This in turn prevents them from realistically assessing how today's development affects future evolution and spatial distribution of economic capacities and hence the long-term process of spatial distribution of development and incomes.

5. Methods

Another issue that separates EG from NEG and NNUE is that the two latter use more econometrics than the former. Econometrics is good at detecting repeated, large number influences on outcomes, but notoriously bad at detecting the causes of major shifts in development patterns, the 'shocks' mentioned above. Change can sometimes be detected using such methods, but it requires a lot of luck and 'art' to identify the points at which the structural determinants shifted. It is easy to miss them unless there is a clear orientation to change and dynamics, and to some kind of possible causal hierarchy built into the research effort in advance and as a matter of substantive theoretical commitment. Moreover, once huge numbers of factors are thrown into a causal mix and it is assumed that their respective rates of change reveal their mutual 'causal' cross-elasticities, then even excellent econometrics can obscure the detection of major events, shocks, and causal sequences that may lie underneath more continuous subsequent adjustments. Assuming that most change is the incremental outcome of knife-edge and widely available arbitrages of preferences

and factors is an overbearing and largely unproven assumption about the world and how it works.

6. Economic geography

The response of much work from economic geographers is to go in the opposite direction from the spatial economists, emphasizing that spatial economic development is the result of unique, context-driven, place-specific combinations of forces that, as a consequence, can neither be modelled nor even subject to large-scale causal inquiry. At best, they can sometimes be captured in descriptive typologies. This extreme has its own weaknesses, to which we now turn, but we shall also observe that EG has developed some key insights into spatial development that could fruitfully be absorbed into a revitalized NEG framework.

EG refers here to a diverse body of work on spatial economics, from geographers, empirically-minded regional economists and certain branches of innovation and organization studies. There is less agreement on theory and methods than in the NEG and NNUE, and hence EG is more difficult to characterize succinctly.

Since the 1980s, EG scholars have made substantial progress in understanding how agglomerations function. They have investigated their local labour markets, the geography of technological spillovers, and regional contexts of formal rules and informal norms and conventions. All these appear to shape the ways that agglomerations function and their ongoing development, decline or resilience in the face of challenge.

Within this field, economic geographers have made considerable efforts—paralleling those of the economists—to see local and regional processes in the wider context of open economy forces. This has spurred considerable work on local versus long-distance relationships (mirroring the concern with trade costs in the other literature), and how the division of labour and fragmentation of production systems develops. The geographers' approach to trade costs stresses the divide between innovative or uncertainty-dominated activities and routinized ones, noting their tendency to have different organizational structures (scale and scope of firms, degree of intermediate transactions) and hence trade costs (Scott, 1988; 2004).

Both NEG and EG are interested in dynamic aspects of these forces. Lower transport costs could lead to spatial concentration, not just in the simple case of serving big markets from large-scale firms and plants, but through more subtle dynamic reorganizations of sectors, such that there is an endogenous component to resulting unit trade costs. In this view, lower trade costs seem to heighten competition in far-flung markets (eliminating old spatial monopoly effects) and lead firms to speed up the introduction of new products, which in turn introduces new forms of complexity and uncertainty into their strategies, hence raising unit trade costs upstream, as they fragment production systems to be more flexible and reactive, and this in turn leads to spatial concentration (Duranton and Storper, 2008).

Theories of spatial economic development face an additional challenge: how to characterize the pattern of growth and decline of places at a high-enough level of resolution. This task is not the same as understanding the average characteristics of growing and declining places, or the average trajectory of agglomeration or dispersion. Many economists choose to smooth the pattern around means and trends in means and

to consider that cases that are far off the means are there because of 'random effects' or 'noise'. Geographers tend to call these same effects 'complex, specific development processes,' and believe that they should be taken seriously and even theorized. These differences reflect conflicting underlying convictions about spatial-economic development along the lines I have been tracing. Thus, for most economists, dispersions around the mean are just a matter of time, because 'history matters' only for a while following random shocks (until everything reverts to the mean, in spatial equilibrium theories). In the NEG, the new local equilibria around agglomerations can temporarily involve some fairly important spatial price differentials between dense and big regions and less dense or small regions. Unlike the out-of-equilibrium view explained above, however, such differences must strictly reflect spatial productivity differentials, not alternative spatially-differentiated auctioning of factors (with rents and the accumulation of endowments and path dependencies from them). For most other economists, the noise is simply dealt with by introducing controls. Geographers are less persuaded about noise, and more tempted to believe that local, agglomeration-based growth processes have unobserved specificities that are masked by economists' theories and methods; as noted above, some geographers also believe that the notion of general spatial equilibrium actually obscures the durable economic differences between places and prevents accurate mapping of prices, quantities and endowments.

In contrast to spatial economists, geographers have spent considerable effort on explaining the 'where' of agglomeration, and have been partially successful at this. First, they have argued—even before NEG existed—that important new innovations create 'windows of locational opportunity' in the sense that the input structures and labour skills of previously-existing industries may be of little relevance to new sectors (Scott and Storper, 1987). If such existing activities are also institutionally-organized, they may even hinder new development by monopolizing resources, attention and skills (Chinitz, 1961). Such regions may not be able to capture new industries in spite of the greater size of their home markets as compared to alternative less-developed locations. It is important to note that this is a basic—'Schumpeterian'—addition to spatial economics, but it has not been taken up by NEG theorists.

Second, geographers have made efforts to theorize the generation or capture of innovative activities by regions, or at least how regions 'pair up' with such activities. Evolutionary economic geographers argue that in well-developed regions, the previously-existing routines and competences of firms, labour and institutions shape their ability to capture new activities, what kinds of innovations they generate, and what kinds of blockages to resilience they may exhibit (Boschma and Frenken, 2006; Boschma and Martin, 2007). They draw directly on evolutionary economics' notion that technology is path-dependent. Thus, regions will tend to develop or capture activities that are technologically-related ('related variety'). This is a fundamental insight, but it is probably more adapted to 'European' regions' tendency to move up quality ladders in related areas. 'American style' innovation, however, has a higher proportion of more radical innovations, which are not well-explained by the related-variety concept. In the spectacular case of the US semiconductor/IT industry, pre-existing agglomerations of telephone and broadcast equipment on the US East Coast did not prevent the sector from clustering in Silicon Valley, 4600 km distant. Silicon Valley itself had no strongly identifiable technological precursors of the IT industry. In the developing world, the rapid ascension up the technology ladder by South Korea cannot be explained as an evolutionary process of related variety; the

Koreans, like Silicon Valley, pulled off radical spatial and economic ruptures with the past. This is also what happened when the US aerospace industry located in Los Angeles in the 1920s, or the film industry did in the 1910s. There are additional examples in 19th century Europe.

Geographers have also actively explored institutionalist approaches to the origins of agglomerations, the spatial sorting of economic activity, and the matching of regions to activities (Farole et al., 2010). If such things as: technology, trade costs, agglomeration and migration give rise to broad tendencies in the sorting of activities to places, they are nonetheless potentiated or blocked by regional social forces, which are also unevenly distributed among places. The 'dark matter' of regional development is the specificity of human interactions from one place to another, its context. 'Context' refers to the microeconomic ways that the behaviours of agents are structured through institutionalized, spatially-differentiated forces (Storper, 2009).

'Institutions' can be considered at two different scales. At the sectoral or activity scale, any agglomeration will have such features as entrepreneurial and labour market networks; business associations; specially-adapted relations with local government; technology and R&D networks; and possibly even local or regional circuits of finance (such as venture capital today). There is abundant recent literature on such phenomena, at the borderline of geography, sociology and management studies (Saxenian, 1994; Powell and Sandholtz, 2011; Powell et al., 2011). Another scale is that of the region—its formal public-sector institutions, the institutions of economic life that cross-over sectors, and the institutions of civic life that involve actors from many different sectors and interact with the economic and public-sector institutions. Some promising cases studies have been carried out along these lines (Safford, 2009). Others have tried to theorize such institutions, their micro-economic properties, pathways of change and relationship to regional development and resilience (Storper, 2005, 2008; Rodriguez-Pose and Storper, 2005). All in all, then, some pieces of the 'where' puzzle are in place, but geographers need to relate them more strongly to the NEG and NNUE frameworks of open-economy spatial development; at the same time, economists would do well to try and incorporate the dispersion of growth processes in their frameworks and the causes of such dispersion through selection and matching of places and activities, rather than through largely vacuous statistical control strategies.

7. Elements of a framework

In light of the above, a framework for the field should deal with both allocation and adjustment of spatial patterns 'within' a set of parameters, as well as major causes of growth and change. Stated differently, it should deal with both large-scale, continuous or 'regular' growth processes, and be able to incorporate 'shocks' and 'events.' It should, moreover, incorporate endogenous regional dynamics that are potentially non-linear and that themselves generate further sudden change and local selection processes. In order to do this, it may be time to revisit some of the basic tasks of the field. What are the most fundamental behaviours and processes that could be considered building blocks of a renewed framework?

Research on economic growth has long struggled with the relationship between structures and events. Formal economic models, as well as inductively empirical

approaches, are structural, in the sense that they seek the parameters that are associated with growth across a wide panel of regions. But decisive singular events are often cited when the medium-run fates of particular economies are cited: Detroit declines 'because' of the restructuring and decline of the American car industry; Los Angeles has declining per capita income 'because' of a massive wave of low-skilled immigrants in the 1980s and the downsizing of the Cold War military-industrial complex in the 1990s; Washington, DC becomes richer 'because' of rapid growth in the high-wage lobbying industry. Often, the latter are treated as exogenous shocks in formal models, but if such events are frequent and important shapers of growth, then much that is essential to explanation is being arbitrary excluded from the construction of the framework.

This ambiguity is particularly potent with respect to agglomeration-generated specialization. Theories of agglomeration show that localization economies have a structural character, in that once started, they can be strongly self-reinforcing; but as noted, explaining their origins may require recourse to one-off events as well as structural factors. Two future strategies to analyse the causes of growth of regions can now be suggested. The first would consist of structured, in-depth comparisons of the growth trajectories of different areas, using a set of standardized categories such as sorting, selecting and shaping. The result would be rich in detail and 'on the ground' accuracy, but it would have few degrees of freedom.

A second approach would require large-scale data. As noted, the problem with most existing econometric analysis of urban and regional growth is that it may identify certain structural determinants of growth, but does so in a time-invariant manner, and has difficulty separating fixed effects from the forces that shape pathways and select cities into different growth experiences. An ideal way out of this dilemma would be to be able to estimate the sources for growth over a wide panel of cities for different time periods. High quality, sufficiently disaggregated data on the nature of specialization, human capital and institutions would be required, as well as those for a wide set of controls. Once the structural determinants for different periods are estimated and compared, then a further stage of research would consist in estimating them for individual city-regions.

Finally, research on the geography of development would benefit from assembling rigorous data on events/shocks, not simply on panels of fixed effects, structural determinants or controls. For example, having data on the establishment and lock-in of localizations, or on other time-dependent shocks such as technologies that shock trade costs of existing sectors, or of changes in institutions that alter comparative advantages, and then estimating their effects jointly with structural determinants over different time periods would get us much closer to measuring sequential (possibly path-determining) forces and structural determinants in a single model.

This would represent a significant departure from most of the research that is reported today, in that it would join approaches that are radically separated in academia: large-scale quantitative analysis of structural determinants as favored by economists, and large-scale comparative development, as in the methodological tradition of Barrington Moore (1966) or Charles Tilly (1984), with the substantive contemporary concerns of comparative growth economics, and the 'on the ground' sensibilities of geographers. Though the challenge is daunting, only an approach that melds structure, events and processes, and hence can tackle directions of causality, is likely to advance us significantly in understanding the complex problem of differential

regional specialization and development and the large-scale transformation of urban and regional systems.

References

Black, D., Henderson, V. (1999) A theory of urban growth. *Journal of Political Economy*, 107: 252–284.

Boschma, R., Martin, R. (2007) Constructing an evolutionary economic geography. *Journal of Economic Geography*, 7: 537–548.

Boschma, R. A., Frenken, K. (2006) Why is economic geography not an evolutionary science? Towards an evolutionary economic geography. *Journal of Economic Geography*, 6: 273–302.

Chinitz, B. (1961) Contrasts in agglomeration: New York and Pittsburgh. *American Economic Review, Papers and Proceedings*, 51: 279–289.

Duranton, G., Puga, D. (2004) Micro-foundations of urban agglomeration economies. In J. Henderson, J.F. Thisse (eds) *Handbook of Regional and Urban Economics*, vol. 4, pp. 2063–2117. Amsterdam: Elsevier.

Duranton, G., Rodriguez-Pose, A. (2005) When economists and geographers collide, or the tale of the lions and the butterflies. *Journal of Economic Geography*, 5: 545–566.

Duranton, G., Storper, M. (2008) Rising trade costs? Agglomeration and trade with endogenous transaction costs. *Canadian Journal of Economics*, 41: 292–319.

Farole, T., Rodriguez-Pose, A., Storper, M. (2010) Human geography and the institutions that underlie economic growth. *Progress in Human Geography*, (forthcoming, online version published 12 July 2010).

Fujita, M., Thisse, J.-F. (2009) *Paul Krugman's New Economic Geography: Past, Present and Future*. Mimeo: CORE-UC Louvain Belgium.

Fujita, M., Krugman, P., Mori, T. (1999) On the evolution of hierarchical urban systems. *European Economic Review*, 43: 209–251.

Gabaix, X., Ioannides, Y. (2004) The evolution of city-size distributions. In J.F. Thisse, J. V. Henderson (eds) *Handbook of Urban and Regional Economics*, . vol. 4. Chapter 53, pp. 2341–2378, Amsterdam: Elsevier.

Glaeser, E. L. (2008) *Cities, Agglomeration, and Spatial Equilibrium*. Oxford: Oxford University Press.

Grossman, G., Helpman, E. (2001) *Special Interest Politics*. Cambridge, MA: MIT Press.

Hall, P. (1998) *Cities in Civilization*. London: Pantheon.

Henderson, J. V. (1974) The sizes and types of cities. *American Economic Review*, 64: 640–656.

Hirschman, A. O. (1958) *The Strategy of Economic Development*. New Haven: Yale University Press.

Krugman, A. (1991) Increasing returns and economic geography. *Journal of Political Economy*, 99: 483–499.

Krugman, P., Venables, A. (1995) Globalisation and the inequality of nations. *Quarterly Journal of Economics*, 110: 857–880.

Krugman, P. (2010) The new economic geography, now middle-aged. Paper presented to the Association of American Geographers Annual Meeting, Washington, DC, 16 April.

Martin, Ph, Ottaviano, G. I. P. (1999) Growing locations: industry location in a model of endogenous growth. *European Economic Review*, 43: 281–302.

Martin, R. (1999) Critical survey. The new 'geographical turn' in economics. Some critical reflections. *Cambridge Journal of Economics*, 23: 65–91.

Muth, R. F. (1971) Migration: chicken or egg? *Southern Economic Journal*, 37: 295–306.

Myrdal, G. (1957) *Economic Theory and the Underdeveloped Regions*. London: Duckworth.

Powell, W., Packen, K., Whittington, K. (2011) 'Organizational and institutional genesis: the emergence of high-tech clusters in the life sciences. In J. Padgett, W. Powell (eds) *The Emergence of Organization and Markets*. Chapter 14. Stanford: Stanford University, forthcoming.

Powell, W. W., Sandholtz, K. (2011) Chance, Nécessité, Naiveté: ingredients to create a new organizational form. In J. Padget, W. Powell (eds) *The Emergence of Organizations and Markets*. Chapter 12. Stanford: Stanford University Press, forthcoming.

Pred, A. R. (1977) *City Systems in Advanced Economies*. London: Hutchinson.
Rodriguez-Pose, A., Storper, M. (2005) Better rules or stronger communities? On the social foundations of the institutional change and its economic effects. *Economic Geography*, 82: 1–25.
Rossi-Hansberg, E., Wright, M. (2007) Urban structure and growth. *Review of Economic Studies*, 74: 597–624.
Scott, A. J. (1988) *Metropolis: From the Division of Labor to Urban Form*. Berkeley and Los Angeles: University of California Press.
Scott, A. J. (2004) A perspective of economic geography. *Journal of Economic Geography*, 5: 479–499.
Scott, A. J., Storper, M. (1987) High technology industry and regional development: a theoretical critique and reconstruction. *International Social Science Journal*, 112: 215–232.
Storper, M. (2005) Society, community and economic development. *Studies in Comparative International Development*, 39: 30–57.
Storper, M. (2008) Community and economics. In A. Amin, J. Roberts (eds) *Organising for Creativity: Community, Economy and Space*, pp. 37–68. Oxford: Oxford University Press.
Storper, M. (2009) Regional Context and Global trade. *Economic Geography*, 85: 1–21.
Storper, M. (2010) Agglomeration, trade and spatial development: bringing dynamics back in. *Journal of Regional Science*, 50: 313–342.