1. Theories of agglomeration and regional economic growth: a historical review

*Philip McCann and Frank van Oort*

1. INTRODUCTION

In the 1990s and 2000s many observers argued that the then modern treatment of agglomeration economies and regional growth really represents a re-discovery by economists of well-rehearsed concepts and ideas with a long pedigree in economic geography. Advocates of this position doubted the validity or originality of much of this recent research. After many plenary and panel debates on the identification and integration potentials of geographical and economic perspectives of local economic development in recent years at conferences of economic geography, geography of innovation, regional studies, urban economics and regional science, a mixed picture emerged in 2018. There is some ongoing mutual criticism, with some economic geographers tending to ignore or forget more economic-based insights and findings, and vice versa, but there are also new lines of research aiming to build on the different insights emerging from the different fields. Yet economic geography as a discipline seems to be fervently looking for rights to exist, with observations that economic geography departments are de-prioritized, emasculated and increasingly out-placed at business schools (Martin 2018). Also, urban and regional policy on many scales increasingly demands economic geographical insights and applications. Many also share worries on the applicability of the still dominant Anglo-Saxon conceptualization in human and economic geography to situations and dynamics in transition and developing countries (Robinson and Roy 2015). Meanwhile, the new economic geography (NEG) debates have also fallen silent. After some years with little significant new developments in the NEG research field, an area that was thriving a decade ago, the focus of many economists has shifted more to urban economic explanations of development and to identification issues. These insights have become important in recent years owing to the burgeoning number of regional shocks which are evident, along with the greater shift towards more localized policies. The urban economics discipline has developed more accurate methods and modelling techniques than other disciplines for assessing both these shocks and the effectiveness of policy responses, but this does not imply a monopoly of knowledge over these issues. The more we understand these issues, the more we appreciate their complexity, and simple parsimonious model frameworks are rarely sufficient for explaining them. Moreover, after almost three decades of analytical and empirical debates, it is still not clear whether or when regional diversity or specialization are growth-enhancing phenomena (Beaudry and Schifflauera 2009; Melo et al. 2009; De Groot et al. 2015), thereby questioning some of the claims of urban economics to have made significant strides forward (Ellison et al. 2010; Kerr and Kominers 2015; Faggio et al. 2017). Indeed, as we show in this chapter, explaining causalities and mechanisms of resilience and local economic growth needs integration of insights.
Theories of agglomeration and regional economic growth

from various disciplines more than ever (Storper 2011) as the underlying complexity of the issues becomes increasingly apparent.

Several criticisms of the monopolistic modelling logic underpinning NEG have come from economic geography schools of thought as well as from both orthodox and heterodox schools of economics. These critiques focus variously on the immeasurability of some of the notions of increasing returns inherent in the NEG frameworks, the static nature of some of its assumptions, the specific focus on the representative firm, the presence only of pecuniary economies and the absence of either human capital or technological spillovers as externalities. Alternatively, advocates of the new economic approaches argue that their analyses provide insights into spatial economic phenomena which were previously unattainable under the existing analytical frameworks and tool kits. In this chapter we reflect on these developments, putting equal weight on both evolutionary and institutional economic geographical conceptualizations as on geographical economic conceptualizations, set off against a historical review of agglomeration and regional economic growth theories. We show that the modern concepts and modern treatment of agglomeration and regional growth do indeed build upon previous classical insights, while also introducing new insights. This fusion of new and old ideas also raises new questions, poses new challenges and opens up new directions for future research. The currently experienced fragmented pluralism in economic geography may be a blessing instead of an obstacle towards integration with more economic approaches, with increasingly good examples of studies integrating causality and impact identification into agglomeration and economic growth empirics.

2. CLASSICAL AND NEOCLASSICAL INSIGHTS INTO REGIONAL GROWTH

The major developments in spatial economics and economic geography from the late nineteenth century up until the 1960s came from a variety of different traditions, and from a variety of different analysts. As to the location of economic activities, major insights were provided by, among others, Weber (1909), Lösch (1954), Isard (1956) and Christaller (1933 [1966]). Also, related work on the causes and regional growth consequences of the spatial clustering of economic activities was being undertaken by Lichtenberg (1960), and Vernon (1960) and Chinitz (1961) whose work focused specifically on issues relating to growth and agglomeration. In particular, the focus of their work was on the features of different types of agglomeration economies, and their analyses were undertaken within the traditional analytical framework of agglomeration phenomena, which had emerged as a fusion of the insights of Marshall (1890) and Hoover (1948). Marshall (1890) focused on the role of local knowledge spillovers, and the existence of non-traded local inputs and a local specialist labour pool, while Hoover (1948), Ohlin (1933) and Isard (1956) allocated the sources of agglomeration advantages into internal economies of scale and external economies of scale in the form of localization and urbanization economies. Internal increasing returns to scale may occur for a single firm owing to production cost efficiencies realized by serving large markets; so, there is nothing inherently spatial in this concept other than that the existence of a single large firm in the space implies a large local concentration of factor employment. However, external economies are qualitatively very different.
Whether owing to firm size or a large initial number of local firms, a high level of local factor employment may allow the development of external economies within the group of local firms in a sector. These are termed localization economies. The strength of these local externalities is assumed to vary, so that these are stronger in some sectors and weaker in others (Duranton and Puga 2000; Combes and Gobillon 2015). The associated economies of scale comprise factors that reduce the average cost of producing outputs in that locality. The theories on localization economies can be further enhanced by explicitly taking market form into consideration (Gordon and McCann 2000). Externalities characterized by knowledge spillovers between firms in a spatially concentrated industry are generally known as Marshall–Arrow–Romer (MAR) externalities. The MAR theory in a dynamic context (Glaeser et al. 1992; Henderson et al. 1995) predicts, as Schumpeter (1934) did, that local monopoly is better for growth than local competition, because local monopoly restricts the flow of ideas to others and so allows innovator-internalization. Porter (1990) agrees with the importance of localization economies, also arguing that knowledge spillovers in specialized, geographically concentrated industries stimulate growth. Alternatively, urbanization economies reflect external economies passed to enterprises as a result of savings from the large-scale operation of the agglomeration or city as a whole, and which are therefore independent from industry structure. Relatively more populous localities, or places more easily accessible to metropolitan areas, are also more likely to house universities, industry research laboratories, trade associations and other knowledge generating institutions. It is the dense presence of these institutions, which are not solely economic in character, but are social, political and cultural in nature, that support the production and absorption of know-how, stimulating innovative behaviour and differential rates of interregional growth (Harrison et al. 1997). The diverse industry mix in an urbanized locality therefore improves the opportunities to interact, copy and modify practices and innovative behaviour in the same or related industries. In her well-known theory on urban growth, Jane Jacobs (1969) defines diversity as a key source of agglomeration economies, and unlike the MAR theory, believes that the most important knowledge transfers come from outside the own industry.

Quigley (1998) describes four features of agglomeration economies. The first factor concerns scale economies or indivisibilities within a firm, that are the historical rationale for the existence of productivity growth in agglomerated industries in the first place (Isard 1956; Brakman et al. 2009). Without the existence of scale economies in production, economic activities would be dispersed to save transportation costs (Palivos and Wang 1996; Fujita and Thisse 2002). In consumption, the existence of public goods leads to urban amenities. Cities function as ideal institutions for the development of social contacts corresponding to various kinds of social and cultural externalities (Florida 2002).

The second factor, namely, shared inputs in production and consumption, encompasses the economies of localized industry described by Marshall. The use of shared inputs to produce more differentiated consumption goods in agglomerations associated with variety, fashion, culture and style is well known (Katz and Shapiro 1985).

A third possible reason why agglomeration economies may provide greater economic efficiency growth arises from potential reductions in transaction costs (Martin and Ottaviano 1999). Western economies in general have developed primarily into services-based economies. Business and consumer services now make up most of urban employment, and most of these urban activities are characterized as a knowledge-based
information society. A logical outcome of the interaction between urban economies and knowledge-based service industries is the growing importance of transactions-based explanations of local economic productivity growth (Gottmann 1983; Castells 1989). The Californian school of economic geography emphasizes transactional costs in explaining agglomeration economies (Scott 1988), and the survival of local firms and the lower search costs of workers (Kim 1987; Helsey and Strange 1990; Acemoglu 1996) demonstrates that in a matching-context returns to human capital accumulation can be shown to exist, even when all output in a city is produced with constant returns to scale and with no technological externalities. Again analogous to production, better matching may occur in consumer functions (shopping).

The fourth set of potential economies identified by Quigley (1998) relates to the application of the law of large numbers to the possibility of fluctuations in the economy. Fluctuations in purchases of inputs are usually as imperfectly correlated across firms as the sales of outputs are across buyers. Less inventory holding is required owing to the greater possibilities for pooling supplies.

Each of these aspects of agglomeration economies provides a possible rationale as to why regions characterized by agglomeration will generally exhibit higher growth than regions without such features. In addition to these features of agglomeration economies, there are also two additional features of cities, which contribute to the growth potential of a city-region.

First, the structure of a regional or urban economy can be considered in a manner analogous to corporate diversification in product portfolios. Regional variety can be considered a portfolio strategy to protect regional income from sudden asymmetric sector-specific shocks in demand (Mills 1972; Attaran 1986; Dissart 2003). This will especially protect labour markets, and thus prevent sticky unemployment occurring (Diodato and Weterings 2015). Even if interregional labour mobility is high, asymmetric shocks reduce economic growth as agglomeration economies and the tax base deteriorate (Krugman 1993). Following this reasoning, industrial variety at the regional level would reduce regional unemployment and promote regional economic growth, while specialization would increase the risk of unemployment and a growth slowdown. For firms, a central question is whether related or unrelated diversification is most rewarding for stability and growth (Baldwin and Brown 2004). Related industries can more often (though, again, not as a rule) be expected to have correlated demand shocks. Therefore, spreading risk over unrelated sectors is likely to be preferred from the viewpoint of a portfolio strategy. However, we should also take into account the possible benefits from related diversification. Analogous to economies of scale at the firm level, we expect knowledge spillovers within the region to occur primarily among related sectors, and only to a limited extent among unrelated sectors. In agglomeration theory, Jacobs’s externalities are expected to be higher in regions with a related variety or cognitive relatedness of sectors than in regions with an unrelated variety of sectors (Frenken et al 2007; Neffke et al. 2011).

Secondly, technological development and the diffusion of knowledge and innovation are regarded as central to the modern concept of regional growth. However, the concept of knowledge diffusion across space in the economic geographical literature dates back some 60 years, beginning with the growth pole theory of Perroux (1950) which was subsequently embedded in geographical space by Boudeville (1966). Its main assumption is that economic growth, manifested in the form of innovations, is spread throughout a
growth centre’s hinterland to lower-order cities and localities nearby. Innovations and knowledge once generated in a particular central location are expected to spread among regions from one locality to its neighbours (Richardson 1978; Parr 1999). Hirschman (1958) distinguished two types of spillover effects associated with growth pole theory: backward linkages and forward linkages. The former effects are associated with activities that provide inputs to economic activities, drawing towards the location where the clients are. The latter concern activities that use outputs by new activities or expanding existing activities that draw them towards locations where these existing activities are already (over-)represented. This can turn into backwash effects that are usually unanticipated, occurring when the growth pole attracts so much attention and cumulative growth that it drains the surrounding areas. Migration of workers towards the pole and the concentration of investment capital in the initial centre of innovation initiate the emergence of high-level urban services in the growth pole. This can then lead to a further polarization of economic growth, restricting growth elsewhere (Richardson 1978). The existence of spread effects is based on the belief that the ongoing growth of the core location (the growth pole) will eventually lead to diseconomies of scale owing to congestion and the appreciation of factor costs. A parallel stream of work also emerged from Vernon (1960) and Chinitz (1961) in which the role of cities as incubators of new firms and new ideas was regarded as critical. More recently, this theoretical framework has been applied in agglomeration studies of Henderson (1997), Feldman and Audretsch (1999), Rosenthal and Strange (2001) and Carlino and Kerr (2015) on innovation intensity, and Henderson et al. (1995), Van Oort and Atzema (2004) and Eriksson and Hane-Weijman (2017) on employment growth and resilience. All these papers argue that there is an urban product cycle notion in that new products are more easily developed in large diverse metropolitan areas with a diversified industrial structure and skill base, and particularly those with many corporate headquarters (Pred 1977), whereas mature products eventually are decentralized to hinterland or peripheral areas.

After the period of rapid analytical developments up to the late 1960s associated with the quantitative revolution in economic geography and the microeconomic-based breakthroughs in regional science (Isard 1956), outside the specialist research field, widespread interest in spatial economic issues to a large extent waned in both economics and geography for two decades. Therefore, it was another 20 years until a major resurgence of interest in spatial and regional economic issues was witnessed. This resurgence of interest was associated with the work of Paul Krugman (1991) and Michael Porter (1990), and both of these commentators not only borrowed from the existing insights, but also added new insights to analyses.

3. THE 1990S REVOLUTION: NEW ECONOMIC GEOGRAPHY AND NEW GROWTH THEORY

Prior to the development of new trade theory, traditional international trade theory was largely unable to explain intra-industry, intra-national or intra-regional trade. Gravity models suggested that most trade tended to be localized. The development of new trade theory based on the Dixit–Stiglitz modelling framework (Dixit and Stiglitz 1977), subsequently led to renewed interest in both localized and intra-industry trade. These
Theories of agglomeration and regional economic growth

developments in international trade theory, in turn, led to a renewed modelling interest in spatial economics in the form of NEG, and regional economics as a whole subsequently experienced a resurgence via a combination of the developments in both NEG and also new growth theories.

New economic geography is based on the insights and analytical approaches that are common to new growth theory and new trade theory. As both new growth theory and new trade theory predate NEG, it is worth recapping on the basic features and insights of NEG’s two antecedent literatures. In both of these strands of literature the dominant analytical approach is the modelling of imperfect competition and increasing returns to scale within the monopolistic competition framework of Dixit and Stiglitz (1977), in which utility is a function of variety. New trade theories now allowed for the modelling of inter- as well as intra-industry trade flows within a general equilibrium framework in which the structure of demand and supply is endogenously determined.

Krugman (1991) first applied this modelling framework to the issue of geography under conditions of economies of scale and labour mobility, and reinterpreted Marshall’s principles of externalities as stemming from the benefits of the pooling of the local labour supply and the demand for specialized non-tradable inputs. In these models, spatial concentration and dispersion were seen to emerge as a natural consequence of market interactions involving economies of scale at the level of the individual firm, with many of the results generated by these models being reminiscent of the results of central place theory and the rank-size rule (Fujita et al. 1999). Indeed, the cumulative causation characteristics of these models is in many ways akin to the processes described by, among others, Pred (1977) and in this respect the Krugman–Fujita–Venables work builds on most of the standard location theory (Krugman 1993; Dymski 1996).

This spatial version of the Dixit–Stiglitz monopolistic competition theory has since become a very important element of all spatial economists’ models on the location of economic activities (Abdel-Rahman 1988; Fujita et al. 1999) and several key insights have emerged from this literature. First, if internal economies of scale are strong and transportation costs are low, this induces a circularity that tends to keep geographic concentration in existence once established (compare Pred 1977 and Myrdal 1957 on their notions on cumulative causation). The reason is that manufacturers in the larger economic agglomerations have an advantage, since the size of local demand allows them to profit more from internal economies of scale, and therefore they can afford higher nominal wages. A higher local demand for goods induces a greater range of variety of goods, which induce real income effects that attract new workers, consumers and firms. These developments are manifested in a greater range of local forward linkages (the supply of a greater variety of goods increases the worker’s real income) and local backward linkages (a greater number of consumers attracts more firms) as pecuniary externalities create scale economies at the individual firm level that are transformed in increasing returns at the level of a location as a whole (Gianmarco et al. 2001). In general, this effect will be stronger as local demand is greater and internal economies of scale are higher.

Meanwhile, this observation of spatial industrial concentration is also consistent with the observation that some producers survive in peripheral locations. One reason is that peripheral producers exhibit local advantages outside the large agglomeration owing to higher transportation costs, which mean that they face less competition for their local demand. A second reason is that negative externalities such as congestion and high land
Handbook of regional growth and development theories

rents in the larger agglomerations (Quigley 1998) may eventually lead to decreasing returns to scale in cities (Moomaw 1985; Glaeser et al. 1995). If the industrial sector itself constitutes a principal source of demand for industrial products, and if transportation costs increase with distance, then firms will cluster because they produce under increasing returns. The existence of sufficiently high transportation costs therefore ensures that multiple clusters will exist instead of one monocentric city. The pull of Krugman’s pecuniary externalities balances the push of transportation costs. The ultimate equilibrium depends on the initial point of departure and the extent of economies of scale, and the level and structure of transportation costs (McCann 2005). Equilibrium no longer automatically means that spatial units of observation converge with regard to regional growth (Kubo 1995).

A second and recent body of literature related to geography and space has been developed on the basis of the new- or endogenous-growth theories. These theories themselves are built on similar foundations to new trade theory and NEG (Barro and Sala-i-Martin 1995), although they are different in that they do not treat time in a comparative static manner, but take growth over time and its determinants as the principal subjects of the analysis. According to this view, when individuals or firms accumulate new capital, they inadvertently contribute to the productivity of capital held by others. Such spillovers may occur in the course of investment in physical capital or human capital (Lucas 1988). As Romer (1986, 1990, 1994) demonstrated, if the spillovers are strong enough, the private marginal product of physical or human capital can remain permanently above the discount rate, even if individual investments would face diminishing returns in the absence of external boosts to productivity. These model approaches also became widely known as endogenous growth theory, because technological change is also seen to be endogenously determined in these models (Romer 1994; Solow 1994).

When applied to regions and geography, these models all assume that the notion of increasing returns is spatially embodied in agglomeration economies. Endogenous regional growth models are similar to NEG models in that their effects can only operate within an environment of imperfectly competitive monopolistic competition. However, these regional growth models are also different to mainstream NEG models in that in the endogenous growth framework, local external economies may not only be associated with market-size or pecuniary external economies, but can also be related to information or technological externalities and spillovers (Englmann and Walz 1995; Caragliu and Nijkamp 2015). Martin and Ottaviano (1996), Baldwin and Forslid (1997) and Coe et al. (2009) show that by incorporating research and development (R&D) activity into models reminiscent of Krugman (1991) and Krugman and Venables (1996), local factor accumulation can play a similar role to that of either labour migration (Krugman 1991) or input–output linkages (Puga and Venables 1996; Venables 1996) in fostering agglomeration via local demand linkages. However, whereas agglomeration in new trade theory and NEG is the geographic outcome of modelling, in new growth theory it forms an endogenously determined explanation of growth. These types of arguments therefore provide some additional possible explanations for systematic variations in competitive advantage (Porter 1998) across regions and why it is that certain regions are able to maintain and even reinforce their advantages over other regions, once certain locations have taken a lead in a particular activity (Krugman 1991; Arthur 1994; Thissen et al. 2016).

Lucas (1993) emphasizes that the most natural context in which to understand the
mechanics of dynamic knowledge externalities and economic growth related to (knowledge embedded in) labour is in metropolitan areas where the compact nature of the geographic unit facilitates communication and human capital accumulation. He argues that the only compelling reason for the existence of cities would be the presence of increasing returns to agglomerations of resources that make these locations more productive. This view of human capital as social input that induces productivity gains in cities has been further explored by others (Henderson 1986; Bostic et al. 1997; Cheshire and Duranton 2004; Rosenthal and Strange 2004; Puga and de la Roca 2017) who all argue that the microeconomic foundations of the external effect of human capital is the sharing of knowledge and skills between workers that occurs through both formal and informal interactions. The distinction between tacit and implicit knowledge bases as against explicit knowledge bases is deemed to be crucial here in terms of the ways that knowledge externalities are embodied in growth (implicit) and innovation (explicit) externalities. Intuitively it seems clear that the higher the average level of human capital (knowledge) or the more spatially concentrated are the numbers of agents, the more ‘luck’ these agents will have with their meetings and the more rapid will be the diffusion and growth of knowledge (Rauch 1993). Storper and Venables (2005) used the concept of ‘buzz’ to denote that much communication between decision-makers is accidental and happens in various non-organized meetings. Other authors also emphasize the importance of accidental meetings (Charlot and Duranton 2004; Fu 2007), whereby complex information transmission via face-to-face contacts plays a crucial role, in addition to the provision of specialized services and labour supply. These features are argued to be dominant in cities (Duranton 1999; Feser 2002). This all therefore points to metropolitan areas being the major locations where the productivity-enhancing effects of human capital primarily operate (Gaspar and Glaeser 1998; Glaeser 1999), and this pure agglomeration argument (Gordon and McCann 2000) provides a natural explanation for higher wages as well as higher land rents in cities.

These observations, which emphasize the role played by the city as a knowledge and information environment, also largely accord with many of explanations employed by the economic geography, institutional and evolutionary approaches. The original behavioural arguments generally pointed to large urban agglomerations as being superior incubator locations (Chinitz 1961) compared with other places. This thinking has also heavily influenced contemporary economic geography thinking. The difference, however, is in terms of the emphases. The evolutionary–institutional approaches stress institutions and policy-makers (Amin and Thrift 2002) on the assumption that in each observed case, the outcome of these externalities on productivity remains heavily dependent on the historical economic context (Bostic et al. 1997), the industrial structure (Moomaw 1988; Glaeser et al. 1992) and the specific role played by face-to-face contact in local production processes (McCann 2007). Therefore, when behavioural and evolutionary explanations for interregional economic development are taken seriously, primary attention is paid to the behavioural and entrepreneurial causes of agglomeration. The concept of externalities in this schema is therefore also related to the nature of information transmission mechanisms between actors in firms and the cognitive and interactive characteristics that determine the construction of locational preferences. Identification of mechanisms of knowledge transfer and causalities are of prime importance in this, and in recent years the methodological element of urban economics has entered many areas of empirical research in economic geography. This is particularly so in the sub-fields of evolutionary
and institutional economic geography where causality is now increasingly becoming a central research agenda.

4. ECONOMIC GEOGRAPHY AND EVOLUTIONARY ECONOMICS

Several criticisms of the monopolistic modelling logic underpinning NEG have come from economic geography schools of thought (Martin and Sunley 1996; Martin 1999) as well as both orthodox (Neary 2001) and heterodox schools of economics (Peneder 2001). These critiques focus variously on the immeasurability of some of the notions of increasing returns inherent in these frameworks, the static nature of some of the assumptions, the specific focus on the representative firm, the presence only of pecuniary economies and the absence of either human capital or technological spillovers as externalities, and the problems associated with the iceberg transport costs assumption (McCann 2005; Fingleton and McCann 2007). Other evolutionary critiques (Martin and Sunley 2003) also question the originality and validity of the Porter (1990) concept of clusters. Many of these criticisms actually relate to specific models and specific papers, rather than to the whole field. However, the most fundamental critique of these fields in general relates to the question of institutions, and the relationship between knowledge and institutions. Within economics, institutions are regarded as being important in explaining economic growth (North 1990; Aghion and Howitt 1998; Helpman 2004). However, for economic geographers and heterodox economists working within the evolutionary and institutional economics arenas, the role played by institutions in economic development is seen to be paramount. In this institutional–evolutionary schema, regions and countries that have more efficient institutions, formal or informal, are therefore superior in both the generation and diffusion of knowledge, and consequently have better prospects for economic growth (Cortinovis et al. 2017). While NEG and new growth theories are mathematically complex, they are still regarded by these analysts as being philosophically too simplistic. This is because they aim to produce generalizable predictions based on a representative model, whereas the counter-argument implies that the appropriate investments, favourable institutional arrangements and entrepreneurial dynamics which allow regions to grow are features of regions which have emerged for historically contingent and spatially contingent reasons, rather than for generalizable reasons. For economic geographers, as well as institutional and evolutionary economists working in this tradition, cultural and cognitive proximity are therefore deemed to be just as important as is geographical proximity in the transmission of ideas and knowledge (Boschma 2005). Boschma and Lambooij (1999) further argue that the generation of local externalities are also crucially linked to the importance of selection in terms of ‘fitness’ of a local milieu, the sociological dimensions of which can be institutional, cultural, legal and historical. According to these perspectives, it is these specific historically contingent and geographically contingent features, rather than simply space as a dimension, which is crucial in determining the geography of entrepreneurship and growth (Audretsch et al. 2006).

The original behavioural geographical literature (Webber 1964; Pred 1966) focused on incomplete information, the limited cognitive capacities of entrepreneurs and the differences in information absorption abilities of firms at different stages in their life
Theories of agglomeration and regional economic growth

cycles (Alchian 1950). However, institutional structures are now regarded as being much more than simply the aggregation of individual choices, but rather the result of many interactive processes. Economic geography research has always emphasized the untraded interdependencies (Storper 1997) that function as externalities and spillovers, and this has led to calls for research to focus on institutional issues (Amin and Thrift 2002). Evolutionary economic geography theory focuses primarily on the creation of new spatial structures, rather than on explaining equilibrium states (Boschma and Frenken 2006). Within the same spatial and institutional context, firms and entrepreneurs may arrive at different location behaviour either by means of chance occurrences or by fundamental processes of neo-Schumpeterian, creative destruction. Alternatively, different spatial and institutional contexts will mean that firms and entrepreneurs may arrive at either different or similar locational outcomes, but for a variety of different reasons. The initial states which determine allocations may vary significantly, although the future trajectories of these initial outcomes are determined primarily by path-dependency phenomena, which themselves are underpinned by local externalities and spillovers. In turn, these path-dependent phenomena subsequently give rise to localized regional clustering.

Evolutionary economic theory, as originally developed by Nelson and Winter (1983), emerged from economics as a result of dissatisfaction with many of the equilibrating notions of neoclassical economics. In many ways these evolutionary theories are inspired by Darwinian processes of biological change (Boschma and Lambooy 1999) and embody within them a very particular set of behavioural and environmental heuristics. First, uncertainty provokes firms into routinized, risk-averse behaviour which determines to a large extent the available options and probable outcomes of searches. This implies that technical, technological and human capital issues generally exhibit path-dependency behaviour characterized by lock-in processes (David 1985). Under these conditions, Arthur (1994) shows that the notion of increasing returns provides an explanation for why technology is able to maintain and reinforce its competitive advantage once it has taken the lead in the market, irrespective of whether the lead was taken due to superiority, coincidence or luck. Secondly, physical capital investments are a source of locational inertia. History, in the form of sunk costs resulting from the operation of many firms at a site, creates a first-mover disadvantage that can prevent relocation (Arthur 1989; Rauch 1993). Thirdly, the selection environment functions as a filtering mechanism that ultimately decides which of the innovations will thrive or fail. This selection environment consists of a number of consumer and financial markets as well as a set of non-market institutions such as regulations, values, norms and customs (Storper 1997). Evolutionary theory therefore implies that there may be a multiplicity of future spatial outcomes, many of which cannot be hypothesized on the basis of current observations. This argument is reflected in some of the NEG-type frameworks employed by several authors (Cronon 1991; Rauch 1993; Bostic et al. 1997; Ottaviano and Puga 1998; Berliant and Konishi 2000), who also conclude that there is a strong tendency toward path-dependency based on historical contingency. However, because of the long-lasting geographical history of cumulative causation from the 1950s onward, economic geographers claim that this type of thinking and this approach to the treatment of history is theirs, in contrast to the ‘newly’ discovered nature of this subject in the mainstream economics literature.

The current weakness, however, with evolutionary and institutional approaches to regional growth, is still that it is empirically primarily an ex post analytical framework.
More complicated, however, are the questions of the role played by institutions and economic networks. The complexity here arises from the many areas of neoclassical economics that have employed a minimalist definition of institutions in property rights and firms, whereas institutional approaches variously allow for a whole array of social, legal, political, historical, geographical and cultural phenomena to be characterized as institutions. There is currently no agreed parsimonious definition of institutions on the part of institutional approaches, and interactions appear to be complex (Boschma and Capone 2015). Urban economics and NEG approaches have tended to ignore institutional issues and this may well be a major weakness in their formulation as, increasingly, mainstream economics itself acknowledges the fundamental importance of institutional set-ups and systems (Rodrik et al. 2004; Rodrik 2007). Yet, the challenges set by economic geography and evolutionary approaches to explicitly incorporate institutional issues in their frameworks are daunting, and the reason for this is that it is currently very difficult, despite recent advances in data availability, identification and analysis (for example, Cortinovis and Van Oort 2018), to determine which observable growth outcomes can be more widely generalized or predicted on the basis of \textit{ex ante} evolutionary or institutional observations. In this sense, while many aspects of the growth processes can be described in detail, the ability to extrapolate is still currently limited.

5. COMMON GROUND?

Although the differences between the formal modelling approaches of NEG, new growth theory and the evolutionary–institutional approaches to regional growth at first may appear to be irreconcilable, common ground between these different competing theories can be found on several key issues.

First, in each of these different literatures, as we have already seen, the role of agglomerations is regarded as being a crucial element of regional performance, and the common element here is the issue of local knowledge generation, accumulation and spillovers (McCann and Ortega-Argiles 2016).

Secondly, and related to the first point, is the issue of the level of connectivity; specifically, the number of connections between local regional nodes to other key international nodal points in the global economy is regarded by all of these theories as being important (Saviotti 1996; Scott 2006). Recent work on global cities (Sassen 2001, 2002; Taylor and Derudder 2015) suggests that particular cities that are well connected via international hub airports, in particular, are consistently at an advantage to other locations for acquiring relevant knowledge spillovers.

Thirdly, the geographical scale over which knowledge spillovers operate is regarded as a critical issue and, again, most of the apparently competing theories are largely in agreement. On this third point, one of the features which neither the NEG nor the new growth theory explicitly models is the geographical scale over which any knowledge spillover mechanisms operate (Cortinovis and Van Oort 2018). As Jaffe et al. (1993) conclude, we know very little about where such spillovers go, although we can acquire some information regarding this by studying the geographic location of patent citations. Jaffe et al. (1993) therefore test the extent to which knowledge spillovers are geographically localized. Their measured effects were particularly significant at the local Standard
Metropolitan Statistical Areas (MSA) level, indicating that localization fades over time, but only very slowly. Further research by Audretsch and Feldman (1996), Acs (2002), Van Soest et al. (2006) and Feldman (1994), among others, provides corroborating evidence that knowledge spillovers tend to be geographically bounded within the location where the new economic knowledge was created. Recent European scaled research though, such as Crescenzi and Rodríguez-Pose (2011), Greunz (2003) and Moreno et al. (2005), show that within Europe, knowledge externalities have a significant impact within a range of 200–300 kilometres, dying out once this distance threshold is crossed. In the evolutionary tradition introduced previously, Boschma (2005) offers a general critique of the role of spatial proximity as the major catalyst for knowledge spillovers, suggesting that together with spatial closeness other forms of networked proximity may facilitate knowledge spillovers. Connections with cognitively similar actors, even if located far away, can provide access to valuable information for firms and individuals (Frenken et al. 2007).

Building on endogenous growth and evolutionary arguments, Huggins and Thompson (2014, 2017) have developed the concept of ‘network capital’ and Cortinovis and Van Oort (2018) that of ‘network-relatedness’ that theorizes a tight conceptual link between local economic performance and the ability to access economically valuable knowledge through network linkages.

Fourthly, the issue of governance has now become central to many analyses of the relationship between agglomeration and economic development. The Organisation of Economic Co-operation and Development (OECD) research agenda (Ahrend et al. 2014b) has demonstrated that fragmented governance systems can limit any potential benefits associated with agglomeration. Recent worldwide evidence suggests that, in advanced economies at least, the scale of agglomeration benefits are typically slightly less than was previously understood a decade ago (Ahrend et al. 2014a) and that such benefits can be reduced or even undermined by fragmented urban governance systems. There are different approaches to metropolitan governance (Ahrend et al. 2014b) and the nature and composition of urban governance systems are key determinants of urban performance. Similarly, on the regional scale, the nature and quality of governance is also increasingly argued to be critical in helping to foster development (Barca et al. 2012) and the emerging empirical literature linking the quality of governance (Charron et al. 2014) to local development processes is proving to be fertile ground for current economic geography research.

Finally, the rise of populist politics in many western countries, in which support for nationalist movements is strongly associated with economically weaker regions, has recently led to widespread doubts regarding the space-blind policy logic so strongly advocated by many urban economists (World Bank 2009). During the decade following the New Millennium, the dominant thinking in urban economics focused on efficiency and almost entirely ignored distributional matters. In contrast, a Kaldor–Hicks type logic, whereby efficiency is underpinned by the principle that the winners (both people and places) could potentially compensate the losers, is increasingly seen as being wholly unrealistic in modern political economy terms. Whether the compensation actually takes place, rather than whether it could take place, is now seen to be critical, and many areas of urban economics ignored the warnings of economic geographers regarding such an overly narrow perspective. Potentially, the policy responses to regional shocks can now build more effectively on lines of enquiry and research which derive insights from various different approaches, rather than narrow silos of analysis.
6. CONCLUSIONS

The NEG and new growth approaches rightly argue that their analyses provide insights into spatial and economic phenomena which were previously unattainable under the existing analytical frameworks and tool kits. The conceptualizations of endogenous growth, monopolistic competition and increasing returns to scale triggered a new phase of development in economic modelling. By accepting that in reality spatial and firm-level heterogeneity are much greater than these current general equilibrium models allow for, the phenomena of path-dependency and heterogeneous sectoral development trajectories emphasized by evolutionary economic arguments do not necessarily contradict the analytical outcomes of NEG and new growth models. Indeed, they provide profoundly new insights as to the nature of transition processes and the relationships between technology, firms, diversity, specialization and geography. However, the plethora of social, legal, political, historical, geographical and cultural phenomena variously defined under the umbrella notion of ‘institutions’ makes more traditional economic geography and evolutionary approaches more difficult to operationalize than urban economic-type frameworks. This does not necessarily imply that they are less insightful, but that a tighter specification and definition of the analytical elements is required to make the models tractable. Therefore, while the analyticals associated with economic development are always very complicated, the problems posed by these definitional issues may be more problematic for the evolutionary–institutional approaches than they are for the NEG approaches. The reason for this is that the NEG and new growth theories assume that the dominant growth mechanisms are economic in nature and determined by pricing and allocation outcomes, whereas the evolutionary–institutional arguments assume that these other institutional and network phenomena are dominant. So, until evolutionary–institutional approaches develop ways of clearly defining the nature, characteristics and behavioural features and outcomes of institutions and network relatedness, any real-world observations will always suffer from the inherent methodological problem of observational equivalence (McCann 2007). Therefore, while there is already much common ground between these various approaches to regional growth, much interesting work remains to be done to fully reconcile these different analytical approaches.

REFERENCES

Theories of agglomeration and regional economic growth


20 Handbook of regional growth and development theories


Theories of agglomeration and regional economic growth

22 Handbook of regional growth and development theories


Theories of agglomeration and regional economic growth


