Modern Urban and Regional Economics

Second Edition

Philip McCann

Oxford University Press

2013
2.1 Introduction

In Chapter 1 we discussed the theoretical issues which affect the location behaviour of the individual firm. Each of the models presented provides us with a way of analysing the particular microeconomic effects on firm location behaviour, of various influences such as transport costs, local factor prices, production and substitution possibilities, market structure, and competition. In reality the actual location behaviour of the firm is a result of a complex mix of each of these influences. Therefore this leaves us with the problem of determining which particular influences are the dominant influences in which situations. Unfortunately, without information on the individual firm and industry, however, the various microeconomic models above do not lead to any systematic conclusion as to whether optimal firm location behaviour is more likely to result in industrial clustering or in industrial dispersion. Yet, in describing the generally observable features of industrial location behaviour, two particular features do stand out.

The first generally observed feature of industrial location behaviour is that most industrial activities tend to be clustered together in space, and the reasons for this clustering behaviour are examined in this chapter. Most productive and commercial activities do take place in the immediate vicinity of other such activities, and such clusters may take a variety of forms, including industrial parks, small towns, or major cities. However, as we have discussed in Chapter 1, firms make different location choices based on their production and cost functions, the spatial delineation of their markets, and their competitive strategies. Moreover, firm clustering obviously often leads to congestion costs and higher land prices. Yet, in spite of these different location-production choices and congestion costs, the general observation that industrial clustering is widespread is indeed valid. As such, this observation raises the important question of why it is exactly that activities are generally grouped together geographically, and the reasons why firms are often clustered together in space is the central issue discussed in this chapter.

The second generally observed feature of industrial location behaviour, which will be discussed in detail in Chapter 3, is that there appears to be a size distribution of spatial clusters, with different ranges of activities taking place in different clusters. Some activities
are typically dispersed over large areas, with goods being shipped generally over large distances, while others are more geographically concentrated. These differences in the degree of dispersion also give rise to regularities in the patterns of clusters. In particular, within an individual country or market area, there will usually be a single largest city cluster which exhibits almost all types of activities, followed by larger numbers of other smaller clusters which increase in number as their individual size falls. The smaller clusters will tend to exhibit a smaller range of activities taking place within them than the larger clusters. These observations are collectively known as an 'urban hierarchy', and the particular reasons for the development of such a system of cities are discussed in Chapter 3. Instead, this chapter will focus specifically on the issues explaining why activities cluster together in space.

In sections 2.2 to 2.5 of this chapter, we will discuss in detail the various arguments and explanations as to why industrial activities are often observed to be clustered geographically. The arguments in sections 2.2 and 2.3 centre on the role played by knowledge spillovers, labour markets, and economies of scale in fostering agglomeration. Moving beyond simply agglomeration, section 2.4 examines the transactions costs relationships underpinning a broader range of different types of clusters, and section 2.5 discusses the role played by both creativity and consumption in fostering urban agglomeration and clustering, in particular. Yet these explanations all suggest a level of information availability, awareness, and rationality on the part of individual firms and people, such that collectively their individual decisions give rise to agglomerations and clusters. For this reason section 2.6 raises the question of how clusters do actually arise in an environment of limited information, conflicting goals, and uncertainty, and here we see that evolutionary processes become central to the discussion. Not surprisingly, these evolutionary processes also reappear as a critical element in Chapter 7 where we discuss regional growth.

### 2.2 Industrial clustering: returns to scale and geography

Industrial clustering refers to the observation that all types of commercial activities—manufacturing, services, resource-based industries—are frequently observed to be grouped together in space. In attempting to explain this observation, it is necessary to employ the notion that economies of scale can be place specific. To see this, we can consider the spatial outcomes of the alternative hypothetical case where all firms achieve constant returns to scale. If, for whatever reason, a large group of such firms in the same sector or a range of sectors ends up being located in the same place, the result of this clustering will be a large level of investment at that particular location. These firms will require land and space for their activities, and the high demand for land at the particular location will force up its price. If everything else is unchanged, and if the firms all achieve constant returns to scale, the increase in the price of land will reduce the profitability of all the firms at that location. Similarly, the increase in the local land price will mean that the living costs of the labour employed will also go up. In order to maintain the local labour supply the firms will also have to increase wages. Once again, this will reduce the profitability of all the firms in the area. The reduced profits will mean that firms located here will be less competitive than their competitors located elsewhere and will struggle to survive in the market. Some firms will move away to locations where they can still achieve lower costs and higher profits.

One implication of this is that firms which are still able to achieve internal economies of scale will now require much higher profits to make up for the increased local costs. These increased profitability requirements will mean that many firms will become constrained, and if they are not able to achieve profits at the levels required, then the urban growth will slow down.

This situation of high local costs and low profitability means that some firms will be forced to move to locations where the local costs are lower. These firms will typically be able to achieve internal economies of scale in their production, and they will be able to maintain profitability in these locations. As these firms move to new locations, the process will start over again, and we will see that the process of agglomeration will continue, although at a slower rate than before.

### 2.3 Location and agglomeration

Location and agglomeration are both important factors in the development of industrial clusters. Weber (1909) argued that there are three main factors which influence the location of industrial activities: the cost of raw materials, the cost of labour, and the cost of transport. These three factors are all interrelated, and it is often the case that they are all involved in the location of industrial activities. Weber also argued that the location of industrial activities is determined by the costs of the raw materials, the cost of labour, and the cost of transport. The location of industrial activities is determined by the costs of the raw materials, the cost of labour, and the cost of transport. The location of industrial activities is determined by the costs of the raw materials, the cost of labour, and the cost of transport.
move away to alternative locations while others will simply go out of business. Eventually
the cluster will disappear. This hypothetical example is, however, inconsistent with the gen-
eral observation that most activity clusters continue to exist.

On the other hand, let us imagine a situation where each of the firms in the same locality
achieves significant external economies of scale precisely because of the large number of
firms located in the area. In this situation, the high level of investment in the local area will
still imply high local land prices and high labour prices as before. However, the difference
now is that these increased factor prices may be more than compensated for by the
increased efficiency on the part of each firm. The result of this will be even higher profitabil-
ity for all the local firms, even though the local factor prices may be higher than else-
where. Other firms from other areas may now also consider moving into our area,
contributing to a further growth in the levels of local investment, factor prices, and firm
profitability. This in-migration of new firms will lead to a cumulative process of local
growth.

This hypothetical example is consistent with the observation that most activity clusters
continue to exist. However, in reality, the growth of local clusters tends not to be a process
which is continuously cumulative, as in the latter example. The reason for this is that this
would imply that all activity would end up at one location! Therefore, in order to discuss the
existence of spatial industrial clusters, it is necessary to employ the notion that place-
specific economies of scale do exist, but also to acknowledge that there may be limits to such
effects.

2.3 Agglomeration economies

Location-specific economies of scale are generally known as agglomeration economies. The
existence of agglomeration economies was acknowledged by classical authors such as
Weber, but it was Alfred Marshall who first provided a detailed description of the sources of
these economies. In Marshall’s (1890, 1920) schema, these economies are generally under-
stood to be external economies, which are independent of a single firm, but which accrue to
all of the firms located in the same area. A good description of the Marshall approach is
given by Krugman (1991). However, as well as Marshall’s description of the sources of
agglomeration economies, there is also Hoover’s (1937, 1948) classification of types of
agglomeration economies. Hoover’s description is a somewhat different characterization to
that of Marshall, but it shares some elements, and given that both approaches are frequently
adopted, it is necessary to be explicit about their differences.

2.3.1 The sources of agglomeration economies

Marshall (1890, 1920) observed that firms often continue to cluster successfully in the same
locations. From our example above, this implies that increasing returns to scale must be
achieved by the firms in the cluster. Marshall provided three reasons why such economies
of scale might be achieved. In other words, he identified three possible sources or origins of
such economies of scale. These are knowledge spillovers, local non-traded inputs, and a
local skilled labour pool.
(i) Knowledge spillovers

If many firms in the same industry are grouped together in the same location, it implies that the employees of any one particular firm have relatively easy access to employees from other local firms. This easy access can be either through the facility to have frequent direct face-to-face contact in business meetings, or alternatively though frequent informal contacts such as lunch meetings, sports activities, or other such social occasions. The important point about such informal meetings is that they allow tacit information to be shared between the participants. Tacit knowledge is knowledge or information which is incomplete and which is shared on a non-market basis, and can relate to issues such as new products, personnel, technology, and market trends. The participants in such meetings will each give information which is partial in order to acquire other information which is also partial. This process of the mutual trading of information allows each of the market participants to build up a more coherent picture of the overall market environment, thereby improving their ability to compete in the market. The more such participants there are in the local area, the more complete a picture can be assembled by each participant. The advantage of spatial clustering in this case is therefore that proximity maximizes the mutual accessibility of all individuals within the cluster, thereby improving the knowledge and information available to all local participants. In market environments characterized by rapidly changing information, such clustering affords the agglomerated firms an information advantage relative to all other firms, and the extent of this advantage depends directly on the number of such firms which are located in the same area.

Good examples of this are the international financial markets, which are centred on highly concentrated areas such as Wall Street, New York, the City of London, and the Marunouchi district of Tokyo. In this sector, international financial market information is changing by the minute. Managers have to make important decisions on a daily basis, and these decisions often involve rapid negotiations within the banking syndicates of which each financial institution is a part. Immediate access to market participants is essential.

(ii) Non-traded local inputs

In situations where many firms in the same industry are grouped together in the same area, there will be possibilities for certain specialist inputs to be provided to the group, in a more efficient manner than would be the case if all the firms were dispersed. These inputs are described as 'non-traded' in order to distinguish them from consumed inputs of the type described in the Weber and Moses models. For example, if we once again take the case of the financial markets, areas such as Wall Street and the City of London have many specialist legal firms and software firms, whose only role is to provide specialist services to the international financial sector. The provision of such specialist services is very expensive. However, where there are many firms within the industry which are located at the same place, the average cost of this service provision to each market participant will be low. The reason is that the costs of setting up such services will be spread over a large number of local customer firms. Similarly, in automotive engineering clusters in city-regions such as Detroit, Michigan, Nagoya, Japan, Stuttgart, Germany, and Turin,
Italy, there are specialist testing firms. Their only role is to test the accuracy and safety of industrial components, using highly specialized and expensive equipment. The cost of employing such equipment would be prohibitive to most market participants, but the fact that a large number of firms within the same sector that require such testing services are also located at the same place allows the costs to be spread across the group. A second type of non-traded local input is that of specialist local infrastructure. In the City of London there is a specialist dedicated wide-band fibre-optic cable system which is designed to allow the maximum possible flows of data between the local financial institutions, while excluding the general public. Access to the system comes about only through location in the City. All the local market participants benefit from the specialist infrastructure, and the cost of it is spread across all the beneficiaries. As above, the costs of the non-traded local inputs to each firm within the group will fall as more firms join the cluster.

(iii) Local skilled labour pool

The third source of agglomeration economies is the existence of a specialized local labour pool. This allows firms to reduce their labour acquisition costs, and there are two aspects to this. The first is that firms require sufficient quantities of labour to respond to market conditions. Therefore, if market demand conditions improve rapidly, a firm will wish to expand its labour force quickly, and will need to undertake a search process to acquire the workers. Secondly, the firm will also need to ensure that the employees are able to carry out the tasks correctly. In many sectors the costs of training labour and skills acquisition can be extremely high. This is because workers will need to be provided with specialist courses and instruction. Also, the opportunity costs involved with the time involved in these training activities can be extremely high. However, if a firm is located in an area which already has a large local pool of workers with the specialist skills required by the particular industry, the costs to the firm of expanding its workforce will be relatively low. This is because the firm will have to undertake little or no retraining activities. For industries in which skills-acquisition costs are high, or in which the opportunity costs of time are significant due to rapidly changing market conditions, a local pool of skilled workers will therefore be of great benefit. The labour acquisition costs on the part of the firms, which include both the search costs and the retraining costs, will be reduced relative to firms in dispersed locations.

These three sources of agglomeration economies have been succinctly captured by Duranton and Puga (2004), who describe them as processes of learning, sharing, and matching. This very neat formulation allows us to see agglomeration not as a static phenomenon, but as a dynamic phenomenon of simultaneous processes of interaction. Together, these three sources of agglomeration economies can allow firms within a cluster to experience economies of scale which are external to any single firm, but which are internal to the group. The key feature of each of these sources of agglomeration economies is that spatial clustering reduces knowledge and information transactions costs. Clustering therefore increases the likelihood that the appropriate information will be transmitted, that the specialist requisite services will be provided, and that the appropriately skilled labour will be available, at that location, relative to other more dispersed locations.
2.3.2 The types of agglomeration economies

The sources of agglomeration economies described above allow firms within the same industry which are clustered together in space to achieve localized external economies of scale. However, in many areas, groups of firms in different industries can be clustered together geographically. For example, major cities may contain hundreds of industrial clusters. The exact nature of the agglomeration economies may therefore be different in different locations. In order to describe the particular nature of agglomeration economies in any particular area, economists often adopt a classification which was first employed by Ohlin (1933) and Hoover (1937, 1948). This classification splits agglomeration economies into three types, namely internal returns to scale, localization economies, and urbanization economies.

(i) Internal returns to scale

Some firms achieve significant economies of scale in their production simply by reason of their size. These economies of scale are regarded as being internal to the firm, in that the efficiency gains are explicitly deemed to be a result of the size of the individual firm. As such, these internal economies do not concur with the Marshall description above of economies of agglomeration as being external. Yet the notion of economies of scale here is explicitly spatial in that it is assumed that the internal economies of scale are generated because a large level of investment takes place at one particular location, rather than across a range of different locations. A large factory, such as the Fiat automobile plant in Turin, or the Boeing Everett hangar in Seattle, requires a large quantity of capital and a large labour force to be located at the same place. These internal production economies of scale are therefore associated with a high spatial concentration of both investment and people. As such, large stocks of factors are clustered together in space, although in this case they are within the definition of a single firm. However, the point is that the internal economies of scale are location specific.

(ii) Economies of localization

Localization economies are the agglomeration economies which accrue to a group of firms within the same industrial sector located at the same place. For example, in the case of Seattle, there are many firms that produce specialist aerospace components supplied directly to Boeing. Similarly, in automobile clusters such as in Detroit, Michigan, Stuttgart, Germany, and Nagoya, Japan, there are many firms producing specialist supplies for the major automobile-producing firms. There are several ways in which the local supply firms can benefit from close proximity to their major customer firms, which are the firms achieving internal returns to scale. According to Marshall's first source of agglomeration economies, the supply firms may benefit from frequent information exchanges with the customer firms, thereby increasing the mutual understanding and familiarity of these firms at different stages within the production process. Sometimes this will also involve exchanges of personnel and consultants. These activities can facilitate product development in markets where the risks are high. Also, as with Marshall's two other sources of
agglomeration economies described above, the firms in the same sector can benefit from specialist non-traded local services and a skilled local labour pool. Each of Marshall's sources of agglomeration can therefore contribute to localization economies, the definition of which is therefore that they accrue within a particular industrial sector.

(iii) Economies of urbanization

Urbanization economies are those economies of agglomeration which accrue to firms across different sectors (Jacobs 1960). For example, in the case of the cities mentioned above, namely Seattle, Detroit, Stuttgart, and Nagoya, the economy of each of these cities is centred around a single plant or a group of very large plants, each of which exhibits internal returns to scale. Around these plants are many supplier firms, and the group of customer and supplier firms together achieve localization economies. However, each of these cities is much larger than simply the single sector of aerospace in Seattle, and the automobile sector in Detroit, Stuttgart, and Nagoya. In order for other activities to continue to be clustered in these cities, they must also experience economies of scale. For example, all the people who live and work in the sectors achieving localization economies will require legal, real-estate, retail, educational, health care and leisure services. Similarly, the firms themselves may require services such as marketing, advertising, catering, packaging, transportation, real estate, and security, among others. These various activities, although not directly related to the sector experiencing internal returns to scale and localization economies, will still cluster in the local economy in order to provide services for the firms and employees of this sector. This clustering is in response to the large local market possibilities which exist. However, as before, these firms will experience increased local factor prices, which must be compensated by economies of scale if the clustering is to continue. The agglomeration economies experienced by these other sectors are termed urbanization economies.

Following on the from the seminal papers of Glaeser et al. (1992) and Henderson et al. (1995), there have been many discussions using urban and regional data from many countries regarding the extent to which diversity or specialization is more advantageous for growth. Many studies support the advantages of diversity over specialization, although, as a whole, the evidence is still rather inconclusive (de Groot et al. 2009). Part of the problem is that many knowledge activities and transactions transcend individual industries, sectors, or technologies, and therefore identifying and classifying these effects is very difficult (Mameli et al. 2008).

In the Hoover typology, internal returns to scale are firm-specific economies of agglomeration, localization economies are industry-specific economies of agglomeration, and urbanization economies are city-specific economies of agglomeration. As we see in Urban and Regional Example 2.1, the difference here between the three classes of agglomeration economies therefore depends on the definition of the boundaries of the firms and the sectors.

These definitional changes do not always simply reflect changes in firm boundaries or changes in firm ownership. In some cases, the change in definition of agglomeration economies may also indicate more fundamental changes in the nature of the interactions taking place between firms, and in many cases these changes can actually be driven largely by the
changes of ownership. Indeed, ownership reorganization is a key feature of most forms of industrial restructuring, and if the ownership changes are associated with beneficial changes in the competitive and cooperative relationships between firms (Saxenian 1994), these ownership changes will be associated with genuine agglomeration impulses. We will deal with the features of different types of agglomeration and clustering and the relationships between firms in section 2.4. Before this however, Box 2.1 reviews a range of other descriptions of industrial clusters, all of which play an important role in contemporary discussion if industrial clustering.

**BOX 2.1 Alternative descriptions of industrial clusters**

As well as the classic Marshall and Hoover descriptions of agglomeration economies, there is a range of other models which discuss particular aspects of industrial clustering. Each of these models is also regularly adopted by analysts to describe various types of clusters, so we will briefly discuss them here. They are the three long-established models known as the growth pole model, the incubator model, and the product cycle model, as well as two more recent models, the Porter model (1990, 1998) and the new industrial areas model (Scott 1988).

The growth pole model

The growth pole model was originally associated with the work of Perroux (1950), although it employed some of the ideas of Schumpeter (1934). Perroux described economic relationships in terms of monetary space, in which there are certain polarities regarding financial transactions. In other words, the decisions made by key large firms have major financial implications for other firms which are linked to the key firm.
through customer-supplier relationships. Perroux described these key large firms as 'growth poles', and the decisions they make which affect other firms relate primarily to issues such as innovation. The concept was translated into spatial planning terms primarily by Boudeville (1966), in which it was argued that the spatial behaviour of an area will be affected by the location behaviour of certain major firms or plants. For example, if large firm investments or public investment projects are implemented at a particular location, this location-specific investment can act as a focus for local growth. Other local firms may be able to use the advantage of proximity in order to increase their local sales to the new investment, thereby generating localized growth in the hinterland of the investment. These beneficial effects, however, will take time to develop. For example, as we saw in section 3.1, the immediate effect of a large location-specific investment may be an increase in local factor prices. This may reduce local efficiency in the short run and also lead to some localized crowding-out effects. In the growth pole literature these negative local effects are collectively known as 'backwash effects'. However, as the positive economic aspects of the growth pole begin to take effect, the beneficial spread effects are assumed to dominate the negative backwash effects within the hinterland region, thereby engendering positive local growth over the long run. The major point raised by the growth pole model is that large innovative firms or investment activities will tend to have significant impacts on the local development of industrial clusters. Innovation alone may not be of such significance, nor will size alone, although a combination of the two may engender significant local growth effects. The weakness of the approach, however, is the lack of any coherent analysis of the costs and benefits of such investment schemes. Good reviews of this literature are provided by (Richardson 1978; Parr 1999a,b).

The incubator model

The incubator model is associated with the work of Chinitz (1961, 1964) and was derived from observations of the industrial structure of Pittsburgh and New York. At the time Pittsburgh was dominated by a small number of very large firms and a small number of sectors—primarily coal and steel. In contrast, New York contained a wide variety of different sectors, and was less dominated by particular large firms. The argument of Chinitz is that industrial clusters which are highly diversified, and which contain a range of types of industries and firm sizes, will act as superior 'incubators' for the development and growth of new firms. The reason for this is that in such an environment, there will be a variety of local business services available to these small firms which will facilitate their growth. On the other hand, in industrial clusters dominated primarily by large firms, many of these requisite services will not be available because the large firms will be able to supply such activities internally, or on the basis of long-term contracts with a limited number of suppliers and customers. These arguments have been extended more recently by Duranton and Puga (2001), who examine the incubator advantages of 'nursery' cities. Chinitz's argument is important because it suggests that the issues of firm ownership structure, as discussed in section 3.2.2, may also play an independent role in affecting the growth of the cluster. In particular, a larger diversity of local sectors, firm sizes, and firm types may be important for the growth of the cluster, as it provides greater potential opportunities for firms to find new market niches. The Chinitz argument therefore suggests that diversified regions offer greater growth potential for newly emerging technologies and firms.

The product cycle model

The product cycle model is associated with the work of Vernon (1960, 1966) and is one of the approaches most frequently used to describe the qualitative aspects of spatial investment patterns. The most common use of the product cycle model has been in discussions of international investment flows (Vernon 1966), although the origins of the theory are actually found in observations of city clusters (Vernon 1960). Vernon's argument is that firms will separate activities by location according to the stage in the life cycle of the product, which in turn is reflected by the activities of the particular plant. For example, in high-level industrial clusters, such as in dominant central cities, firms will tend to locate
information-intensive activities such as research and development and high-level decision-making, which together relate to the early stages of the life cycle of the product. The reason is that these activities require large knowledge inputs, and, in particular, knowledge which is generally non-standardized. The non-standardized nature of the product or process is precisely due to the newness of the product or technique. Following the Marshall argument above, these activities may therefore benefit from the informal knowledge and information spillovers associated with localized clustering. Meanwhile, highly skilled employees will also be required to carry out such activities, and, once again, from the Marshall argument, it may be consistently easier to find these workers in such clustered areas. On the other hand, once a product and the associated production process have been designed, tested, and developed, the firm will be able to issue a blueprint or template which documents all the aspects of the production of the product or service and also the detailed specification of its delivery mechanisms. This information will now be available to other branches of the firm’s organization. The information regarding the product is thus standardized and, as production increases, the product becomes ‘mature’; in other words, it is no longer so novel. Over time, the production techniques tend to become better understood and relatively easier to carry out. As this happens, the requirement for knowledge inputs and highly skilled labour tends to fall, with the result that the firm can move the production process to other lower-cost and lower-skilled areas. The product-cycle argument therefore implies that more geographically peripheral areas, which tend to exhibit lower labour costs and lower labour skills, will also tend to have plants producing more mature, less novel, and more standardized products (Markusen 1985). The result is that there will tend to be a clear separation of activity types between central city-regions and more peripheral areas.

The important observation of the product-cycle model is that there may therefore emerge a qualitative distinction between the types of activities which take place at the economic centre or at the periphery of any geographical area. As we will see in Chapter 9, the evidence of the recent era of globalization suggests that the spaces over which these differences often emerge is now global as well as regional.

The Porter model

Within the business management community there has been much interest in the arguments of Porter (1990, 1998a,b). Porter focuses on the concept of competitiveness; a concept which is broader than just profitability, because it relates to all the elements and manifestation of the processes of profit creation. Porter’s arguments on competitive advantages were initially developed within a firm and industry-based approach (Porter 1985), but his arguments were subsequently translated into an urban and regional setting. In terms of urban and regional issues, Porter (1990) argues that clustering may act as an alternative organizational form to the standard markets and hierarchies dichotomy associated with the work of Williamson (1975). He argues that clustering provides individual firms with another way of organizing their transactions in an environment of rapidly changing information and technology. This particular form of spatial industrial organization maximizes the transfer of technology, knowledge, and information flows between the firms, and is particularly important in the case of small firms which rely mainly on external sources of information and technology. However, the key point of Porter’s argument is that proximity also engenders mutual visibility between competitors. In other words, firms are able to observe the competitive developments of each other, and this visibility itself acts as a spur to all firms to continue to improve their own individual competitiveness. The proximity of many suppliers, customers, and related institutions or organizations also helps to create the cluster, but Porter’s (1990) crucial point is that transparency drives competition and competition drives innovation.

Innovation is not invention. Innovation is the translating of new ideas, new designs, new inventions, and new concepts into new marketable and commercially viable products and services. The three fundamental features of all innovations are newness, improvement, and reduction of risk (Gordon and McCann 2005). Firms attempt to overcome long-term risk by trying to develop monopoly positions over

BOX 2.1 Continued
competitors, and this is the motivation for innovation. In order to do this they strive to develop novel products or services which are commercially viable, but the determination as to whether a new product or service is commercially viable depends on whether the customers in the market perceive the product or service to be an improvement on existing offerings, given the prices being charged. As such, it is not only novelty that drives growth, but the flows of knowledge between potential suppliers and customers. The mutual transparency associated with clustering fosters both the striving for novelty and the generation of the requisite knowledge flows for market selection processes to operate efficiently. The Porter (1990) approach closely relates to the Alchian (1950) adapting-adopting issues discussed in Section 2.6. However, the Porter (1990) arguments represent a much more dynamic approach, emphasizing that localized competition drives innovation, which in turn increases the competitiveness and growth of the cluster as a whole. Porter’s arguments are rather different in origin to many of the other urban and regional economic approaches examined in this book, in that he makes almost no assumptions about the nature of the representative firm, the properties of market equilibrium conditions, or the properties of production factors such as diminishing marginal returns and factor substitutability. His arguments derive primarily from numerous case studies of the behaviour of firms and industries.

The new industrial areas model

The new industrial areas model, or new industrial spaces model, derives primarily from a series of observations within the field of urban planning and geography (Scott, 1968). Certain industrial clusters, such as the electronics cluster in Silicon Valley (Saxenian, 1994), the electronics and biotechnology cluster in Cambridge, UK (Keeble and Wilkinson, 1999), and the small-firm manufacturing industry of the Emilia-Romagna region of Italy, have shown themselves to be major centres of innovation (Breschi and Malerba, 2005; Becattini et al., 2009). These observations have led to suggestions from many observers that industries which are made up of spatial networks or clusters of small firms tend to be more highly innovative than industries comprising mainly large firms (Saxenian, 1994), because such environments provide the appropriate ‘milieux’ for such innovations to take place (Aydalot and Keeble, 1988). In particular, as well as Marshall’s classic sources of agglomeration economies, it is also argued that networks of important business relationships operate between local decision-makers which allow for risk-taking. These relationships are perceived to depend on strong mutual trust between the participants, and are assumed to have developed partly due to geographical proximity. The reasons why these arrangements are argued to work is that the people trust each other not to engage in opportunistic behaviour, but rather to lean towards cooperative types of relationship. There is clear evidence in favour of these arguments from a small number of European countries (Becattini et al., 2009). At the same time, however, there is little or no consensus among observers concerning the applicability of these arguments to other countries or regions (Suarez-Villa and Walrod, 1997; Simmie, 1998; Arita and McCann, 2000). What is evident is that the relationships between firm size, innovation, and industrial clustering highlighted by this literature have led to a great deal of interest among public policy planners. The primary motive for this is the belief that understanding this relationship may improve public policy initiatives to foster such developments elsewhere (Castells and Hall, 1994).

2.4 Clusters, firm types, and the nature of transactions

A problem with discussing the nature of, and reasons for, industrial clustering is that it is very difficult to distinguish between each of the above reasons for, and descriptions of, industrial clustering. From an empirical point of view, distinguishing between urbanization economies and localization economies is notoriously difficult (Glaeser et al., 1992; Henderson et al., 1995), because many industrial clusters such as cities
will contain all types of clusters both within and across a range of sectors. From a theoretical perspective, the Marshall description is important in allowing us to understand the sources of agglomeration economies within an individual industry sector, while the Hoover approach is important from the point of view of identifying the particular firms and sectors which experience these agglomeration economies. However, it may appear that the distinction between the three Hoover classifications is rather arbitrary, given that mergers and acquisitions mean that firms are frequently changing ownership and sectors. Moreover, the relationship between the Marshall and Hoover classifications may be further complicated by each of the issues raised by the growth pole, the incubator, the product cycle, the Porter, and the new industrial area models. Yet there are clear differences between the types of industrial clusters which exist in reality, according to the nature of relations between the firms within the individual clusters and the particular features they exhibit (Gordon and McCann 2000).

In order to understand these relations, it is necessary for us to focus on the characteristics of firms which exist in the cluster, and the transactions which take place within the cluster. Adopting this approach, we see that there are three broad typologies of spatial industrial clusters, as defined in terms of the features they exhibit (Gordon and McCann 2000). These are the pure agglomeration, the industrial complex, and the social network. The key feature which distinguishes each of these different ideal types of spatial industrial cluster is the nature of the relations between the firms within the cluster. The characteristics of each of the cluster types are listed in Table 2.1, and as we see, the three ideal types of clusters are all quite different. In reality, all spatial clusters will contain characteristics of one or more of these ideal types, although one type will tend to be dominant in each cluster. Therefore, from an empirical point of view or from a public policy perspective which seeks to influence the behaviour of the cluster, it is necessary to determine which of these particular ideal types of industrial cluster most accurately reflects the characteristics and behaviour of any particular cluster.

Table 2.1 Industrial clusters

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Pure agglomeration</th>
<th>Industrial complex</th>
<th>Social network</th>
</tr>
</thead>
<tbody>
<tr>
<td>firm size</td>
<td>atomistic</td>
<td>some firms are large</td>
<td>variable</td>
</tr>
<tr>
<td>characteristics of</td>
<td>non-identifiable</td>
<td>identifiable</td>
<td>trust</td>
</tr>
<tr>
<td>relations</td>
<td>fragmented</td>
<td>stable trading</td>
<td>loyalty</td>
</tr>
<tr>
<td></td>
<td>unstable</td>
<td></td>
<td>joint lobbying</td>
</tr>
<tr>
<td>membership</td>
<td>open</td>
<td>closed</td>
<td>joint ventures</td>
</tr>
<tr>
<td>access to cluster</td>
<td>rental payments</td>
<td>internal investment</td>
<td>non-opportunistic</td>
</tr>
<tr>
<td></td>
<td>location necessary</td>
<td>location necessary</td>
<td>partially open</td>
</tr>
<tr>
<td>space outcomes</td>
<td>rent appreciation</td>
<td>no effect on rents</td>
<td>partial rental capitalisation</td>
</tr>
<tr>
<td>notion of space</td>
<td>Urban</td>
<td>local but not urban</td>
<td>local but not urban</td>
</tr>
<tr>
<td>example of cluster</td>
<td>competitive urban</td>
<td>steel or chemicals</td>
<td>new industrial areas</td>
</tr>
<tr>
<td></td>
<td>economy</td>
<td>production complex</td>
<td></td>
</tr>
<tr>
<td>analytical approaches</td>
<td>models of pure</td>
<td>location-production theory</td>
<td>social network theory (Granovetter)</td>
</tr>
<tr>
<td></td>
<td>agglomeration</td>
<td>input–output analysis</td>
<td></td>
</tr>
</tbody>
</table>
In the model of pure agglomeration, inter-firm relations are inherently transient. Firms are essentially atomistic, in the sense of having no market power, and they will continuously change their relations with other firms and customers in response to market arbitrage opportunities, thereby leading to intense local competition. As such, there is no loyalty between firms, nor are any particular relations long term. The external benefits of clustering accrue to all local firms simply by reason of their local presence, the price of which is the local real-estate market rent. There are no free-riders, access to the cluster is open, and consequently it is the growth in the local real-estate rents which is the indicator of the cluster's performance. This idealized type is best represented by the Marshall model, and the localization and urbanization economies of Hoover, but also contains elements of the Porter, Chinitz, and Vernon models. The notion of space in these models is essentially urban space in that this type of clustering only exists within individual cities.

The industrial complex is characterized primarily by long-term stable and predictable relations between the firms in the cluster. This type of cluster is most commonly observed in industries such as steel and chemicals, and is the type of spatial cluster typically discussed by classical (Weber 1909) and neoclassical (Moses 1958) location-production models, representing a fusion of locational analysis with input-output analysis (Isard and Kuenne 1953). Component firms within the spatial grouping each undertake significant long-term investments, particularly in terms of physical capital and local real estate, in order to become part of the grouping. Access to the group is therefore severely restricted both by high entry and exit costs, and the rationale for spatial clustering in these types of industries is that proximity is required primarily in order to minimize inter-firm transport transactions costs. Rental appreciation is not a feature of the cluster, because the land which has already been purchased by the firms is not for sale. This ideal type of cluster more closely reflects the internal returns to scale argument of Hoover and aspects of the growth pole model of Perroux than the other cluster types. The notion of space in the industrial complex is local, but not necessarily urban, in that these types of complexes can exist either within or outside an individual city.

The third type of spatial industrial cluster is the social network model. This is associated primarily with the work of Granovetter (1973, 1985, 1991, 1992), and is a response to the hierarchies model of Williamson (1975). The social network model argues that mutual trust relations between key decision-making agents in different organizations may be at least as important as decision-making hierarchies within individual organizations. These trust relations will be manifested by a variety of features, such as joint lobbying, joint ventures, informal alliances, and reciprocal arrangements regarding trading relationships. However, the key feature of such trust relations is an absence of opportunism, in that individual firms will not fear reprisals after any reorganization of inter-firm relations. Inter-firm cooperative relations may therefore differ significantly from the organizational boundaries associated with individual firms, and these relations may be continually reconstituted. All these behavioural features rely on a common culture of mutual trust, the development of which depends largely on a shared history and experience of the decision-making agents. This social network model is essentially aspatial, but from the point of view of geography, it can be argued that spatial proximity will tend to foster such trust relations, thereby leading to a local business environment of confidence, risk-taking, and cooperation. Spatial proximity is necessary but not sufficient to acquire access to the network. As such, membership of the network
is only partially open, in that local rental payments will not guarantee access, although they will improve the chances of access. The social network model therefore contains elements of both the Porter model (1990, 1998) and the new industrial areas model (Scott 1988), and has been employed to describe the characteristics and performance of areas such as Silicon Valley and the Emilia-Romagna area of Italy. In this model space is once again local, but not necessarily urban.

When we consider the fundamentally different nature of each of these cluster types, there are two major issues which immediately arise. Firstly, there is the question of whether there are any risks associated with locating in a cluster which might deter firms from doing so, and secondly, there is the question as to how to identify empirically the dominant features and logic of the cluster.

On the first point regarding the potential risks associated with locating in a cluster, the major issues here relate to the risks of experiencing what are known as unintended knowledge outflows (Grindley and Teece 1997). From the agglomeration arguments we know that one of the reasons why firms might wish to co-locate is to benefit from knowledge spillovers. From the perspective of the individual firms the positive and beneficial aspects of knowledge spillovers are knowledge inflows from other firms. However, where firms undertaking R&D have privately developed novel ideas or techniques on which they wish to build new products or services, in general their wish will be to keep such knowledge secret until it can be protected by patents, licences, copyrights, and the like. Such firms will want to avoid unintended knowledge spillovers to other firms, or in other words unintended knowledge outflows. Otherwise the potential competitive advantages afforded by this proprietary knowledge and proprietary assets will be lost. Tight firm boundaries and long-term legal contracts are the standard technique for avoiding such unintended knowledge outflows, exactly in accordance with the markets and hierarchies argument of Williamson (1975) underpinning the organizational logic of the industrial complex. The industrial complex model therefore allows for large-scale R&D to be undertaken with the risk of such unintended knowledge outflows occurring. In the case of the social network model the risks associated with unintended knowledge outflows are controlled by the shared norms and values of the network, the foundations of which are built on a code of non-opportunism. In the case of the pure agglomeration model, there are really no means by which such unintended knowledge outflows can be limited. The point about this argument is that it is not automatic that firms will wish to cluster together in space (see Urban and Regional Example 2.2).

What we see is that whether firms are clustered together depends on the knowledge advantages of clustering versus the knowledge disadvantages of clustering, and these issues are just as pertinent to small companies as they are to large companies. The only proprietary asset that many small companies have is a novel idea, product, or technology. For these firms the risks of clustering can be very real because if knowledge about the novel idea, product, or technology unintentionally leaks out to competing firms, the raison d'être of the firm disappears. However, these firms are often too small to fully pursue their idea alone and therefore they must take on board the risks associated with sharing knowledge and information with potential suppliers, customers, or collaborators, all of whom could become competitors. As such, the risks of clustering are not just limited to large firms, but also arise for small firms. What becomes clear is that while there are advantages associated
Urban and Regional Example 2.2  To cluster or not to cluster?

Whether a firm wishes to co-locate with other firms depends on the balance between the positive benefits from received knowledge inflows, which are a public good, and the potential losses associated with unintended knowledge outflows. In markets where product life cycles are very short, such as in international finance or in some of the Silicon Valley components of the semiconductor industry (Saxenian 1994), for many firms the only way to acquire the relevant market knowledge is by very frequent face-to-face contact, which can only be afforded by proximity (McCann 2005). Where the need for such high-frequency interactions is less, firms have much more flexibility to move away from each other if they wish.

Where secrecy is absolutely paramount due to the enormous costs of R&D, as is the case in pharmaceuticals or in many non-Silicon Valley parts of the semiconductor industry (McCann and Arita 2006), a very different location approach is often adopted by firms. In some cases firms will choose to pursue internal R&D within a closed environment while developing little or no linkages with the surrounding regional economy, or even in effect becoming ‘islands’ of innovation (Simmie 1988). Alternatively, firms may even choose to move away from other firms in the same industry. In both cases, the aim of the firm is to remove the risks of unintended knowledge outflows, as these are regarded as being too costly relative to the benefits of knowledge inflows.

with agglomeration, as outlined above, there are also disadvantages associated with clustering that go well beyond the issues of congestion or rising land prices.

On the second point regarding empirical observation, as we have already noted, all industrial clusters will contain features of one or more of these ideal types. However, determining the major features of any particular cluster will require us to consider which of these particular types is dominant. The reason is that empirically each of these cluster types will have different manifestations. For example, in the pure model of agglomeration, the dominant feature will be an appreciation in local real-estate values, with no particular purchasing or decision-making linkages being evident between local firms. Discussions of the strength of local purchasing or decision-making linkages, or the types of local alliances undertaken by firms, will not indicate agglomeration behaviour. On the other hand, in the case of the industrial complex there will be no real-estate effects, but the dominant feature will be a stability of both purchasing and decision-making linkages. However, measuring the scale of such linkages is less important than measuring the duration of the linkages. Finally, in the network model, although the market environment will be highly competitive, the dominant feature of the firm relations will be a willingness to undertake a variety of informal collective and cooperative activities which cut across organizational boundaries, such as joint lobbying, or inter-firm credit availability. Measuring real-estate appreciation or the strength of local purchasing linkages will not tell us very much. Industrial clusters are therefore of a variety of types, the observation and measurement of which is also a complicated topic.

Linking these individual types of clusters to economic growth and development is also complex. It is clear from the data that both innovation and entrepreneurship are related to agglomeration (Acs 2002; Van Oort 2004). However, the arguments here also imply that innovation may just as easily be related to industrial complexes and social networks as to agglomeration, because each of these three broad types represents a different way of solving
the problems related to the generation, exchange, acquisition, and exploitation of knowledge. In order to examine how these cluster types are related to innovation processes, Iammarino and McCann (2006) develop a cluster-innovation taxonomy which explicitly links a slightly extended version of the Gordon and McCann (2000) cluster typologies outlined here to a famous innovation-system classification scheme known as the Pavitt (1984) taxonomy. While the detailed issues arising from the Iammarino and McCann (2006) approach are beyond the scope of this book, the important point to note here is that the different forms of clusters are all related to innovation, but the ways in which we can empirically observe these relationships are very different. Simple observations of the spatial clustering of firms cannot be taken at face value as evidence of agglomeration, because their reasons for locating in the same place may have little or nothing to do with knowledge spillovers (McCann 1995a). In order to identify the relationships between firm location behaviour, knowledge, and innovation, it is nearly always necessary to complement the use of secondary empirical data sources with the use of primary survey and case-study type evidence on the nature of the industrial transactions and interrelationships operating between the firms, in a manner akin to the Porter (1990) approach.

2.5 People clustering: creativity and urban consumption

As well as the various industrial clustering arguments described here, over recent years two other closely related explanations of the advantages of urban clustering have also emerged. However, these explanations are somewhat different from those described so far, because they focus on the advantages of people clustering in urban areas, rather than firm clustering, and these explanations are the creative class hypothesis (Florida 2002, 2005) and the consumer city hypothesis (Glaeser et al. 2001; Glaeser and Gottlieb 2006).

The creative class hypothesis was originally developed by Richard Florida (2002, 2005), who sought to explain how and why clusters of certain types of people, rather than firms, were so important for regional growth. Florida’s argument in its original form (Florida 2002) examined the performance of places populated with diverse and often rather unconventional and unorthodox communities, and his initial observations (Florida 2002) centred on places such as San Francisco, a city with a long history of social innovation and social tolerance. The central tenet of the creativity hypothesis is that places which are tolerant of cultural diversity and cultural differences are also environments which are ideally suited for fostering unconventional approaches to the development of novel ideas, systems, products, or services. This unconventionality itself is argued to enable and encourage experimentation, innovation, and entrepreneurial behaviour of all different forms. Florida (2002) argues that the natural outcome of this culture of tolerance is an environment of creativity, which is manifested in the form of various types of creative outputs in art, theatre, and media, as well as in terms of novel business concepts and new technology and investment ideas. As such, his original argument is that in these types of environments artistic creativity will also be associated with commercial dynamism, and the creativity of the community will therefore be manifested in different ways.

Florida also argued (Florida 2002, 2005) that people with these requisite creative personalities and skills will increasingly move to culturally tolerant places, in search of both
creative employment and commercial opportunities but also in search of a particular lifestyle, which is tolerant of unconventionality. The in-migration of creative people also drives further innovation in the tolerant and creative regions, both via local competition effects and also via innovation-creation effects. In contrast, regions that are intolerant of cultural diversity will increasingly lose these creative people, and consequently they will also lose the commercial dynamism associated with these people.

In many ways Florida’s arguments echo some of the explanations for the success of the Dutch Republic during the seventeenth-century ‘Golden Age’, which posit that the flowering burst of economic ingenuity in an environment of religious tolerance allowed for inflows of Jews, Huguenots, and Catholics seeking refuge from religious intolerance in other countries. Even explanations for the nineteenth-century growth of some of the US coast cities contain some of this logic. Moreover, it will become clear that the Florida arguments are also in some ways related to the bridging and bonding (Putnam 1996) dimensions of social capital, as discussed in Chapter 7. Indeed, as we will see in Chapter 7, some authors actually refer to creative capital as a distinct entity in its own right.

A related argument is that of the consumer city hypothesis (Glaeser et al. 2001; Glaeser and Gottlieb 2006), which posits that high-skilled and high-income people will increasingly migrate towards cities offering high-quality amenities, such as opera houses, theatres, museums, culinary outputs, and so on. This argument conceives of the modern city as a place of leisure as well as work, where high-income workers consume highly income-elastic goods and services produced in these particular types of urban environments. This argument is related to the amenity-migration arguments discussed in Chapter 6, although the emphasis of the amenity-migration models is on the importance of natural environmental amenities while the consumer city hypothesis here emphasizes the importance of human-produced urban amenities. The consumer city hypothesis is also closely related to the urban gentrification arguments discussed in Chapters 4 and 10.

A point of overlap between these two models comes from the fact that artistic and culinary outputs produced by creative people concentrated in urban areas are exactly the types of highly income-elastic human-produced goods consumed by high-income workers in urban areas. Various papers (Comunian et al. 2010; Abreu et al. 2012) have demonstrated recently that the average wages earned by creative workers working in artistic activities tend to be lower than those of equivalently skilled workers working in other activities. In part this reflects a more skewed wage distribution among creative-artistic occupations than other activities, but, even allowing for this, it appears that creative workers are willing to accept lower wages, presumably because some of the rewards to their labour are in the form of job satisfaction. However, this observation suggests that the major beneficiaries of these creative processes are not the creative workers themselves, but the high-income consumers working in the city and enjoying the high-quality creative services and amenities on offer in the city, exactly as argued by the consumer city hypothesis.

Our understanding of these issues is still developing and there are both supporting and dissenting voices (Caves 2000; Scott 2000; Markusen 2006) on these matters. What is important for our purposes, however, is to recognize that urban clustering is not simply about firms, nor where it relates to labour is it purely about matching. It also appears to be related to interactions of people, and in particular to social aspects of interactions, which facilitate the generation and the spreading of new ideas.
2.6 Limited information, uncertainty, and the evolution of clusters

The models discussed in Chapter 1 and also in the earlier parts of this chapter rely largely on the assumption that firms and individuals are basically rational, and the behaviour of firms and individuals predicated on this rationality is sufficient to generate clusters. Yet this argument itself has three implicit dimensions. First, it is assumed that firms and individuals either know, or the price mechanism will quickly reveal, the locations where the potential profitability advantages are greater for their particular activities. Second, it is assumed that firms and individuals will act on this information and use their location behaviour in order to maximize their profits. Third, the cumulative effect of such individual choices is that agglomerations and clusters will naturally emerge.

In reality, however, the information available to firms and individuals is often rather limited, and different firms will often have different information available to them. For these reasons, some commentators have argued that firms cannot and do not make rational decisions in order to maximize their profits. Rather, they argue that firms make decisions in order to achieve goals other than simply profit maximization. Therefore, from the perspective of both individual and aggregate group location behaviour, this critique might suggest that the underlying motivation of our location, clustering, and agglomeration models would need to be reconsidered. In particular, the ways in which agglomerations and clusters arise in environments of uncertainty and limited information may need to be reconsidered. This critique, which is important to understand, has three themes: bounded rationality, conflicting goals, and relocation costs. The first two themes discussed in Box 2.2 can be grouped

**Box 2.2 Behavioural theories and bounded rationality**

Where firms face limited information, Baumol (1959) has argued that they will focus on sales revenue maximization as the short-run objective of their decision-making. One reason for this is that sales revenue maximization implies the maximum market share for the firm in the short run. Where information is limited, current market share is deemed by many observers to be the best indicator of a firm's long-run performance, because it provides a measure of the monopoly power of the firm. The logic of this approach is that the greater the market share of the firm, the greater the current monopoly power of the firm, and the greater will be the firm's long-run ability to deter potential competitors through defensive tactics such as limit-pricing and cross-subsidizing. From the perspective of location models, this may imply that the firm will make location decisions primarily in order to ensure maximum sales revenue rather than maximum profits. In the Hotelling model described in Chapter 1, these two objectives coincide. However, if the costs of production or transportation faced by the firm were to vary with location along the line OL in Figure 1.17, as they can do in the Weber and Moses type models, the two objectives of sales maximization and profit maximization may not coincide at the same location point in the Hotelling model. The eventual location result will therefore depend on which particular performance measure the firm adopts and chooses to maximize.

The second critique of profit maximization as the decision-making goal of the firm is that of 'conflicting goals'. This critique is most closely associated with the work of Cyert and March (1963). The argument
here is that in a world of imperfect information, the separation of ownership from decision-making in most major modern firms means that business objectives are frequently pursued which are different from simply profit maximization. Only shareholders have a desire for maximum profits in the short run. On the other hand, in modern multi-activity, multi-level, multi-plant and multi-national firm organizations, corporate decisions are the result of the many individual decisions made by a complex hierarchy of people, each with particular business objectives, and many of which are different from profit maximization. The reason is that the performance of different employees within a company is measured in different ways. For example, the directors' performance may be evaluated primarily by the firm's market share, whereas the sales manager's performance may be evaluated by sales growth. Similarly, the production manager's performance may be evaluated primarily by inventory throughput efficiency, whereas the personnel officer may be evaluated according to the number of days lost through industrial disputes. Given that each of these different decision-makers is evaluated on different criteria, the success, promotion, and consequently the wages earned by each of these workers will be evaluated differently. Therefore the objectives pursued by different employees may be quite different from profit maximization. Under these conditions, the 'conflicting goals' critique suggests that firms will aim to 'satisfice.' In other words, the firms will aim to achieve a satisfactory level of performance across a range of measures. In particular, the firm will initially aim to achieve a level of profit sufficient to avoid both shareholder interference in directors' activities and also to avoid the threat of a takeover. Once this objective is achieved, the other various goals of the firm can be satisfied. For example, the firm may aim to achieve market share levels as high as possible without jeopardizing the efficiency cost gains associated with production and logistics operations. Equally, employees' pay may be increased in order to encourage firm loyalty. The point here is that the overall objective of the firm can be specified in various ways.

In Figure 2.1, the firm's total profit function $\Pi$ is constructed as the difference between the total revenue function $TR$ and the total cost function $TC$. The firm may choose to produce at output levels $Q_1$, $Q_2$, and $Q_3$, which represent the minimum cost output, the maximum profit output, and the maximum revenue output levels, respectively. All these levels of output produce a profit level sufficient to maintain a firm's independence $\pi$, but only one of these output levels $Q_2$ is the profit maximization level of output.

![Figure 2.1 Profit maximizing, revenue maximizing, and profit satisfying](image)

Figure 2.1 Profit maximizing, revenue maximizing, and profit satisfying.
under the general heading of Behavioural Theories, and these arguments are not fundamentally geographical in nature, nor were they originally directed at location models in particular. In contrast, the third theme is essentially an explicitly spatial question.

The arguments concerning 'bounded rationality' are most closely associated with Simon (1952, 1959). This critique concerns the fact that firms in the real world face limited information, and this limited information itself limits firms' ability to be 'rational' in the sense assumed in microeconomics textbooks. These arguments are a more general critique of rationality within microeconomics as a whole. However, they have been argued to be particularly appropriate to the question of industrial location behaviour. The reason is that information concerning space and location is very limited, due to the inherent heterogeneity of land, real estate, and local economic environments. Therefore, when considering location issues, it would appear that the ability of the firm to be 'rational' is very much 'bounded' by the limited information available to it. In these circumstances, decisions guided by straightforward profit maximization behaviour appear to be beyond the ability of the firm. Therefore location models based on this assumption seem to oversimplify the location issue. Location behaviour may be determined primarily by other objectives than simply profit maximization, as discussed in Box 2.2.

The behavioural arguments imply that firms do not necessarily have the ability or the desire to make decisions which are explicitly aimed at maximizing short-term profits. Applying this argument to geography means that if we are faced with a set of spatial total cost and revenue curves, such as those described by Figure 2.2, the firm will make different location decisions, according to whether it is aiming to maximize profits in the short run or whether it is aiming to earn satisfactory profits in the short run along with achieving some other goals. For example, in Figure 2.2, if the firm is aiming to maximize profits in the short run it will locate at point $P$. On the other hand, if it is aiming to maximize sales it will locate at $S$, and if it is aiming to minimize production costs and to maximize production efficiency it will locate at $C$. If the firm had perfect information regarding these different spatial cost and revenue curves we can argue that the firm would always move to point $P$. However, behavioural theories assume that information is imperfect. Given the limited information

Figure 2.2 Spatial cost and revenue curves
available and the conflicting goals within the organization, the actual location behaviour of the firm will depend on which is its particular dominant objective.

The third critique of the classical and neo classical location models comes from the question of relocation costs. Relocation costs are the costs incurred every time a firm relocates. The models described above all assume that location is a costless exercise. However, relocation costs can be very significant, comprising the costs of the real-estate site search and acquisition, the dismantling, moving, and reconstruction of existing facilities, the construction of new facilities, and the hiring and training of the new labour employed. These significant transactions costs, along with imperfect information and conflicting goals, will mean that firms are unlikely to move in response to small variations in factor prices or market revenues. In Figure 2.2, the areas in which positive profits are made, i.e. where $TR > TC$, are known as 'spatial margins of profitability' (Rawstron 1958), and are represented by the areas between locations a and b, c and d, and e and f. The relationship between marginal location change and the profitability of the firm in these areas is given by $\frac{\partial (TR - TC)}{\partial d}$, and this is represented by the differences in the slopes of the spatial revenue and spatial cost functions as location changes. In the spatial margins of profitability in which the slopes of the spatial revenue and spatial cost functions are very shallow, the marginal benefit to the firm of relocation will be very low. Therefore, in the presence of high relocation costs the firm will not move to a superior location even if it knows which alternative is superior. In conditions of imperfect information and bounded rationality, conflicting goals, and significant relocation costs, the behavioural approach would argue that once a firm has chosen a location, it will tend to maintain its location as long as profits are positive, and not use relocation as a competitive weapon. Rather the firm will attempt to reorganize its factor allocations and activities among its current set of existing plants. At the same time, the firm will focus primarily on other price and non-price issues as competitive weapons, and the relocation of a plant, or the reorganization of multi-plant activities which involves either the closing or opening of a plant, will only be a last resort strategy. On the other hand, where relocation costs are insignificant, the firm will take advantage of spatial revenue and spatial cost differences and will be able to move to superior locations as a competitive strategy.

One obvious weakness of the behavioural critiques is that, unlike the classical and neo classical location models discussed in Chapter 1, the behavioural theories do not of themselves indicate why a firm chooses a particular location in the first place. In this sense the behavioural approach is not prescriptive. However, the applicability of the location behaviour insights offered by the classical models and theories to real-world situations does need to be interpreted in the light of the behavioural critique, because uncertainty, bounded rationality, imperfect information, conflicting goals, and relocation costs are all features particularly characteristic of the spatial economy. Indeed, as a whole, the application of the behavioural critique to spatial behaviour suggests that observed spatial patterns are not necessarily reflective of optimum location behaviour, but rather sub-optimal adjustments to restricted alternatives. This critique provides an explanation as to why firm location behaviour may not be so responsive to the available optimization possibilities, and why aggregate spatial investment patterns in general may be very slow to adjust to the emerging profitability opportunities. This obviously also puts into question the validity of many of our models which assume that clusters and agglomerations will arise naturally in response to price signals. This is important because we implicitly assume that firms and individuals
will voluntarily move and congregate where agglomeration advantages are on offer. Indeed, if the behavioural arguments provide a powerful critique of the ability of firms in the real world to act rationally in terms of location behaviour, then it would also appear *prima facie* to question our assumptions as to how agglomerations and clusters arise in the first place.

These difficulties can be overcome by considering the evolutionary argument of Alchian (1950). Alchian’s argument is that the behaviour of firms in conditions of uncertainty can be understood by discussing the relationship between a firm and its environment, whereby a firm’s environment is understood to encompass all the agents, information, and institutions competing and collaborating in the particular set of markets in which the firm operates. In Alchian’s argument, we can characterize the uncertain economy by two broad types of environments. One is an ‘adoptive’ environment and the other is an ‘adaptive’ environment. These two classifications are not mutually exclusive, but serve as the two extreme stylized types, between which the real economy will exist.

In the ‘adoptive’ environment, all firms are more or less identical in that no firm has any particular or systematic information advantage over any other firm. However, they will have differences in their ability to survive, but these differences are often unrevealed or unknown *ex ante*. The results of the competitive process will imply *ex post* that some firms will be successful while others will not, although *ex ante*, no firms had any *a priori* knowledge that their products or techniques would be superior to those of their competitors. This characterization of the economy is Darwinian, in that the environment ‘adopts’ the firms which were better suited to the needs of the economy, even though the firms had no particular knowledge beforehand that this was the case. In statistical terms, in any given time period in the ‘adoptive’ environment, the probability of a particular single firm making a successful strategic decision is identical to that of all the other individual firms.

On the other hand, in the ‘adaptive’ environment, some individual firms are able to gather and analyse market information, simply by reason of their size. Large firms in general are able to utilize resources in order acquire and process information relating to their market environment, and the purpose of these information-gathering activities by the firms is to subsequently use the information to their own advantage, relative to their competitors. In statistical terms, therefore, in any given time period in the ‘adaptive’ environment, the probability of a particular firm making a profitable strategic decision is increased by reason of its size.

In the real world of heterogeneous firms and imperfect information, smaller firms will tend to perceive themselves to be at an information disadvantage relative to larger firms. Therefore they will tend to make decisions which mimic or dovetail with those of the larger firms, in matters such as styles, protocols, formats, and technology. In part this is because they perceive the market leaders to be the best barometers of market conditions, and also because the behaviour of the market leaders itself often contributes significantly to the overall economic environment simply by reason of their size. By copying the behaviour of the larger firms the small firms therefore perceive that they will maximize the likelihood of their own success. The result is that large firms tend to overcome uncertainty by information gathering and analysis, and small firms tend to overcome uncertainty by imitation.

This type of leader–follower behaviour is common in models of oligopoly and uncertainty. However, this behaviour is particularly pertinent to questions of location. In environments
of uncertainty, larger firms will generally have the information and financial resources to make more considered location decisions than small firms. Major firms will be able to make location decisions more akin to those described by the Weber, Moses, and Hotelling models of Chapter 1, given that they will generally have sufficient resources to evaluate the cost and revenue implications of their location choice. These large firms will attempt to make rational and optimal decisions, and the results of their location choices can be analysed by the types of classical and neo classical models described above. On the other hand, small firms will generally be located where their founders were initially resident. There will have been no explicit initial location decision as such when the firm began operating. Often such small firms are entrepreneurial start-ups, whereby founders who were previously working for large firms decide to set up a new business in a related field, in many cases selling goods and services back to the large firm for which they previously worked, as well as accessing new markets. The geographical distribution and also the technological profile of such 'spin-offs' therefore tends to closely mirror that of the established firms, and gives rise to the evolution (Arthur 1994; Boschma and Martin 2010) of localized clusters of small and large firms, often in related technological fields (Frenken et al. 2007). As we will see in Chapter 7, this technological relatedness is a fundamental aspect of regional growth.

Over time, increasing competition means that location will eventually become a decision-making issue even for small firms as they develop and grow. As such, in subsequent location decisions, many small firms will tend to choose to open new establishments close to other large firms located in different market areas. Similarly, there is much evidence to suggest that large multi-plant firms grow by means of the establishment of new ventures, in particular already well-established spatial concentrations of firms (Delgado et al. 2010). This itself favours further concentration. Therefore, for small firms which are risk averse, these clustering strategies are particularly good strategies, because as we see from the Hotelling model, locating close to competitor firms ensures that an individual firm's market share is no lower than that of an equivalent firm. Moreover, the Salop (1979) argument suggests that clustering acts as also partial defence against the instability associated with price movements. The clustering of small firms around major firms is therefore very commonly observed. As such, imitation also takes place in terms of the foundation or relocation of new establishments, as well as the location behaviour of established firms.

### 2.7 Conclusions

This question of industrial clustering is the topic of Chapter 3, in which we discuss agglomeration economies, the growth of cities and urban hierarchies, and centre-periphery relationships.

Many activities are clustered together in space, giving rise to the formation of cities. This process of industrial clustering, however, typically leads to an increased demand for local land and consequently local real-estate prices will tend to increase, as will local labour prices. These increases in the prices of local factor inputs will therefore reduce profits, *ceteris paribus*, thereby reducing the attractiveness of the area as a location for the firms, unless certain countervailing features exist that more than compensate for the increased local factor costs. These countervailing features are generally understood to be agglomeration economies.
Following on from Marshall's (1890) insights as to the sources of agglomeration and the Ohlin-Hoover approach to classifying agglomeration effects, various other related reasons for the growth of cities have been proposed by a range of authors. The common element of all of these approaches is that the generation, acquisition, and transfer of knowledge is a key component of all aspects of industrial clustering. However, as our transactions costs approach has also demonstrated, the actual mechanisms by which these are achieved differ between different types of cluster. Understanding the types of cluster is critical for identifying the knowledge spillover processes which are economized on by the formation of a cluster, and vice versa. Yet, while these various analyses explain why clusters may arise, they do not explain exactly how they may arise in an environment of limited information, risk, and uncertainty. Here, a discussion of the behavioural critique suggests that the leader-follower behaviour typical of many industries will tend to encourage small firms to cluster together in space close to larger firms. Such behaviour underpins an evolutionary process of cluster formation, in which clusters and cities are seen to arise naturally, even in the context of less-than-perfect information. That is not to say, however, that agglomeration processes are linear and indefinite. In contrast, we also observe not only that many activities are geographically dispersed, but also that many clusters and cities are also dispersed. Indeed, the spatial economy appears to be characterized by both geographical concentration and geographical dispersion, and the balance between these two features produces a system of cities, an urban hierarchy. It is to these issues that we turn in Chapter 3.

Discussion questions

2.1 What are the three major sources of agglomeration economies and how do they operate?
2.2 Explain the three major types of agglomeration effects. What difficulties are there in identifying these different classifications?
2.3 What other descriptions and mechanisms of industrial clustering do we have?
2.4 Using a transactions costs framework, explain the role and contribution of knowledge, uncertainty, and trust in different types of industrial clusters.
2.5 Which examples of industrial clusters from your country best reflect each of the different types of clusters you have discussed?
2.6 What role do creativity and consumption play in clustering?
2.7 What insights are provided for industrial location analysis by behavioural theories of firm behaviour?
2.8 Explain the ways in which evolutionary processes of adaptation and adoption to the competitive environment are important for firm location behaviour.