INDIVISIBILITIES, SYNERGY AND PROXIMITY:  
THE NEED FOR AN INTEGRATED APPROACH TO AGGLOMERATION ECONOMIES

Roberta Capello

Department of Management, Economics and Industrial Engineering, Politecnico di Milano

SOMMARIO

The aim of the paper is to present a review of the literature on agglomeration economies in a light such that it is possible to uncover the building blocks for use in a multi-dimensional approach to agglomeration economies. The paper claims that the different conceptual dimensions – industrial, socio-cultural/cognitive and geographic – build upon the three micro-foundations of agglomeration economies – indivisibility, synergy and proximity – and explain their nature, scope and intensity. The paper stresses two main issues. The first is that a socio-cultural/cognitive perspective still today ignored by mainstream approaches to agglomeration economies is an additional important conceptual dimension on which to understand the complex mechanisms whereby exchanges of knowledge, labour or intermediate goods take place in agglomerated areas. The second is that a multidimensional approach comprising the three different dimensions in the study of agglomeration economies makes it possible to overcome a deterministic approach to agglomeration economies and move instead towards a stochastic interpretation which opens the black box of this complex phenomenon.
1. Introduction

Since Marshall (1919), much work has been done on the external advantages of agglomerated activities and environments. The aim of this paper is to present a review of the existing literature on agglomeration economies in a light such that it is possible to uncover the building blocks for use in a multidimensional approach to agglomeration economies.

To this end, the paper presents a review organised around the different conceptual dimensions – industrial, socio-cultural/cognitive and geographic – applied in the literature to interpret agglomeration economies. These three dimensions build upon three main micro-foundations of the concept of agglomeration economies: indivisibility, synergy and physical proximity. Indivisibilities occur when the supply of services and goods requires the presence of a certain demand threshold in order for production to take place under efficiency principles. Synergy, in the form of network relationships among economic agents, allows outsourcing and flexibility in production and is therefore another important efficiency source. Both indivisibilities and synergies display their efficiency effects on production processes most markedly in concentrated environments, i.e. in the presence of physical proximity among agents. A concentration of activities in one particular place enables the achievement of demand thresholds, both for input and for output, and facilitates synergies among actors: agglomeration economies are, that is to say, advantages which arise from the interaction (often involuntary) among economic agents made possible by the lesser amount of spatial friction in concentrated locations.

Building on the existing literature, the paper seeks to show that, over time, the attention in the literature has shifted from one of the three dimensions to another, in which light it provides a particular interpretation of the causes of agglomeration economies. The paper shows how this historical uni-dimensional approach has missed opportunities to merge complementary perspectives in the interpreting the causes of the phenomenon. By doing so, the paper demonstrates that local network behaviours – which are greatly emphasised in a socio-cultural/cognitive perspective but largely ignored in mainstream approaches to agglomeration economies – place the emphasis on understanding the complex mechanisms whereby exchanges of knowledge, labour or intermediate goods are facilitated in agglomerated areas. An integrated and multidimensional approach comprising all three strategic perspectives is achieved when agglomeration forces are interpreted by means of different approaches that are complementary rather than substitutes. This approach should be able to disentangle the complex phenomenon of agglomeration economies and yield in-depth and greater understanding of the concept, of its causes, nature, scope and advantages.

The paper is structured as follows. It first highlights the structure with which the comparison of approaches is developed throughout the paper (Section 2). It then provides insights into each approach to the understanding of the nature, scope and intensity of agglomeration economies, stressing the interpretative power and the limits of each approach (Sections 3-6). Section 7 envisages new research directions in both the logic with which the analysis of the literature is conducted, and the structure of the comparison among blocks of literature. Section 8 contains some concluding remarks.

2. A Comparison of Approaches to Agglomeration Economies

This paper reviews the literature on agglomeration economies according to the logical dimensions – industrial, socio-cultural and geographic – that are inherently part of this concept. These three dimensions build upon three main micro-foundations of the concept of agglomeration economies: indivisibility, synergy and physical proximity.
Indivisibilities occur when the scale of agglomeration activities adds to productivity by causing shifts in a firm’s production or cost curve, i.e. more output for a given input, or lower input costs for a given output (Rosenthal and Strange, 2001; Cohen and Paul Morrison, 2009). Indivisibilities prevail when an industrial perspective is taken: some sectors are more dependent on large scale production processes than others, and some sectors derive more advantage from the presence of other sectors than others. In light of these sectoral peculiarities, a large body of literature measures the degree to which the presence of a mix of industries or of the same industry generates greater agglomeration advantages. Synergies relate to the socio-cultural dimension: trust, sense of belonging, cultural and religious homogeneity, which are typical features of agglomerated and specialised areas. They heighten the intensity of local cooperation and market interactions, thus giving rise to increasing returns on production factors via both transaction and production cost minimisation for a given output. Proximity is by definition linked to the geographic dimension of agglomeration economies: if interaction (information and transportation) costs were nil, in the absence of scale economies there would be no reason to concentrate activities, because doing so would not produce ‘economies’. In this sense, agglomeration economies are ‘proximity economies’.

These micro-foundations explain why the causes of agglomeration economies, their nature, scope and intensity, have been explained from a particular perspective: the achievement of productivity increases can derive from local scale effects, or from easier market interactions, or from limited distance friction. Over time, these different dimensions and causes have been conceptualised in specific approaches to agglomeration economies; historically, each approach has emphasised one single dimension, taking the others for granted. The result has been the enrichment of interpretation of the phenomenon from different perspectives.

This historical development of the approaches to agglomeration economies is presented in Fig. 1. Each approach is characterised by a particular dimension, highlighted in grey in the figure. They range from the industrial dimension in the early 1970s, to the socio-cultural dimension in the mid-1970s, to the cognitive dimension in the early 1980s, to the first approach integrating two dimensions, the industrial and the cognitive ones, in the mid-1980s, to the pure geographic dimension in the early 1990s, to the most recent second attempt to take an integrated approach comprising two dimensions, the geographic and industrial ones, in the late 1990s.

Fig. 1 - here

The results of all these studies are clearly displayed in Fig. 1; the overarching framework of analysis of the phenomenon is rich and diversified, adding to each new approach new insights in the interpretation of agglomeration economies (Fig. 1.e). The claim put forward here is that an additional effort can and should be made to exploit all the existing perspectives at the same time in a multidimensional framework of analysis which yields the most comprehensive interpretation of the phenomenon (Fig. 1.f).

To explain the main idea behind this paper, each approach will be analysed in detail in the next sections, in terms of framework of analysis, conceptualisation of space, assumptions behind each dimension characterising agglomeration economies, the perspective through which the causes of agglomeration economies are studied, and the new insights achieved. In this way, it will be possible to uncover the building blocks of a multidimensional approach to agglomeration economies.
Fig. 1. Conceptual Approaches to Agglomeration Economies: A Diachronic Perspective

- **a) The seventies: the Industry Dimension**
  - Static localization vs. urbanization economies
  - Geographic dimension
  - Industrial districts

- **b) Mid-seventies: the Socio-cultural Dimension**
  - Static localization vs. urbanization economies
  - Geographic dimension
  - Cultural/cognitive dimension

- **c) The eighties: the Cognitive Dimension**
  - Dynamic location vs. urbanization economies
  - Geographic dimension
  - Cultural/cognitive dimension

- **d) Early nineties: the Geographic Dimension**
  - Spatial spillover theory
  - Static localization vs. urbanization economies
  - Geographic dimension

- **e) late nineties: the Geographic/Industry Interaction**
  - Dynamic location vs. urbanization economies
  - Static localization vs. urbanization of economy theories
  - Geographic dimension

- **f) 2000 onward: an Integrated Approach**
  - Functional approaches to spatial concentration and/or Geographical approach to location/urbanization effects
  - Static localization vs. urbanization of economy theories
  - Geographic dimension
Table 1. Approaches and causes of agglomeration economies

<table>
<thead>
<tr>
<th>Approaches</th>
<th>Indivisibilities (industrial dimension)</th>
<th>Static synergies (cultural dimension)</th>
<th>Dynamic synergies (cognitive dimension)</th>
<th>Synergies and Indivisibilities (industrial/cognitive dimension)</th>
<th>Physical proximity (geographic dimension)</th>
<th>Indivisibilities and physical proximity (industrial/geographic dimension)</th>
<th>A multi-dimensional approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input sharing</td>
<td>Scale economies in input production</td>
<td>Higher and less expensive recourse to the market:</td>
<td>Specialised shared knowledge among workers</td>
<td>Shared common or complementary knowledge among workers</td>
<td>Lower transport costs for intermediate goods</td>
<td>Factor sensitivity to shipping costs and</td>
<td>Shared common or complementary knowledge among workers</td>
</tr>
<tr>
<td>Home market effects</td>
<td>Large and diversified market for output</td>
<td>Lower transport costs for final goods</td>
<td>Goods sensitivity to shipping costs</td>
<td>Goods sensitivity to shipping costs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge spillovers</td>
<td>Adequate technical knowledge for choice of suppliers</td>
<td>Widespread entrepreneurial expertise</td>
<td>Socialization of the risk associated with innovative activity</td>
<td>Higher collective learning processes due to:</td>
<td>High mobility of tacit information</td>
<td>Nature of knowledge transmitted (codified vs. non-codified knowledge)</td>
<td>Codified vs. non-codified knowledge and degree of collective learning processes</td>
</tr>
<tr>
<td>Matching (reduction of transaction costs)</td>
<td>Availability of skilled labour</td>
<td>Informal contracts, i.e. flexible, non-bureaucratized relationships among firms</td>
<td>Specialised shared knowledge in labour searching</td>
<td>Highly qualified shared knowledge in labour searching</td>
<td>Easier labour demand/supply match</td>
<td>Geographical scope for labour sharing and shared knowledge in labour searching</td>
<td></td>
</tr>
</tbody>
</table>
Particular attention will be paid to the causes of agglomeration economies. Generally speaking, the latter – emphasised since Marshall – can be grouped into four types: input sharing, home market effects, knowledge spillovers and labour market pooling. However, the way in which these causes are interpreted changes considerably according to the interpretative perspective adopted: each approach in fact furnishes additional insights into this general typology of agglomeration causes. Table 1 provides a schematic representation of the different interpretations given to each cause of agglomeration economies by the different approaches. They will be discussed in detail in the next sections.

3. The Early 1970s: Indivisibilities and the Industrial Dimension

During the 1970s, the first studies on agglomeration economies, which were of an empirical nature, sought to test whether the scale of agglomerated environments added to productivity. The well-known dichotomy drawn between urbanisation and localisation economies reflected these attempts (Jacobs, 1969). The framework of analysis for this approach was identification of whether local scale economies are related to the scale of local own industrial activity or to the cross-fertilisation enhanced by the scale and diversity of activity outside the own local industry (Fig. 1a).

Approaches of this kind built upon a particular conceptualisation of space, which was interpreted as dichotomous and ‘absolute’. This was a dichotomous space because these studies envisaged a concentrated vs. dispersed space, the former being expected to add to factor productivity. It was ‘an absolute space’ because it could be analysed only through endogenous industrial specificities, leaving aside any economic, social, physical and/or environmental interactions with other areas. These studies, in fact, used standard a-spatial techniques in which the proximity of the geographical units was not taken into account.

Interestingly, because this approach analysed the industrial dimension in depth, it postulated the geographic and socio-cultural dimensions: concentration and agglomeration in urban areas were largely assumed to be acceptable starting points for empirical analysis, whilst the socio-cultural aspects behind synergy mechanisms and market interactions were left to pure probability contacts characterising densely populated areas.

Within this conceptual interpretation of space, agglomeration economies were analysed in terms of their industrial scope, i.e. their capacity to generate and expand their advantages by exploiting indivisibilities within either a specialised environment or a diversified industrial one (Segal, 1976; Henderson, 1985; Shefer, 1973; Carlino, 1980; Sveiskauskas, 1975; Moomaw, 1983; Hoch, 1972; Mera, 1973; Mills, 1970, to cite only some authors).

In the industrial approach, the causes of agglomeration economies were identified in the production cost minimisation achieved through scale economies (Table 2): input sharing was interpreted in terms of increasing returns to scale in production; the labour market as the size which allowed a better match to be achieved between employers’ needs and workers’ skills and reduced risks for both; home market effects and scale of pecuniary externalities (later developed in more detail by Krugman (1991)), whose concentration of demand involved economies of scale; adequate technical knowledge achieved on a large scale; and highly specialised labour markets (Rosenthal and Strange, 2001).
The new insights, with respect to Marshall’s first theoretical approach to the concept, yielded by this approach stemmed from the attempts made to determine empirically whether the scale effects are subject to industry specific effects.

4. The Mid-1970s: Synergies, Local Networks and the Socio-Cultural Dimension

During the mid-1970s a theoretical model of regional development in which the economic aspects of development were reinforced by increasing returns and self-reinforcing mechanisms of development was conceptualised for the interpretation of regional dynamics (Becattini, 1989). On this approach, mainly known as the “industrial district theory”, endogenous causes of agglomeration economies became causes of local growth, and therefore required thorough theoretical reflection and convincing arguments for their existence.

The main concern of this approach was to identify the local conditions that generate and support market interaction and synergies among economic agents, and the effects produced by these synergies on firms’ productivity. For the first time, network advantages were associated with highly agglomerated and specialised areas, and they became the main explanation for agglomeration processes. According to this approach, the capacity of economic actors to cooperate is rooted in the socio-cultural environment in which firms operate, and it generates increasing returns – or more precisely, localization economies, or again ‘district economies’.

The link among a socio-cultural environment, synergies among actors, and local efficiency was the driver of agglomeration forces according to this approach. Socio-cultural proximity – defined as a shared code of behaviour and a set of common values penalizing opportunistic behaviour – penetrates the market, structures it around clearly-defined rules, and generates efficiency.

The socio-cultural dimension reinforced the interpretation of agglomeration economies as driven by an agglomerative force never previously analysed in detail: that of synergy. This element shed further light on, and further emphasised the causes of, agglomeration economies as already stressed by an industrial approach; these economies being based on indivisibilities as their main agglomerative forces: trust, sense of belonging, social proximity in general, reinforce and facilitate input sharing, give rise to knowledge sharing, and enable an easier match to be made between the supply and demand of inputs and intermediate goods. In this approach, the genetic conditions of agglomeration economies were locally embedded elements difficult to copy and to reproduce in other specialised and densely populated areas (Table 2).

Whilst the geographic dimension was still left aside – space was again considered to be dichotomous and absolute – the interest of this branch of literature was mainly directed at the socio-cultural dimension in agglomeration economies. The industrial dimension was also taken for granted: conceptual reflection in fact concentrated on highly specialised areas called “industrial districts” (Fig. 1.b). Interestingly, this approach treated the importance of large-scale advantages, in the form of indivisibilities, by stating that a high concentration of small firms in an area enables them to exploit indivisibilities without having to forgo the advantages that small size gives them. Marshall (1919) argued that the advantages of large scale may in general be achieved either by grouping a large number of firms into a single district, or by constructing a few large-scale factories. This approach was backed by rigorous quantitative analyses (Capello, 1999; Pietrobelli, 1998; Rabellotti, 1997) which moved away from the simplistic “observational equivalence” approach (Overman, 2004; McCann, 2007).
An important theoretical step forward was made possible by this approach, because it warned that agglomeration effects are subject to social specificities, which perform crucial functions in various respects: they ensure the smoother operation of market mechanisms, more efficient and less costly production processes, the accumulation of knowledge in the local market, and a more rapid pace of innovation – all of which are factors fostering agglomeration and local development. Thus discarded was the simple deterministic “physical proximity/scale advantages” dichotomy, and a stochastic element was introduced to explain the circumstances under which physical proximity generates scale advantages. This approach explained, in fact, why territories similar in terms of density, size and industrial specialisation may be characterised by different degrees of agglomeration forces and regional economic growth.

Despite the steps forward made possible by the socio-cultural dimension, its scientific recognition is still rather limited. Orthodox reviews on agglomeration economies often neglect this specific dimension, and it is never clearly recognised as a source of agglomeration economies. The same applies to its more modern definition, namely the cognitive dimension.

5. The 1980s: Synergies and Knowledge Networks

5.1. The Cognitive Dimension

A subsequent stage of reflection along the socio-cultural dimension concentrated on identifying a “cognitive dimension” in the causes of agglomeration economies (Fig. 1.c). Never explicitly mentioned by mainstream economists as a source of agglomeration economies, this perspective instead placed great emphasis on the “synergy” element as an agglomeration force, as the socio-cultural approach had done.

Like the approach based on the socio-cultural dimension, this one took the industrial and the geographic dimensions of agglomeration economies for granted. Space was abstract and dichotomous, while territories were diversified in terms of the “cognitive capability” of actors, by which was meant the ability to manage information in order to identify and solve problems; or more precisely in the economic sphere, the ability to transform information and inventions into innovation and productivity increases through co-operative or market interaction (Lundvall, 1992; Lundvall and Johnson, 1994; Asheim, 1996; Maskell and Malmberg, 1999). This cognitive capability increases in highly specialised and more densely populated environments (Aydalot and Keeble, 1988; Camagni, 1991; Rallet, 1993; Rallet and Torre, 1995; Ratti et al., 1997).

This approach focused mainly on the mechanisms that construct knowledge through cooperative learning processes nourished by spatial proximity (“atmosphere” effects), network relations (long-distance, selective relationships), interaction, creativity and recombination capability: what has been called “relational proximity” (Camagni, 1991). Shared behavioural codes, common culture, mutual trust and sense of belonging act not only on the capacity of firms to engage in market interactions but also on their cognitive capacity. They develop and enhance collective learning processes by means of specific territorial channels through which knowledge flows by virtue of: i) the huge mobility of professionals and skilled labour – among firms but internally to the local labour market defined by the district or the city, where mobility of this kind is maximal, and ii) intense co-operative relations among local actors, and in particular customer-supplier relationships in production, design, research, and finally knowledge creation (Camagni and Capello, 2002).

On this approach, agglomeration economies were dynamic in nature; they stemmed from processes of collective learning, from tacit knowledge continuously created, exchanged and utilised, and from
the ability to organise local knowledge into continuously innovative production processes and products. Agglomeration economies were no longer sources of increased efficiency; rather, they became sources of increased innovation capability.

This approach highlighted new and particular elements among the causes of agglomeration economies: input sharing and matching between the demand and supply of production factors and intermediate goods are facilitated by shared common knowledge, while the socialisation of risks associated with innovative activity increase innovation activities within local firms (Table 2); common knowledge is, in its turn, explained by relational proximity in the form of networking capacity (Camagni, 1991; Keeble and Wilkinson, 1999; Lawson and Lorenz, 1999). The important step forward made by this approach was that it identified the way in which local networks act on the exchange, creation and diffusion of knowledge: this was not confined to pure contact probability in limited geographical areas but was explained through specific territorialised networks among local economic actors like intense cooperation among local customers and suppliers, a high internal mobility of the labour force, and spin-offs of local firms.

As in the case of the socio-cultural element, despite the interest that developed around the cognitive approach, its clear identification in orthodox literature reviews on agglomeration economies has surprisingly been very limited, and never explicitly mentioned. It is also true, however, that an approach such as this takes the existence of concentrated and specialised areas for granted. Far from being a substitute approach for the industrial one, it is indubitably a rich complement to the study of agglomeration in space. It warns that agglomeration effects are subject to cognitive processes embedded in the social specificities of each area. For this reason, agglomeration forces can exert different forces on local growth in areas of similar population density, industrial specialisation and socio-cultural atmosphere. This theory, mainly known as the “milieu innovateur” theory, is able to explain the success and failures of industrial districts, an interpretation which is impossible using the socio-cultural approach.

5.2. A First Multidimensional Approach: the Industrial/Cognitive Approach

In the late 1980s, a first multidimensional approach to the study of agglomeration economies was developed: the cognitive approach merged with the industrial approach, and empirical studies mainly sought to understand whether collective learning mechanisms are more intense in specialised or diversified areas, in milieus or/and in cities (Fig. 1.c).

Theoretical reflections in the late 1980s, in fact, were centred on the idea that - despite many differences between the milieu and the city, especially ones related to the physical element with which the city is generally associated - there are numerous similarities between these two concepts: they share a common genetic principle, that of agglomeration, through which they may develop elements, processes and effects which range from the development of a common identity and sense of belonging to the socialised production of human capital and know-how (Camagni, 1999). The conclusions reached were that in all contexts characterised by a plurality of agents – like cities or industrial districts – knowledge evolution “is not the result of individual efforts in R&D within individual firms, but rather the combination of complementary capacities and of widespread interactive learning processes, which involve many ‘customers’ and ‘suppliers’ along a well-defined filière or supply chain” (Cappellin, 2003, p. 307). Collective learning processes were interpreted as forming the core of both the innovative milieu and the growth of the city, influencing their innovative capacity and growth.
The interpretation of the city as a place where milieu mechanisms can be put in place raised an interesting conceptual question: whether the innovative dynamics of firms located in urban areas are influenced more by externalities stemming from their urban location (urbanisation economies) or by localisation, or “milieu economies” as they are termed in a modern version. Quantitative empirical analyses were conducted to measure the degree to which urban areas are able to register advantages in their innovative activity stemming from collective learning processes (Capello, 1999).

Like the socio-cultural approach, also these studies were based on the idea that space is dichotomous (densely vs. non-densely populated) and absolute, with no relations being assumed among different geographical areas. The latter differ in terms of social structures and industrial composition, and these differences explain the scope of agglomeration advantages.

Causes of agglomeration economies were interpreted from a new and additional perspective: the question was not just whether input sharing is facilitated by a socialised knowledge shared among workers, but whether it is facilitated by common or complementary knowledge. Knowledge spillovers were analysed in terms of either common or complementary knowledge spillovers, and mechanisms behind labour market pooling were studied to identify whether a better match of workers could be achieved through complementary or common knowledge flows.

With respect to a cognitive or socio-cultural approach, a new step forward was achieved by interpreting the different scope and intensity of agglomeration economies through social and industrial specificities at the same time. The scope of agglomeration economies depends on different aspects simultaneously: socio-cultural and cognitive elements act differently in different industrial realities.

6. The 1990s: the Geographic Dimension and a Second Multidimensional Approach

6.1. The Geographic Dimension: Spatial Spillovers

Another important revolution in the study of agglomeration economies started at the beginning of the 1990s, when interest shifted from the industrial and cognitive dimension to the geographic one as a result of advances in statistical, and especially econometric, techniques able to capture, in more sophisticated manner, the role of space in empirical analyses of agglomeration economies. In that period, econometric techniques moved from a-spatial approaches built on the idea of ‘an absolute space’ to spatial ones mostly interested in analysing regions as parts of a larger system, and which therefore sought to capture interactive effects among regions and hence the relative position of one region vis-à-vis the others (Fig. 1.d).

The change brought about by these new techniques was that an observation in space was not analysed on its own, but rather as part of a system with which it interacted. Space became a continuous space, and physical distance among observations importantly explained their degree of interaction, with a consequent shift away from the absolute space of previous techniques.

A first application of this concept of space to agglomeration economies was made by the spatial spillover theory devised during those years: this theory viewed space as a driver of knowledge from R&D clusters (Fig. 1.d) (Acs et al., 1994; Audretsch and Feldman, 1996 and Feldman and Audretsch, 1999; Anselin et al., 1997). Interestingly, the other two dimensions explaining sources of agglomeration economies – industrial and socio/cultural cognitive – were ignored by this approach. In the original studies on spatial spillovers, industrial specificity in advanced functions of the local area was an acceptable starting point for empirical analysis. The presence of advanced
sectors and advanced functions like R&D and higher education – not interpreted but taken for
granted – guaranteed knowledge spillovers, while the capacity to absorb knowledge was assumed to
be homogenously distributed over space and based on pure probability contacts.

The aim of this approach was to stress the importance of geographic proximity as a condition for
the diffusion of local knowledge: firms located in areas characterised by a high number of scientific
activities register increasing returns in innovation processes thanks to the knowledge that they
acquire from the external environment.

Whilst a step forward was achieved in measuring geographical proximity by moving away from the
simplistic assumption of a postulated concentration of economic activities in a point in space, one
criticism brought against a pure geographic approach was that other mechanisms were taken for
granted, so that theoretical reflection inevitably moved a step backwards. In particular, synergy
elements were left to pure probability contacts that increased as distance decreased. This simplified
approach also impoverished the interpretations of the mechanism by which knowledge flows are
spread, which was reduced to a simple epidemic process. Thus ignored was one of the most crucial
aspects of the innovation process: how people (or the context) actually learn, and how
agglomeration can support learning processes through interaction, networks, and exchange of
labour and knowledge (Capello and Faggian, 2005). Also lost was identification of the territorial
channels through which knowledge spreads; for this reason, the pure geographic perspective can
easily be interpreted as the reason why one step forward and one step back were taken in the
complex process of disentangling the black box of agglomeration economies.

6.2. Geographic/Industrial Interaction

Interestingly, this approach influenced the most traditional and oldest studies on agglomeration
economies: in recent times the geographic dimension has been merged with the industrial
dimension in order to provide evidence on whether or not agglomeration economies attenuate
rapidly across space, and whether this occurs more among establishments operating in the same or
in different sectors (Fig. 1.e).

Studies on agglomeration economies revised the original empirical works of the 1970s in light of
geo-referenced observations, augmenting the specialisation vs. diversification dimension with the
geographic perspective. The existence of geo-referenced data on establishment on the one hand, and
of spatial econometric techniques on the other, opened the way for a new multidimensional
approach, which at present is the most advanced, and the one most in fashion (Ellison and Glaser,
1997; Ciccone, 2002; Ciccone and Hall, 1996; Rosenthal and Strange, 2001 and 2003; Henderson,
2003).

Recent technical developments in spatial econometrics have given rise to two new interrelated
research directions. The first consists in investigating whether the new techniques corroborate the
results achieved, correcting for the problems that space can generate in estimation. On applying new
spatial econometric techniques, the results improve in terms of robustness when spatial proximate
dependence in the form of spatially lagged versions of explained and explanatory variables is
introduced into the econometric models (see, for example, van Oort, 2007).

In relation to the first direction, the second one consists in the application of new techniques for a
different theoretical purpose: to test whether agglomeration economies attenuate across geographic
space. Spatial econometric techniques are able to prove the existence of interdependence among
local actors, and can therefore provide evidence of the importance of physical proximity as an agglomeration force.

This new multidimensional approach provides evidence that agglomeration economies are subject to distance decay effects. This result varies according to factor sensitivity to shipping costs, to the nature of factors (non-material vs. material), and to the nature of agglomerative forces. The geographic scope of agglomeration economies is at the state level for factors sensitive to shipping costs, at local level for knowledge factors, at both national and local level for labour, and of wider geographic scope for input sharing with respect to knowledge spillovers (Rosenthal and Strange, 2001).

Related to the idea that agglomeration economies are subject to distance decay effects is the notion that the identification of agglomeration economies in some areas depends closely on the geographic scale of analysis. Empirical results on the existence of agglomeration economies are more clear-cut when these are analysed on lower spatial scales (van Oort, 2007; van Soest et al., 2006).

Theoretically speaking, this multidimensional approach is important for providing insights into the spatial dimension of the phenomenon, but it does not add much to the conceptual framework already known. It is a first attempt to adopt an integrated approach, since it considers two main micro-foundations of agglomeration processes at the same time. However, in this framework no role is given to synergies, to cooperation networks for innovation and learning, or to the socio-cultural and cognitive aspects behind cooperative attitudes, all of which are elements already envisaged in the literature to explain agglomeration forces but nevertheless omitted from this framework.

7. Future Challenges: Towards an Integrated Approach

This is where theory and empirical analyses currently stand. The question that naturally follows is this: can we envisage some further important research directions in this field? The reply is in my opinion ‘yes’, and hereafter are presented some future challenges that need to be addressed in this important area of analysis.

A step forward in interpreting agglomeration economies would be to reposition the socio-cultural/cognitive dimension in the debate on agglomeration economies. This dimension is vital for distinguishing between a territory and a pure geographic space: whilst the latter is a geographic container of specific functions and economic activity, whose economic dimension is easily measurable in physical terms, the former is a relational space where functional and hierarchical, economic and social interactions and networks are facilitated by social rules and common behaviours embedded in geographic space.

Why should the socio-cultural/cognitive dimension be re-discovered in this debate? The reason is that this dimension requires one to move away from the simple idea that pure geographic proximity, enhancing both indivisibility and synergy, is at the basis of agglomeration forces. Whilst this might be true for indivisibilities, when proximity allows average production costs minimisation, the idea that pure geographic proximity influences synergies and exchange of knowledge among actors is too simplistic, because it is based solely on the higher probability of contacts in a reduced space. This reasoning entails some very simplistic assumptions on a homogeneous capacity to cooperate, to exploit knowledge, and to learn across space. The pure geographic approach is able to explain knowledge diffusion mechanisms through spatial epidemiological diffusion processes of information and knowledge, but it is unable to interpret knowledge creation mechanisms or to
identify channels of direct knowledge interaction, and learning (Capello, 2008). But this criticism also applies to other causes of agglomeration economies, like input-sharing and labour market pooling. In traditional approaches these are explained by the presence of scale economies in input production when the production is concentrated; in reality they also depend on trust, cooperative attitude and common values – important elements for increasing outsourcing by firms.

Possible solutions to this limitation can be found by adopting a multidisciplinary and integrated approach (Fig. 1.e). A step forward in the analysis of agglomeration economies, of their source, scope and nature should in fact be the result of the three approaches used jointly. Spatial proximity to diversified/specialised establishments should also take account of cognitive proximity between two establishments, measured in terms of absorptive capacity, common knowledge background, cooperation attitude, openness to new opportunities and needs of market interactions; all these aspects have different weights in different industries. In this way, the three main micro-foundations jointly underpinning agglomeration economies would be considered all together and given their correct weights in explaining this phenomenon.

In theoretical terms, this would mean expecting agglomeration economies to attenuate across geographic space, across sectors and across socially heterogeneous environments at the same time. This may be the case of some agglomeration forces more than others (e.g. input sharing more than home market effects) and of some micro-foundations more than others (synergy and knowledge sharing more than indivisibility). It would also enable explanation of why, with the same physical distance between two establishments in the same industry (at micro-territorial level) or between regions (at the macro-territorial level), agglomeration economies may emerge in some cases and not in others, the difference being due to the different absorptive capacities of firms, at micro territorial level, or to the different socio-cultural/cognitive conditions of regions, at the macro-territorial level. In other words, the advantage intrinsic to an integrated approach is that it makes it possible to overcome the deterministic assumptions typical of traditional approaches like the spatial and industrial ones, in which proximity means agglomeration advantages. The cognitive approach instead enables inclusion of a stochastic element among the causes of agglomeration economies – an element related to the capacity of actors, firms and regions in general – so as to grasp the advantages that proximity creates.

This research programme has three main obstacles to overcome, which are now presented in increasing order of magnitude. The first is data availability on non-material elements, such as attitudes towards cooperation, level of trust and sense of belonging. Some attempts (like the European Values Study) have already been made to collect information of this kind both in America and in Europe, and use has been made of studies on a number of proxies (such as crime rates or election participation rates for social capital, and the presence of exhibitions, cultural events and graduates in specific disciplines for creativity), either directly or in the form of indexes, which bears witness to an interest in solving the problem of the lack of data.

The second obstacle to be overcome is the development of quantitative methodologies able to treat different kinds of proximities simultaneously. The simplest approach would involve measurement of interaction among economic agents – regions, firms or consumers – over physical space. The use of distance matrices based on social or cultural distance together with physical distances or travel times would enable regional scientists to validate their hypotheses in a way almost fully consistent with their theoretical assumptions. Some steps have already been taken in this direction, for spatial weights matrices can now also be calculated using non-spatial features. For example, at country level, bilateral trade flows can easily replace physical distance. This, however, requires the weight matrix to be the same throughout the estimation period. While physical distance can easily serve this purpose, other kinds of distance, like bilateral trade flows, are impossible to use as constant
variables over a period of time (Abreu et al., 2004). From the methodological point of view, this is the main challenge facing econometricians: how to move from ‘spatial’ to ‘territorial econometrics’.

The third, and most difficult, challenge is making mainstream approaches sensitive to the role of non-tangible elements in explaining the sources, nature and scope of agglomeration economies, and recognising this role explicitly. As far as one can tell from the existing literature reviews in this field, this situation is still a long way away. However, I remain rather optimistic in this regard. Just as we would have never thought of finding studies in mainstream economic journals (American Economic Review, Quarterly Journal of Economics, Journal of Economic Literature) dealing with the role of trust (Glaeser et al., 2000; McCloskey and Klamer, 1995), social capital (Glaeser et al., 2002; Knack and Keefer, 1997; La Porta et al., 1997; Beugelsdijk and van Schaik, 2005), sense of belonging to a society (Bowles et al., 2001; Lazear, 1999; Alesina and La Ferrara, 2000) in national and regional economic dynamics, so we may be confident that analysis in this direction will emerge in the agglomeration economies theory as well, enriching the already wide and profound reflections on this issue.

8. Conclusions

Building on what already exists in the literature, this paper has highlighted new insights and limits in the field of agglomeration economies, and it has indicated further research directions in that field. In particular, the main claim of the paper is that the different approaches to agglomeration economies should be seen as complements rather than as substitutes: they should merge into a “multi-dimensional approach”. This would mean expecting agglomeration causes to attenuate across geographic space, across sectors and across socially heterogeneous environments at the same time. It would also require overcoming the simplistic and deterministic idea that physical proximity means by definition agglomeration advantages, as well as the idea that density (of activities, of people) means agglomeration economies: agglomeration advantages are the result of an intertwined set of socio-economic, geographic and industrial conditions that, once they are all present, display the highest intensity in agglomeration advantages. An approach like this can be useful in explaining why territories with the same size, density and industrial specialisation can display different agglomeration advantages; a condition which finds no explanation in the present interpretative approaches.

References


