

CREA

Discussion Paper

2011-01

Center for Research in Economic Analysis
University of Luxembourg

Economic Geography

available online : http://www.wfr.uni.lu/recherche/fdef/crea/publications2/discussion_papers/2010

Jacques-François Thisse, Université du Luxembourg, Université Catholique de Louvain

January, 2011

For editorial correspondence, please contact : elisa.ferreira@uni.lu

Economic Geography*

Jacques-François Thisse
CORE, Université catholique de Louvain
CREA, Université du Luxembourg

1. Introduction

Economic geography – or geographical economics – aims to explain the existence of peaks and troughs in the spatial distributions of population and wealth. Ever since the emergence of civilization, human activities and standards of living have been unevenly distributed among both the continents and their territories. Economic life is concentrated in a fairly limited number of human settlements (cities and clusters), which are gathered under the heading of “economic agglomerations.” Furthermore, there are large and small agglomerations with very different combinations of firms and households.

The purpose of this chapter is to provide a bird-eye overview of the main contributions made by economists and regional scientists in understanding how the space-economy is organized. There is a wide agreement that the space-economy may be viewed as the outcome of a trade-off between different types of scale economies in production and the mobility costs of goods, people and information. Although it has been rediscovered many times (including in recent periods), this trade-off is at the heart of economic geography ever since the work of Lösch (1940). It implies that the location of economic activities is the result of a complicated balance of forces that push and pull consumers and firms in opposing directions. Yet, it was not until Krugman (1991a) that this idea became widely spread within the economics profession.

I want to stress from the outset the influence of three major scientists who epitomize the main questions raised in economic geography: Johann Heinrich von Thünen (1783-1850), Harold Hotelling (1885-1973) and Paul Krugman (1953-). Their contributions have paved the way to large flows of high-quality research. Thünen (1826) is the founding father of land use theory and his work as served as the corner-stone for the development of modern urban economics. Hotelling (1929) deals with a very different but equally fundamental issue, namely the nature of competition across space and the way firms choose their location in a strategic environment. Last, Krugman (1991b) has highlighted the microeconomic underpinnings of both spatial economic agglomerations and regional imbalances at national and international levels. He has achieved that by building a full-fledged general equilibrium model, which is able to explain why, how and when the economic activity may be agglomerated in a few places.

Ironically enough, none of those three authors is a “spatial economist” per se. They turned their attention to spatial issues for reasons that are not directly related to the location of economic activities. Thünen was interested in the allocation of resources and the determination of prices. He emphasized space because land was an essential production factor in the main sector of his time. Hotelling aimed to build a theory of product selection by oligopolistic firms. To achieve his goal, he used space as a metaphor. As for Krugman, he was mainly interested in the interplay between increasing returns and imperfect competition in globalizing markets, commodity trade and production factor mobility being the two fundamental ingredients.

A last warning is in order. Even though Krugman gave new life to the field of economic geography, many ideas and concepts have been around for a long time. However, they were fairly disparate and in search of a synthesis ranging from the small to the large, which Fujita and Thisse (2002) have

* This paper has been prepared for the *Handbook of the History of Economic Analysis* edited by Gilbert Faccarello and Heinz D. Kurz. I thank the editors, Luisito Bertinelli and Michel de Vroey for their comments and suggestions. I am also grateful to Masahisa Fujita with whom I have discussed and shared many of the ideas presented here.

endeavored to develop. To a large extent, the history of economic geography may be viewed as a process that has gradually unified various bodies of knowledge, as shown by the different names given to the field (regional and urban economics, location theory, and spatial economics), within a theoretical framework in which the focus has shifted from perfect competition to imperfect competition and various types of market failures. In what follows, I will only discuss the fundamental contributions made in economic geography through the lenses of modern economic theory. This choice will lead me to put aside a wide range of contributions that have not passed the test of time. It is fair to say, however, that some of them - think of Weber (1909), Fetter (1924) or Predöhl (1928) - have contributed to the field through the impact they had on several different scholars.

2. Thünen and land use theory

2.1 The location of agricultural activities

(a) Thünen (1826) sought to explain the pattern of agricultural activities surrounding cities in preindustrial Germany. Each location in space is characterized by various factors such as soil conditions, relief, geographical position, and the like. Both land rent and land-use vary across locations depending on these characteristics. Among them, the most important for location theorists is the transport-cost differential over space. Whereas Ricardo concentrated on fertility differences in his explanation of the land rent, Thünen constructed a theory focusing on the transport-cost differentials across locations. To this end, he used a very simple and elegant setting in which space is represented by a plain on which land is homogeneous in all respects except for a market town in which all transactions regarding agricultural goods must occur. The location of the market town is supposed to be given and the reasons for its existence are left outside of the analysis. By allocating an acre of land near the town to some crop, the costs of delivering all other crops are indirectly affected as they are forced to be grown farther away. Hence, determining which crops to grow where is not an easy task. Though simple, this setting is rich enough to show how a competitive land market can structure the use of land across space by perfectly divisible activities.

The principles underlying his model are so general that Thünen can be considered the founder of marginalism (Samuelson, 1983). In addition, the importance of Thünen's analysis for the development of location theory is twofold in that space is considered as both an economic good and as the substratum for economic activities. In his framework, the allocation of land to the different economic activities is shown to emerge as the equilibrium outcome of a perfectly competitive land market. The assumption of a competitive land market can be justified on the ground that land in a small neighborhood of any location belonging to a continuous space is highly substitutable, thus making the competitive process for land very fierce. Very ingeniously, Thünen imagined a process in which each farmer makes an offer based on the surplus he can generate by using one unit of land available at any particular location. This has led him and his successors to develop the concept of bid rent function, which describes the maximum price an agent is willing to pay to occupy each location.

This approach is probably what makes Thünen's analysis of land use so original. In a sense, it rests on the idea that land at a particular location corresponds to a single commodity whose price cannot be obtained by the textbook interplay between a large number of sellers and buyers. Specifically, land at any point is allocated to an activity according to a bidding process in which the producer offering the highest bid secures the corresponding lot. A farmer's bid depends on the transportability of his output and the amount of land needed to produce one unit of the good. Land being allocated to the highest bidder, economic activities are distributed according to the famous pattern of concentric rings, each of them being specialized in one crop. The land rent decreases with distance from the market town at a rate which is constant in each ring and decreasing from one ring to the next.

(b) The model can be closed by assuming that all agricultural activities use land and labor while a manufactured good is produced in town by using labor alone, typically under the form of

craftsmanship. Such a specialization of tasks reflects the traditional division of labor between cities and the countryside. Workers are perfectly mobile and landlords reside in town; they all have identical preferences. The solution to such a general spatial equilibrium model, in which the real wage common to all workers as well as the prices of agricultural and manufactured goods are endogenous, has been obtained by Samuelson (1983).

Yet, despite his monumental contribution to economic thought, Thünen's ideas languished for several decades. Blaug (1985, ch.14) attributes the subject's neglect to Thünen's cumbersome style. Indeed, one had to wait for Launhardt (1885, ch.30) to have a formal treatment of his ideas in the special case of two crops. The first model with n crops is due to Dunn (1954), while Schweizer and Varaiya (1976) have provided the complete solution to the general model with a Leontief technology in which goods may be used both in the final and intermediate sectors. Whatever their use, goods are either shipped toward the marketplace or used locally.

It took even more time to explain how and when a market town, which imports agricultural goods from and exports manufactured goods to its rural hinterland, may emerge as an equilibrium outcome. More precisely, the key-question that has been at the heart of economic geography for decades may be stated as follows: what binds together manufacturing firms and workers within the city? Using the new economic geography framework discussed in Section 4, Fujita and Krugman (1995) have identified sufficient conditions for a monocentric economy to emerge as an equilibrium outcome. Specifically, when (1) the transport cost of the agricultural good is low relative to that of the manufactured good and (2) when the total population is small enough, all manufacturing firms, which operate under increasing returns, agglomerate within a single district together with their workers, while farmers are dispersed across the agricultural hinterland. When one of these two conditions does not hold, shipping produces from and to a single market town is so costly that several cities emerge and generate an urban system.

2.2 Urban economics

(a) In his doctoral dissertation, Alonso (1964) succeeded in extending Thünen's central concept of bid rent curves to an urban context in which a marketplace is replaced by a Central Business District (CBD). In this context, the only spatial characteristic of a location is its distance from the city center, while the land available for raising crops is now used for housing, plants, offices, and infrastructure. The main objective of urban economics is to explain the internal structure of cities, that is, how land is distributed among various activities and why cities have one or several CBDs. The basic concept of urban economics is the land market, which serves to allocate both economic agents and activities across space. Alonso (1964) and Mills (1967) may be considered as the founders of this field. Treading in these authors' footsteps, several economists and regional scientists have developed the model of the monocentric city. The main focus is on the households' trade-off between housing size and the accessibility to employment centers, typically the CBD. Ever since the 1970s, urban economics has advanced rapidly. The reason for this success is probably that the canonical model can take leverage on the competitive paradigm of economic theory.

In equilibrium, identical consumers establish themselves within the city so as to equalize utility across space. In such a state, no one has an incentive to change location, the land rent at a particular location being equal to the largest bid at that location. Building on this idea, urban economists have endeavored to explain the internal structure of cities, that is, how land is distributed across activities and economic agents around the central business district. Though very simple, the monocentric city model has produced a set of results consistent with the prominent features of cities. In particular, it explains the decrease in the urban land rent with distance away from the city center as well as the fall in the population density as one moves away from the center. The model also explains how the development of modern transportation means (cars and mass transportation) has generated both suburbanization and a flattening of urban population densities, an evolution known as urban sprawl. The today-best

synthesis of the results derived within the monocentric framework remains the landmark book of Fujita (1989).

(b) Very much as in the Thünen model which does not say why transactions take place in a given market town, the monocentric city model is silent on the reasons that would explain the existence of a district where jobs are available. So, we are left with the following question: why do city centers exist? Or, more generally, why do cities exist? This question has haunted economic geography for decades.

Beckmann (1976) views personal relations as the essence of societies, even though the consequences of relations are often double-edged. The propensity to interact with others has a gravitational nature in that its intensity increases with the number of people living in each location and decreases with the distance between two locations. Beckmann then focused on the trade-off between the desire of an individual to interact with others and her need to consume a large plot of land. Under such preferences, the spatial equilibrium exhibits a bell-shaped population density supported by a similarly shaped land rent curve, and thus face-to-face contact supports urbanization. This provides an illuminating explanation for the existence of cities, which combines the natural gregariousness of human beings together with their desire to consume more space.

Although very suggestive, this approach does not explain the existence of an employment center because firms are left aside. Thus, beyond the standard market transactions in which firms are involved, one may wonder what the interactions that would foster their concentration are. The reason here is very different from what Beckmann assumed in that it refers to the role of information as a basic input in firms' activities, a kind of information difficult to codify because it is tacit and which can be collected through face-to-face communications only. The exchange of information between firms generates externality-like benefits for each of them. Provided that firms own different pieces of information, the benefits of communication generally increase as the number of firms rises. The quality of the information is also better when firms are gathered in that the number of intermediates is smaller. Because communications typically involve distance-decay effects, the benefits are greater if firms locate within the same district.

The seminal contribution in this respect is due to Ogawa and Fujita (1980) who explored the implications of spillovers, the intensity of which is affected negatively by a distance-decay effect. Specifically, the agglomeration force finds its origin in the existence of the exchange of information which allow companies to learn from each other how to do things better. The transmission of tacit knowledge and information often requires face-to-face communication between agents, which typically involve distance-sensitive costs. Hence the benefits of information are larger when firms locate closer to each other. On the other hand, the clustering of many firms into a single area increases the average commuting distance for their workers which, in turn, leads to higher wages and land rent in the area surrounding the cluster. Such high wages and land rents tend to discourage the agglomeration of firms and acts as a dispersion force. Consequently, the equilibrium distributions of firms and households/workers are determined as the balance between these two opposite forces.

Ogawa and Fujita showed that high commuting costs lead to the completely-mixed configuration, that is, a pattern with no land specialization and no commuting. As commuting costs fall while the intensity of communication between firms rises (two fairly general trends observed since the development of the Industrial Revolution), one moves from backyard capitalism to a monocentric city with completely specialization of land. In other words, low commuting costs and/or strong spatial externalities foster the emergence of a monocentric city in which firms gather to form a CBD.

During the last two decades, urban economics has explored a much broader range of issues. It is worth mentioning here that modern models of neighboring effects and segregation originate in the pioneering work of Schelling (1969, 1971). Urban economics now has strong links to theories of social networks and other forms of local interactions, the urban neighborhood being the place where many nonmarket

relationships are developed (Durlauf, 2004). It is also closely related to new growth theories through the study of the microeconomic foundations of agglomeration economies (Duranton and Puga, 2004).

Despite such progresses, the most enduring problem in urban economics, that is, the existence of an urban hierarchy involving large and medium-sized cities as well as towns and villages, remains unsolved. Although Christaller (1933) has forcefully argued that the number of goods supplied in a city rises with its size, with the manufactured goods supplied in a low-rank city being also supplied in cities of higher rank, there is still no comprehensive microeconomic model explaining the urban hierarchy. So far, the most elegant proposal to describe how cities having different sizes emerge has been provided by Henderson (1974, 1988). In each city, there is again a tension between two forces. On the one hand, there are external economies associated with the agglomeration of firms at the city-center. On the other hand, there are diseconomies generated by the need to commute to the city-center. Hence, in equilibrium, each city has a well-defined size that depends on the type of firms it accommodates. As cities vary in their industrial mix, they have different sizes because industries differ in the external economies they are able to create. The setting remains incomplete, however. Cities are like floating islands because nothing is said about their relative locations. Furthermore, the model does not explain why a few cities are diversified. Last, Henderson's model is silent about the power law on city sizes.

3. The nature of competition in space

The debate about whether the general competitive equilibrium model is comprehensive enough to fully reflect the working of the spatial economy has a long history. Arrow and Debreu have made an interesting attempt to integrate space within general competitive analysis. Specifically, they assume that a commodity is defined not only by its physical characteristics but also by the place where it is made available. This implies that the same good traded at different places is treated as different economic commodities. In addition, the above approach integrates spatial interdependence across markets into general equilibrium in the same way as other forms of interdependence: location choices are contained in the specification of the production or consumption plans selected by firms and households. Hence, the Arrow-Debreu model seems to obviate the need for a space-specific theory of prices and markets. And, indeed, standard general equilibrium theory has proven to be very useful for the study of commodity flows across space provided that both firms and households have exogenously given locations. However, as highlighted by Krugman (1995), things become more problematic once agents are free to choose their locations.

3.1 The spatial impossibility theorem

(a) I begin the discussion by considering the assignment problems introduced by Koopmans and Beckmann (1957). Assume that n firms are to be assigned to n locations. Each firm is indivisible, and the amount of land available at each location is such that a single firm can be set up there. Hence, every firm must be assigned to a single location, and every location can accommodate only one firm. Each firm produces a fixed amount of goods and uses one unit of land. Suppose further that the technology used by each firm is not affected by the chosen location. Koopmans and Beckmann first considered the linear assignment problem in which firms receive revenues from the rest of the world, which are location-specific. They showed that this problem can be expressed as a linear program, the solution of which is given by integer numbers. Since the shadow prices generated by the dual of this program are location-specific, these prices have the nature of land rents. Thus, a competitive equilibrium exists since the optimal solution may be decentralized through a competitive land market, very much as in Thünen.

Koopmans and Beckmann then turned to the quadratic assignment problem in which each firm uses the goods produced by the others and bears the corresponding transportation costs. Because of the exchange of goods, this problem cannot be expressed as a linear program anymore. When locations

generate similar revenues, Koopmans and Beckmann showed that no feasible location pattern of firms can be sustained as a competitive equilibrium, thus implying that there exists no competitive equilibrium. Revisiting the quadratic assignment problem, Heffley (1972) showed that decentralization is possible when sites have very different comparative advantages. Hence, as Hamilton (1980, p.38) put it: "Stability is lent to the system by having plants differ from one another in their preferences for the sites *qua* sites, and instability arises from a large volume of trade among plants."

(b) In the long debate concerning the comprehensiveness of general equilibrium theory for the spatial economy, Starrett (1978) has made the fundamental contribution. The question is whether the competitive price mechanism is able to explain the endogenous formation of economic agglomerations and the existence of large trade flows. Because they are not perfectly divisible, agents are not ubiquitous and, therefore, must choose an "address." Space is then said to be homogeneous if (1) the utility function of each household is identical no matter what his location and (2) the production function of each firm is independent of its location. In other words, the location choice made by a consumer or a producer does not affect her preferences or the technologies that are available. The spatial impossibility theorem may then be stated as follows:

Consider an economy with a finite number of locations. If space is homogeneous, transport is costly, and preferences are locally non-satiated, then there exists no competitive equilibrium involving the transport of goods between locations.

Consequently, the perfectly competitive price mechanism alone is unable to deal simultaneously with cities and trade. This has a fundamental implication for economic geography: if the purpose is to build a theory explaining the formation of economic agglomerations, then such a theory must depart from general competitive analysis. What is the meaning of this result? Whenever economic activities are perfectly divisible, the spatial impossibility theorem implies that the mobility of production factors is a perfect substitute for trade. Such a result is hardly surprising because every activity can be carried out on an arbitrarily small scale in every possible place, without any loss of efficiency. Firms and households are then induced to suppress all distance-related costs by producing exactly what they need where they are. In contrast, as pointed out by Starrett (1978, p.27), "so long as there are some indivisibilities in the system (so that individual operations must take up space) then a sufficiently complicated set of interrelated activities will generate transport costs." In this case, the spatial impossibility theorem tells us something really new and important: whenever agents have to choose an address, there is no competitive equilibrium (hence the term "impossibility" in the name of the theorem) such that trade across space occurs. In other words, factor mobility and interregional trade are incompatible in the standard neoclassical world. This result is especially meaningful insofar as it is internal to the theory itself.

Intuitively, the reason for this is that the only location factor that matters to an agent is its position with respect to the others. In this case, the price system must play two different roles: (1) it must allow trade between locations while guaranteeing that all local markets clear and (2) it must give firms and households the incentives not to change location. Once the economy is competitive and space homogeneous, the spatial impossibility theorem tells us that it is impossible to kill two birds with one stone: prices that sustain commodity flows between places send incorrect signals from the point of view of the stability of locations, and vice versa. The fundamental reason for the spatial impossibility theorem is the non-convexity of the set of feasible allocations caused by positive trade costs and the fact that agents have an address in space, even though the individual land consumption is endogenous. Hence, absent external factors that drive firms' and households' locations, such as the existence of a market town or of spatial externalities, a sound spatial economic theory cannot be built within the competitive general equilibrium framework by differentiating goods through their locations and adding land as a new commodity.

3.2 Spatial competition theory

(a) Because consumers are dispersed across space, they differ in their access to the same firm. In such a context, firms anticipate accurately that each consumer will buy from the firm posting the lower full price, namely the price at the firm's gate, called mill price, augmented by the travel costs that consumers must bear to go to the firm they patronize. As a consequence, firms have some monopoly power on the consumers located in their vicinity, which enables them to choose their price. Of course, this choice is restricted by the possibility that consumers have to supply themselves from competing firms. This process of competition among spatially dispersed firms has been described by Launhardt (1885) who proposed a model of price formation, in which he anticipated the concept of Nash equilibrium. In particular, he was the first to show what became to be known as the Principle of Differentiation in industrial organization: "the improvement of means of transport is dangerous for costly goods: these lose the most effective protection of all tariff protections, namely that provided by bad roads." (p.150 of the English translation). In other words, firms want to be separated to relax price competition.

Launhardt's contribution remained ignored outside the German-speaking scientific community until recently. Hotelling (1929), who came up more than 40 years later, has had more impact although the path-breaking nature of his paper has been fully recognized when economists became aware of the power of non-cooperative game theory. The value and importance of Hotelling's contribution was brought to light in the 1980s by showing that its use exceeds the original geographical interpretation to accommodate various dimensions that differentiates firms and consumers. To be precise, the spatial framework may serve as a powerful metaphor for dealing with issues involving heterogeneity and diversity across agents in a host of economic, political and social domains. In addition, Hotelling's paper may be viewed as one of the prototypes of the modern economic literature: it is self-contained and focuses on a specific problem, which is studied by means of a simple and elegant model.

Because any single consumer is negligible to firms, Hotelling assumed that consumers are continuously distributed along a linear and bounded segment - think of Main Street. Two stores, aiming to maximize their respective profits, seek a location along the same segment. Each firm being aware that its price choice affects the consumer segment supplied by its rival, spatial competition is, therefore, inherently strategic. This is one of the main innovations introduced by Hotelling who uses a two-stage game to model the process of spatial competition: in the first stage, stores choose their location non-cooperatively; in the second, these locations being publicly observed, firms select their selling price. The use of a sequential procedure means that firms anticipate the consequences of their location choices on their subsequent choices of prices, thus conferring to the model an implicit dynamic structure. The game is solved by backward induction. For an arbitrary pair of locations, Hotelling starts by solving the price sub-game corresponding to the second stage. The resulting equilibrium prices are introduced into the profit functions, which then depend only upon the locations chosen by the firms. These functions stand for the payoffs that firms will maximize during the first stage of the game. Such an approach anticipates by several decades the concept of sub-game perfect Nash equilibrium introduced by Selten in the 1960s.

Hotelling's conclusion was that the process of spatial competition leads firms to agglomerate at the market center. If true, this provides us with a rationale for the observed spatial concentration of firms selling similar goods (e.g. restaurants, movie theaters, fashion cloth shops). Unfortunately, Hotelling's analysis was plagued by a mistake that invalidates his main conclusion: when firms are sufficiently close, the corresponding sub-game does not have a Nash equilibrium in pure strategies, so that the payoffs used by Hotelling in the first stage are wrong (d'Aspremont et al., 1979).

This negative conclusion has led d'Aspremont et al. to modify the Hotelling setting by assuming that the travel costs borne by consumers are quadratic in the distance covered, instead of being linear as in Hotelling. This new assumption captures the idea that the marginal cost of time increases with the length of the trip to the store. In this modified version, d'Aspremont et al. show that any price sub-

game has one and only one Nash equilibrium in pure strategies. Plugging these prices into the profit functions, they show that firms choose to set up at the two extremities of the linear segment. Firms do so because this allows them to relax price competition and to restore their profit margins. Therefore, the slight change made by d'Aspremont et al. leads to conclusions that completely differ from those obtained by Hotelling.

(b) In his review of Chamberlin's book, *The Theory of Monopolistic Competition*, Kaldor (1935) forcefully argued that, once it is recognized that firms operate in space, each one competes directly with only a few neighboring firms regardless of the total number of firms in the industry. The very nature of competition in space is, therefore, oligopolistic, thus casting serious doubt on the relevance of monopolistic competition as a market structure. Beckmann (1972) has developed a full analytical treatment of spatial competition in a well-crafted paper that went unnoticed, probably because it was published in a journal having a low visibility in the economics profession. In addition, Beckmann's main results were rediscovered by Salop (1979) in a paper that became famous in industrial organization. These two authors show how free entry may determine the equilibrium number of firms operating under increasing returns and competing oligopolistically with adjacent firms. Among other things, their analysis shows in a very precise way how the market solves the trade-off between increasing returns (internal to firms) and transport costs.

Building respectively on Kaldor and Hotelling, Eaton and Lipsey (1977) and Gabszewicz and Thisse (1986) have provided syntheses that help to clarify what spatial competition theory is about and what it can accomplish. This work was timely. Indeed, Salop was not aware of the contributions made by his four predecessors (Launhardt, Hotelling, Kaldor, and Beckmann), who all had a clear understanding of the nature of competition in space. This list of unrelated contributions, which cover almost one century, provides evidence of the very dispersed and fragmented nature of research in spatial economics until the emergence of new economic geography, which has served as a catalyst.

Increasing returns and strategic competition are, therefore, the basic ingredients of a relevant theory of spatial equilibrium. The difficulty of the task has put off more than one scholar. Exaggerating a little, we may say that the ability of the competitive model to tackle various issues as well as the absence of alternative models have generated a lock-in effect that economists had a lot of trouble escaping. It is, therefore, not totally surprising that the surge of new economic geography took place a few years after the revival of monopolistic competition and industrial organization, from which it borrows many ideas and concepts.

(c) As seen above, when firms sell a homogeneous good, they want to avoid spatial clustering because price competition has devastating effects upon them. It should be kept in mind, however, that this result is based on an extreme price sensitivity of consumers: If two firms are located side by side with identical prices, a small price reduction of one firm will attract all the customers. Such an extreme behavior seems unwarranted. When the product is differentiated and when consumers like product variety, the aggregate response to a price cut will not be so abrupt because the quality of product match matters to consumers. By turning the picture around, this observation suggests that firms selling differentiated goods may want to gather at some central market location, because price competition is now weakened (de Palma et al., 1985).

When transport costs are low, the benefits of geographical separation are reduced and prices are lower. Firms then choose to reconstruct their profit margins by differentiating their products along some non-geographical characteristics that are tangible or intangible. Stated differently, product differentiation is substituted for geographical dispersion. In this case, firms no longer fear the effects of price competition and strive to be as close as possible to the consumers with whom the matching is the best. Because these consumers are spread all over the market space, firms set up at the market centre and, therefore, minimize their geographical differentiation.

4. New economic geography

The existence of interregional inequalities has long attracted the attention of economists, especially in the area known as “regional economics.” For a long time, however, regional concepts, models and techniques were a mere extension of those used at the national level, with an additional index identifying the different regions—think of interregional input-output matrices. Despite valuable earlier contributions, no one before Krugman (1991b) had been able to show how regional imbalances could arise within the realm of economic theory.

In the 1950s, several development theorists put forward a principle that allowed them to uncover the underpinnings of unequal development—a principle that has been ignored, however, for several decades—that of *circular or cumulative causation*. Myrdal (1957, p. 13) sums up these ideas in the following paragraph: “The idea I want to expound in this book is that...there is no such tendency towards automatic self-stabilisation in the social system. The system is by itself not moving towards any sort of balance between forces, but is constantly on the move away from such a situation. In the normal case a change does not call forth countervailing changes but, instead, supporting changes, which move the system in the same direction as the first change but much further. Because of such circular causation a social process tends to become cumulative and often to gather speed at an accelerating rate.” Applied to economic geography, this principle says that regional disparities are driven by a “snowball” effect, which results in its continuous reinforcement once it is set in motion. Krugman (1991b, p.486) states the same idea when he writes: “manufactures production will tend to concentrate where there is a large market, but the market will be large where manufactures production is concentrated.”

4.1 The core-periphery structure

It is by marrying the Dixit and Stiglitz (1977) model of monopolistic competition with the iceberg transport technology that Krugman (1991b) may find out when and why Myrdal’s prediction materializes. The Dixit-Stiglitz model of monopolistic competition relies on product differentiation and increasing returns at the firm’s level, as in spatial competition. Unlike spatial competition, however, monopolistic competition involves weak interactions among firms, which respond to aggregate market statistics only. The iceberg cost means that only a fraction of good shipped between two places reaches the destination, the missing share having “melted” on the way. This ingenious modeling trick allows one to integrate positive shipping costs without having to deal explicitly with a transport sector. As for the other ingredients of his model, Krugman considers the standard setting of international trade theory, which involves two regions, two sectors, and two types of labor. The former sector produces a homogeneous good under constant returns and perfect competition, using one type of labor which is spatially immobile. The latter sector supplies a horizontally differentiated good under monopolistic competition and increasing returns, using the other type of labor which is mobile across space.

When footloose workers move to a new region, they bring with them both their production and consumption capabilities. As a result, their movements affect the size of labor and product markets in both the origin and destination regions. These effects have the nature of pecuniary externalities because migrating workers do not take them into account in their decisions. Moreover, such externalities are of particular importance in imperfectly competitive markets, where prices fail to reflect the true social value of individual decisions. Hence, the effects of migration are best studied within a general equilibrium framework, where one can capture not only the interactions between spatially separated (product and labor) markets, but also the dual role of individual-as-worker and individual-as-consumer. The great accomplishment of Krugman (1991b) was to integrate all these effects within a single framework and to determine precisely the conditions under which the

cumulative process predicted by Myrdal occurs or not. Turning next to the specific conditions for agglomeration, Krugman has shown that the value of transport costs is the key-determining factor.

If transport costs are sufficiently low, then workers will concentrate in a single *core* region, whereas the *peripheral* region supplies only the standardized good. In this way, firms are able to exploit increasing returns by selling more goods in the larger market without losing much business in the smaller market. It is worth stressing here that the core-periphery structure is the involuntary consequence of decisions made by a large number of economic agents pursuing their own interests. By contrast, if transport costs are sufficiently high, then interregional shipments of goods are discouraged. Hence the economy displays a symmetric regional pattern of production in which firms focus on local markets. The core-periphery model thus allows for the possibility of convergence or divergence between regions, whereas the neoclassical model, based on constant returns and perfect competition in the two sectors, would predict convergence only. Consequently, it is fair to say that Krugman has presented a synthesis of the polarization and standard neo-classical theories. Note, in passing, that Krugman has achieved this by using a model of monopolistic competition à la Chamberlin, that is, the model criticized by Kaldor because it does not provide a good description of competition across space!

4.2 The bell-shaped curve and regional growth

The core-periphery model has triggered a huge flow of extensions, which have contributed to make new economic geography one of the most lively research topics of the 1990s. Fujita et al. (1999) is the first place where to go to, whereas Neary (2001) remains the best critical review of the canonical model. One of the main criticisms is that the model ignores the congestion costs generated by the gathering of people and firms within the same territory. In particular, armchair evidence shows that a human settlement of a sizable scale almost inevitably takes on the form of a city, as in Section 2.2. As a consequence, a growing concentration of people intensifies competition for land and, therefore, leads to higher housing costs and longer commuting. In other words, even when nominal wages increase with employment density, housing and commuting costs, as well as pollution and crimes, make such large agglomerations less attractive.

As transport costs steadily decrease, the spatial economy now moves through three phases instead of two: dispersion, agglomeration, and re-dispersion of the mobile sector (Helpman, 1998).

Agglomeration arises in the second phase for the reasons highlighted in the core-periphery model. The dispersion in the first and third phases emerges for very different reasons. In the first phase, firms are dispersed because shipping their output is expensive whereas, in the third phase, dispersion occurs because housing and commuting costs are too high for the agglomeration to be sustainable. Put differently, beyond some threshold congestion prompts firms and workers to re-disperse in order to alleviate the corresponding costs. In the limit, high commuting costs are sufficient to prevent the formation of a large city and guarantee the continuation of industrial activities within several small cities, a situation fairly characteristic of preindustrial economies.

Another major shortcoming of the core-periphery model is that it overlooks the importance of intermediate goods. Yet, the demand for consumer goods does not account for a very large fraction of firms' sales, being often overshadowed by the demand for intermediate goods. Therefore, in making their location choices, it makes sense for intermediate-goods producers to care about the places where final goods are produced; similarly, final-goods producers are likely to pay close attention to where intermediate-goods suppliers are located. Giving intermediate goods a prominent role is a clear departure from the core-periphery model, which allows one to focus on other forces that are at work in modern economies. To this end, note that, once workers are immobile, a higher concentration of firms within a region translates to an increase in wages for this region. This gives rise to two opposite forces. On the one hand, final demand in the core region increases because consumers enjoy higher incomes. As in Krugman, final demand is an agglomeration force; however, it is no longer sparked by an increase in population size, but by an increase in income. On the other hand, an increase in the wage

level generates a new dispersion force, which lies at the heart of many debates regarding the deindustrialization of developed countries, i.e., their high labor costs. In such a context, firms are induced to relocate their activities to the periphery when lower wages more than offset lower demand (Krugman and Venables, 1995).

Last, the vast majority of economic geography models rely on a fairly naive assumption regarding migration behavior: individuals care only for real wages. Leaving aside migratory movements triggered by wars, people are heterogeneous in their perception of the non-economic attributes of the different regions, and this heterogeneity affects the nature and intensity of migration flows, as in Section 3.2 for shopping behavior. Because labor mobility is also driven by non-economic variables, workers do not react to economic inequalities in the same way. In such a context, the bell-shaped curve emerges again: workers move to the core when spatial inequalities are large but stay put when they are small (Tabuchi and Thisse, 2002). This is because workers bestow increasing relative weight on noneconomic factors affecting the quality of their life once they have achieved a sufficiently high material welfare. If this premise is correct, both economic growth and the development of the welfare state combine to slow down individuals' mobility, by allowing them to satisfy their needs for socializing and/or their attachment to a certain environment.

The benefits of using the Dixit-Stiglitz model of monopolistic competition are reap when we turn to regional growth because new economic geography and endogenous growth theories have been built on the same setting, thus making it easier to combine these two bodies of research within a unified framework. The contributions reviewed by Baldwin and Martin (2004) emphasize the possible geographical concentration of the innovation sector. Innovation being one of the main sources of the long-term growth of the economy, this innovation-driven concentration supplements the core-periphery effect to generate long-run patterns characterized by persistent and sizable income differences. In other words, the predominant centers would retain the high value-added activities, whereas the routine activities would be relocated into the periphery. This challenges the unfolding of the bell-shaped curve and keeps open the debate on the spatial diffusion of economic development. However, thanks to new economic geography, we understand much better the various forces at work.

5. Concluding remarks

The (relative) absence of space in economic theory lies in the attempt made by economists to develop a rigorous theory of markets and prices. This attempt has led them, through a series of simplifications and shortcuts taken long ago, to focus on the combination “constant returns and perfect competition” with consequences for economic geography that are comparable to those for growth theory. While the lack of interest manifested by many economists about spatial issues is regrettable, the opposite attitude (disinterest in economic theory as a whole on the grounds that it is a-spatial) is untenable. This attitude long characterized traditional regional economists, and it largely explains the stagnation of this field (Krugman, 1995).

The present state of the art in economic geography is the outcome of a two-sided process. The first side involves regional scientists who felt the need to root their field in economic theory. For a long time, because the emphasis was on city and regional planning, regional science focused more on optimization techniques than on equilibrium analyses. Despite some mutual ignorance between economists and regional scientists, most of the main contributions in regional science have been incorporated into the realm of economic theory, often after some delay. Among the community of regional scientists, one scholar stands out, namely Masahisa Fujita whose contributions have been acclaimed by the economics profession. The second side involves the few economists who faced the challenges posed by the introduction of space in economic theory, such as Martin Beckmann and Edwin Mills. Their task was not an easy one since space is plagued with all the difficulties that one may encounter in standard economic theory: non-convexities, externalities and imperfect competition.

For a long time, the contributions to economic geography made by economists have been confined to a small circle of isolated specialists. The situation has vastly changed. Whereas most old contributions to economic geography were often poorly related to mainstream economic theory, Krugman's work has drawn space from the periphery to the center stage of economic theory, making new and already existing ideas more amenable to both theoretical and empirical scrutiny. While new economic geography is closely related to trade theory, it is also very much connected to industrial organization. It is, therefore, not totally surprising that the surge of new economic geography took place a few years after the revival of monopolistic competition and industrial organization, from which it borrows many ideas and concepts. There are also strong links to new growth theories, where many scholars see cities as the engine of growth. Thus, it is fair to say that new economic geography has contributed to the development of a new and large flow of high-quality research and to the gradual emergence of a unified field. All of this highlights the integrative power of modern economic theory.

References

- Alonso, W. (1964), *Location and Land Use*, Cambridge, MA: Harvard University Press.
- Baldwin, R.E. and Ph. Martin (2004), 'Agglomeration and regional growth', in J.V. Henderson and J.-F. Thisse (eds.) *Handbook of Regional and Urban Economics, volume 4*, Amsterdam: North Holland, pp. 2671-2711.
- Beckmann, M.J. (1972), 'Spatial Cournot oligopoly', *Papers and Proceedings of the Regional Science Association*, **28**, pp. 37-47.
- Beckmann, M.J. (1976), 'Spatial equilibrium in the dispersed city', in Y.Y. Papageorgiou (ed.) *Mathematical Land Use Theory*, Lexington, MA: Lexington Books, pp. 117-125.
- Blaug, M. (1985), *Economic Theory in Retrospect*, Cambridge: Cambridge University Press.
- Christaller, W. (1933), *Die zentralen Orte in Süddeutschland*, Jena: Gustav Fischer Verlag.
- de Palma, A., V. Ginsburgh, Y.Y. Papageorgiou and J.-F. Thisse (1985), 'The principle of minimum differentiation holds under sufficient heterogeneity', *Econometrica*, **53**, pp. 767-781.
- Dixit, A.K. and J.E. Stiglitz (1977), 'Monopolistic competition and optimum product diversity', *American Economic Review*, **67**, pp. 297-308.
- Dunn, E.S. (1954), 'The equilibrium of land-use pattern in agriculture', *Southern Economic Journal*, **21**, pp. 173-187.
- Duranton, G. and D. Puga (2004), 'Micro-foundations of urban increasing returns: theory', in J.V. Henderson and J.-F. Thisse (eds.) *Handbook of Regional and Urban Economics, volume 4*, Amsterdam: North Holland, pp. 2063-2117.
- Durlauf, S.N. (2004), 'Neighborhood effects', in J.V. Henderson and J.-F. Thisse (eds.) *Handbook of Regional and Urban Economics, volume 4*, Amsterdam: North Holland, pp. 2173-2242.
- Eaton, B.C. and R.G. Lipsey (1977), 'The introduction of space into the neoclassical model of value theory', in M. Artis and A. Nobay (eds.) *Studies in Modern Economics*, Oxford: Basil Blackwell, pp. 59-96.
- Fetter, F.A. (1924), 'The economic law of market area', *Quarterly Journal of Economics*, **34**, pp. 520-529.
- Fujita, M. (1989), *Urban Economic Theory. Land Use and City Size*, Cambridge: Cambridge University Press.
- Fujita, M. and P. Krugman (1995), 'When is the economy monocentric? von Thünen and Chamberlin unified', *Regional Science and Urban Economics*, **25**, pp. 505-528.
- Fujita, M., P. Krugman and A.J. Venables (1999), *The Spatial Economy. Cities, Regions and International Trade*, Cambridge, MA: The MIT Press.
- Fujita, M. and J.-F. Thisse (2002), *Economics of Agglomeration. Cities, Industrial Location and Regional Growth*, Cambridge, Cambridge University Press.
- Gabszewicz, J. J., and J.-F. Thisse (1986), 'Spatial competition and the location of firms', in: J.J. Gabszewicz, J.-F. Thisse, M. Fujita and U. Schweizer. *Location Theory*, Chur: Harwood Academic Publishers, pp. 1-71.

- Hamilton, B.W. (1980), 'Indivisibilities and interplant transportation cost: do they cause market breakdown?', *Journal of Urban Economics*, **7**, pp. 31-41.
- Heffley, D.R. (1972), 'The quadratic assignment problem: a note', *Econometrica*, **40**, pp. 1155-1163.
- Helpman, E. (1998), 'The size of regions', in D. Pines, E. Sadka and I. Zilcha (eds.) *Topics in Public Economics. Theoretical and Applied Analysis*, Cambridge: Cambridge University Press, pp. 33-54.
- Henderson, J.V. (1974), 'The sizes and types of cities', *American Economic Review*, **64**, pp. 640-656.
- Henderson, J.V. (1988), *Urban Development. Theory, Fact and Illusion*, Oxford: Oxford University Press.
- Hotelling, H. (1929), 'Stability in competition', *Economic Journal*, **39**, pp. 41-57.
- Kaldor, N. (1935), 'Market imperfection and excess capacity', *Economica*, **2**, pp. 35-50.
- Koopmans, T.C. and M.J. Beckmann (1957), 'Assignment problems and the location of economic activities', *Econometrica*, **25**, pp. 1401-1414.
- Krugman, P. (1991a), *Geography and Trade*, Cambridge, MA: MIT Press.
- Krugman, P. (1991b), 'Increasing returns and economic geography', *Journal of Political Economy*, **99**, pp. 483-499.
- Krugman, P. (1995), *Development, Geography, and Economic Theory*, Cambridge MA: MIT Press.
- Krugman, P. and A. J. Venables (1995), 'Globalization and the inequality of nations', *Quarterly Journal of Economics*, **110**, pp. 857-880.
- Launhardt, W (1885), *Mathematische Begründung der Volkswirtschaftslehre*, Leipzig: B.G. Teubner.
- Lösch, A. (1940), *Die Räumliche Ordnung der Wirtschaft*, Jena: Gustav Fischer.
- Mills, E.S. (1967), 'An aggregative model of resource allocation in a metropolitan area', *American Economic Review*, **57**, pp. 197-210.
- Myrdal, G. (1957), *Economic Theory and Underdeveloped Regions*, London: Duckworth.
- Neary, J.P. (2001), 'Of hype and hyperbolas: introducing the new economic geography', *Journal of Economic Literature*, **39**, pp. 536-561.
- Predöhl, A. (1928), 'The theory of location in its relation to general economics', *Journal of Political Economy*, **36**, pp. 371-390.
- Ogawa, H. and M. Fujita (1980), 'Equilibrium land use patterns in a non-monocentric city', *Journal of Regional Science*, **20**, pp. 455-475.
- Salop, S.C. (1979), 'Monopolistic competition with an outside good', *Bell Journal of Economics*, **10**, 141-156.
- Samuelson, P.A. (1983), 'Thünen at two hundred', *Journal of Economic Literature*, **21**, pp. 1468-1488.
- Schelling, T. (1969), 'Models of segregation', *American Economic Review*, **59**, pp. 400-493.
- Schelling, T. (1971), 'Dynamic models of segregation,' *Journal of Mathematical Sociology*, **1**, 143-186.
- Starrett, D. (1978), 'Market allocations of location choice in a model with free mobility', *Journal of Economic Theory*, **17**, pp. 21-37.
- von Thünen, J.H. (1826), *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*, Hamburg: Perthes.
- Weber, A. (1909), *Über den Standort der Industrien*, Tübingen: J.C.B. Mohr.