



**HAL**  
open science

# Alternative explanations of hierarchical differentiation in urban systems

Denise Pumain

► **To cite this version:**

Denise Pumain. Alternative explanations of hierarchical differentiation in urban systems. Pumain D. Hierarchy in natural and social sciences, 3, Springer, pp.169-222, 2006, Methodos series. halshs-00145933

**HAL Id: halshs-00145933**

**<https://shs.hal.science/halshs-00145933>**

Submitted on 13 May 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Alternative explanations of hierarchical differentiation in urban systems

Denise Pumain

Chapter 7 in Pumain D. (ed). Hierarchy in natural and social sciences. Springer, Methodos series 3, 169-222

Cities<sup>1</sup> always have been the locus of power and social organisation. They were also places of social and technological innovation, accompanying an increasingly complex division of labour and growing economic accumulation. In the history of societies, cities emerged at different moments on different parts of the earth, several thousand years ago, but always after agricultural techniques had been invented (about 3000 years later) and coinciding with new forms of political organisation in society (Bairoch, 1985). The maximum city size increased through history, following demographic and economic growth and technological progress. As recalled by Michael Batty in the preceding chapter, “there are always many more smaller cities than larger ones”. A less known fact is that all through historical times, as well as today in every country, in wider regions such as Europe or even in the entire world, city sizes differentiate in a surprising regular manner: the number of cities follows an inverse geometric progression in relation to their size. It is therefore not surprising if the notion of hierarchy seems almost intrinsic in urban systems, and in the two acceptations mentioned in the introduction to this book. The hierarchy of cities is a complex architecture of social relations of power or influence, inextricably mixed within a huge variety of political, economic or cultural networks, and within an apparently simpler statistical ordering of population or economic values, producing a highly skewed distribution of city sizes. Is the essence of urban hierarchy social and political, or is it the product of physical constraints generating the same kind of statistical distributions in urban systems as are found in natural systems? A big challenge for urban research is both to disentangle the two possible sources of explanation for the existence of urban hierarchies, and to discover how they can be articulated in formalised theories enabling further predictions.

The interpretation of urban hierarchy is enriched by consideration of the spatial organisation of human habitat across the globe, according to a geographical approach. Urban systems (and more generally, settlement systems) can be represented as the *hierarchical organisation* of human activities into three levels, on three different geographical scales: the elementary units (urban actors, housing units, factories or offices buildings, transportation networks etc), the city as a whole, and the system of cities belonging to a given territory. This third approach of urban hierarchy was first formalised in systemic terms by the American geographer Brian Berry (1964), who used the famous phrase “cities as systems within systems of cities”. Although in this chapter reference will be made to this view of urban hierarchy (a kind of inclusive hierarchy), we shall mainly develop ideas that have set out to explain urban hierarchy according to the second meaning of the expression. In this case, the concept of hierarchy pinpoints the strong *differentiation in the sizes of cities belonging to any system of cities* (usually today from  $10^3$  inhabitants to  $10^6$  for medium size countries, up to  $10^7$  in the

---

<sup>1</sup> The word « city » is used throughout in the general acceptance of town or city (urban area including its suburbs).

most populated countries). The differentiation in size is usually highly correlated to an ordering by other mass indicators, such as the gross urban product, the total number of businesses, or the spatial range of the influence of the cities. Of course this correlation only holds inside systems of cities which have been well connected for long periods within political territories, ensuring relative homogeneity of social rules and economic conditions. Hierarchy of size also corresponds in more subtle ways to a hierarchy of complexity, as approached by the variety of urban functions, or the diverse levels of skill in the labour force, and the unequal ability to adopt innovation. To give an example, it may be noticed that it takes a very long time for a conurbation (juxtaposition of cities that have grown independently and become spatially contiguous) to develop a functional level equivalent to that of an urban area of the same size but that has grown as a single entity (for instance, the total population of the Ruhr conurbation in Europe is larger than the *agglomerations* (*the French term for large urban areas including town or city and suburbs, distinct from conurbations*) of London or Paris, but its economic weight and its cultural influence are much smaller). Therefore, the notion of *hierarchical differentiation* in a system of cities is not merely a statistical curio. It raises questions about organisational features or evolutionary properties of these city systems. As such, it indicates a path of research to explain the metastability of the differentiation in city sizes over time and its universality among different social, economic, cultural and political systems.

The regular hierarchical differentiation of urban systems, usually summarised by a Pareto-like or lognormal distribution of city size, was noticed a long time ago, and various explanations have been suggested since. Among these: an intentional functional organisation for the purpose of controlling a territory; the application of a spatial economic equilibrium principle; the statistical addition of Pareto-like elementary phenomena; a “purely” random growth process; self-organisation of settlement subsystems under a space-time optimisation principle; co-evolution of competing subsystems under territorial and social constraints, and so forth. We shall review these explanations and related methods of analysis, trying to assess their relevance and exploring the possible similarities between urban dynamics and other types of hierarchical complex systems.

## **1. Two acceptations for urban hierarchy**

When applied to urban systems, the concept of hierarchy has two meanings. We suggest using the term “hierarchical organisation” to refer to the view of urban systems in three nested levels of analysis, according to the scale of observation, and the term “hierarchical differentiation” for the more classic notion of urban hierarchy, consisting in cities of different size, power, and influence within a given regional, national or even broader territory. Although this chapter focuses on the second meaning, we shall use the concept of hierarchical organisation to construct a more complete theoretical representation of urban systems.

### *1.1 Hierarchical organisation*

The notion of *hierarchical organisation* is linked to the scale of observation, and recognizes three main levels of more or less distinct and autonomous entities, following a possible analogy with the hierarchical organisation of living organisms (figure 1). The first level is made up of elementary entities, like individual persons, firms or institutions, which can make decisions in terms of locating and organising activities, building housing, offices or monuments, travelling on foot or by car, and so on. Their multiple interactions, among themselves as well as with the more or less altered components of the natural environment which forms the site of the city, occurring over different scales of time (typically, from one

day to a life-time, see Lepetit, Pumain, 1999), contribute to the definition of the meso-level entity: the city. At this level, new properties emerge and characterize the city as a collective entity. These properties cannot be attributes of the individuals living in the city: for instance, an urban landscape or morphology, an urban density gradient, an industrial portfolio including particular specialisations, as well as more subjective features belonging to social cognition, such as the urban images that are exploited by urban marketing. Some of these new properties can be directly related to the intention of some institution, but most of the time they are the unexpected (and sometimes unwanted) result of collective interaction. However, a “city”, when properly defined (see below section 2.3), is an easily recognisable entity which has its own history and a specific trajectory over long periods of time, much longer than the life span of any individual person or even any firm present in the city at a given moment. The typical life span of a city within a particular trajectory is generally a few decades, sometimes more than a century (see below section 5 and figure 9).

### *1.2 Hierarchical differentiation*

A regular feature that is usually less well perceived is that such cities together form another level of organisation in urban systems, which is the network, or *system of cities*. It has long been observed that towns and cities always develop links with their surrounding environment: a settlement specialised in non-agricultural activities carries out central functions for its complementary region (Reynaud, 1841, Christaller, 1933). Unlike villages (or mining settlements) that exploit resources at their *site*, or in their close neighbourhood, towns and cities make a living from the wealth created by their *situation*. They capitalise on their position in trade networks, the spatial range of which depends upon the size and specialisation of the city (Reymond, 1981). For this reason, a city can hardly be conceptualised as an isolated entity, it always belongs to a network or system of cities. The city system acts as a constraint on or regulation of the dynamics of every individual city, through competitive and cooperative interaction (Pumain, 1992). City systems have their own regular properties which are not only defined by the sum of the individual cities composing them. They constitute a meaningful level for the analysis of urbanisation (Pred, 1977). City systems arise from the interactions between individual cities (interactions are produced through exchanges of persons by travel and migration, trade of goods, phone calls, circulation of information and knowledge, etc). Systems of cities are defined according to new emerging properties, such as a fairly regular spatial organisation, marked differentiation in size (figure 2), and specific features of urban co-evolution corresponding to the multiple interdependencies between the cities. The main structural features of such city systems have a much longer life span than the specific attributes (in terms of economic specialisation, rank in the hierarchy of size, or social image for instance) of the individual cities composing them. Their global configuration can indeed remain fairly stable for several centuries.

### *1.3 Complexity in urban systems*

Because of their obvious structure into various organisational levels, and the non-linear effects characterising their organisation, urban systems belong to the category of complex systems. Hence to analyse urban hierarchies, it is tempting to make use of the tools elaborated by the sciences of complexity. This approach however requires rather strict definitions and accuracy in measurement, which are difficult to establish in the case of urban entities. Compared to physical or biological systems, the handling of the concept of urban hierarchy is delicate because the separation between the levels is not always easy to determine. While the urban phenomenon as a scientific object for investigation can be envisaged using three levels of spatial organisation, i.e. the individual actors, the city itself and the system of cities, in the real world some intermediate levels, such as neighbourhoods or quarters inside cities, or

regional subsets of cities inside a country, can sometimes also be considered as more or less autonomous subsystems. Although the identification of the three levels of organisation is clear in a conceptual framework, it is often difficult to recognize them in reality. City quarters may have clear-cut edges but most of the time several different partitions of an urban space can be considered equally relevant. The boundaries of the city itself have become blurred, due to the spread of suburbs and of long-distance commuting, which has broken up the spatial continuity of daily urban systems. It is also sometimes difficult to differentiate clearly between a city and a network of cities, as in the case of large conurbations or very densely connected regions like the Randstad in Holland, the Ruhr area, the Italian and French Riviera, Spanish coastal resorts or Megalopolises (such as the North East of the United States between Boston and Washington, or central Japan from Tokyo to Osaka).

Systems of cities are also difficult to isolate as scientific objects of study. In theory, they can be defined as sets of cities having more interactions among themselves than with any others, or, in an evolutionary perspective, as sets of cities where any significant change in one city has consequences on other cities (Pred, 1977). In practice, systems of cities are generally defined within the limits of a single country, since international borders greatly reduce the intensity of spatial interactions. However, as is becoming clear with the modern globalisation process, interurban interactions and their consequences on urban change are significant even between cities belonging to different and sometimes distant countries. Moreover, the degree of openness of the city system varies according to the situation of each city within the urban system. The largest cities, or the ones that are specialised in international activities, are likely to become more open to external interactions than the other towns. It is therefore very difficult in practice to define the actual boundaries of a city system in a precise way.

Another source of complexity in urban systems is the many time scales which operate simultaneously in any city. One only has to consider how the timing of daily life adjusts (for instance by commuting) to the more stable pattern created by the location of jobs and housing facilities in the city. When longer time intervals are considered, similar time-scale differences occur, the life span of the buildings being generally longer than the duration of stay of their users or inhabitants, or even than people's life expectancy (Whitehand, 1987). This leads to the well-known pattern of residential moves, from central locations to the periphery and back to the centre, which is linked to the successive stages in the life cycle of individuals. But in the life of a city other time-scales, which may have decisive and sometimes catastrophic effects on its inhabitants, are involved: the duration of a cycle of economic specialisation (i.e. the adoption of a large set of innovations) may vary from a few decades to one century or more, leading to successive periods of rapid growth, stability and slow decay. Even if cities succeed in adapting to several successive innovation waves, the speed of change in economic and social function is generally greater than the speed of change in the town layout and infrastructure. Momentary non-adaptation of form to function, congestion phenomena, time lags in adjustment to change, mismatch in facilities and infrastructure, discrepancies between real needs and the objectives of policies can all arise from the unequal intrinsic duration of the life-cycle of each component in an urban system. The analytical processing of such a large variety of time-scales is very difficult and a main source of problems for urban theory and modelling. However, without being too deterministic, or resorting to finalism, we can suggest that the articulation of these different temporalities, which is a part of urban complexity, is also a characteristic explaining the survival and the persistence of cities over very long periods of time. The same structures can respond to very different social needs and economic activities, and this adaptability and flexibility ensures the sustainability of the whole urban system.

The problem of the blurred boundaries of systems of cities mentioned earlier is also linked to the diversity of scales in time and space governing the interactions between cities. Two types or orders of interaction can be distinguished. First order interactions correspond to the flows (of people, goods, money, or information) which circulate between the cities and can be quantified. Second order interactions emerge from these concrete exchanges, and generate the diversity of situation of each city within the network, constraining further exchanges. A city's situation is conceptualised in a relative way, in terms of symbolic or political position, economic comparative advantage, or range of influence, and very often these attributes are linked with its rank in the urban hierarchy. Of course these different situations have been progressively constructed by the first order interactions and their asymmetries or dissimilarities. But these same situations also act as constraints or regulations on the dynamics of the system. Individual actors may have a certain knowledge about urban situations, either accepting to adapt to them or trying to use them to redirect first order interactions. Very often however, there is a time lag between the actual evolution of urban situation (loss of influence or loss of market shares) and the moment when it becomes perceptible and changes the image of the city. Thus complexity in urban systems arises from the interconnections between spatial levels and time scales in urban dynamics. A decision made by one actor may be decisive, not only for the dynamics of a single city, but also (even if this is rare) for the dynamics of the system of cities as a whole. However, although entirely made up of intentions and decisions on the part of social actors, most of the dynamics of urban systems can be described without taking into account the detail of these intentions. This aspect of complexity in social systems highlights the difficulty of setting out the issues of urban hierarchy in simple political terms.

#### *1.4 Urban hierarchy as a social problem*

Why do we live in cities so different in size? We know that there are many urban entities of very unequal sizes all over the world, from the small towns of a few thousands inhabitants to the gigantic megapolises of 10 million and more (the largest today may be the agglomeration of Tokyo with about 30 million). The hierarchy of urban size is very regular: according to F. Moriconi-Ebrard (1993), there were in the world around 1980 about 26 000 towns and cities of more than 10 000 inhabitants, among which about 2500 had more than 100 000 inhabitants, 220 more than one million and about 10 more than 10 million. Within the European Union (25 members), there are at the beginning of the 21st century roughly 4000 agglomerations of more than 10 000 inhabitants, 390 of more than 100 000 and 38 of more than 1 million (Rozenblat, Cicille, 2003). Why is there such regularity? Will it be maintained once the total population of the planet is urbanised, and what will happen when the trend to overall demographic growth that has prevailed in recent centuries slows down or even reverses? During the urbanisation process, there was obviously some ratio between demographic growth and urban expansion, as well as between the total population of a country and the size that its largest city had already reached or could reach. The most impressive urban growth rates today are occurring in the developing countries: whereas the majority (54 out of 86) of urban agglomerations larger than one million inhabitants were located in the developed world in 1950, from 1990 on they are concentrated mainly in developing countries (174 out of 298, Moriconi-Ebrard, 1993). Is this rift between the location of the largest concentrations of population and the location of available resources and the ability to control them sustainable? Can cities grow indefinitely?

There is a contradiction between the appearance of larger and larger urban entities and the frequently mentioned aspirations toward a “rurban” way of life or repeated calls for keeping

towns at to a manageable size in human terms. There seems to be an opportunity for a new dispersion of settlement on account of the more ubiquitous means of distributing information provided by new communication technologies. In contrast, there is a challenge for some cities to try to remain or become part of a “top list” of “global cities” in an era of globalisation (Friedmann 1986, Hall, 1984, Sassen, 1991 ). To what extent are power and influence linked to urban size and rank in urban hierarchies? Are some cities too large, are others too small, is it possible to imagine something like an optimal size for a city (Bairoch, 1978)? What would be the cost of seriously applying the policy recommended by the European Union in the ESDP report (1999) aiming at more “polycentric urban development”? Answering such questions, or trying to make predictions about the future size of cities, requires a better understanding of the actual dynamics of urban hierarchies.

## 2 Cities, city size and city systems

To consider that the size of a city, generally measured by the number of inhabitants, is an indicator of its situation in a hierarchy of power or potentiality enabling a comparative ranking with other cities, is an idea that should not be assumed to hold true in all cases. Neither can the identification of urban entities, or the actual measurement of their size, be regarded as an easy question. The necessity for comparison in space and time explains the choice of population numbers in urban agglomerations as a universal simple measure. The question of the identification of a system of cities also requires discussion, even if the most usual framework for studying systems of cities is generally the territory of a state.

### 2.1 *Qualitative urban hierarchies in the past*

The historian Marco Folin (2003) recalls that for a very long time, Italian settlements were considered as having unequal importance according to their official status, connected with the ecclesiastic hierarchy (only the places of residence of a bishop could be called *città*.) This was not the case in other European countries, where the presence of walls was sufficient to ensure the greatness and privileges associated with the status of a city. In fact, the Italian practice derived from politico-administrative autonomy acquired sometimes long before by the *civitates*, which were places where a group of citizens could be self-administered via the presence of officials and a power of jurisdiction. According to different authors, there were about 300 “*città*” of this sort at the beginning of 16th century in Italy. Other criteria (presence of walls, urban or rural character) contributed to the emergence of a typology of settlements including places of lesser importance as *terra*, *castello*, *villa*. It remains that, until the 18th century, in various forms all over Europe, the classification of settlements was of a political nature, and the status of a place was rooted in a long history and interrelated with the nobility of its residents. (The status was in fact granted to the people rather than to the place where they lived). The significance of urban hierarchy was, at the time, mainly dependant on a symbolic ordering in society.

This conception evolved with the detailed observation of other indicators of the “*grandezza e magnificenza delle città*”, as quoted in the title of a remarkable book by Giovanni Botero (1588). He was the first to suggest measuring the size of cities by “the number of their inhabitants and their properties”. Here “objective”, “material” indicators of the importance of cities challenged the former considerations of nobility in establishing an urban hierarchy. But it was only at the time of the French Revolution that the former nominal classifications were abandoned, with the emergence of a new social order, and replaced by quantitative measures (Folin, 2003: 24). For instance in France a survey was executed around 1810 to determine the status of small settlements with 1000 to 2000 inhabitants (in three ordered categories of

decreasing importance: *ville, bourg* or *village*). The results indicated no clear correspondence, at that lower level of the urban hierarchy, between the number of inhabitants and the economic function or social structure of the settlement. But the main objective of the survey was to determine a quantitative threshold allowing the distinction between villages and towns, which was fixed at 2000 inhabitants for the 1856 census in France and is still in use for the purpose of distinguishing rural *communes* from the urban ones (despite the considerable changes in urban population and function, which probably alter the significance of comparisons over such a long period of time).

## 2.2 *Delimitation of urban entities for the purpose of quantifying city size*

Although some countries, for instance Germany or India (for some places only), still use the political status conferred by the federal government to identify cities, and while a wide range of criteria are used worldwide to define what a city is (United Nations, 2004), the criteria of the number of inhabitants for the purpose of measuring the size of cities is today universally accepted. This measurement is by no means a simple operation. In the course of time, cities have grown and the new buildings aggregated to the original urban node have spilled over the former walls, or the formerly established administrative boundaries, expanding the urban agglomeration into several *communes* or contiguous administrative areas (in some countries the territorial limits of the authority of municipalities are periodically revised, but this process is not very frequent and never exactly follows physical expansion). Alternatively, as today in many countries, the new urban development sprawls into the surrounding countryside, functionally dependant upon but no longer contiguous to the original urban core, and blurring the limits between rural and urban areas (Dureau et al., 2003, Champion, Hugo, 2004). The use of the number of inhabitants living in a continuously built-up agglomeration was for a long time recommended by United Nations (1978) to facilitate international comparisons. But nowadays the spatial continuity of the built-up area previously used for the definition of an agglomeration (usually no more than 200 or sometimes 500 meters between two groups of buildings in a constructible area) has lost its significance with the wider range of spatial interaction made possible by the automobile, leading to alternative definitions of statistically defined urban areas (like the US SMSAs since the 1950s or the more recent French *aires urbaines*) which are based on a given threshold of labour force resident in an administrative subdivision (county, commune, district etc) and commuting to the centre. However, not all countries define more realistic or functional urban entities of this sort (commuting data are not available everywhere), and most of the existing urban data bases that can be used for comparative purposes (Moriconi-Ebrard, 1993) are based on the definition of urban agglomeration, especially for historical times (Bairoch and al., 1988, de Vries, 1984, subsequent to the previous classic work by K. Davis or Chandler and Fox),

The results presented below have been mainly obtained by using the GEOPOLIS data base on world agglomerations prepared by F. Moriconi-Ebrard (1994). This very exhaustive data base has been too rarely used and quoted as a powerful instrument for international comparisons using the best definition for the purpose of comparison and reliable methods of delimitation of urban entities (Pumain, Moriconi, 1997). Delimitating urban agglomerations and measuring their size is a huge task that could now be undertaken by statistical institutes using GIS and satellite images, if they would agree to use a common concept and procedure, as illustrated for instance by the generation of a CORINE Land Cover view of north-western France (figure 3), even if it remains difficult to delimitate coherent, meaningful urban entities in some very dense, intensely connected settlement systems (but here, it should be possible to propose more than one definition). In this respect, we regret that the most recent attempt by Eurostat (a programme known as Urban Audit II, 2003 which includes a very carefully designed survey



(more than 300 variables), will provide urban data that are perfectly comparable in their statistical definition but absolutely not so in the spatial framework of the urban entities under consideration: the delimitation of the urban entities for the different European states varies in this document from political agglomeration (France), to NUTS3 (Spain) or NUTS4 (UK) regions... There is a considerable need for development and implementation of more comparable urban data bases.

Measuring the size of cities by the number of inhabitants remains the easiest way to compare their rank within an urban hierarchy. However, even if this method of evaluation may be motivated by a feeling that there is some sort of equality among human persons, one should be aware of the sometimes huge differences in the economic weight and political power of cities when making cross-national comparisons. For example, several orders of magnitude separate the economic size of Lagos or Manilla and Los Angeles, despite their comparable population size. Indicators of urban product would provide a more realistic view of the importance of a city in economic terms. However, production statistics are usually collected in business headquarters which may not be located in the place where the added value is actually produced. Personal income could be used as a proxy, but such statistics are also very rarely produced in relation to urban agglomerations. Thus most studies on urban hierarchies use the demographic indicator to measure city size. When the objects of study are areas where the levels of economic development and personal income are fairly homogeneous (usually within the limits of a nation-state), they have a broader significance than when they cross boundaries between areas presenting wide differences.

### *2.3 A geographical ontology for urban systems*

Why do we prefer a geographical delimitation for the purpose of identifying a “city”, rather than any other criteria that could be relevant for the definition of an autonomous urban entity? For instance, it could be argued that a political definition of a city as a group of residents living on a territory governed by a single administrative power (or co-operating local governments as in some large modern metropolises) would ensure greater coherence and autonomy to the urban entity, from the point of view of collective decision and management. Many studies on urban hierarchies (and even a recent data base that was compiled by Eurostat for “comparative” purposes, Urban audit I, 1999) are based on the acceptance of a “city” as a single administrative unit, in which the core of an urban agglomeration is located. However, despite more or less recent (Dahl, 1961, Freire and Stren, 2001, quoted by Gaudin, 2004) call for an “urban governance” on the basis of relevant territorial delimitations, for which there are many possibilities, the definition of urban entities according to political boundaries is rarely designed to enable coherent management of all the coordination problems that arise from the spatial juxtaposition of administrative areas (such as communes, counties, or districts) that have been progressively urbanised into a continuity. Moreover, whereas administrative units are very variable in surface area in different countries or even within one country, therefore putting an arbitrary limitation on the possible extension of an urban unit, there is on the contrary marked consistency in the spatial organisation of the urban agglomerations, at least in regions like Europe. The existence of an “urban field” (like a gravitational field) that structures the spatial pattern of densities and land prices according to a steep gradient of land use intensity from the centre towards the periphery has been illustrated by a recent study using comparable CORINE Land Cover data for European metropolitan areas (Guérois, 2003). There is a fairly regular steep gradient of population and built-up densities within a radius of 30 km around the historical centre of any European metropolitan area, which corresponds to the continuously built-up urban entity (defining an urban agglomeration). Another less steep but still visible gradient also structures the less dense new peri-urban spaces within a radius of

40 to 100 km from city centres (figure 4). Accordingly, a definition of urban entities based on the continuity of the built-up space frequently gives a better picture of the coherence of an entity where all elements necessarily interact, because they share a common environment, they are competing for the same space, they have to negotiate and cooperate for infrastructures and they jointly construct a sense of place. They jointly exploit the advantages of a site, a situation and an urban history. This value attributed to place and space around a given central location is specific to the urban character of a location and is well expressed by the concept of the urban field.

Of course, it is only by approximation that the inhabitants (and their “properties”, as Botero termed it) are allocated to their place of residence and counted as a quantitative attribute of the urban agglomeration. People move, they can have several temporary residences, they invest elsewhere and abroad, or they make decisions for other places than their own. It should however be recalled that for centuries, these spatial interactions were highly constrained by short distances and that even today, a majority of the population living within an urban agglomeration also works there, because the temporal constraint of the duration of the day for organising human activities has remained invariant through history. The concept of a “day-based urban system” is still valid for a large number of people, even if a few individuals no longer come under this constraint. “Space matters”, still.

Space (in the sense of continuity and proximity of relationships) also matters at the other level of organisation of urban systems, in the definition of the subset of urban agglomerations that are considered as belonging to the same “system of cities”. Such systems can be defined inside a given political territory, a region, usually a nation, or sometimes even on a larger scale, such as Europe or even the whole world (in these cases, only subsets of cities, usually among the largest, sometimes among the most specialised in international functions, are considered to belong to global city networks). Strictly speaking, city systems should be defined as subsets of cities having more intense interactions among themselves than with any other cities. However, we have seen that the kind of interaction envisaged matters: first order interaction may still be constrained by the proximity factor, and develop with greater intensity between cities that are close to each other, whereas second order interactions, which are more important in influencing the dynamics of each city relative to the others, may come from very distant urban locations. Most of the literature on urban hierarchies positions itself at the level of nation states to identify relevant city systems, because they offer a certain homogeneity of social rules in the development of the urban system, but there is a general awareness that a national territory is too narrow for the purpose of defining the boundaries of the space in which cities function (their range of action), especially the largest ones. Some studies have also demonstrated the validity of other emerging city systems, for instance at European level at least from 1950 on (Cattan et al., 1994) and many authors are also trying to assess the relevance today of a concept of a world network of cities. So for the sake of clarity in this chapter we shall mainly refer to a simple ontology of nested urban systems made up of sets of urban agglomerations, each grouping a number of inhabitants, inside national territories, for the study of urban hierarchies, but we shall return to the problems of definition when discussing the results of observations made in this simplified framework.

### **3 Static functional explanations**

The spatial patterns of urban hierarchies are so regular that they were rapidly interpreted either as the result of political intention, or as the product of a universal spatial constraint on economic behaviour. These logical explanations help to understand the regular spacing

between cities of similar importance and their differentiation into distinct functional levels. But the discontinuous distribution into levels of city size that they theoretically generate is challenged by the observation of continuous distributions, that are better described by a simple statistical model, such as Zipf's *rank size rule*. These two quite different types of interpretation are both *static explanations*, which consider urban hierarchy as a state of equilibrium, but do not really explain how that momentary equilibrium is generated and then maintained over time.

### *3.1 Regular spacing and hierarchy to control a territory*

The co-existence of towns and cities of different sizes was first interpreted as resulting from an intentional organisational process, expressing the will of a political power to control and administrate a territory. This explanation can be seen as a simple transposition, to the level of a political territory, of the old conception of the rank of a city in a hierarchy of dignity, although completed by a sense of efficiency of hierarchy for functional organisation. This type of interpretation is typical of the military who conceived the spacing of fortified towns for the protection of a territory, for instance Vauban (1707), or characteristic of far-seeing political governments, as reported by Marco Polo in relation to the three levels of cities governing the administrative division of the Chinese empire by Khoubilai Khan. This way of thinking is quite natural, as it reflects the two main processes of "creation" or "emergence" of cities, that were well known and shared in the historical culture of these authors. In Ancient times, colonisation proceeded by dissemination of new settlements derived from the metropolis, the "mother city", firstly for trading purposes, as was the case for the Phoenicians or Greeks establishing counters and ports along the Mediterranean coasts; secondly, when consolidating kingdoms, the political powers posted administration units, at different levels, according a more or less geometric partitioning of their territory, as was the case in the Roman empire. Most of the time, they did it in a hierarchical way, installing a main capital (even when the residence of the prince was partially nomadic), and different levels of dignities associated to the ordered status of urban places. These similarities in choices made by different societies could be interpreted as the product of the empirical observation that the social communication is made easier by hierarchical organisation. The circulation of information, both top-down and bottom-up, costs less energy when it is organised according to a pyramidal construction of responsibilities (it grows in a linear manner, proportionally to the number of elements instead of increasing to the square in the case of fully connected networks). Most armies in the world have adopted this structure, and its use was systematic in the conception of the Catholic Church, to the point that in the French 19th century, the word "hierarchy" was understood as ecclesiastical hierarchy alone (as noted in this book by N. Verdier). For a long period of time, during the 19th and most of the 20th centuries, business management has favoured hierarchical organisation for reasons of efficiency of command and decision making.

According to these views, urban hierarchy could be interpreted as a simple reflection of the social order in the spatial organisation of societies. Political or religious functions are indeed important in explaining the creation of many cities and part of their further development. Without mentioning the obvious case of national capitals, it has been demonstrated that, size being equal, the attribution of an administrative role has a high probability of boosting further growth of the selected cities, as for instance in the case of the *préfectures* after the French Revolution, or more recently in France the regional capitals (Bretagnolle, 1999). However, not all urban development relies upon these political decisions alone. Botero (1588) already quoted, among the factors explaining the greatness of cities, the commodity of the site (its accessibility), the fecundity of the soil, the efficiency of trade and industry, together with the

social and cultural policies of the princes aimed at making them attractive. He thus introduced new explanations for the differences in urban sizes (Pumain, Gaudin, 2002).

### *3.2 Central place theory: economic optimisation under distance constraints*

The economic activity of cities, like trading, industrial production and services, is also widely responsible for their differences in size. During the 19th century, following the theories of Cantillon and von Thünen (1826), many observations were made concerning the regularities in the spatial distribution of city sizes according to their economic functions. For instance, the German geographer Kohl (1841) designed geometric models for the purpose of optimising flows of circulation between places. The French engineer Lalanne (1863) formulated laws for the spatial configuration of administrative centres at different levels (equilateral triangles and separating distances that were multiplied by the same factor from one level to the next). Elisée Reclus (1895) distinguished four levels in the size and spacing of French cities reflecting their role as steps on itineraries, according to “a sort of natural cadence relating the progress of men, horses and carriages”, while the Frenchman J. Reynaud (1841) was the first to describe the main principles of the central place theory that were formalised later by the German geographer W. Christaller (1933). Christaller set out to explain the number, size and spacing of cities. He defined them as “central” places, that is, locations where goods and services are supplied for a surrounding market area (including rural areas and other towns of lower order). Two postulates are essential to the theory: 1) consumers use the nearest centre to acquire what they need (principle of minimum cost, since transport is added to the price of the goods); 2) the goods and services at a given level (corresponding to the same spatial range) cluster in the same centres. Christaller takes for granted the principle of the hierarchy of economic goods, according to their temporality of use and degree of availability (or scarcity), as first mentioned in 1871 by the economist Carl Menger who with L. Walras co-invented the theory of marginalism. He uses a kind of marginal spatial reasoning himself to explain how the supply of each commodity in a given spatial range is located in a limited number of places. His demonstration leads to a kind of spatial equilibrium, where the number of places supplying a commodity is constrained both by the maximum distance that the consumer can accept and by the minimum number of consumers that ensures the viability of the firm supplying the commodity (fixed costs). In conditions of homogeneity (of income and population density), he derives spatial models for a hierarchy of centres dominating market areas in nested hexagons. The spatial models designed by Christaller illustrate three types of urban hierarchies that correspond to different principles, each of them presenting a typical ratio  $K$  between the market shares of two successive levels in the hierarchy: maximising the number of centres and hence the accessibility to the consumers (market principle,  $K=3$ ), reducing the length of transportation networks connecting the centres (transportation principle,  $K=4$ ), and non-competition between centres by including all lower order centres inside the sphere of influence of a higher order centre (administrative principle,  $K=7$ ). According to Christaller, the observed hierarchies, such as those in his own empirical study of Southern Germany, reflected combinations of these theoretical models (figure 5).

Central place theory gives an interesting insight into the explanation for inequalities in city size and how they correlate to many aspects or indices of a hierarchy of urban economic activities. Although it has been tested and fairly well verified in a wide variety of territorial and economic environments (even on systems of periodic markets, for instance by Skinner), and is still valid in many respects, this theory has lost some of its explanatory power with increasing urbanisation, mainly because cities are no longer just central places for surrounding rural populations. Nor does the theory take into account the presence of non-service urban activities, that are determining factors for understanding a city’s dynamics, even if services to

the resident population may still concentrate more than half and often two thirds of a city's labour force. The theory also has been criticised because the postulate of a consumer minimising the purchase distance for each type of commodity is not acceptable: nowadays, more than 40% of the purchases of low order goods are made in higher order centres that are preferred because they offer a wider choice and allow multi-purpose shopping trips. Thus, contrary to the position adopted by Fujita et al. (1999), the problem is not the absence of a micro-economic theory in Christaller's book, but the lack of sophistication of the micro-economic hypothesis, leading to an underestimation of the attractivity of larger centres. Christaller may have suspected this, since he was surprised to find fewer small towns than expected from the theory in his empirical results.

### *3.3 The rank size rule: a statistical model*

Parallel to the classification of central places after the scope of their supply of services and the range of their sphere of influence, other representations of urban hierarchies have been developed, in terms of the statistical distribution of city size, usually as measured by the number of inhabitants. The first mention of a mathematical relationship is provided by the German geographer Auerbach (1913) who noticed that the product of the rank of a city by its population is approximately a constant and he used this value as an index of concentration (he made plots for Germany, Great Britain, the United States, France, Austria, and Russia). The statistician Lotka (1924) applied this regularity to US cities and introduced a graphical representation of city populations as a function of their rank on a double logarithmic plot. The sociologist Goodrich (1926) from the Chicago school, also mentioned this statistical regularity. The economist Singer in 1936 remarked on the similarity between the distribution of city sizes and the law that Pareto adjusted to the distribution of income, while in France, the statistician Gibrat (1931), in a thesis on economic inequalities, suggested another statistical model, the lognormal distribution.

Despite this early interest from various disciplines, the benefit of the "invention" was to be carried off by Zipf who gave a systematic form to the "rank-size rule" in 1941. He suggested that when the cities of a country are ranked from the largest to the smallest, the population  $P_i$  of a city is linked to its rank  $r_i$  by the relationship:  $P_i = K / (r_i)^q$ , where  $q$  and  $K$  are constants. In his first book, Zipf introduced himself as the discoverer of the law, not mentioning any pioneer, arousing an immediate reaction from Lotka who pointed out the similarities with the Pearson's type XI or VI laws. In his later book in 1949, Zipf only briefly mentioned his predecessors. Probably the success of his formulation lies in its capacity for a descriptive qualitative analysis of the shape of the distributions (Zipf plotted series of curves at different dates for many countries), whereas the statisticians of the time were mainly aiming at computing concentration indices. However, the unusual method he used for representing the statistical distributions led to recurrent misunderstandings of the significance of the law (for instance some authors presented it as trivial since a good fit between population and rank is to be expected, because they are necessarily linked variables, even though the question is the distribution of the *number* of cities as a function of their size). It is also useful to note that the  $q$  parameter has a value which is the reverse of the parameter of a Pareto distribution which would usually be formulated as the number of units  $R_i$  having a size larger than  $P_i$ :  $R_i = A / P_i^a$ ; hence, contrary to the corresponding Pareto parameter  $a$ , the higher the absolute value of  $q$ , the greater the size inequalities within the observed distribution.

The explanation given by Zipf is very general. Having applied his model to different socio-economic or linguistic distributions, he suggests that they are all constrained by the "law of least effort", as a universal expression of human behaviour. For cities, he considers that two

opposite forces act on the spatial distribution of human activities: a force of concentration tends to bring together production and consumption, whereas a force of dispersion is linked to the scatter of natural resources that are necessary for production. If the distribution of city sizes appears as a straight line on the double logarithmic plot, it is because these two forces balance! However, he proposes no demonstration. There is a need for a better understanding of the origin and persistency of the statistical regularity of the distribution of city size. The more recent attempts to deduce an urban hierarchy as the expression of a spatial equilibrium are not very convincing. Why should this structural feature in any urban system represent some form of optimisation, and for what purpose? Who decides? Why do such similarities in organisation of urban systems emerge despite the diversity of political, administrative, economic organisation in the different countries, and who or what ensures that they persist over time?

### *3.4 The difficulties of static explanations*

Cities are neither businesses nor simple institutions. They do not have any general aim or function to fulfil and, even if subsets of interested actors or certain specialised institutions can be identified to represent a sector of activity or a group of citizens, there is no omnipotent body responsible for supporting the general development of the city in all its dimensions. This could have been the case in historical times, for instance in Europe, when cities were governed by a prince or a bishop having full authority over the population and the territory. Such cities should in fact be considered as “states” and are indeed identified as “city-states” by historians. Their rivalries and events in their development can be related to well-identified “actors” who made decisions according to their representation of what their city should become. We have however shown, on the issue of the difficulty of providing reliable urban statistics, the usual mismatch between coherent physical or functional urban entities and the administrative boundaries on which they are built. Today the real political power that controls a city’s development is no longer unified on the scale of a single city. Nor is there, on the higher scale, namely systems of towns and cities, any decisional body that could make decisions for keeping the size of cities within the general model of city size. Of course towns and cities are connected by a multitude of links corresponding to a variety of social networks. A few networks can be identified that might have the incentive to lobby for the cities that belong to them. But no single institution, nor even a few competing institutions, can be taken as directly responsible for the persistence of the general model of city size distribution. What is required is a very general process operating beyond individual intentions and decisions to produce at a higher level of organisation the ordering of cities into hierarchies of a given statistical shape. Moreover, if cities retain relatively constant shares in the general economic, demographic and cultural or technical developments, it is probably by a deconcentrated process composed of many decisions made at a micro level (to invest in a given activity, to migrate to a city, to adopt an innovation...), that drives the transformation of each town or city, in an incremental way, in the general direction of change.

Another static explanation of the rank-size rule has been suggested by describing cities as the aggregation of a variety of activities. As each of these activities follow Pareto laws of distribution (for the size of firms), and since a sum of Pareto laws is still a Pareto law, this could explain the shape and regularity of the distribution of city size (Roehner, Winiwarter, 1985). Of course this leaves the problem of how to account for the Pareto distribution of firms within an industry. And it does not explain how cities come to concentrate variable amounts of business and activity. Moreover, the growth process of businesses is not at all the same as that of cities. Firms can merge or divide, they can also collapse, many are created or disappear over a short period of time. If we leave aside the rare cases of cities merging (forming

conurbations) and of cities dying (this seems to have happened very rarely in modern historical times) cities grow or decline in population from natural increase, or by migration. They can grow more or less rapidly, and towns entering the system are usually former villages which reach the threshold for urban definition.

In an earlier book (Pumain, 1982) we have demonstrated that a static explanation using methods from statistical mechanics was not a satisfactory explanation either. Leslie Curry (1964) suggested that the city size distribution observed was the most probable state in a statistical process allocating cities into size groups, according to a principle of maximisation of entropy. His interpretation was supported by B. Berry (1964). But the constraint that is added to the mathematical model, to derive a distribution of size of exponential type, actually adds a strong hypothesis to the random process, since it defines a mean value for city size, and consequently fixes the number of cities for a given total urban population. The level of concentration of urban population is then *a priori* and implicitly introduced into the model.

Economic theory is still endeavouring to provide explanations for this structure of the urban system in terms of optimisation. However, urban economic theory mainly conceives cities as places where agglomeration economies can be produced to attract business (Derycke et al., 1996; Huriot, Thisse, 2000). City size is theorised as resulting from the compromise between agglomeration economies and congestion costs (both quantities are very rarely empirically measured or estimated). Usually a theory of this sort leads to the notion of an optimal city size (Bairoch, 1978). In order to explain urban systems and to account for their hierarchical structure, micro-economic theories make the assumption that returns increase with the size of the urban centre. For example, Fujita *et al.* (1994) worked on the hypothesis that the system optimises its operation by establishing equilibrium between supply and demand for services at the meso level of the city, while individuals optimise their localisation by maximising their utility. This means that the towns and cities which offer the widest range of services will be more attractive, and the influx of migrants will in turn cause an increase in the range of services they can offer. Large cities are thus more attractive and grow faster than smaller ones, but the theory does not explain why differences in city size exist, and why the urban hierarchy acquires a particular form or why this form is stable over time. The economic models that try to produce a rank size distribution (Cordoba, 2003, Gabaix and Ioannides, 2003, Fujita, 2000) are merely theoretical and have not been empirically tested. Economic theory cannot yet explain why there should be a Zipf's law, and it is highly significant that in their last book on spatial and urban economy, Fujita, Krugman and Venables (1999) devote a whole chapter (chapter 12) to this question, entitled: "an empirical digression: the size of cities"! They quote a suggestion by Krugman to develop an analogy between cities and river networks, forgetting to recall that this analogy has already been suggested by Chorley and Haggett in their book on geographical models in 1967. Moreover, Krugman suggests that the "percolation theory" of physics could be used for modelling urban hierarchies as resulting from a diffusion process. However, it is well known that diffusion processes in urban systems are very often "hierarchical", in the sense that the adoption of innovation "jumps" from a large city to a very distant large one before occurring in those closest (Pred, 1977) and that this cannot be simulated by a passive and contiguous diffusion process.

We would suggest that the somewhat tautological hypothesis used by economists to justify the concentration of economic activities in urban centres, which presupposes the existence of agglomeration economies, (or increasing returns with increasing city size) is perhaps not necessary to explain the differences in the size and evolution of towns and cities. An evolutionary theory can account for this without having to accept the idea that large cities are

more productive and more efficient in economic terms than smaller ones, to explain the existence of urban networks in their present form. Indeed, the greater economic efficiency of large urban centres, if this is actually proved to be the case (Rousseau, 1998), and thus the existence of agglomeration economies, could be interpreted equally as the *consequence* and as the *cause* of their success.

### 3.5 Variations and evolution in urban hierarchies

Zipf's model however remains useful for descriptions and comparisons and has been fitted many times to more or less correctly measured population data sets for towns and cities (as in figure 2). When adjusted to the population of the urban agglomerations (towns and cities over 10 000 inhabitants) for states across the world (including all those having at least 30 urban agglomerations in the Geopolis data base), the estimated values for the parameter  $q$  range from 0.7 to 1.3 (Moriconi-Ebrard, 1993). The variation among different countries is rather small: the standard deviation of measured  $q$  values is only 0.138. The fit of the model is rather good, even if in many cases better fits can be found with other types of asymmetrical distribution (Quandt, 1964, Guérin-Pace, Lesage, 2001). Usually, when settlements of smaller sizes are included, the lognormal distribution provides a better fit than the Pareto model (Baker, 1969, Robson, 1973, Pumain, 1982). Very often, the upper part of the size distribution, corresponding to the largest urban settlements, does not fit any model very well: these cases of urban primacy (one to up to eight cities per state whose size exceeds the expected values) were detected a long time ago (Jefferson, 1939) and seem to be a generality rather than an exception. For two thirds of the world's states, the mean value of the ratio  $P1/P2$  is significantly higher than that which would correspond to the model. When this "primacy index" is computed, as the ratio between the population of the largest and second largest city, it is found that in most states in the world it is much larger than the value of two which would roughly correspond to Zipf's rank size rule (for a Pareto distribution with a value of 1.3 for the parameter  $q$ , the expected ratio would be 2.5) and the mean value for all countries of the world taken together is 5.2 (Pumain, Moriconi, 1997). In some countries, it is not only the largest city but a few large metropolises that create a discontinuity in the distribution of city sizes, and "macrocephaly" indices have been proposed for measuring their pre-eminence (Moriconi-Ebrard, 1993).

Several remarks are required here, because the literature on Zipf's law is full of ill-founded conclusions, mainly due to small observation samples and a lack of accuracy in empirical data (for instance, Rosen and Resnick, 1980, returned to by Alperovitch, 1993). It is true that whatever the part of the world and the period of observation, over the 10 000 years since towns first emerged, the model of settlement size distribution has always been reasonably well approximated by a Pareto or lognormal distribution. This has been demonstrated by Fletcher (1986), based on available data on early urban settlements as determined by archaeologists. During historical times, inequality in city sizes has been increasing (Roehner, 1991). Empirical evidence from historical data (Bairoch, Batou, Chèvre, 1988, de Vries, 1984) shows a clear evolution from values for the  $q$  parameter of around 0.7 in many countries before the 19th century to significantly higher values (at least 0.9 and often larger than 1) for distributions observed since the middle of 20th century. However, the direction of the most recent evolution is by no means clear: between 1950 and 1990, in the states with at least 30 urban agglomerations (Geopolis database) the value of the  $q$  parameter has steadily increased in 19 countries, steadily decreased in 9 others, while it followed no regular evolution in the remaining 21. There is a rather general trend towards a lesser contrast in city sizes in the more developed countries, but there are exceptions (USA, France, Japan, Russia,



Hungary and Greece). The diversity in evolutionary paths is still greater in third-world countries (Moriconi-Ebrard, 1993).

Over time, there is no indication either of any convergence towards regularity in the shape of the size curve. This contradicts a hypothesis made by Berry (1964) and reiterated many times since (for instance, Haggett, 2001). According to these authors, the existence of a primary city could reflect more primitive state of urbanism, and the size distribution should evolve towards a more regular pattern over time. However, the persistence of systematic deviations from the Pareto model seems to be the rule, especially in the upper part of city size distribution. This can be explained by noting that urban systems never are completely isolated from their environment, and that the larger the cities are, the broader is the range of their relationships. Perhaps the largest cities in each country should be considered as being parts of wider territorial systems or networks, which would make their frequently exceptional size more understandable (see below). However, some very large urban systems, as in United States, have very regular size distributions. Another type of regularity should be mentioned. Although it may be rather loose, there is a definite relationship between the magnitude of the size of the largest city in a country (P1) and the total urban population of this country (PT). The proportion of the urban population which is concentrated in the major city varies between 10 and 30%. This proportion tends to be larger in smaller countries than in the large ones, according to the well adjusted non-linear relationship:  $P1 = k PT^{0.8}$  (Moriconi-Ebrard, 1993). This is a direct consequence of the general Pareto shape of all national city size distributions (Gibbs, 1963).

The observations made by using statistical models such as Zipf's law or lognormal distribution, clearly maintain city size distribution within the sphere of attraction of Levy's stable laws, whereas other distributions of inequalities in social systems have evolved towards the attraction domain of the normal law. M. Barbut (2004) has demonstrated that the distribution of income shifted towards the attractor of the Gaussian model in about the middle of the 1930's in France and about the same time in other developed countries which had undertaken policies of social redistribution. Similarly, the inequalities in the size distribution of firms were reduced by antitrust regulations. Since in the case of urban hierarchies, the inequalities persist, can this be interpreted as the absence of any intentional control or regulation policy? Either the right policies have not yet been invented, or perhaps any regulation, for instance the European Commission's call for "polycentric development", is doomed to remain inefficient.

There is obviously a need for a better understanding of the universality and persistency of the statistical regularity and marked inequalities in size within urban hierarchies. Instead of developing static explanations, other authors have suggested examining temporal processes. In a first step, urban hierarchy is no longer understood as the result of a social intention or as the necessary outcome of an optimal use of geographical space for economic or transactional purposes, but on the contrary it is interpreted as the outcome of a purely random process.

#### **4 The outcome of a random growth process**

As urban systems involve large numbers (of settlements and persons), their regularities have been conceived to be a consequence of general statistical effects or stochastic processes. In fact, the power laws that are observed in so many natural and social complex systems could suggest a universal statistical explanation for hierarchical structures. Models of distributed growth generate of power law distributions. This was shown as early as 1922 by Willis and

Yule in a deterministic model explaining the linear relationship between the logarithms of the numbers of genera and species; the model was applied to cities by Steindl who again demonstrated in 1965 that two hypotheses are sufficient to generate a rank-size distribution: all towns and cities have the same growth rate, and there is a constant ratio over time between the urban population growth rate and the rate of appearance of new towns in the urban system. Earlier, other authors in an attempt to take into account the numerous fluctuations of observed urban growth rates from one city to another and over time had designed stochastic processes involving similar principles: H. Simon (1955), transposing a model first elaborated by Yule (1924), showed how a Pareto-like distribution, which is characteristic of contagious processes, can be generated if the probability for a migrant to reach a city of size  $i$  is proportional to the total number of people already living in cities of size  $i$  and if the individual probability of migrating to a new town remains constant over time; R. Gibrat (1931) demonstrates that the “law of proportional effect” as a stochastic model is sufficient to explain the emergence of a lognormal distribution of city sizes. We shall discuss this model in greater detail because it makes predictions that can be empirically tested. Despite the quality of fit with empirical observations, the model requires further exploration as to its hypotheses. More generally, it seems unlikely that the similarity in formal processes should remove the need to search for an explanation specific to each kind of complex system.

#### *4.1 Gibrat and the law of proportional effect*

The simplest alternative to the static point of view is the approach which treats urban hierarchy as the product of a stochastic process distributing population growth between cities. This involves transferring a *statistical model* for the dynamic description of an urban system. In his book: *Les inégalités économiques* (1931), Gibrat demonstrated that when cities are growing at the same average rate but with fluctuations or growth inequalities, the distribution of the city sizes will consistently take a lognormal form. He explained that whereas an additive process of growth would lead to a normal distribution, multiplicative growth, which he calls “the law of proportional effect”, will result in a lognormal distribution of sizes. Let us consider a set of localities with a certain size distribution (they can even be all the same size at the start of the process) whose evolution over a long period (several centuries, for example) is modelled, using a large number of short time periods. It is assumed that in each time interval (for example, one year, or ten years) the population  $P_i$  of each locality grows on average (though with fluctuations) by an amount  $dP_i$  which is proportional to the population  $P_i$  (and low in relation to  $P_i$ ). This is the same as saying that the proportional rate of change of the population  $dP_i/P_i$  (which is the measure usually expressed in terms of percentage and used to describe and compare the growth of cities and regions) has the same average value for all settlements, whatever their size at the beginning of each time period. If in addition the distribution of these rates between localities is independent from one time period to the next, the growth process it defines will always result in a distribution of settlement sizes that is a lognormal distribution, highly skewed, characterised by a large number of villages and small towns and a geometric decline for the numbers of cities according to their size. If the rate of growth and its variations are known, it is even possible to predict (for a known probability) by how much the size disparity will increase.

Gibrat adjusted the lognormal model to the distribution of size of European cities (larger than 100 000 inhabitants) at two dates, and he also computed the evolution of an index of concentration for the French system of cities for the period between end of 19th century and first third of the 20th century, noting an increase in the concentration of urban population. The lognormal distribution was taken as a descriptive model by B. Berry who plotted the distribution for a number of countries on Gaussian logarithmic graphs. But it is Robson

(1973) who was the first to check the hypothesis of the model in terms of the growth process. He demonstrated that Gibrat's hypothesis about the distribution of urban growth rates could approximately hold in the case of British urban areas all through the 19th century, but with a bias since the variance of growth rates in the largest size classes of cities was smaller than in the case of small towns. More empirical testing of this model, using the statistical observations of urban growth rates at regular intervals over long periods, have confirmed its relevance for many different countries and periods (Pumain, 1982; de Vries, 1984; Guérin-Pace, 1993; Moriconi-Ebrard, 1993). An illustration of this process of "spatially distributed growth" is provided by comparing two maps of the sizes of European cities between 1850 and 1990: despite two centuries of intense urbanisation, technological innovation, population migration and economic growth, the multiplication of the urban population by a factor 7 and of the number of cities by a factor 4, the relative size of the cities and the spatial configuration of the urban system seems not to have changed (figure 6).

According to this process, the "attractivity" of cities does not have to be assumed to increase with their size (according to an "increasing returns" hypothesis for instance) to explain a highly uneven size distribution, yielding a small number of very large cities. A more satisfactory "explanation" for the shape of the distribution of city sizes is thus obtained when we view it as resulting from a dynamic process for the distribution of urban growth rather than simply as the expression of a static equilibrium. Compared with these static interpretations, Gibrat's model is in fact the only one to show why this form of organisation is unchanging over time, and to demonstrate a gradual concentration of the population in increasingly large cities at the top of increasingly skewed distributions. The model thus provides an "explanation" for the gradual differentiation in size between cities that are involved in the same evolution, though with an uneven effect produced by the "accidental" (or random, in the sense that they are not determined by the model) repetition in some places of incremental increases that are larger or smaller than those experienced by the set of other cities. As a first approximation, therefore, Gibrat's model provides both a good description of growth in an urban system and an explanation for its hierarchical structure. It can be used to predict the evolution of city sizes over periods of up to several decades.

It is tempting to be satisfied with the overall quality of the fit of a random growth model to the development of a system of cities, and to quickly conclude that the urban hierarchy that it produces is another example of the effects of the law of large numbers, or of the unexpected self-organised collective structure that emerges under a "process without intent", since it occurs in many other domains of knowledge, which would constitute yet another case of convergence between structural processes in social and natural sciences, whether or not this is welcome. If a random process like the "law of proportional effect" can explain the lognormal distribution, the stochastic nature of such effects could at first sight be allowed, and it is indeed sometimes asserted in more recent applications of random growth models (Halloy, 2002). Is urban hierarchy a simple consequence of "the law of large numbers", hence trivial, since it is common to so many complex systems? Indeed, transferring models always requires a careful transposition of concepts and measures between the domains of knowledge: the statistical mean growth trend itself should be interpreted in social terms, in its form of a random proportion of the previous mass. According to Halloy, "possibly the primal feature of complex systems is greed (or more euphemistically, resource attraction) and competition is its secondary outcome. It is this resource attraction and competition which in turn determines the primary interactions between agents, as well as the adaptive nature of agents changing rules to outcompete others" (p.2). Therefore, deeper insight into the growth generating process and in the way it is distributed among cities will lead to a less simple explanation. What social

processes can maintain urban growth  $dP$  as an average proportion of the initial size of a city and why should the growth rates ( $dP/P$ ) be randomly distributed over each short time interval? By trying to substantiate the general processes that are involved in the definition of Gibrat's model and to explain the main deviations that were observed in its applications to real systems of cities, it can be seen that an explanation in terms of a generic stochastic process responsible for any type of hierarchy of sizes is far too superficial.

#### *4.2 Implicit specifications in a stochastic model*

First, why should the population growth of a city be proportional to its size for any given short time interval? If we set aside for a moment the question of the spatial expansion of the urban agglomeration, we have two sources of variation in population, which are natural increase and migration. Natural increase of population, without limits relating to resource availability, is clearly a variable of the multiplicative kind, on biological grounds; this is confirmed by the universal use of quotients that always relate the numbers of births or deaths to the magnitude of present population in demographic comparisons. The same is true for migration quotients, but in this case one has to explain why the numbers of people that enter or leave a city during a given period of time are proportional to its population. Clearly, this relates to the laws of spatial interaction as they were summarised in a variety of models, all of the gravitational type. The proportionality between resident population and inward or outward migratory flows which is derived from the multiplication of the population at origin by the population at destination in the numerator of the model can be seen as merely an application of a random interaction process. It is in this sense that A. Wilson (1970) suggested interpreting the spatial interaction model in geography as an expression of the entropy maximisation principle. But the gravity model itself includes the decaying effect of distance on the number of flows in its denominator, and this is usually explained by a principle of minimisation of distance in social spatial interactions, because of constraints of effort, transportation costs, or travel time, or because of availability of information acting on individual behaviours in space.

Another reason for giving more explicit consideration to the effect of geographical space when applying Gibrat's model to city growth emerges if one tries to explain why the urban growth process of a given territory should exhibit, in each short time interval, a general trend represented by the statistical average, which applies to all cities with certain variations. From a social science perspective, this means that all cities belonging to a given territory are expected to share the same temporal trends in relation to growth, and that they are a part of a common trend of urban development. This relates to the mode of definition of a "system of cities" already mentioned in terms of homogeneity, either because the cities are located in a territory where socio-economic conditions are similar, usually because of the similarity in rules in the context of a single political control, or because of many exchanges and interdependencies between the cities. In other words, there is a network where the cities are mutually informed about what is happening in the others, in a context of competition among these interconnected cities for the adoption of innovations. The urban growth process is a more or less continuous adaptation to urban change, either by creating new products or services and imposing them on, or trading them to, other places, or by imitating those that have been created elsewhere. So it is not surprising that in the first book applying Gibrat's model to the British urban system, B. Robson (1973) established a direct connection between the urbanisation process and the diffusion of innovations among towns and cities during the nineteenth century.

In this process, the variance of urban growth rates, as well as the almost random redistribution of growth among cities at each period of time, requires explanation. The fluctuations observed in the growth rates have been related to the process of adaptation to urban change, which is not automatic or deterministic, but has an incremental random character, since cities participate more or less rapidly and intensely in the qualitative ongoing social and economic changes (Pumain, Saint-Julien, 1978). It has been noted that urban change (new techniques, but also new economic activities, professions, physical infrastructures, as well as fashions, cultural practices and so on) is adopted very quickly in cities, nowadays within a time lapse of only a few years. Cities adapt to the changes (that they also contribute to creating) by small discontinuous adjustments: the deviations from the mean change are generally not found in the same cities in two successive time intervals. Local micro-cycles of advance and delay on the scale of each city form a global cycle of innovation in the system of cities overall. Qualitative urban changes (and corresponding quantitative growth) thus diffuse rapidly within the entire urban system and therefore do not alter the initial structure of the system: the relative situations (in terms of size or economic specialisation for instance) remain the same. It has been demonstrated *a contrario* that when these fluctuations cease to be random and become amplified, they can lead to local or general branching in the structure of the urban system, for instance via the emergence of new economic specialisation. For example, new urban specialisations associated with very rapid urban growth have been observed in connection with the corresponding product cycles (Paulus, Pumain, 2002). Therefore, within a system, cities are in competition for the same thing (to attract population and activities, to capture investment or the benefits of innovations). In a sense, they “behave” as if they were “greedy” (Halloy, 2002). Thus behind a process that could apparently be reproduced by a purely stochastic model, there may be a relevant explanation in social terms. If so, is this process totally free, or is it controlled, constrained, or regulated?

#### *4.3 Hierarchical selection, hierarchical diffusion of innovations and space-time contraction*

Even if satisfactory as a simple model that provides a fairly good proxy for the process of urban growth in systems of cities, Gibrat’s model fails to explain a few important features in the process. Systematic deviations have been observed, which invite the enrichment of the model rather than its total rejection.

First, the model expects a progressive reinforcement of the urban hierarchy, determined by the value of the average and variance of the growth rates: the higher these parameters, the larger are the contrasts between city sizes in the resulting urban hierarchy. We have demonstrated however, from various experimentations, that in reality the hierarchies become more contrasted over time than one would expect from the strict application of Gibrat’s hypothesis. A first observation is that the correlation between urban growth rates and city sizes does not vary randomly around zero but always tends to have slightly positive values, even if very low (Pumain, 1982, Guérin-Pace, 1993, Moriconi-Ebrard, 1993, Bretagnolle et al. 2000). A second observation by Robson as early as 1973, is that despite the very low value of the correlation between city size and the growth rates, there is a significant increase in the mean value of growth rates of cities classified according to increasing size, and this observation has since been confirmed many times (figure 7). We have suggested that the historical trend towards a reinforcement of urban hierarchies could be linked with a systematic process that we call “*hierarchical selection*”. This process operates in two ways. The first process is as follows: the emergence of innovations or their early adoption is more likely in large cities (because they generate a higher probability of social interaction, and also because their higher level of social complexity increases the ability to innovate or the likelihood of early adoption); also, the economic growth associated with a new product or service confers the

maximum “initial advantage” of the innovation in its first locations (although this may not occur at the time of the first initiators who run the risk of failure, but rather in the early stages of development); there is thus a dual advantage for the already largest urban agglomerations. As a result the concept of “self sustained urban growth” has been advanced for large cities. Thus the hierarchical diffusion of innovations (a process already noticed by Hägerstrand, 1953) is a first explanation of the trend towards more contrasted distribution of city size in the urban hierarchies. The second explanation is a simplification at the bottom of the urban system, which is associated with an apparent contraction of geographical space: as the speed of transportation increases, travelling time from one places to another is reduced, the spatial range of the sphere of influence of larger cities is increased, and the smallest towns are left out: they lose their market share. As they are also reached later by most innovations and sometimes not at all (for instance railways, and later airports, did not connect all urban nodes), there is an irreversible trend towards the relative and even absolute decay of the smallest towns in urban systems. The trend has been partly masked by different processes: the demographic and economic increases made necessary the increased number of service facilities and sustained the expansion of small towns in absolute terms (but not in relative terms: for example in Europe during the last two centuries, the population increased sevenfold, the number of cities by a factor 4, the income by a factor 14 but the speed of transportation by a factor 40).

Although there were periods in history that hampered the development of large cities in favour of the development of smaller towns (or the emergence of many new towns through increases in the population of villages) as attested by De Vries (1984) for Europe in 17th century for instance, we have observed that the trend towards a reinforcement of urban hierarchies has been dominant over the last two centuries of intense urbanisation and proliferation of innovations. Table 2 brings evidence for France, Europe, and also for India. We have shown that an appropriate delimitation of urban entities is necessary to observe this trend (Bretagnolle et al. 2002). For instance, it is difficult to decide, on the basis of the contradictory results obtained by Batty (2001) for Great Britain, if the decrease in inequalities in city size that he observes is a real exception in Europe (perhaps confirming the efficiency of the policy of “containment of urban England”) or if it merely reflects the fact that the geographical entities in use for measuring urban growth are not normally expanding urban agglomerations but fixed administrative subdivisions.

The historical trend towards a reinforcement of urban hierarchies corresponds to a period of intense urbanisation. However the question is what will happen after the “urban transition” has been completed, when demographic growth and urbanisation rates slow down? It is still possible to imagine a reversal in the trend towards more contrasted urban hierarchies, which would lead to a decrease in the inequalities in city sizes. This might be for instance a statistical consequence of a negative growth rate, as shown in figure 8, where two lognormal curves have been fitted to the distribution of French settlements: whereas the sizes of the growing urban agglomerations are highly contrasted, the range of inequalities (as measured by the slope of the curve) is much narrower among the villages that have been losing population almost continuously for more than one century. However, if urban populations are stabilising, the urban product is likely to continue its growth and the inequalities in economic concentration could still increase. It has to be remembered that during the 1970s, a reversal in urbanisation trends, called “counter-urbanisation” had been predicted (for instance by Berry, 1976), but this was disproved by the further evolution from the 1990’s on, when a new phase of concentration of innovations, investments and skilled jobs in large agglomerations was observed, and termed “metropolisation”. This trend is in fact not new and it characterizes the

start of every new innovation cycle in the urban systems, before the subsequent phase of diffusion.

Of course innovations do not all diffuse equally within the systems of cities. The development of functionally specialised cities is usually linked with different economic cycles which have favoured the growth of particular places : before the industrial revolution, there are references to similar “generations” of cities, that were once driven by the textile industry, trade with the colonies, or since the end of nineteenth century by mass tourism. A recent version of this specialisation process can be seen for instance in the cities engaged in “high-tech”, or finance activities. Even if they were successful during the boom of the specialising activities, such places may have difficulty continuing to adapt further to new cycles of products, or new modes of production, unless they continuously innovate. Some of them may follow the course of small towns that had short-lived success, but were subsequently able to readapt to some innovation and grow again, even if the probability that they would ever challenge the largest metropolises is very small. The largest metropolises usually have very long histories of successful adaptation. For instance, Paris and London in Europe, even if they were state capitals, both benefited from the industrial revolution of the 19th century – probably because they had for a long time been the largest city in their country, and were thus more likely to be open to innovation of any kind (figure 9). The industrial revolution can be seen as a major upheaval for the European urban system because specialised cities, sometimes very large, like Birmingham, or Manchester, or cities in the Ruhr area, were boosted from a status of small towns to the upper ranks of the urban hierarchy. But despite this, it should be recalled that the urban hierarchy in Europe was not totally upset by the industrial revolution: the correlation between the ranks of cities in the middle of eighteenth century, before it started, and their ranks in 1950, long after its end, is very high (coefficient around 0.8! ). The capacity of this system of cities to absorb large upheavals was already attested in the 15th century by its recovery (with respect to both city size and rank) one century after the Black Death had decimated half the population. This is another indication of the number and strength of the links that have for centuries tightly linked the network of European towns and cities, making their co-evolution a fully competitive process long before the continent was equipped with direct and rapid transport connexions throughout. The major transformation that occurred between the 16th and 17th century, transferring the core of the system from the Mediterranean coast to the North Sea (de Vries, 1984, after Braudel) was a very slow process.

Our conclusion is that even if Gibrat’s model remains generic and universal, it can by no means be accepted as a “purely” stochastic process requiring no further explanation. On the contrary, it should be enriched by reference to historical context and trends. First, historical considerations are needed to validate the homogeneity of criteria in relation to the territory that defines the systems of cities under consideration (if in most cases the “territory” is a continuous portion of the earth surface delimited according to long-standing political boundaries, it can also be defined as a network of intensely connected cities, as for instance in the case of today’s “global cities”). Second, the intensity and timing of urban growth should be specified, and also the conditions of spatial interaction (speed of transportation) in order to understand the accentuation of contrasts in city size in relation to the apparent “space-time contraction”. For instance, contrary to a frequent hypothesis, the global level of inequalities in city size, as reflected by the values of parameter  $q$ , are not correlated with the level of economic development. The same average is found for industrialised and developing countries, approximately 1.05. Despite rather large intra-group variations, a significant difference can be found between “old” and “recent” urban systems, the former including those mainly in Europe and Asia, the latter those in America and Australia. This, when related

to average spacing between cities (13 km in Europe against 48 km in USA), can be explained by differences in the age of the settlement systems (Moriconi-Ebrard, 1993). Spacing between towns and cities has been determined by the time necessary for connecting them in the course of development. This is one argument in favour of including the speed of the means of inter-urban transportation in the theoretical conception of urban systems. Moreover, the specific function of some large capitals as gates of communication between different levels of urban systems or as centralising forces for many important functions, also needs to be integrated to understand the phenomena of urban primacy or macrocephaly.

#### *4.4 New types of stochastic models*

Many new models of distributed growth that can generate power laws have been developed recently. The novelty is that they are not mathematical models of aggregated growth, they are an attempt to generate global distributions from agent-based models or multi-agent systems that define rules of interaction at a micro-level. A large variety of applications can be found, often in journals of physics, from stock market and wealth distribution in a population (Solomon, Richmond, 2001), to the size of firms in a nation state (Axtell, 2001) or lengths of words in languages (Cancho, Solé, 2002). However, hypothesis and results should not simply be transferred from one discipline to others, because the selection of relevant variables and processes is very specific, if one wants to represent even in a simplified way the empirical knowledge associated with each field. Below are a few examples among the most interesting suggestions.

S.E. Page (1988) proposed an agent-based model to simulate the emergence of cities, from very simple assumptions about location behaviour of agents, conditioned by a preference for agglomeration and an average distance in relation to other agents. However, in this model an agent's utility is defined, in reference to the distribution of agents on the lattice, and it is not very plausible that real agents could possess this sort of information. Axtell and Florida (2000) provide a more detailed microeconomic multi-agent model of endogenous formation of business firms, allowing agents to move between firms and between clusters of firms (thus assimilated to cities). Under the hypothesis of increasing returns from clustering at the level of the firm, they simulate a size distribution with constant returns (average growth rate) at the aggregate level. A stationary macro-structure is generated from a non-equilibrium microeconomic process. However appealing, because it reconciles two apparently contradictory but observed processes (search for increasing returns at the individual level, no decisive increasing returns at the aggregate level), this model has not been validated from empirical observations.

Anderson et al. (2003) use an algorithm generating « scale-free » networks. This corresponds to a class of growing networks whose node degrees are power-law distributed (Barabasi, 2002). In their model, the nodes of the network represent pieces of land which over time become more and more connected by edges representing exchanges of goods and services (the result of this trade is in fact simulated by a trade benefit or financial investment directed from one node to another). The model proceeds by adding new links between already developed nodes, with a probability of this occurring that is proportional to the relative size of the node in the total number of nodes, and by selecting new nodes. The mean probability of developing existing nodes is significantly higher than that relating to the development of new nodes. Spatial rules are added to specify this selection process, according to hypotheses about a distance-decay interaction model. Thus it is not quite clear whether this model is designed to simulate the urbanisation process on the intra-urban scale, or the formation of urban hierarchies, or both (Pumain, 2004). In any event, the concept of “scale-free” networks, or



“small worlds”, seems well adapted to the simulation of urban systems, since they reveal the hierarchical structure that emerges in progressively constructed networks.

## 5 Evolutionary theory of urban hierarchies

As an alternative approach to static urban theories with their intractable problems of logic, as well as to the over-simplification of the too markedly generic stochastic or agent-based models, our solution would be to conceptualise the urban system as an evolutionary system, which at once is self-adapting to the change that is generated by human societies, and contributes to that change. The urban system is an “invention” the technical nature of which is usually not apparent in collective representations. For centuries we have been using the single word: “cities” to refer to urban agglomerations that have increased their size by several orders of magnitude and have integrated many qualitative changes into their morphology, their social composition and their economic production, or even their symbolic cultural references. A “city” is an admirable territorial adaptor to social change! Like other social systems, urban systems are the product of historical self-organising processes that mix deliberate actions and involuntary outcomes of social interactions. The evolutionary specificity of urban systems cannot be totally dissociated from the intentionality of social actions. Their dynamics are driven by a general expansive trend, which is rooted in social practices aiming to increase symbolic power, available resources and space for action; in cities, this trend is converted into invention intended to reduce the local uncertainties that constrain the development of a site and to search further afield for complementary resources, either in the surrounding territory or in more distant networks; it follows that interurban interactions contribute in turn, by emulation, to hastening the process of globalisation through urban networks and to enhancing the complexification of human activities through the division of labour and specialisation; this trend has already generated a major bifurcation, known as the *urban transition*, which has transformed the way we inhabit the planet by converting a set of settlements whose original function was the agricultural use of a territory, into a much more concentrated, hierarchised and qualitatively differentiated system of towns and cities, as soon as the increase in productivity made it possible to relocate a significant share of the labour force previously engaged in agricultural production. However, even if the expansion of urban entities has connections with intentional processes at individual level, the resulting regular urban hierarchies that emerge from the interactions between cities are not produced by conscious design. Their structure is however constrained by the competitive process of growth, which explains their form, similar to the size distribution of elements in other complex systems, and it is also constrained by the available technological means for connecting human activities in geographical space – and this process is specific to urban systems. The historical trends towards greater speed of communications known as *space time convergence* (Janelle, 2002) has certainly contributed to reinforcing the inequalities in city size, while on a lower scale it has widened the perimeter of urban areas. The historical coincidence between the development of urban systems and the speed of available means of transportation has also created some irreversible differences in urban hierarchies all over the world. Besides these general mechanisms, that are relatively easy to model, there are other systematic variations in urban hierarchies that are related to the political and administrative management of territories, according to the more or less intense centralisation of different powers, which can explain some discontinuities in the hierarchy of urban sizes.

Specific models of the dynamics of urban systems have explored different kinds of tools to simulate the evolution of urban hierarchies. We can briefly recall a few of them, underlining the points that they solve and the problems they still raise.

### *5.1 Self-organisation Models*

Dynamic models of urban systems have been elaborated in the context of self-organisation theories, where formalisation shows up some forms of interdependence between the elements of the system, such as competition to attract activities or population. Using mathematical models of differential equations, the evolution of a set of central places was simulated from the growth rates and movements of population and employment between the urban centres of a region, in relation to relative local attractivity and an exogenous growth trend (White, 1977; Allen and Sanglier, 1979). Other models have simulated the evolution of city sizes from the migration of individuals between centres (Weidlich and Haag 1988; Sanders, 1992). These models have been related to the macroscopic structure of the urban system, for example by demonstrating that under certain hypotheses, the distribution of city size behaves as an attractor for a synergetic dynamic model of inter-urban migration (Pumain and Haag, 1994).

A general criticism that can be made of these models is that they describe *how* change occurs but not *why*. Although the essential mechanism responsible for the form of the urban system is the competition between its component geographical units in attracting and accumulating the product of different resources, and although this competition is made possible by the circulation of information between the units which thus constitute a network, the fact remains that the motor of change is the continuous creation of new products and wealth in the system. Yet the emergence of innovation remains a stumbling-block in dynamic modelling, one that even the most sophisticated efforts at simulation have failed to resolve (Allen, 1991).

### *5.2 Innovation as a motor of hierarchical differentiation*

Mathematical models have proved inadequate to account for creative change, i.e. the intentional transcending of the limits of a dynamic process, which is characteristic of social evolutionary processes, which is why models based on a biological analogy have been no more successful than those borrowed from physics. The transposition of the Lotka-Volterra predator-prey models to the study of competition between regions or cities (Dendrinos and Mullaly, 1985) is only possible if the focus is restricted to the observation of relative variation, in other words if the urban dynamics are reduced to a struggle to secure market shares, or a zero sum game (see also the 'technology substitution' model by Marchetti, 1979). Yet this overlooks, first that the limiting factor present in the Lotka-Volterra ecological competition models is continually challenged by human societies, and second, that temporarily abandoned territories can always find a different use in a new cycle of innovation, thereby invalidating the idea of an analogy with technological substitution between products.

Here can be seen the limitations of analogies with physics and biology for a theory of urban systems. Although models based on physical systems allow two possible processes of change in the structure of the systems, i.e. bifurcation due either to the amplification of an internal fluctuation or to the occurrence of some external disturbance, they have great difficulty in accounting for the crucial phenomenon of innovation. Not only is this to a large extent endogenous to urban systems, but the continuous renewal that it induces means that innovation has a fundamental role in the genesis of the system's structure, through the diversification and complexification of its elements. Consequently it seems inappropriate to consider it simply as a fluctuation or external disturbance. The process has received

considerable attention from economists seeking to replace the general equilibrium theory by an evolutionary approach.

Some authors have also suggested incorporating into the theory not only aspects of physical dynamics but also evolutionary processes based on biological theories. Allen (1991) has argued that towns and cities belong to the types of system in which new forms and functionalities are created. The appearance of innovation in the systems is not due to some optimisation of their functioning at a given time, but results from the practices, “discoveries” and “inventions” of non-average individuals. This is in fact a social interpretation of the notion of *diversity* from the biological theory of evolution. The models based on this theory require very powerful computers to simulate the endogenous emergence of innovation. However, it is not certain that social innovation, and the collective learning processes that it implies, can be modelled in quite the same way as a biological phenomenon. Future models will probably have to include cognitive processes, both for their role in the appearance and selection of innovation, and as additional regulators in evolution. It may then be possible to explain why bifurcations or “chaotic” behaviour are less common in urban systems than in other types of self-organised systems.

### 5.3 *A governance for urban hierarchies?*

We have experienced a computational model using multi-agent systems to generate urban hierarchies. The first version of this model was written in Smalltalk and published in *Geographical Analysis* in 1996 (the first application of multi-agent systems to geography). SIMPOP (Bura et al., 1996) is a multi-agent model that is designed to simulate the emergence, speciation and further evolution (over a period of 2000 years) of a system of towns and cities from a former set of rural settlements. The model is both a dynamic version of central place theory and an extension of the theory to include manufacturing activities. The agents are the individual towns and cities (defined as urban agglomerations), that compete for the acquisition of more and more complex urban functions (mainly trading activities, administration and industrial production). Among each type of function, different levels are distinguished according to the degree of specialisation or the range of service activities. Interactions between agents consist in exchanges of goods, persons and information, that are constrained by distance. Detailed interactions are only partially represented, through market trade (information on demand and supply and mean market price for different types of goods), or summarised by a global balance (measured in terms of wealth and population) in the other cases, providing a variance in urban growth rates for each short period (ten years). In this model, the properties of the distribution of urban size were used only as a test to validate the relevance of the rules, which generate a fictitious urban system. Three major results were established through simulation: 1) the model is able to simulate the emergence of a hierarchy of urban entities, even when the initial conditions include a uniform or normal distribution of settlement size and resources; 2) the model can generate different types of urban systems (according to their hierarchical differentiation and spatial pattern) through slight changes in the rules or the parameters; 3) the (exogenous) appearance of innovation, taking the form of new urban functions, is necessary to maintain the dynamics of hierarchisation within the system.

A new version of the model, SIMPOP2, is now being developed, and includes a number of improvements. The urban entities are no longer only reactive agents that can acquire a representation of their environment as a set of resources, and neighbouring agents that have demand and supply for a variety of goods. They are also cognitive agents, able to choose among different strategies for their development (by investing in new functions or in access

infrastructures for instance) and they produce innovation from emulation (exchange of information, co-operation and competition) of other agents. In this sense, the innovation process, which is essential to the dynamics of urban systems, is rendered partially endogenous. A new function of “urban governance” is created to integrate cognitive and decisional processes, which simulate a variety of possible strategies in the competition process. The strategies can be imposed on certain places or chosen at random. This new version of the model is written on a SWARM platform. The objective is to produce a generic model that can be adapted to a variety of national histories and economies (developed/developing countries; old/recent settlement systems; long term – thousand years- or shorter term – last fifty years- evolution). We also want this model to be able to reproduce coherent relative trajectories for the individual cities (similar to those observed in a variety of examples shown in figure 9), and also to generate a diversity of regional hierarchies such as the ones that were identified in Europe by an analytical survey (see figure 10).

Even if we succeed in incorporating into the model the main processes that we think decisive in the formation of urban hierarchies, and even if this makes it possible to model a few important empirical observations about their evolution and variations, we are well aware that the set of rules selected is open to discussion: other rules could perhaps do as well. The problem of finding a middle pathway between too marked generality and *ad hoc* modelling is not easy to solve. In addition, despite efforts to detail certain social cognitive and decisional processes, we cannot claim that our agents behave in an autonomous and creative way! It therefore seems that classic models and empirical statistical checking remain as important as sophisticated simulation models to accumulate knowledge about urban hierarchies.

In the future, one could imagine that multi-level models could be built to combine processes explaining hierarchical organisation and hierarchical differentiation in a unified way, perhaps through fractal behaviour. M. Batty (2001) used cellular automata to generate poly-nucleated urban areas. We have also attempted to investigate in this direction, by making the spatial process of urban growth more explicit, including centre-periphery effects in a simulation model less detailed than SIMPOP but more powerful (Page et al., 2001). One solution could be to include a model of growth with similar characteristics to the already existing models, which that can generate skewed distribution of size aggregates, but on two different scales according to the intensity of activities enabled by speed of transportation, and according to changes in speed and range over historical time. The fractal dimension of the generated aggregates could be used to validate the model.

## **Conclusion**

The general structure of urban systems, including hierarchical organisation and differentiation, is interpreted in terms of a social evolutionary process : as in biological sciences, one can identify effects of mutation, competition, cooperation and selection, but in this case evolution is also partly driven by the cognitive activity of inventing technical and social artefacts. Improvement in accessibility, directed towards more and more sophisticated activities, in order to reduce uncertainties of life (ecological and social), can be considered as the main constraints acting on the evolution of the system, on both scales of analysis: the city itself, and the system of cities. The action of this organising principle on the spatial structure of the urban systems is almost always indirect : especially at the level of the system of cities, there is no conscious will nor any responsible institution to organise and adapt the system to ensure this increasing accessibility. The global structure and its more or less continuous adaptation emerge from interurban competition.

In this interpretation, the accessibility constraint is viewed as the systemic ordering principle of the spatial structure of urban systems. There is a kind of collective “rationale”, distinct from the actual individual intentions of urban actors. Generally, individual actors try to make a better living by adding value to the urban “heritage” (global value of urban estates, production capital of firms, human capital of resident population, development potential of activities and urban development potential, symbolic values generating urban attractiveness, and so on). This very general aim, at the level of urban actors, produces, when aggregated, an apparent “greediness” at the level of each city, which explains their incremental competitive adaptation to change and the process of creation and capture of innovation which characterises their interactive dynamics within the system of cities. It is this historical competitive process which explains the persistence of a continuum of towns and city sizes. This continuum bears the mark of individual histories of towns in their success or failure in adapting at different stages in their evolution. City size (and the correlated variety and complexity of the activity portfolio and social and cultural sophistication) is the cumulated product of a history. Therefore, the evolution of each city is also constrained by the feedback effects of the organisation of cities into a hierarchical structure within systems of cities.

## References

- Allen, P., 1991: Spatial models of evolutionary systems, in *Spatial analysis and population dynamics*, D. Pumain, ed., :John Libbey-INED, *Congresses and Colloquia* 6, Paris.
- Allen, P., and Sanglier, M., 1979, A dynamic model of growth in a central place system, *Geographical Analysis*. **2**: 256-272.
- Alperovich, G., 1993, An explanatory Model of City-size distribution: Evidence from Cross-country Data, *Urban Studies*. **9**: 1591-1601.
- Amson, J.C., 1972, Equilibrium models of cities, an axiomatic theory, *Environment and Planning A*. **4**: 429-44.
- Anderson, C., Hellervik, A., Hagson, A., and Tornberg, J., 2003, The urban economy as a scale-free network. *ArXiv:cond-mat/0303535* v2.
- Auerbach, F., 1913, Das Gesetz der Bevölkerungskonzentration, *Petermans Mitteilungen*. **59** (1): 74-76.
- Axtell, R.L., 2001, Zipf Distribution of U.S. Firm Sizes, *Science*. **293** (5536): 1818-1820.
- Axtell, R.L., and Florida, R., 2000, Emergent cities: a microeconomic explanation of Zip’s Law.
- Bairoch, P., 1978, *Taille des villes, conditions de vie et développement économique*, EHESS, Paris.
- Bairoch, P., 1985, *De Jericho à Mexico*, Gallimard, Paris.
- Bairoch, P., Batou, J., and Chèvre P., 1988, *La population des villes européennes. Banque de données et analyse sommaire des résultats: 800-1850*. Centre d’Histoire Economique International de l’Université de Genève, 339 p.
- Baker, A.R.H., 1969, Reversal of the rank-size rule : some nineteenth century rural settlement sizes in France, *The Professional Geographer*. **XXI** (6): 386-392.
- Barabasi, A.L., 2002, *The new science of networks*, Cambridge University Press, Cambridge.
- Barbut, M., 2004, A family of distributions : from Paretian to the « contra-paretian ». Application to the study of urban concentration, *Cybergeo*. **266**:30. (<http://www.cybergeo.presse.fr>)
- Batty, M. , 2001, Polynucleated urban landscapes, *Urban studies*. **38** (4): 635-655.

- Batty, M., and Kim, K.S., 1992, Form follows function: Reformulating Urban Population Density Functions, *Urban Studies*. **29** (7)1043-1070.
- Beckmann, M.J., 1958, City hierarchies and the distribution of city sizes, *Economic Development and Cultural Change*, **6** : 243-248.
- Beckmann, M.J., and MacPherson, J.C. , 1970, City size distribution in a central place hierarchy, *Journal of Regional Science*.**10** (1): 25-33.
- Berry, B.J.L., 1964, Cities as systems within systems of cities, *Papers and Proceedings of the Regional Science Association*. **13** :147-163.
- Berry, B.J.L., ed., 1976, *Urbanization and Counter-Urbanization*, Sage Publications, vol. 11, Urban affairs annual review, Beverly Hills, London..
- Botero, G., 1588, *Della cause della grandezza e magnificenza delle città*. Tre libri. New edition in 1598 with « *La Ragion di Stato* » (seventh edition) in Venice, and a cura di Luigi Firpo by Unione Tipografico Editrice Torinese in 1941.
- Bretagnolle, A., 1999, *Les systèmes de villes dans l'espace-temps : effets de l'accroissement des vitesses de déplacement sur la taille et l'espacement des villes*. Université Paris I, thèse de doctorat.
- Bretagnolle, A., Mathian, H., Pumain, D., and Rozenblat, C., 2000, Long-term dynamics of European towns and cities: towards a spatial model of urban growth, *Cybergeo*.**131**, 17 p.
- Bretagnolle, A., Paulus, F., and Pumain, D., 2002, Time and space scales for measuring urban growth, *Cybergeo*. **219**, 12 p.(<http://www.cybergeo.presse.fr>)
- Bura, S., Guérin-Pace, F., Mathian, H., Pumain, D., and Sanders L., 1996, Multi-agent systems and the dynamics of a settlement system, *Geographical Analysis*. **2** :161-178.
- Cancho Ferrer I, R., and Solé, R.V., 2003, Least effort and the origins of scaling in human language, *PNAS*, February, **100** (3): 788-791.
- Cattan, N., Pumain, D., Rozenblat, C., and Saint-Julien, T., 1994, *Le système des villes européennes*, Anthropos, Paris.
- Champion, T., and Hugo, G., eds, *New Forms of Urbanization, Beyond the Urban-Rural Dichotomy*, Ashgate, Aldershot.
- Chorley, R. and Haggett, P., 1967, *Models in Geography*, Methuen, London.
- Christaller, W., 1933, *Die Zentralen Orte in Süddeutschland : eine Ökonomisch-Geographische Untersuchung Über die Gesetz Massigkeit der Verbreitung und Entwicklung der Siedlungen mit Städtischen Funktionen*, Fischer Verlag, Jena, (English translation: *Central places in Southern Germany*, Prentice Hall, Englewood Cliffs, N.J.)
- Cordoba ,J.C., 2003, *On the Distribution of City Sizes*. Working Paper, Rice University, 31 p.
- Curry, L., 1964, The random spatial economy: an exploration in settlement theory, *Annals of the Association of American Geographers*. 138-146.
- Dahl, R., 1961, *Who governs?*, Yale University Press, New Haven.
- Dendrinios, D., and Mullaly, H., 1985, *Urban evolution*. Oxford University Press, Oxford.
- Derycke, P.-H., Hurriot, J.-M., and Pumain, D. ,eds, 1996, *Penser la ville*, Anthropos, Paris.
- Dureau, F., Dupont, V., Lelièvre, E., and Lévy, J.P., 2003, *Métropoles en mouvement*, Anthropos, Paris.
- ESDP, 1999, *European Spatial development Perspective. Towards Balanced and Sustainable Development of the Territory of the European Union*. Approved at the Informal Council of Ministers responsible for Spatial Planning in Posdam, May,European Community, Luxemburg.
- Fletscher, R., 1986, Settlement in Archaeology: world-wide comparison, *World Archaeology*. **18** (1): 59-83.
- Folin, M., 2003, Hiérarchies urbaines/hiérarchies sociales : les noms de ville dans l'Italie moderne (XIVe-XVIIIe siècles), *Genèses* **51** :4-25.
- Freire, M., and Stren, R., 2001, *The Challenge of Urban Governmen*, World Bank Institute, Washington.
- Friedmann, J., 1986, The world city hypothesis, *Development and Change*. **17**: 69-84.

- Fujita, M., Krugman, P. and Mori, T., 1994, *On the evolution of hierarchical urban systems.*,: North American meeting of the Regional Science Association, Niagara Falls.
- Fujita, M., Krugman, P., and Venables, A.J., 1999, *The Spatial Economy*, The MIT Press, Cambridge.
- Gabaix, X; Ioannides, Y.M., 2003, The evolution of City size distributions, in :*Handbook in Urban and Regional Economics*, V. Henderson and J.F. Thisse, eds, vol IV, North Holland Pub., Amsterdam.
- Gibbs, J.P., 1963, The evolution of population concentration, *Economic Geography*. 119-129.
- Gibrat, R., 1931, *Les inégalités économiques*. Sirey, Paris.
- Goodrich, E.P., 1926, The statistical relationship between population and the city plan, in : *The urban community*, E.N. Burgess, ed., Chicago, pp. 144-150.
- Guérin-Pace, F., 1993, *Deux siècles de croissance urbain*, Anthropos, Paris.
- Guérin-Pace, F., and Lesage, X., 2001, Des différentes mesures de l'inégalité de distributions de type paretien à partir de l'étude du système urbain français, *Histoire et Mesure*. **XVI**(1-2) : 157-183.
- Guérois, M., 2003, *Les formes des villes européennes vues du ciel. Une contribution de l'image CORINE Land cover à la comparaison des grandes villes d'Europe occidentale*. Université Paris I, thèse de doctorat.
- Jefferson, M., 1939, The law of primate city, *Geographical Review*. **29** : 226-232.
- Haggett, P., 2001, *Geography, a global synthesis*, Pearson Education, Prentice Hall.
- Hall, P., 1984, *The World Cities*, Weidenfeld and Nicholson, London.
- Halloy, S., 2002, The lognormal is a universal descriptor of unconstrained complex systems. New Zealand, Crop and Food research.
- Huriot, J.M., and Thisse, J.F., eds, 2000, *Economics of Cities*. Cambridge University Press, Cambridge.
- Kohl, J.G., 1841, *Der Verkehr und die Ansiedelungen der Menschen in ihrer Abhängigkeit des Gestaltung der Erdoberfläche*, Arnold, Dresden/Leipzig.
- Lalanne, L., 1863, "Essai d'une théorie des réseaux de chemin de fer, fondée sur l'observation des faits et sur les lois primordiales qui président au groupement des populations", *Comptes rendus des séances de l'Académie des sciences*. **57** : 206-210.
- Lepetit, B., and Pumain, D., 1999, *Temporalités urbaines*, Anthropos, Paris (first ed. 1993).
- Lotka, A.J., 1924, *Elements of physical biology*. Baltimore.
- Lotka, A.J., 1941, The law of urban concentration, *Science*. **94**:164.
- Lösch, A., 1940, *Die Räumliche Ordnung der Wirtschaft*, Fischer, Iena.
- Mackinder, H.J., 1902, *Britain and the British Isles*, W. Heinemann, London.
- Marchetti, C., and Nakicenovic, N., 1979, The dynamics of energy systems and the logistic substitution model. IIASA, Research report 79-013.
- Moriconi-Ebrard, F., 1993, *L'urbanisation du monde depuis 1950*, Anthropos, Paris.
- Moriconi-Ebrard, F., 1994, *Geopolis, pour comparer les villes du monde*, Anthropos, Paris.
- Page, S.E., 1998, On the emergence of cities. *Santa Fe Institute, Research in Economics*, 98-08-075a.
- Page, M., Parisel, C., Pumain, D., and Sanders, L., 2001, Knowledge-based simulation of settlement systems, *Computers, Environment and Urban Systems*. **25** (2): 167-193.
- Parr, J.B., 1970 Models of city size in an urban system, *Papers and Proceedings of the Regional Science Association*. **25**:221-253.
- Parr, J.B., 1985, A note on the size distribution of cities over time, *Journal of Urban Economics*. **18**:195-212.
- Paulus, F., and Pumain, D., 2002 Répartition de la croissance dans le système des villes françaises, *Revue d'Economie Régionale et Urbaine*. **1**:35-48.
- Pred, A., 1966, *The spatial dynamics of US industrial growth 1800-1914*, MIT Press, Cambridge.

- Pred, A., 1973, Systems of cities and information flows, *Lund Studies in Geography*, Serie B. **38**.
- Pred, A., 1977, *City systems in advanced societies*, Hutchison, London.
- Pumain, D., 1982, *La dynamique des villes*, Economica, Paris.
- Pumain, D., 1992, Les systèmes de villes, in : *Encyclopédie de géographie*, A. Bailly, R. Ferras and D. Pumain eds, Economica, Paris, chap. 34, pp.645-664.
- Pumain, D., 1997, Vers une théorie évolutive des villes, *L'Espace Géographique*. **2** :119-134.
- Pumain, D., 2004, *Scaling laws and urban systems*. Santa Fe Institute, Working Paper n°04-02-002, 26 p.
- Pumain, D., and Gaudin, J.-P., 2002, Systèmes de villes et pouvoir. L'analyse de Giovanni Botero à l'époque de la Renaissance, *Cybergeo*. **227** : 19 p.
- Pumain, D., and Moriconi-Ebrard, F., 1997, City Size distributions and metropolisation, *Geojournal*. **43**(:4) : 307-314.
- Pumain, D., and Haag, G., 1994, Spatial patterns of urban systems and multifractality, in *Evolution of Natural Structures*, 3rd International Symposium of the Sonderforschungsbereich 230, Universität Stuttgart, 4-7 october, Natürliche Konstruktionen, Mitteilungen des SFB 230, Heft 9: 243-52.
- Pumain, D., and Robic, M.-C., 1996, Théoriser la ville, in : *Penser la ville*, P.-H. Derycke, J.-M. Huriot and D. Pumain eds., Anthropos, Paris, pp. 107-161.
- Pumain, D., and Saint-Julien, T., 1978, *Les dimensions du changement urbain*, CNRS, Paris.
- Quandt, R.E., 1964, Statistical discrimination among alternate hypothesis and some economic regularities, *Journal of Regional Science*. **5** (2): 1-23.
- Reclus, E., 1895, "The evolution of cities", *The Contemporary Review*. **67** (2): 246-264.
- Reymond, H., 1981, 'Une problématique théorique de la géographie: plaidoyer pour une chorotaxie expérimentale', in : *Problématiques de la Géographie*, H. Isnard, J.-B. Racine and H. Reymond eds, P.U.F., Paris.
- Reynaud, J., 1841, Villes, in *Encyclopédie nouvelle*, Gosselin, Paris, vol. VIII, pp. 670-687 (discovered by Robic, 1982).
- Robic, M.-C., 1982, Cent ans avant Christaller, une théorie des lieux centraux, *L'Espace Géographique*. **1** : 5-12.
- Robson, B.T., 1973, *Urban Growth, an approach*, Methuen, London.
- Roehner, B.M., 1991, The long-term trend toward increased dispersion in the distribution of city sizes, *Environment and Planning*. **23**:1725-1740.
- Roehner, B., and Winiwarter, P., 1985, Aggregation of independent Paretian random variables, *Advances in Applied Probability*. **17** :465-469.
- Rosen, K.T., and Resnick, M., 1980, The size distribution of cities: an examination of Pareto Law and Primacy, *Journal or Urban Economics*. **8**:165-186.
- Rousseau, M.P., 1998, *La productivité des grandes villes*, Anthropos, Paris.
- Rozenblat, C., and Cicille, P., 2003, *Les villes européennes : analyse comparative*, DATAR-La documentation Française, Paris, 97 p.
- Sanders, L., 1992, *Système de villes et synergie*, Anthropos, Paris.
- Sassen, S., 1991, *The Global city: New York, London, Tokyo*. Princeton University Press.
- Singer, H.W., 1936, The "courbe des populations": a parallel to Pareto's law, *Economic Journal*. **46** : 254-263.
- Simon, H., 1955, On a class of Skew distributions, *Biometrika*. **42**: 425-440.
- Solomon, S., and Richmond, P., 2001, Power laws of Wealth, Market Order Volumes and Market Returns, *ArXiv:cond-mat/0102423 V2*, 2 Apr.
- Steindl, J., 1965, *Random processes and the growth of firms*. Hofnir, New York.
- Tornqvist, G.E., 1973, Systems of Cities and Information Flows, *Lund Studies in Geography*, Serie B. **38**.



- Ullman, E., 1954, Geography as spatial interaction, in *Interregional linkages. The proceedings of the Western Committee on Regional Economic Analysis*. Berkeley, California, 63-71,
- United Nations, 1978, *Universal Instructions for Population Censuses and Households*. New York
- United Nations, 2004, *Urban hierarchy*. Annual Report, Population Division, World Urbanization Prospects, 75-87.
- Vauban, 1707, *La Dîme Royale*. Paris.
- von Thünen, J.H., 1826, *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*, Perthes, Hamburg.
- de Vries, J., 1984, *European urbanization, 1500-1800*, Methuen, London.
- Weidlich, W., and Haag, G., 1988, *Interregional Migration*, Springer Verlag, Berlin.
- White, R.W., 1977, Dynamic central place theory, *Geographical Analysis*. **10**: 226-43.
- Whitehand, J.W.R., 1987, *The changing face of cities*, Basil Blackwell, Oxford.
- Willis, J., and Yule, G., 1922, Some statistics of evolution and geographical distribution in plants and animals, and their significance, *Nature*. **109** :177-179.
- Wilson, A.G., 1970, Entropy in urban and regional modelling, Pion, London.
- Yule, G.U., 1924, A mathematical theory of evolution, *Philosophical Transactions of the Royal Society*, Serie B. **213** : 21-87.
- Zipf, G.K., 1941, *National unity and disunity*, Principia Press, Bloomington (Indiana).
- Zipf, G.K., 1949, *Human behaviour and the principle of least effort*, Addison-Wesley Press, Cambridge (Mass.).