METHODOS SERIES

VOLUME 3

Hierarchy in Natural and Social Sciences

Edited by
DENISE PUMAIN

Springer

HIERARCHY IN NATURAL AND SOCIAL SCIENCES

METHODOS SERIES

VOLUME 3

This Book Series is devoted to examining and solving the major methodological problems social sciences are facing. Take for example the gap between empirical and theoretical research, the explanatory power of models, the relevance of multilevel analysis, the weakness of cumulative knowledge, the role of ordinary knowledge in the research process, or the place which should be reserved to "time, change and history" when explaining social facts. These problems are well known and yet they are seldom treated in depth in scientific literature because of their general nature.
So that these problems may be examined and solutions found, the series prompts and fosters the setting-up of international multidisciplinary research teams, and it is work by these teams that appears in the Book Series. The series can also host books produced by a single author which follow the same objectives. Proposals for manuscripts and plans for collective books will be carefully examined.
The epistemological scope of these methodological problems is obvious and resorting to Philosophy of Science becomes a necessity. The main objective of the Series remains however the methodological solutions that can be applied to the problems in hand. Therefore the books of the Series are closely connected to the research practices.

Hierarchy in Natural and Social Sciences

edited by

Denise Pumain

University of Paris I,
Panthéon Sorbonne, France
Institut Universitaire de France

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN-10 1-4020-4126-8 (HB)
ISBN-13 978-1-4020-4126-6 (HB)
ISBN-10 1-4020-4127-6 (e-book)
ISBN-13 978-1-4020-4127-3 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springeronline.com

Printed on acid-free paper

Printed in the Netherlands.

Table of contents

Contributing Authors

Michael BATTY
Bartlett Professor of Planning
University College London
Centre for Advanced Spatial Analysis (CASA)

University College London
1-19 Torrington Place
London WC1E 6BT
UK

m.batty@ucl.ac.uk

Bruno GAUME
Researcher
Laboratoire IRIT-CNRS, Université de Toulouse

118, route de Narbonne
31062 Toulouse Cedex
France

gaume@irit.fr

David LANE
Professor of Economics
University of Modena and Reggio Emilia
External faculty, Santa Fe Institute, USA

Cannaregio 4392
Venezia 30121
Italy

lane@unimo.it

Alain PAVÉ
Senior Scientist, Director,
CNRS-Guyane, Programme Amazonie
Member of the National Academy of Technologies and of the French Academy of Agriculture

Résidence le Relais
16, avenue André Aron
97300 Cayenne
France

alain.pave@cnrs-dir.fr

Denise PUMAIN
Professor,
Institut Universitaire de France, University Paris I

13 rue du Four
75006 Paris
France

pumain@parisgeo.cnrs.fr

Fabienne VENANT
PhD Student,
Laboratoire Lattice-CNRS, ENS, Paris

1 rue Maurice Arnoux
92120 Montrouge
France

Fabienne.Venant@ens.fr

Nicolas VERDIER
Researcher,
CNRS, UMR Géographie-cités

13 rue du Four
75006 Paris
France

nicolas.verdier@parisgeo.cnrs.fr

Bernard VICTORRI
Senior Scientist,
CNRS, Lattice, ENS Paris

1 rue Maurice Arnoux
92120 Montrouge
France

Bernard.Victorri@ens.fr

Geoffrey B. WEST
President,
Santa Fe Institute

Theoretical Division
T-8, MS B285
Los Alamos National Laboratory
Los Alamos, NM 87545
USA

gbw@pion.lanl.gov

INTRODUCTION

Denise PUMAIN
Professor,
Institut Universitaire de France, University Paris I

Hierarchy is a type of systemic organisation into levels that are ordered with reference to criteria of a normative character, and fully or partially subordinated by relationships of power, influence, or control. Many intermediate situations are observed between strictly designed hierarchies where levels are distinct and communication restricted to vertical top-down command, and more flexible structures where the levels are not easily recognisable and the network circulation of information reveals unequal degrees of accessibility or control only after a detailed analysis. We are interested in hierarchies because they are at the core of many complex systems. Complex systems are not easy to define, and applications of this concept can be found in realms as diverse as computer networks, natural languages, biological systems, business organisation, cities and regions. Complex systems may have more or less simple components, but they always exhibit an overall behaviour that is difficult to analyse and predict from the list of components and the interactions between them. We think that the analysis of the hierarchical organisation of complex systems can provide deep insight for understanding their emergence and evolution. Therefore the purpose of this work is to explore the concept of hierarchy in different fields of science, looking for possible common processes and general explanations.

HIERARCHY IN COMPLEX SYSTEMS

Exploring complexity puts a focus on the objectives of a number of bodies that have been recently created for the purpose of developing a field of knowledge that is thought to be transversal to many sciences. This ambition is mentioned on the website of the earliest of these institutions, the Santa Fe Institute for Complex Systems (created in 1984): “transcending the usual boundaries of science to explore the frontiers of knowledge” A still more ambitious phrasing appears on the website of the European Complex Systems Society (created in 2004): “towards a science of complex systems”.

D. Pumain (ed.), Hierarchy in Natural and Social Sciences, 1-12.

According to Paul Bourgine and Jeffrey Johnson, “there are two kinds of interdisciplinarity within complex systems. The first begins with a particular complex system and addresses a variety of questions coming from its particular domain and points of view. The second begins with questions that are fundamental to complex systems in general. The first leads to domain-specific interdisciplinary fields such as cognitive science. The new science of complex systems belongs to a second kind of interdisciplinarity. It starts from fundamental open questions relevant to many domains, and searches for methods to deal with them. These two kinds of interdisciplinarity are complementary and interdependent: any advance in one is valuable for the other.”

Without deciding at once if a unified science of complex systems of this nature is even possible, in this book we propose an approach of the concept of hierarchy that is based on specific disciplinary objects or methods and at the same time intended to provide answers to common or transversal questions. We think that the widespread occurrence of hierarchical structures in nature and in human society invites a pluri-disciplinary inquiry, to explore the dynamic or evolutionary processes that could explain them. Are the processes leading to a given kind of structural effect common or distinct? The first focus of this book is on alternative explanations for hierarchical structures in different fields of knowledge. As such, it is positioned as the third volume in the Methodos Series, the objective of which is to provide bridges between disciplines by sharing conceptual tools and methods.

Several methods for the comparative study of hierarchies have originated in the field of natural sciences, physics or biology, where mathematical tools have been developed. Of course the point is not to look in the field of “hard sciences” for a justification of any social order which could thereby be viewed as “natural”. As a first methodological principle, it is important to agree that the intentional, political, or more generally cognitive aspects cannot be forgotten when considering social systems. However, alongside discussion of the *raison d’être* of social hierarchies, there is a need for instruments or technical references. Can we learn from the methods and concepts that are used in natural sciences to explore the dynamics of social systems? Conversely, are there aspects of intentionally designed social systems that could provide a better understanding of natural processes? Obviously, only very broad concepts can be shared and they need to be adapted to different contexts. First of all, we need to define a general framework by clarifying the usage and acceptations of the word hierarchy.

HIERARCHY: THE WORD AND THE CONCEPTS

What is meant by hierarchy? Hierarchical structures are part of our common experience, in the physical as well as in the living and social worlds. There are distinct levels in nature between the infinitely small (atoms or elementary particles) and the infinitely large (galaxies and universe); there is a hierarchy of living entities, from cells to organs, organisms, species, populations and ecosystems; many social organisations like firms or administrative services display pyramidal layouts defining unequal degrees in power or competence between the levels; differences in size, wealth and power establish a hierarchical order between cities in a region or between countries in the world, *de facto* if not *de jure,* which may persist over long periods of time; various distinctive attributes define hierarchies of social status among groups or individuals in many societies. What we observe or think is never limited to a single scale of facts, it is not made up of an amorphous distribution of elements belonging to a single scale, but it is instead organised into more or less distinct levels that define a variety of scales in space and time and can in some cases be ordered according to these scales.

In the current usage, three types of meaning are regularly attached to the term hierarchy:

- according to the first, hierarchy is a matter of order and subordination. There is a relationship of subordination between each level composing this type of hierarchy. The etymology of the term (Greek hieros, meaning sacred, and archê, meaning government) refers to the divine order among angels. The term was then transferred to relationships of obedience between positions in the Christian Church administration (simply known as "The Hierarchy" without any further qualification as late as the 19th century). The organisation of armies is still based upon subordination principles. Remnant connotations referring to links of dependency, or controlling power, between decision and execution, employers and employees, or to nested cascades of responsibilities, may still appear in the usage of the word when referring to the internal structure of institutions, such as firms, and public or private administrative bodies, even when the hierarchical organisation is no longer designed according to a general subordination principle.
- A second acceptation appeared possibly as early as 17^{th} century and was consolidated during the 18^{th} century, probably in the field of emerging natural sciences. It is used for taxonomy (for instance in the classification of living species by Linnaeus), and means simply an order between elements which are classified in different nested categories, after criteria of less and less importance. In these criteria, there is an explicit or

sometimes implicit reference to a system of values or a norm that legitimate the ordering of the categories into levels. But there is no longer any explicit relationship of subordination between the categories. This definition applies to many social organisations, where different levels may correspond to various degrees of power, influence, social status or information.

- The third acceptation simply refers to the organisation of a set into an ordered series of elements where each term is superior to the following according to some normative character. Ranking and classification methods demonstrate hierarchical orders of this type.

Following this third usage, a more recent extension of the word defines as "hierarchical" systems which have a large number of elements and can be described at least at three levels of observation: at the micro-level are the elements (particles, individuals), at the meso-level are the subsystems made up of a variable number of elements (grouped after some criterion, so that subsystems have a certain autonomy, even if it may be only partial or temporary) and at the macro level the system itself is a collection of subsystems, which also has some properties making it durable and recognisable. The structure is sometimes defined from the number of subsystems of a given size within the system, size being measured by the number of elements in each subsystem. This structure can be summarised by a statistical distribution, which is generally highly skewed (asymmetrical). Well-known examples are the distribution of firm sizes in a given sector of activity, or the distribution of settlement sizes in a given territory. Other examples can be found in the living world (number of species made up of a variable number of individual organisms in a genus for instance), in the physical world (numbers of stars according to size in galaxies), or in artificial worlds like large computer networks.

However useful it might be to bear in mind the distinction between heavily connoted and more neutral meanings attached to the word hierarchy, it must be recognised that these acceptations cannot be completely separated, even in natural sciences. For instance, according to the progress of knowledge about evolution, the early criteria, mainly based upon morphological characters and considered as "important" as a basis for the branched hierarchy of species, have been replaced by genetic considerations, which question the tree-like pattern of the classification representing biological evolution. In the field of social sciences, hierarchies are not based on "pure" statistical differences in size, since various connotations of value, power, moral authority, ability to harm, or risk are always closely associated, to a variable extent, with size.

Therefore, if we leave aside the question of subordination, the concept of hierarchy retains two main distinct meanings: when applied to a social organisation, it relates to a pyramidal configuration, entailing a top-down or bottom-up circulation of control and information. When applied to an open system, the elements of which are grouped into subsystems, it means that there is an order of some kind between the subsystems, according to a nested or inclusion relationship, or through a regular differentiation among the subsystems classified by size, usually following a Pareto-type or lognormal statistical distribution. Mass inequalities are often correlated with qualitative differences that can be summarised by scaling properties. Sometimes, distinct levels can be identified, justifying the use of multilevel methods of analysis (Courgeau, 2004); in other cases, there is a continuum in the hierarchical order.

HIERARCHY, SOCIAL ORGANISATION AND NETWORKS

As the antonyms to hierarchy are anarchy, disorder, but also equality, there is a paradoxical situation in relation to hierarchy in the modern world: although recognizable everywhere, hierarchy is very often denied, ignored or fought against in current political or social thinking. As a tool for co-ordination or governance in social organisation, it is often rejected on a plea of ideas of democracy, equality or equity. As an organisational pattern for administrations or firms, hierarchy has also been challenged in terms of efficiency and adaptability by a newly-marked preference for networks. While hierarchy is merely a specific type of network, the associated semantic connotations are very different: hierarchy is seen as a rigid organisation, difficult to adapt, not allowing enough scope for initiative and giving too little gratification to the actors at each level, whereas networks are supposed to provide more flexibility and adaptability, by ensuring a better circulation of information, not only top-down or bottom-up but horizontally as well, thus providing a wider participation of members in initiative and decision. Command and authority are associated with hierarchy, while cooperation, consensus and mutual interest are felt to characterize the coordination mode of networks. A third distinct model for coordinating social activities is the market, where the underlying processes depend on competition and prices (Thompson, 2003).

Hierarchy used to be set against market in a global approach in terms of transaction costs. Hierarchy is thus seen to represent a deliberately guided governance system, while market order is seen as spontaneously generated. Ordered outcomes are expected *ex ante* from a hierarchical organisation, while they are the *ex post* consequences in a market system. However, since the work by the Austrian school in economics which put emphasis on the dynamics of markets, conceived as continuously changing and adapting via equilibrium processes, with emphasis on the role of information, there has been a shift of interest in economics from "pure" market forms towards real or simulated networks. Between hierarchy and markets, networks are thought to represent an alternative intermediate structure organised into a number of levels and ruled by a large variety of "institutions". In entrepreneurial management, more and more attention is given to "embeddedness" of actors in social networks (Granovetter, 1985) which may facilitate or hamper innovation (Uzzi and Spiro, 2004). In sociology as well as in political sciences, social network analysis has developed considerably during the last fifty years (White, 2004 and INSNA). The emphasis on networks does not mean however that hierarchy has disappeared. Network analysis focuses on the rules of coordination and governance that allow the functioning of these organisations, but at the same time underlines all the relational asymmetries, and the differential access to power and authority or to information resources that occur in any social organisation. The information theory has developed as a means of demonstrating the circulation of power in political science. Regulation theories focus on the *de facto* hierarchies emerging from various reciprocal learning processes, on asymmetrical structures in information exchanges, and on polarisation in networks (Gaudin, 2002). Within large networks of interactive entities (like Internet for example), there is frequently strong empirical evidence of fractal structures or scaling laws, related to self-organisation processes generating hierarchical structures. The models recently developed by physical scientists and mathematicians for analysing large networks of this type provide new tools for comparing complex hierarchical relations (Watts, 1999, Barabási, 2002).

IN SEARCH OF EXPLANATION

In the preceding books in the Methodos series, Robert Franck (2002) and Daniel Courgeau (2004) have focused on the fertility of abstract transdisciplinary approaches by way of the methodological transfer they provide between disciplines. Robert Franck emphasizes the need to construct better links between theoretical and empirical investigation so as to improve the explanatory power of models used in social sciences. This is certainly a first essential step towards improving our understanding of hierarchies. Most chapters in the present book base their theoretical considerations upon large sets of empirical data. Without totally neglecting qualitative approaches to hierarchies, preference has been given to selecting domains of analysis where quantification is possible, enabling interdisciplinary transfer and comparison of quantitative methods. We have not included precise examination of “enacted” hierarchical organisation, such as administrations, managerial organisations, or political bodies. The option was to explore “spontaneous”, self-organised hierarchical structures emerging from types of interaction that are assumed not to be designed according to the rules of a pyramidal order.

R. Franck suggests that methods of reverse engineering (reproducing the functional structure of a system starting from the observation of its properties), could provide good heuristics for social sciences. In the case of hierarchies, the problem seems at first a little different, since we already know the structures, while we need to understand the processes that build and maintain them. Are they merely the inherited traces of previous less democratic ways of functioning? Or do they reflect deeper and more persistent trends in social organisation? To what extent are they similar to observed types of organisation in some biological or physical systems? Are the underlying processes comparable or totally different? Certainly the methods of observation and measurement as well as the conceptual construction of the underlying dynamics, can be usefully investigated and partially borrowed from natural sciences. We shall not return here in detail to the difficult question of identifying levels of hierarchy and delimitating these levels in social sciences, since this is thoroughly discussed in D. Courgeau’s book about multilevel methods of analysis, but of course this major problem will be scrutinised in relation to each of the various objects considered in the relevant chapters.

Why are hierarchical organisations so frequent in the natural or social systems that we observe? Three hypotheses, which could constitute common answers from various disciplines, may be suggested:

- hierarchies are just our way of perceiving and understanding our environment
- hierarchies are spontaneous attractors in unconstrained dynamic random processes
- hierarchies represent the best solution for many optimisation problems

These three hypotheses are not necessarily mutually exclusive. They will be explored systematically in all chapters of the book. The concept of hierarchy may become part of an explanation when specific constructive or evolutionary processes are related to this type of organisation or structure for a given complex system. Careful review of the explanations which have been suggested, for example in the case of the organisation of firms, or the hierarchical differentiation of urban systems, or the size or number of living species, are provided in different chapters. This can be done by accurate comparison of the characteristics of hierarchical structures with existing explanatory models, both for natural and social phenomena. The discussion and comparison of models used between social and natural scientists can contribute to identifying possible common processes and specific features in systems presenting hierarchical properties.

METHODOLOGICAL QUESTIONS

In the investigation of hierarchies using methods and tools that have been elaborated to explore the dynamics of complex systems, general questions are emerging to which no satisfactory answer has yet been found in any field of research.

Are hierarchies produced by constraints, optimisation, or randomness?

What explanation can be retained for scaling behaviour or fractal structures that are observed in complex systems? Are they mainly produced by random processes? Are they common attractors in non-linear dynamics? Or do they on the contrary reflect the action of some underlying general or specific optimising process? For G.B. West at the Santa Fe Institute, "scaling laws are revealing the universal laws of life". In the social world, it is sometimes claimed that independent decisions could, without any intention or constraint, randomly produce the same kind of structures as those invented by engineers or selected in the course of evolution. But it is possible that hidden constraints do act behind what are described as random processes. What methods can be used to detect the constraints that operate in the formation of scaling laws? Another important question is that, in social sciences, the two acceptations of hierarchy may often be associated in the explanations put forward, but how far is this possible? What is deterministic and what is random in the processes generating hierarchies? In other words, which part of the organisation is constrained, which is intentional, and which is not?

Continuous or discrete hierarchical organisation?

An important issue is the fundamental differences in organisational terms between scaling or fractal structure and multi-level differentiation. Are there two different types of hierarchy? How can the co-existence of the two structuring principles in some systems be explained, for instance in urban systems? What kinds of interaction between elements at a micro level lead to each type of structure at the upper level? The task here is to overcome the difficulties that were encountered in the application of analytical statistical methods, the success and limits of which have been well explored by Daniel Courgeau (2004): "The multilevel approach provides a solution to the problems that occur when working on a single level. It avoids the risks of the ecological and atomistic fallacies by accommodating the effects of characteristics operating at different aggregation levels. By contrast, when we want to examine the set of dynamic, reciprocal, and non-linear relationships that exist inside each level and between levels, the multilevel approach still seems too limited in its present form." Can we find new methods to answer questions of the type: how are the emerging properties at one level related to the existing behaviour at another level?

How do hierarchies evolve?

We want to explore the rules applying to hierarchical organisations in their evolution. What are the processes that maintain them? In particular, what are the constraints that eliminate the elements that do not fit into the hierarchical structure, how are the feedback effects through which the structure controls the elements in the system produced? What are the respective roles of aggregation or concentration, fragmentation, splitting or spread, birth and death, migration, competition, differentiation, innovation, adaptation? Can these appellations be defined and related in the same way in different disciplinary fields?

To sum up, one main objective of the book is to explore the analogies between concepts that are in use for a variety of systems, but which have not always been defined in detail, nor compared. The entropy maximisation principle, fractal structures, allometric scaling, the self-organisation theory, and competitive evolution all provide models for structures and processes; however agreement with theories and observations in specific domains of science needs to be explored. Theories are important not only for improving our understanding, but also because, according to their underlying hypotheses, the types of intervention that are possible for controlling the systems will be very different. Thus our book can contribute to an improvement in governance techniques for complex systems.

ABOUT THE AUTHORS[1]

Specialists from different disciplines review problems and methods in connection with hierarchy in their fields of research. Of course many more disciplines could have been invited to discuss concepts of hierarchy and related methods of analysis. We have selected here contributions that can provide an added value in terms of cross comparisons, because they present approaches that are compatible in terms of the objects they analyse and the methods that are applied. A first approach is historical, via an overview of the social representations present in usage of the word "hierarchy". Nicolas Verdier, a young historian at CNRS (chapter 1) recalls which connotations

[1] We acknowledge here the ISCOM Programme of the European Commission directed by David Lane for facilitating meetings between three members among the authors.

were associated with the word in the history of Western thought, until the beginning of 20th century. By then the usage of the term seems to be fixed in its present-day acceptations, but it underwent a radical change in meaning at the turn of the 18th to the 19th century.

Natural sciences are represented by biology, and by inter-relationships between biology and physics. Alain Pavé is an ecologist, and director of a programme on the quality of environment in French Guyana. In chapter 2 he explores hierarchical organisation in biological and ecological systems. He demonstrates how differences in size of living systems are associated with spatial and temporal scales in their building processes. He discusses self-organisation into networks as a major explanatory mechanism, which needs to be interpreted in terms of cooperation as well as competition. Some of the questions he raises about the role of chance and necessity in hierarchical organisation are answered by Geoffrey West (chapter 3) who gives an explanation for the universal scaling laws in biological systems. Geoffrey West is a physicist belonging to Los Alamos National Laboratory and the Santa Fe Institute, and he has brought new insight into scaling laws in biological systems. He emphasises the development of a methodological chain defining the appropriate measurements and formalism for establishing scaling laws, in order to understand the elementary constraints that can explain the success of hierarchical organisations. In biology as well as in fundamental physics, the question of minimisation of energy losses by generating tree-like transportation systems that have specific branching properties seems crucial. His chapter is in favour of relating hierarchical organisation or differentiation to optimisation processes that are constrained by physical mechanisms.

Social sciences are considered to have a broader variety in their approaches of hierarchies. Four chapters are devoted to sociology, linguistics, urban planning and geography. David Lane, as a statistician and economist, brings a general outline of analogies and differences between hierarchy in natural and social sciences (chapter 4). He underlines important structural differences in the hierarchical organisations of biological and social systems: social organisations are "heterarchies" rather than inclusive hierarchies. For him, tangled hierarchies are generic to complex systems and not all complex systems necessarily display self-similarity across scales. He reviews several amongst the most well-known interpretations that have been suggested to explain the emergence of hierarchy as organisational form in complex systems. He develops a methodology for studying the relationships between levels in hierarchical social organisations. Specific networks that he calls “scaffolding structures” are essential in the evolution of the system, since they are the channels through which social innovation is driven.

Natural languages are somehow “in between” natural and social systems: they were developed by human beings and thus do involve complex cognitive processes as well as social interaction, but their origin is very ancient and most of their evolution seems to have occurred without explicit intentional control. Linguistics is a social science where modelling has been widely developed, mainly in network terms. Bruno Gaume, Fabienne Venant and Bernard Victorri, specialists in linguistics at CNRS (chapter 5) propose a method for exploring the structure of a graph of lexical relationships in natural languages, which are rather similar to “small worlds” in social organisation. They are interested in a positional approach to network structures, which they illustrate by alternative measures of centrality related to different definitions of neighbourhood in semantic graphs.

Finally, two chapters are devoted to urban systems. This is a class of complex systems that have been analysed from the very beginning in terms of hierarchy, by geographers mainly, then by economists. As their hierarchical organisation is universal globally and historically, and because comparable measurements can be provided for their analysis more easily than for many other social systems, much gain in knowledge can be expected from the application of quantitative methods and simulation models to urban systems, for the purpose of improving our interpretation of hierarchies in society. Michael Batty, a specialist in urban planning at University College, London, develops a methodological approach for relating hierarchical organisation to growth processes (chapter 6). He compares a purely statistical view of the problem with observations and demonstrates that the spatial dimension should be integrated into simulation models to reproduce and explain hierarchies in cities and city systems. Urban hierarchies are thus related to a space-filling process, in which specifications have to meet the observed fractal features of the urban settlement patterns. Denise Pumain, a geographer at the Sorbonne, (chapter 7) attempts to review the major questions that are still in debate when applying alternative explanations of hierarchies to urban systems. She suggests how to reconcile two different approaches to urban hierarchies, seen either as a functional organisation into levels according to central place theory, or as a continuous statistical distribution of city sizes according to Zipf’s law. She also tries to go beyond the separation between two conceptions of cities, either as central places in their region or as nodes in global networks of economic activities and infrastructures. The integration of these two conceptions into an evolutionary theory for urban hierarchies is made possible by developing a link between scaling properties of urban systems and the process of urban change through technological and social innovation.

The concluding chapter reviews the gains in interpretation derived from confronting the methodological investigations in the different chapters, and suggests a possible reversal in the conceptual linkage between hierarchy and complexity.

REFERENCES

Barabási, A-L., 2002, *Linked: The New Science of Networks*, Perseus Pub, New-York.

Courgeau, D., 2004, *Multi-level analysis,* Kluwer Academic Publisher, Dordrecht, Methodos Series, vol. 2.

Franck, R., 2002, *The Explanatory Power of Models*, Kluwer Academic Publisher, Dordrecht, Methodos Series, vol. 1.

Gaudin, J.-P., 2002, *Pourquoi la gouvernance*?, Presses de Sciences Politiques, Paris.

Granovetter, M., 1985, Economic Action and Social Structure: The Problem of Embeddedness, *American Journal of Sociology,* **91**: 481-510.

Thompson, G., 2003, *Between Hierarchies and Markets: the Logic and Limits of Network Forms of Organisation,* Oxford University Press, Oxford.

Uzzi, B., Spiro, J., 2004, *Do Small Worlds Make Big Differences*? Harvard Strategic Conference, Boston.

Watts, D.J., 1999, *Small Worlds: the dynamics of networks between order and randomness*. Princeton University Press, Princeton.

White, D.G., 2004, Modeling *the Dynamics of Network formation and evolution*, http://www.insna.org

http://www.santafe.edu/sfi/

http://www.open.ac.uk/ecss/index.html

http://www.complexityscience.org

Chapter 1

HIERARCHY: A SHORT HISTORY OF A WORD IN WESTERN THOUGHT

Nicolas VERDIER
Researcher,
CNRS, UMR Géographie-cités

1. BETWEEN THE 14TH AND 17TH CENTURY, A SLOW SHIFT OF THE WORD OUTSIDE THE SPHERE OF THEOLOGY

Usage of the word *hierarchy* dates back a considerable time. It seems to have been coined by Pseudo-Dionysius the Areopagite (6th century AD). It is made up of *hieros* "sacred" and *arkhia* "rule". The first clear meaning arises from this etymology, since *hierarchy* at that time is "the governance of things sacred". As a theological term, it is used to refer to the "subordination that exists between the different choruses of angels. There are nine choruses of blessed spirits divided into three hierarchies" (Furetière, 1690). The word appears to remain in the field of the description of the order and subordination of the different choruses of angels until the 14th century. The concept of hierarchy then enters the register of the description of the ecclesiastical state, and more generally that of society overall. With respect to the clergy, *hierarchy* is "the subordination that exists between the Prelates and the other ecclesiastics, the Pope, the Archbishops, the Bishops, the Curates and the Priests [who] constitute the hierarchy of the Church" (Furetière, 1690).

D. Pumain (ed.), Hierarchy in Natural and Social Sciences, 13-37.

One of the contexts where the transition occurs is in works with a strongly religious content. Descriptions of the Creation – which should be viewed as being related to universal mythology – use this mode of exposition. It enables the description of a system of interdependent degrees or ranks of beings and things from the most to the least pure. The metaphor involved is that of the ladder (with its rungs) rising vertically. Samuel Ward, in his "Life of Faith" published in 1622, writes of this system as extending "from the mushrome to the Angels".

Another context of transition is the cosmograhic style, which uses the heritage of medieval geography and goes on, at the start of the 16th century, to link the celestial globe to the terrestrial globe. Thus P. Apian's *Cosmographicus Libe*r published in 1524 shifts imperceptibly from the celestial spheres to the terrestrial sphere. Apian states that "this science [cosmography] first considers the Circles, of which we imagine the supreme celestial Sphere to be composed. Thereafter, according to the distinction and distribution of the said circles, it declares the situation of the lands which are below them, and the measure and proportions of the same" (p. 5). Moving on from a merely vertical construction, cosmographic systems point to relationships between the hierarchies of the angels and those of the earth seen as an inhabited world. In this way, *hierarchy*, which incorporated the poles of the highest and best on the one hand and the lowest and most uncouth on the other, is extended to include the largest and the smallest: here the visual representation used in cosmography leads on to the construction of a more complex concept, in some ways related to the armillary spheres.

Figure 1: Vincent de Beauvais, *Miroir historial. Speculum historiae* 1463, BNF. Cliché Bibliothèque nationale de France, Paris

Book 5 of Milton's *Paradise Lost* (1674), which revolves almost entirely around the concept of hierarchy, witnesses a reunion of these two modes of representation. Milton refers to "the scale of nature set from centre to circumference". Here the inter-relationship of the spheres is seen as a relationship of "alimental" dependency in which the coarser elements enable the purer elements to exist:

"The grosser feeds the purer, Earth the Sea,
Earth and the Sea feed Air, the Air feeds those Fires
Ethereal, and as lowest first the Moon;
Whence in her visage round those spots, unpurg'd
Vapours not yet into her substance turn'd.
Nor doth the Moon no nourishment exhale
From her moist Continent to higher Orbes.
The Sun that light imparts to all, receives
From all his alimental recompence
In humid exhalations, and at Even
Sups with the Ocean"

The last context of transition that requires mention is at the heart of the discord between Protestants and Catholics. For Luther, neither the Pope nor the bishops nor any man has the right to impose "even one syllable" on a Christian. In other words, all the faithful are priests by their baptism. The Catholics on the contrary put emphasis on the idea that the clerical hierarchy was established by Jesus Christ himself. It is at this time that tensions focus on the question of hierarchy in the clergy, and Catholics resort to the texts by Pseudo-Dionysius the Areopagite (translation, 1865, chap. III, §1, see also Roques, 1983) denouncing the heresy. According to this author there are two worlds : that of pure intelligence and that of incarnate intelligence. The first constitutes the celestial hierarchy, and the second the ecclesiastical hierarchy. For Pseudo-Dionysius, hierarchy is "a sacred order and science and operation assimilable, as far as attainable, to the likeness of God, and conducted to the illuminations granted it by God, according to capacity, with a view to the Divine imitation". It is at the time of the Council of Trent (between 1545 and 1563) that the word *hierarchy* was officially adopted to describe the different degrees of the ecclesiastical state. Indeed, the 23rd session of the Council pronounced an anathema on those who opposed the idea of a hierarchy among the clergy : "*Si quis dixerit, in Ecclesia catholica, non esse hierarchiam, divina ordinatione institutiam, quac constat episcopis, presbyteris et ministris, anathema sit*" (Dolhagaray, 1925). Behind this defensive declaration certainly lay the fear of a secularisation of the clergy which would *de facto* place the concept of hierarchy outside the religious sphere.

Thus it clearly appears that in Europe at the end of the 17th century the usage of the word *hierarchy* had shifted from a limited theological register to other registers outside the field of theology, at the same time undergoing certain alterations.

2. THE 18TH CENTURY: FROM THE HIERARCHY TO HIERARCHIES

The writings of Spinoza, Descartes, Leibnitz and Hume, when approaching the concept of *hierarchy*, question the older conceptions either by broaching the issue of infinity, or by replacing certain levels by others. However, it seems more fruitful to look at more fundamental questionings of the concept of *hierarchy* as it was seen up to the 17th century, involving on the one hand the presence of God, and on the other the replacement of static conceptions of the universe by dynamic conceptions (Miles, 1991). With respect to the divine presence, it can be noted that in the middle part of the 17th century, Pascal, in a clearly deistic view, was still writing : "I cannot forgive Descartes: he would gladly have left God out of his whole philosophy. But he could not help making Him give one flip to set the world in motion. After that he had no more use for God." [1]. As Spinoza wrote in his *Ethica* (around 1670) : "Whatever is, is in God, and nothing can be, or be conceived, without God" [2]. It is with the *Enclyclopédie* produced by Diderot and d'Alembert, however, that conceptions change radically: hierarchy appears as an essentially human construction:

"*Hierarchy*: is said of the subordination between the different choruses of angels that serve the Almighty in heaven. Saint Denis differentiates nine, which he divides up into three *hierarchies* […]; also refers to the different *orders* of the faithful that make up Christian society, from the Pope, who is the head, down to the mere layman […]. In civil society there are different orders (ranks) of citizens rising one above the other, and the general and particular administration of things is distributed in portions to different men or classes of men, from the sovereign who rules everyone down to the mere subject who obeys."[3]

Thus, according to the *Encyclopédie*, it is St Denis who distinguishes the different hierarchies, it is Christian society that makes up the hierarchy, and it is the citizens that build the hierarchy of civil (lay) society.

[1] Pascal, *Pensées*, 194.
[2] Spinoza, *Éthique*, Proposition XV.
[3] Encyclopédie, vol. 8, p. 205.

On the issue of the arrival of dynamic conceptions of the universe at the start of the 18th century, the work by Noël Pluche can be quoted. In his "*Spectacle de la nature*" (1739) he states that the number of plants and animals created by God is fixed: "The number and origin of organised species, as elementary natures, has therefore not been given over to any movement, nor any blind power. One infinitely careful hand has fixed them and they are unchanging, as is the Almighty who has made them" (vol. 2, p. 362). Here the scale of nature is therefore a continuous line, valid since creation. However in 1765 Voltaire's "*Dictionnaire Philosophique*" sets out a definition of the "chain of beings". He sees it as being full of gaps: "the proof is that there are species of plants and animals that have been destroyed. There are no longer any Murex. Jews were forbidden to eat griffon or *ixion*; these two species have probably disappeared from the Earth […]. Lions and rhinoceros are becoming very rare. If the rest of the world had done as the English did, there would be no more wolves in the world […]. Is there not a visible gap between monkeys and men? Is it not easy to imagine a two-legged animal without feathers, intelligent but without speech, and without our likeness….?"[4]. He also sees this chain of beings as made up of incommensurable elements: "But there is a greater distance between God and his most perfect creatures than between the Holy Father and the dean of the Sacred College. The dean may become a pope, but the most perfect genii created by the Supreme Being cannot become God. Between him and God lies infinity." In other words hierarchy (the word is used by Voltaire) does not exist : "this graduation of beings from the smallest atom to the supreme being, this scale of the infinite, strikes wonder. But if one looks closely, this great phantom vanishes, as all ghosts were wont to flee at the first cock-crow" (Voltaire, 1765, article "*chaîne des êtres*"). While the 16th and 17th centuries saw the shift of the concept of *hierarchy* from the sacred to the profane, the 18th century, because it undermined the concept of a unified hierarchy, widened its application, hitherto restricted to certain domains. However it is important to note the wide variability of timings, since it was only in 1835 that the *Dictionnaire* of that venerable institution, the *Académie Française* added an entry to the definitions relating to the angels and the clergy, extending the concept to other "sorts of power, authority, and rank subordinated one to another: political hierarchy, hierarchy of power, military hierarchy…". In comparison, very early on, the definition of "*gerarchia*" in the *Diccionario de la legua castellana* in 1734 puts forward the lay usage of the concept, the definition of the word beginning as follows:

[4] Voltaire seems to know very well the work of Pluche. According to recent hypothesis, Pluche could have inspired the character of Pangloss in *Candide*. See: E. Palmer, "Pangloss Identified", *French Studies Bulletin*, 84, Automn 2002.

"*Gerachia : El orden y subordinacion que en qualquiera Republica bien ordenada, tienen las diversas classes de sugétos que la componen : comme en la Eclesiastica los Prelados superiores è inferiores, il los Clérigos; en la secular les Principes, senores, nobles y plebeyos*". (Real Academia Española, 1734, Tomo quarto, p. 47).

In this way the Spanish Academy positions the ecclesiastic hierarchy on the side of application rather than on the side of principle. Examples of what must be seen as a secularisation of the notion of hierarchy are more widespread in natural sciences, as for instance in Linnaeus and the taxonomy of organisms, or Buffon who believed in the imperceptible evolution of one being into another (Ehrard, 1963, pp. 181-198). They are moreover not the only instances that can be called upon. Indeed, it is clear that these descriptions of nature served as models – or even metaphors – for numerous descriptions of society in both the 18th and the 19th centuries. References to the "nature" of society are well known. It therefore seems worthwhile considering other instances which, while they belong to the scope of what were later to become the social sciences from the 20th century onwards, nevertheless contribute to understanding how conceptions evolved. Thus for the period of interest it is possible to focus on the issue of towns and cities, and the hierarchies applied to them which, for a period, organised social hierarchies.

At the end of the 15th century, or even into the 16th century, cities[5] are immobile. Throughout Europe from one country to another, two or three founding myths are superposed to explain their origins. The first, which is omnipresent, is antiquity. The city is suspended in time and its origins are virtually immemorial. Thus as late as the start of the 18th century the city of Nîmes in France is presented by Piganiol de la Force as having been founded in the "year of the world 2715"; for Chartres, also in France, "if the local tradition is to be believed, it dates back to times very close to the Deluge" (Piganiol, 1718, vol. IV p. 139 and vol. V p. 299). In Italy in 1579, Paolo Paruta, reflecting on what makes up a city, suggests a classification of towns and cities in which superiority is given to the *Città* according its antiquity: "Since virtue requires outside witness of honour, one considers dignity, and, dignity being equal, antiquity". The second founding myth concerns the

[5] Translator's note : there is only partial overlap in French and English of the meanings of words used to designate human settlements, (*cité, ville, bourg*, etc versus the English *city, town, borough*, etc) for obvious historical, cultural and chronological reasons. In translation the fairly general acceptations of the words « town » (a settlement of some size) and « city » (suggesting status and a degree of antiquity and privilege) are used. These terms are in no way used to refer to a definite status. (See on this score PJ Cornfield, the introduction to The impact of English Towns 1700-1800, OUP 1982).

walls, and this idea is found in most of Europe except for Italy, for which country a foreword to Ortelius's *Theatrum orbis terrarum* (added by the Italian editor) specifies:

"*Solamente s'avertisce, chen'paesi oltramontani (*Northern Europe*) si tengono per Città tutte les Terra murate, che hanno grandezza, e privilegio di Città tutte le que in Italia non si dà communemente questo iotolo fuor che a quei luoghi; che tengono il Vesequo*" (Ortelius, 1655).

In this foreword to the work by Ortelius, it can be seen that land within walls is essentially city. On this subject, another passage from Piganiol de la Force describes St Etienne (in France) as follows: "It was only a "*bourg*" (small town) when the inhabitants obtained permission from king Charles VII to surround it with walls". Thus it is the royal consent (or charter) that creates the difference, and this indeed is the third criterion: the granting of a privilege. In his *Corografia Portugueza* in 1706, Antonio Carvalho da Costa thus hierarchises human settlements into at least four levels:

"In this book your majesty will see a number of *cidades*, which he has generously endowed with a charter, with many sumptuous buildings, made safe by unassailable fortifications; *villas*, which he has enhanced with privileges; *lugares* which he has raised to the rank of *villas*; and those for which he has enlarged the *termos* (boundaries)…"

This ancient hierarchy among human settlements belongs to the ancient model of the hierarchy in which each level corresponds to a particular sphere. It reaches well beyond the mere issue of urbanity, since it entails an extremely efficient system of social classification. This is even more marked in Italy until the end of the 18th century, and in Hungary probably as late as 1848, where the *citizen* (city-dweller) enjoys a social status that sets him apart from the rest of the population, and can even go as far as to confer statutory nobility (Czoch, 1996). In the words of Bartolo da Sassoferrato in 1570: "It is noteworthy that it is better to be a mediocre citizen of a noble and honourable city than to be a more illustrious citizen of a mediocre town. And this is true for the *populares* of a noble city, who are to be honoured more than the great men of another town".

This first period in the hierarchy of human settlements ends around the start of the 18th century, and new modes of classification gradually appear, retaining rather hybrid forms at least until the early 19th century. It is useful to briefly recall how criteria used to describe cities and towns evolved. Initially, as seen above, it was antiquity, walls and privileges that enabled comparisons between cities and towns; then definitions evolve in and around the 18th century towards definitions that put emphasis on urban functions and population. Underlying this trend, there is basically an undermining of the static approach, in favour of a dynamic, historicist approach incorporating notions of evolution and fluctuation. At the same time an approach based on measurement takes over in the 18th century for comparing human settlements, in connection with the study of economic functions or populations. As Bernard Lepetit (1979) writes: “With movement comes measurement, an absolute need if evolutions in different directions and of variable intensity are to be taken into account”. The result of this evolution can be clearly seen in the following description of the town of Saint Quentin (France): “Its manufacturers, its workshops, the trading activity and the industry of its inhabitants make it one of the best towns in France” (Couedic, 1791, vol. 1, p. 31).

Two examples of attempts to hierarchise towns and cities are worth considering here to highlight classification processes other than those encountered in natural sciences.

The first example comes from Robert de Hesseln who, in his *Dictionnaire Universel de la France*, in the article entitled “France”, differentiates between the first-ranking towns comprising 100 000 inhabitants, second-ranking towns comprising “nearly 40 000 or 50 000 inhabitants”, the “*bonnes villes*” of the third rank with no more than 20 000 inhabitants, “a large number of towns with 8000, 10 000 and 12 000 inhabitants”, and an “infinite number of small towns”. Thus on the one hand the quantification of the number of inhabitants makes it possible to establish a hierarchy among human settlements, while on the other the viewpoint of continuity in the distribution among these different settlements shows that levels of urbanity are not really commensurable. The system at once accepts the principle of a common measure for all types of settlement, and at the same time, as in the past, distinguishes clearly separate levels.

The second example of attempts at hierarchisation of towns and cities, from the work by Charles de Fourcroy in his *Essai de Tableau Poléométrique* dated 1782, compares urban surface areas. This surface area criterion in fact corresponds to the ancient manner of considering walled areas which made it possible to define cities according to an enclosed surface area. The disappearance of city walls, already widespread at the end of the 18th century, rendered this type of measurement problematic by the time Fourcroy was writing, the indicator having become obsolete. Conversely, the possibilities of a continuous graduation appear clearly at this stage, if only in the very form of Fourcroy’s *tableau*, even if he suggests a hierarchical organisation into "very small towns", "small towns", "medium sized towns", "large towns" and "very large towns". The graduation from one category to another is along a continuum which seems quite clearly inherent in the system of representation. At the same time, the use of quantitative demographics de-couples the hierarchy of towns from the social hierarchy. Fourcroy was no longer thinking in terms of a relationship between the quality of the inhabitants and that of the town, he was restricting himself to quantities:

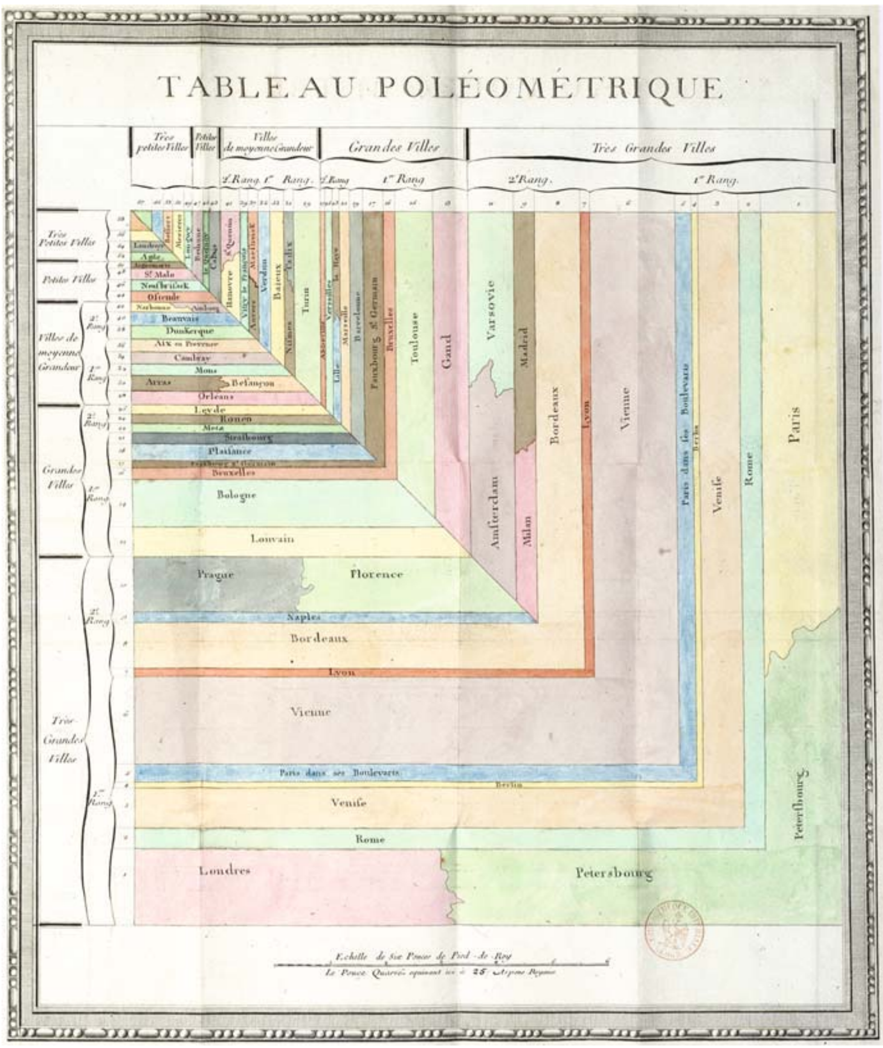


Figure 2: Charles de Fourcroy, *Essai d'une table poléométrique, ou amusement d'un amateur de plans sur la grandeur de quelques villes*, Paris, Dupain-Triel, 1782. BNF
Cliché Bibliothèque nationale de France, Paris

"Between the surface area of a town and the number of its inhabitants there must be a certain ratio that is more advantageous than any other, or which constitutes the most suitable population which question, to my knowledge, has never been treated, although it relates not merely to curiosity" (p. 24, quoted by Palsky, 1996, p. 52).

What then remains of the hierarchical classification of society? In fact, the concept is not affected by this separation, since hierarchy appears as a necessary principle in the organisation of society. As Montesquieu wrote in *L'Esprit des Lois* (1748), "in the state of nature, men are indeed born in a state of equality; but they cannot remain so. Society has them lose this equality, and they only regain it by way of the laws". Even so, this equality before the law is radically altered by inequality in the political sphere. For Montesquieu (who uses the word *hierarchy* to describe society[6]) " monarch rule […] assumes pre-eminence, rank, and even original nobility […], for the prerogatives of the lords, the clergy, the nobility and the towns" (vol. VIII) constitute a rein both to excessive power of the prince and to the passions of the people. The view held by Montesquieu, which was to be taken up ten years later in the *Encyclopédie*, is therefore one of a necessary hierarchy within society. In this way, it is an almost revolutionary conception, since it no longer justifies the existence of the populace by blood rights, but rather by social necessity. This same idea was developed around 1741 by Yves André in his *Essai sur le Beau*, in which he claims the need for a hierarchy in ranks and orders as being essential to the harmony of the social microcosm (see Ehrard, 1963, pp. 516-540). Put in other terms, in another civilisation, the idea is also found in the Koran, according to which "some of us have been raised above the others so that they may take the others into servitude" [7]. In France, this contrast between equality before the law and political equality will be encountered later at the time of the Revolution, which, while it abolished privileges during the night of August 4th 1789, shortly afterwards instated a system of tax-payer vote (Ozouf, 1988).

Overall, it is probably with the disappearance of the *Ancien Régime* that the concept of social hierarchy really developed, the existence of a word making it easier than previously to apprehend in thought and criticise.

[6] Thus, in his "Considérations sur la France", Montesquieu describes how jurys are formed, advocating the fact that "La hiérarchie des mouvances appelait les vassaux du même ordre dans la cour de leurs suzerains respectifs ; de là naquit la maxime que tout homme devait être jugé par ses pairs". *Esprit des Lois*, livre XI, chap. VIII.

[7] *Coran*, sourate 6, verset 165 (French translation by André Jacob (dir.), l'*Encyclopédie philosophique universelle*, Paris, PUF, 1990, vol. 1, p. 1140.

3. PARADOXES OF THE FIRST PART OF THE 19TH CENTURY

One of the features of the start of the 19th century following propositions that, as seen above, were hybrid and sometimes contradictory, is that there are two simultaneous trends, one restricting and one opening up the usage of the word *hierarchy*. There is restriction, in that the meaning of the word is frequently narrowed once more, mainly to the area of theology. Thus in the various editions of the dictionaries of the Royal Spanish Academy there is first a reversal of priorities in the definition between 1780 and 1817, the issues of political organisation previously set out as the first entry becoming the last entry, and the meaning of the order and subordination of angels becoming the first. At the same time, the definition which took up some 15 lines is reduced to four. After 1837 political organisation is only a meaning "by extension" of the word. It is only in 1984 that the definition of the word eventually evolves, returning to a more lengthy definition (8 lines), and applying it to a description of the internal organisation of society. The meaning of "an important person within an organisation" is also added, the word here being used to describe the summit of the pyramid. If however the corpus is widened to works that are more open to the contemporary world, evolutions can be seen at an earlier date. Thus in the *Encliclopedia universal ilustrada Europeo-Americana*, published in Barcelona in 1926, which give translations of words into French, Italian, English, German, Portuguese, Catalan and Esperanto, the word *jerarquia* boasts three entries, one concerning the administrative hierarchy, the second the ecclesiastical hierarchy and the third the military hierarchy.

Despite this, the diversity of usage of the word hierarchy was greatly reduced in 19th century Spain. Along the lines of this narrowing of usage, it can also be noted that the word does not appear to be in use among English-language economists in the 18th and 19th centuries. For instance, if a search is made for the word *hierarchy* in Steuart (1767), Smith (1776, and also the translation into French by Blanqui in 1846), Bentham (1776, 1787, and 1796), Ricardo (1810 and 1817) and finally Mill (1848 and numerous re-editions up to 1886[8]), no references are found. Moreover it can be observed that the 1911 edition of the Encyclopaedia Britannica provides a definition that is restricted to the mere question of theology. After rapid research it seems possible to say that it is only with Keynes (1919) that the word appears in the writings of an English-language economist. It can be added, outside English language usage, that *Das Kapital* by Karl Marx, both in its

[8] About the editions see M.-A. Ellis, "Variations in the Editions of J. S. Mill's *Principles of Political Economy*", *Economis Journal*, vol 16, june 1906, pp. 291-302.

original version and in its French translation, does not use the word *hierarchy*. Thus the word is absent, while authors like Smith or Marx are generally considered as having contributed to originating the concept of social class which can imply the idea of social hierarchy, as is shown for example in the definition of social hierarchy given in the *Encyclopédie Philosophique Universelle* published under André Jacob in 1990 (vol. 1, p.1140). In other words– and the paradox is worth noting – some of the most prominent authors of the nineteenth century with regard to the issues of social hierarchies do not resort to the word to describe society.

What indeed are the usages of the word in the course of the 19th century, during which it appears to have disappeared at least in part? First it is important to avoid oversimplification: even if the word *hierarchy* is not widely used, it is still in existence, on the one hand in the area of theology, which is in no way radically new, and on the other in the description of society, and this is the most relevant aspect here, because it is the most innovatory.

3.1 Alexis de Tocqueville: De la démocratie en Amérique

In this respect, attention can be given first to Alexis de Tocqueville and his book *La Démocratie en Amérique* published in France in 1835 and in the United States in 1839 (translated by Henry Reeve). Tocqueville seems to use the word *hierarchy* in three ways, two with negative connotations and one with mitigated connotations. The first concerns the social hierarchy of the *Ancien Régime* and its aristocracy. Thus, on the subject of the colonies around 1620, Tocqueville writes:

"The population of New England was growing fast, and, while in the mother country the hierarchy of rank still classified men in a despotic manner, the colony increasingly showed the new spectacle of a society homogeneous in all its parts. Democracy, as Antiquity had never dared dream it, emerged large and well armed from within the ancient feudal society" (book 1, chap. II).

The second usage by Tocqueville concerns the administrative hierarchy described as producing perverse centralising effects (book 1 chapter V). The third concerns the natural social hierarchy which arises from the activity of individuals. Here Tocqueville seems indecisive: on the one hand, social hierarchy seems to him to be necessary for civilisation:

"One can conceive of a people in which there would be no cast, no hierarchy, no class; in which the law, recognising no privilege, would share out inheritances equally; and which would at the same time be deprived of enlightenment and freedom. This is not a vain hypothesis: a despot can have motive to render his subjects equal and leave them ignorant, so as to more easily maintain them in slavery. […] Not only would a democratic people of this sort have no aptitude or taste for science, literature and the arts, but one could imagine that they never would come to show any" (book 2, section 1, chap. IX).

In the eyes of Tocqueville, hierarchy is a natural form of society, and absence of hierarchy only appears possible under a rule of tyranny aiming at the ignorance of the people. However, this hierarchy, however natural, is still to be condemned, since it isolates the outsider:

"The North American Indian living in the freedom of the woods was wretched, but felt inferior to no-one; as soon as he wishes to enter the social hierarchy of the white man, he can only occupy the lowest rank; for, ignorant and poor, he enters a society where science and wealth reign. After leading a life of turmoil, full of evils and dangers, but at the same time full of emotions and greatness, he must submit to a monotonous, obscure and degrading existence. The sole result in his eyes of belonging to this much-boasted society is to be earning his bread by drudgery, in the midst of ignominy"(book 1, chap. X).

Finally, what this man (born an aristocrat but conscious of the unavoidable disappearance of his cast) seems to abhor are the superimposed categories with their stereotyped boundaries. There is a last point worth making here concerning usage of the word *hierarchy* in a military context. Indeed, in the first part of his work, Tocqueville draws a comparison between the world of military officers and that of the nobility in order to show that the two hierarchies are not necessarily similar. In so doing, Tocqueville is no doubt one of the first to use the word in the context of the army.

The wide spectrum of usage found in Tocqueville's writings does not appear among other authors in the same period. There are indeed a few uses of the word *hierarchy* by a historian (and politician) such as Guizot. In his monumental *Histoire Générale de la Civilisation en Europe* this author only uses the word when referring to the feudal system, and mainly with respect to the clergy (Guizot, 1838). Usage of the word in these instances refers to the obsolete past. Likewise Marx who, as we have seen, does not use the word in *Das Kapital*, in the Communist party Manifesto in 1948 only uses a restricted meaning based on a military metaphor to describe modern industry: the workers, "simple soldiers of industry are placed under the supervision of a complete hierarchy of officers and commanders" (Marx, Engels, 1848).

3.2 Auguste Comte: hierarchy as a system

The above instances of usage to describe society are not sufficient to fill the relative void in the first half of the 19th century among authors that can in some way be connected with the social sciences of the future. The de-Christianisation of the "republic of letters", as well as the egalitarian ideals that flood across Europe after the French revolution, certainly have considerable influence. But an alternative explanation for this relative absence may reside in a more general manner of apprehending the organisation of formal (disciplinary) areas of knowledge into an all-encompassing system of knowledge. It indeed may be appropriate to reposition the evolution of the term of interest here in the context of the questioning of encyclopaedic knowledge. Indeed, in the 18th century, *hierarchy* relates essentially to a type of approach that is liable to produce a global model for understanding the world. The 19th century is on the other hand an intermediate period in that it is the time of transition between "encyclopaedism" and disciplinary constructs involving restricted fields of knowledge in which hierarchies (rather than any one Hierarchy) will serve as explanatory principles. Thus it is possible to suggest that the usage of the word is in proportion to the degree of autonomy of the discipline to which an author belongs. Conversely, the demise of *Hierachy* should be sought in the last genuinely encyclopaedic experiments, or those aiming at least to think in terms of a rational organisation of knowledge.

From this point of view Auguste Comte and Positivism are very relevant. Indeed Auguste Comte was the creator of the neologism *sociology* (Comte, 1837, lesson 47), and he placed this science at the summit of his classification of knowledge. In his view, it is the science that, on its own, enables understanding of knowledge as a whole, organised in the form of an encyclopaedia. Indeed, Comte points to his use of the term *hierarchy* in his *Cours de philosophie positive*, where he states: "I use this expression purposely to emphasise that I cannot conceive of a genuinely philosophical classification if one has not first managed to apprehend a predominant consideration [hierarchy], common to all instances, and gradually decreasing from one [science] to the other. It [hierarchy] is, in my view, the fundamental condition imposed by the general theory of classifications" (Comte, 1837, lesson 46).

Comte uses the idea of encyclopaedic degrees, or even of an encyclopaedic scale or ladder, emphasising the progression of one science to another. This leads him to "instate moral science proper on the seventh necessary degree in the encyclopaedic hierarchy, to complete the normal progression of complication and specialisation" (Comte, 1851-54, vol. 2) existing among the sciences. This does not prevent him perceiving a genuine continuum among these same sciences. In his view, each science is at least in part at once sovereign and subordinate. "There are undoubtedly today certain methods in chemistry or physiology that it would be useful to transfer to mathematics, and *vice versa*; this is not done, and the question is why? The reason is that each scholar is busy advancing his own particular science, and does not think of extracting and giving assistance to other scholars, nor of looking for it from them"[9].

There should not therefore be any strict divide between the sciences. However the limits to such exchanges should be specified. Thus Comte strongly criticises the "algebraic usurpation" which suggests that mathematics can enable understanding of phenomena that are in fact rendered complex by the wide diversity of approaches required to comprehend them. This is why he refuses to take an interest in Condorcet's "social mathematics" and his use of probabilities in the study of votes (Fédi, 2000, pp. 66-69). It is along these lines and for the same reasons that he criticises the work by Franz Joseph Gall on phrenology which used the shape of the cranium as a basis for a philosophy of the human mind (Gall, 1802).

The conceptions entertained by Comte on the hierarchy of sciences are all the more complex because he initiates a transition between hierarchised knowledge and equality among the sciences. Indeed, for Comte, this hierarchisation is merely functional: it enables him to analyse the general system of knowledge. At the same time, the sciences are seen to be equal one to the other in terms of status: "Although all fundamental sciences do not generate the same interest among vulgar minds, none should be neglected in a study such as the one in hand. As for their importance for human happiness, all are certainly of equal value, when they are examined closely. Indeed, those in which results at first sight present a lesser practical interest recommend themselves clearly, either by the greater perfection of their methods, or by being the essential foundation of all the others" (Comte, 1837, lesson 1).

[9] Lettre à Valat, 24 sept. 1819 (A. Comte, *Correspondance générale*, *op. cit.* n. 2, t. 1, p. 59-60) quoted by Annie Petit, "Des sciences positives à la politique positiviste", in *Auguste Comte. Trajectoires du positivisme*, L'Harmattan, Paris, 2003.

Comte's view of hierarchy is therefore more complex than the views seen hitherto, since it dissociates the value and the function of the elements making up the hierarchy. It seems justifiable here to suggest a genuine "comtian" model for the hierarchical organisation of knowledge. But rather than a branching hierarchy, often derived from his classification of the positive sciences, it seems more appropriate to refer to a pathway model for knowledge. This model, placing the "positive sciences" one after the other in the construction of the ultimate science, i.e. sociology – or the study of man – re-uses the ancient model of the "chain of beings", from the most simple to the most complex, while at the same time attributing the same value of necessity to each component.

4. THE SECOND PART OF THE 19TH CENTURY: SPREAD OR SCATTER?

4.1 The Durkheim hierarchy: a relational social model

One of the most interesting features of Comte's production for this study of the word *hierarchy* probably resides in the idea of a statutory equality among positive sciences, at a time when disciplines were becoming institutionalised, an increasing tendency all through the last half of the 19th century. Indeed, it would appear that in the years following these last attempts at institutionalisation, although the word *hierarchy* certainly does not disappear from usage, there is a relative absence of discussion of the concept, as if the spread of the word rendered it more labile.

The spread of usage in France is indeed quite clear. A rapid review of the catalogues of the *Bibliothèque Nationale* gives around 330 different titles containing the word *hierarchy* between the start of the 16th century and the end of the 20th century. Of the 34 listed between 1800 and 1850, 26 relate to religious questions and 5 to military questions. For the following period, 1851 to 1900, there are 35 references, but only 16 of these concern religion. The 5 pertaining to military matters remain, but there are now 6 references relating to administrative organisation (against 1 in the previous period), 6 relating to the description of society and 2 relating to the hierarchy of laws one to the other. It should however be noted that in the same period in works in English language in the Congress library, usage remains in the religious sphere until as late as 1936.

Thus the spread of the different forms of usage is restricted in the Western world to the “old continent”. Reference can be made here to Emile Durkheim who can be considered as one of the founders of sociology. Durkheim practically never uses the word *hierarchy*. It is found in three instances in his “*Division du travail social* (1893) and five times in *Le suicide* (1897), but it is not found either in *Les règles de la méthode sociologique* (1894) or in *Les formes élémentaires de l’organisation sociale* (1904).

To gain a better understanding of the usage of the word in Durkheim it can first be noted that the word is never defined by the author. It should also be noted that a word such as “graduation” does not occur any more frequently than *hierarchy,* so that it is not a question of choice of term. Durkheim does mention the question in his article in 1900 in the *Rivista italiana di sociologia* (Durkheim, 1900), in which he discusses the idea of demonstrating how hierarchies form by looking for those features that social forms have in common to produce hierarchy. However, he only takes a real interest in the question in the article co-authored with Marcel Mauss on the primitive forms of classification (Durkheim and Mauss, 1901-1902). The authors put emphasis not so much on the act of classification as on the historical nature of the concept and its non-natural aspects, which leads them to consider the issue of hierarchy: “Classification is not a mere matter of forming groups, it is positioning groups according to very special relationships. We see them as being coordinated or subordinated one to the other, we state that these (the species) are included in these (the genera), that the second subsumes the first. There are those that dominate and those that are dominated, and others that are independent one from another. Any classification implies a hierarchical order for which neither the perceptible world nor our awareness can supply a model”.

At the end of the article, the authors conclude: “It is because human groups fit one into the other […] that groups of objects are set out in the same order. Their regularly decreasing extension as we pass from genus to species, from species to variety, etc, arises from the likewise decreasing extension presented in social divisions as one moves away from the widest and most ancient towards the most recent and derived. If all things are conceived as a single system, it is because society is conceived in the same manner. It is […] the single whole to which all else is referred. Thus logical hierarchy is merely another aspect of social hierarchy and unity of knowledge is nothing other than the very unity of the community, extended to the universe”.

Following a functionalistic approach, which sees society as a systemic whole made up of sub-groups which are so many organs involved in reproducing of the whole, Durkheim excludes the hierarchy of key concepts from the reasoning, replacing it by inter-relation between the elements. It can also be added that his determination to remain abstract, in an endeavour to distinguish general social forms – which he intends to construct from shared features of the objects under study – restricts the interest he shows in actual relationships in society as they are practised. The hierarchy of practice is replaced by a functional model of inter-relationships.

The conceptions of Simmel are from this point of view fairly close to those of Durkheim, and this appears clearly in his attempt to define Sociology (Simmel, 1894). The main difference probably lies in the interest Simmel has for forms of socialisation, which he studies in a very subtle manner (Simmel, 1908). But here again it is more a study of functional relationships than a study of hierarchies.

Once again, *hierarchy* remains on the side of generalities and can up to a point be taken to be a synonym of the word "society" as these authors seem to view it. By concentrating on narrower themes, they exclude hierarchy as such from the framework of study, using the word but ultimately avoiding using the actual concept.

4.2 Sigmund Freud: an attempt to go beyond the hierarchy of a functional whole

Even though the work of Freud is mainly positioned in the 20th century, it may be interesting to refer to it here in order to move beyond the often disappointing features of the study of the concept of *hierarchy* in the period extending from the Franco-Prussian war in 1870 to the years between the two world wars in the 20th century. From the point of view of the history of disciplines, it is possible to consider that this period is fairly relevant, since it is during this time that the scientific disciplines of the human and social sciences form and consolidate in Europe, mainly via the creation of university chairs.

Attention will be given here to a single text by Freud published in French in 1921 under the title "*Psychologie collective et analyse du moi*". If it is the only text chosen for analysis it is merely because it is the only one to use the word under consideration[10]. It is also one of the texts taking the most interest

[10] About this, see http://www.uqac.uquebec.ca/zone30/Classiques_des_sciences_sociales/

in collective issues. Freud's attempt to hierarchise the phenomenon of love into degrees on the basis of the analysis of language – ending in a distribution into degrees from the most to the least libidinal – however suggests no hierarchy of the degrees one with respect to the other. The most productive passage from the point of view of the present study appears to be the comparison between the two hierarchies of the army and the Church (which he appears to restrict to Protestantism). For Freud, "a democratic force pervades the Church because all are equal before Christ, since all have an equal right to his love" which he qualifies as fatherly. However, "from the point of view of structure the Army distinguishes itself from the Church in that it is made up of a hierarchy of successive formations : each captain, like the commander in chief, is the father of his company, each officer is the father of his division". What Freud notes here is one of the foundations of the differentiation among hierarchies. In the first instance, we can see the old system of the spheres, on the one hand Christ who loves all men, and men who must love one another in order to love Christ: the two levels are essentially different. In the second instance, in the Army, there is a descending relationship from one level to the next which ensures transfer of the libidinal bond of fatherhood, without this bond changing in nature. Thus Freud shows that with the same explicative principle (the libidinal bond of fatherhood) it is possible to obtain two types of hierarchy that are essentially different. By producing this reasoning, which is in fact fairly close to Durkheim, what Freud does in addition is to change the scale of observation by shifting from a family principle (fatherhood) to social construction by hierarchy. Here hierarchy is not an assumption, it is a product. It would seem moreover that it is this transition, between Durkheim and Freud, from the social to the collective, that enables the shift. It can be noted here that this renewed usage of the collective concept enables Maurice Halbwachs in the 1930s to study the production of memory as a collective form (Halbwachs, 1950).

CONCLUSION

From the above set of examples, although it is neither exhaustive nor adequate to constitute an overview of the usages of the concept of *hierarchy*, in instances where a concept can indeed be identified via the use of a word, it is even so possible to attempt the definition of the four main lines of variation of the concept.

The first point that it seems important to stress is that of the range or scope of the concept: in extreme instances it can explain the whole universe, as with Apian or Milton, or perhaps society as a whole with Durkheim; conversely can be strictly limited to a clearly outlined object, as with Pseudo Dionysius, or with Marx who uses it solely to describe the military system.

The second point to be emphasised is the contrast between continuity and discontinuity. Freud thus appears to view the military system as a continuity of the libidinal bond of fatherhood, where differences in degree or rank in no way affect the nature of the hierarchic bond. Conversely, Voltaire, like Milton, considers that the shift from one level to another constitutes so large a change that the different hierarchical ranks are strictly incommensurable.

The third point to be made – possibly that which is the most relevant to human and social sciences, is that of the status given to the concept. Tocqueville provides two contrasting examples. In the first, hierarchy is synonymous with despotism (and this became a widely shared acceptation in Europe in the 1960s), and in the second it is what endows a civilisation with its intellectual wealth, which is at least partly related to the Koran sura 6 verse 165.

This leads on to the final point concerning the nature of hierarchy. By *nature*, what is meant here is the place given to hierarchy in reasoned thought. The two extremes set the proponents of a hierarchy that is to be discovered (such as Buffon and much of the writing on nature in the 18th and 19th century, or even the exponents of social Darwinism) against the proponents of an elaboration of a hierarchy by researchers in a given area (as in the definition given in *l'Encyclopédie*).

Each of the cases presented above needs to be considered in this four-dimensional system, the positioning within one or other of these dimensions having no impact on the positioning in any of the other three. It can also be recalled that contradictions or different registers in usage in the writings of one and the same author are quite possible, as for instance with Tocqueville.

APPENDIX: DEVELOPMENTS IN THE SECOND PART OF THE 20TH CENTURY

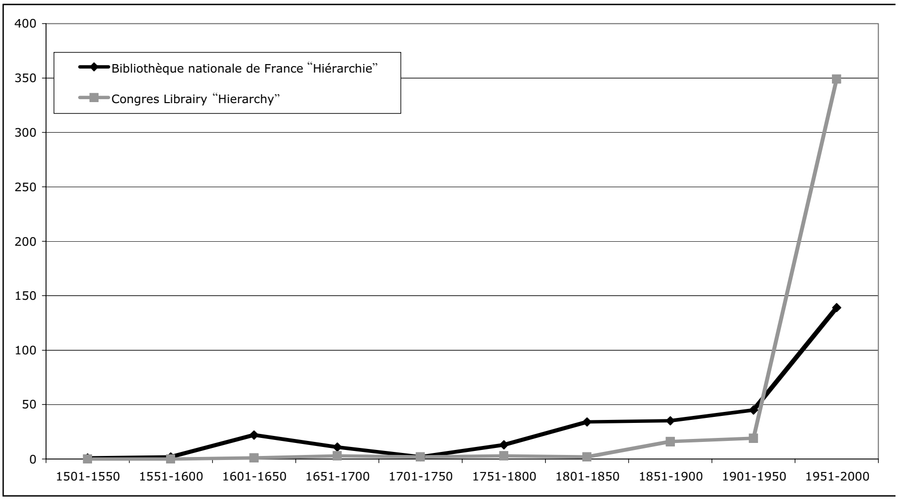


Figure 3: Frequency of the word hierarchy in book titles.

Finally the most recent forms of usage of the word *hierarchy* can be mentioned. In this respect, one of the most noteworthy aspects is its spread at the end of the 1930s, first in relation to questions of society, in some cases resulting in deviations such as those already noted in the work by Franz Joseph Gall, whose successors are to be found in Germany in the ruins of the Prussian military state and elsewhere, whether among geographers like Christaller (1933), or among Nazi theoreticians. This probably also explains the disgrace of the word in the years directly after the 1939-45 war. It is not until the 1960s that the word reappears in social sciences. Thus it appears in American sociology with T. Parsons (1949) and P.Sorokin (1959), but the subject matter is often closer to the notion of stratification than to that of hierarchy. The question of hierarchy is also encountered in French anthropology with the work by L.Dumont (1966) on casts. A less well-known instance should also be noted, on the frontiers between geography (which is a social science in France) and geology, that of Jean Tricart who, on the basis of the new theories on erosion cycles, went on to complexify the concept of hierarchy by linking it to that of scale. Thus this author, linking what he terms "temporo-spatial hierarchies" to scales of analysis, proposes a model to apprehend erosion cycles which links up phenomena occurring at different levels from the continent down to the slope (scree). Thus scale appears clearly as belonging to the question of hierarchy in the 1950s. The same author returns to these principles in analyses of urban geography.

REFERENCES

Académie française, 1835, *Dictionnaire*.

Apian, P., 1581, *Cosmographie universelle*, Anvers, French edition of *Cosmographicus liber*, Landshut, 1524, quoted by J.-M. Besse, *Face au monde, Atlas, jardins, géoramas,* Desclée de Brouwer, Paris, 2003, p. 61.

Bentham, J., 1776, *A Fragment on Government; being an examination of what is delivered, on the subject of Government in General in the introduction of Sir William Blackstone's Commentaries; with a preface, in which is given a critique on the work at large, anonymous,* T. Payne, P. Elmsly & E. Brooke, London.

Bentham, J., 1787, *Defense of Usury; Shewing the Impolicy of the Present Legal Restraints on the Terms of Pecuniary Bargains. In a Series of Letters to a Friend. To which is added a letter to Adam Smith, Esq. LL.D. on the Discouragements Opposed by the Above Restraints to the Progress of Inventive Industry,* T. Payne, London; Lang & Ustick, Philadelphia, 1796.

Carvalho da Costa, A., 1706, *Corografia Portuguesa e Descripçam Topografica do Famoso Reyno de Portugal, com as noticias das fundações das Cidades, villas & lugares que contem…*, Lisboa, vol. 1, Prologue, quoted by Claudia Damasceno-Fonseca, Pouvoirs villes et territoires, Genèse et représentations des espaces urbains dans le Minas Gerais (Brésil), XVIIIe-début du XIXe siècle, Paris, EHESS, 2001, p. 335.

Christaller, W., 1933, *Die zentralen Orte in Süddeutschland. Eine ökonomisch-geographische Untersuchung über die Gesetzmässigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen*, Gustav Fischer, Jena.

Comte, A., 1837, *Cours de philosophie positive*, Paris.

Comte, A., 1851-54, *Système de politique positive*, Paris

Couédic, P., 1791, *Tableau géographique de la puissance industrielle, commerciale et agricole de la Nation Française, par départements, districts, et cantons, suivant l'ordre de la nouvelle distribution du Royaume, dans lequel on trouve la démarcation des limites de chaque département & de ses districts, la nature des productions de son sol, son étendue & sa population….*, rue des Mathurins, Paris, 2 vol.

Czoch, G., 1996, "Réflexions sur les catégories de recensements, l'exemple des bourgeois de la ville de Kassa (Hongrie), 1788-1847)", *Cahiers du centre de recherches historiques*. **17** : 193-198.

de Fourcroy, C., 1782, *Essai d'une table poléométrique, ou Amusement d'un amateur de plans sur les grandeurs de quelques villes,* Dupain-Triel père, Paris.

de Sassoferrato, B., 1570, In tre codicis libros, Giunti, Venezia, quoted by D. Quaglioni, "The legal definition on Citizenship", in A. Molho, K. Raaflaub, J. Emlen, City-states in Classical Antiquity an d Medieval Italy : Athnes, and Rome, Florence and Venice, University of Michigan Press, Ann Arbour, 1991, pp. 97-140.

de Tocqueville, A., 1835, La démocratie en Amérique, Paris.

Dolhagaray, B., 1925, "hiérarchie", in A. Vacant, E. Mangenot and E. Amann, *Dictionnaire de théologie catholique contenant l'exposé des doctrines de la théologie catholique, leurs preuves et leur histoire*, Libr. Letouzey et Ané, Paris, vol. 6.

Dumont, L., 1966, *Homo hierarchicus. Le système des castes et ses implications*, Gallimard, Paris.

Durkheim, E., 1900, "La sociologia e il suo domino scientifique", *Rivista italiana di sociologia.* **4** :127-148.

Durkheim, E., and Mauss, M., 1901-1902, "De quelques formes de classification - Contribution à l'étude des représentations collectives", *Année sociologique*. **VI** : 1- 72.

Ehrard, J., 1963, *L'idée de nature en France dans la première moitié du XVIIIe siècle*, SEPVEN.

Enciclopedia universal ilustrada Europeo-Americana, 1926, Hijos de J. Espasa Editores, Barcelona.

Fédi, L., 2000, *Comte*, Les Belles Lettres, Paris.

Freud, S., 1921, "Psychologie collective et analyse du moi" in : *Essais de psychanalyse*, French translation from German by Dr. S. Jankélévitch, corrected by the author., Éditions Payot, Paris, reprinted in 1968, pp. 83-176.

Furetière, A., 1690, *Dictionnaire universel contenant généralement tous les mots françois…*, Arnout & Reinier Leers, La Haye et Rotterdam.

Gall, F., 1802, *Darstellung der Nerven, auf Untersuchungen desVerrichtungen des Gehirns gegründeten, Theorie des Physiognomik*, Weimar.

Guizot, F., 1838, *Histoire générale de la civilisation en Europe depuis la chute de l'empire romain jusqu'à la Révolution française*, Lacrosse Libr.-Éd., Bruxelles.

Halbwachs, M., 1950, *La mémoire collective*, PUF, Paris.

Jacob, A., dir., 1990, *l'Encyclopédie philosophique universelle*, PUF, Paris.

Keynes J. M., 1919, *The Economic Consequences of the Peace*, Mac Millan and Co, London.

Lepetit, B., 1979, "L'évolution de la notion de ville d'après les tableaux et descriptions géographiques de la France", *Urbi*. **92** : XCIX-CVII.

Marx, K., and Engels F., 1848, *Manifesto of the Communist party*, London.

Miles, S. J., 1991, "From being to becoming : Science and Theology in the Eighteenth Century", *Perspectives on Science and Christian Faith.* **43** (4): 215-224.

Milton, J., 1674, *Paradise lost, A poem un twelve books*, The second edition revised and augmented by the same author, Printed by S Simmons, next door to the Golden Lion in Aldergste-Street, London, book 5, l. 415-425.

Montesquieu, 1748, *L'esprit des lois*. Paris.

Ortelius, A., 1655, *Theatro del mondo : enel quale distinctamente si dimostrano in tavole, tutte la provincie, regni & paesi des mondo*, il Turrini, Ventia.

Ozouf, M., 1988, "Égalité", in François Furet et Mona Ozouf (dir.), *Dictionnaire critique de la Révolution française*, Flammarion, Paris.

Palsky, G.,1996, *Des chiffres et des cartes, la cartographie thématique quantitative au XIXe siècle*, CTHS, Paris.

Parsons, T., 1949, *Essays in sociological theory*, Free Press, Glencoe.

Paruta, P., 1579, "Della perfezione della vita politica", Venetia, Nicolini, Venetia, quoted by Marco Folin, "Hiérarchies urbaines/hiérarchies sociales : les noms de ville dans l'Italie moderne (XIV-XVIIIe siècles)", *Genèses Sciences sociales et histoire*, 2003, **51** : 4-25, (Paruta's text is translated in French as soon as 1582).

Piganiol de la Force, 1718, *Nouvelle description de la France, dans laquelle on voit le gouvernement général de ce royaume, celui de chaque province en particulier et la description des villes, maisons royales, châteaux et monuments les plus remarquables, avec la distance des lieux,…,* T. Legras fils, Paris.

Pluche, A., 1740, *Histoire du ciel, où l'on recherche l'origine de l'idolâtrie et les méprises de la philosophie sur la formation, et sur les influences des corps célestes,* chez la Vve Estienne, Paris, (1st ed. 1739).

Pseudo Denys l'Aréopagite, 1865, *Hiérarchie céleste*, translation J. Dulac, Paris.

Real Academia Española, 1734, *Diccionario de la lengua castellana, en que se explica el verdadero sentido de las voces, su naturaleza y calidad, con las phrases o modos de hablar, les proverbios o refranes, y otras cosas convenientes al uso de la lengua…* Imprenta de la Real Academia Española, por los herederos de Francisco del Hierro, Madrid.

Ricardo, D., 1817, On the Principles of Political Economy and Taxation, John Murray, London.

Ricardo, D., 1810, *The High Price of Bullion, A Proof of the Depreciation of Bank Notes*, John Murray, London.

Roques, R., 1983, L'univers dionysien, structure hiérarchique du monde selon le Pseudo-Denys, Cerf, Paris (1st ed. 1954).

Simmel, G., 1894, “Le problème de la sociologie”, *Revue de métaphysique et de morale*. **6** : 497-504.

Simmel, G., 1908, *Soziologie. Untersuchungen über die formen der vergesellschaftung*, Duncker & Humblot, Leiprzig.

Smith, A., 1776, An Inquiry into the Nature and Causes of the Wealth of Nations, W. Strahan & T. Cadell, London, 2 vol.

Sorokin, P., 1959, *Social and Cultural Mobility*, Free Press, Glencoe.

Steuart, J., 1767, An Inquiry into the Principles of Political Economy, London.

Stuart Mill, J., 1848, Principles of Political Economy, with some of the their Applications to Social Philosophy, Parker, London, 2 vol.

Voltaire, 1765, Dictionnaire philosophique.

Ward, S., 1622, The life of Faith in death, quoted by C. A. Patrides, « hierarchy and order », in: Dictionary of the history of ideas, Ph. P. Wiener ed., 1973, Charles Scribner's sons, New York.

Chapter 2

BIOLOGICAL AND ECOLOGICAL SYSTEMS HIERARCHICAL ORGANISATION

Natural mechanisms and human driven processes' roles in the emergence of living systems' organisation levels and properties

Alain PAVÉ
Senior Scientist, Director
CNRS-Guyane, Programme Amazonie (French National Centre of Scientific Research)
Member of the French National Academy of Technologies and of the French Academy of Agriculture

The concept of hierarchy bears different meanings in Biological and Ecological Sciences. It can stand for knowledge's organisation as well as for a model of how living systems are self-organized and functioning. Although our interest is in the latter meaning, it appears important to say a few words about the former one first.

A rough exploration of databases (mainly Nature and Science ones, from 1992 to 2003) leads to the word following uses in scientific literature:

1. Hierarchy at the genome level: genes are distributed along the DNA sequence; their expression being necessary to initiate a set of other ones some genes have a peculiar role (e.g. the well known "lactose operon", that controls the sequences of genes involved in lactose metabolism in bacteria, or sequential genes expression during the development of organisms). Thus we can define hierarchies of genes in regards to their role and the moment of their expression.
2. Cerebral and intellectual processes (recognition, meaning, thought, etc.): in "human brain", knowledge seems to be hierarchically organized. Or more precisely, when trying to understand how intellectual processes work, such a representation is more convenient than a flat organisation found in many other species such as birds.
3. Social (dominance): some populations are structured in societies, hierarchically organized (termites, ants, bees, some fishes such as clownfish, etc.). Individuals in these populations play a particular role (e.g. a reproductive one for ants or bees queens), ranking them at the top

D. Pumain (ed.), Hierarchy in Natural and Social Sciences, 39-70.

of the corresponding societies (we can admit, when we speak about populations, that we refer mainly to basic demographic properties, and when we use the word “societies”, that we consider an internal “social” structure: within the population, groups of individuals play different roles).

4. Living systems’ natural organisation, from genes to ecosystems: living systems are organized following an nested hierarchy of units (e.g. cells are elements of organisms, organisms are elements of populations, populations are elements of communities, …). This organisation spontaneously emerged during the evolution of life. In this case we speak of “self-organisation”. This organisation seems to correspond to an increase in complexity (Carroll, 2001).
5. Nowadays, we also have to take into consideration the role of human societies in living systems organisation, at least for upper levels (from populations to ecosystems, see Science special folder: “Human-Dominated Ecosystems”, 1997). It is not classical to consider human action as an actual and essential factor of organisation in classical biological or ecological literature. We think that we now have to transgress classical frontiers between disciplines, between Natural Sciences and Human Sciences, and to integrate this human dimension.
6. Systematic, phylogeny and evolution: the need of a comprehensive view of life, to bring out the diversity of organisms and proceed with it, led to classifications. These classifications were based on analysis of similarities between sets of organisms and for a long time morphological similarities. The main goal was recognition: in order to decide whether an organism belongs to a particular class. Then concepts of species, genus, families, … emerged, and led to a hierarchical organisation of knowledge (a tree with at the upper level, the “kingdoms”, at the lower one, the “species”). And with the progress of genetics and the Evolution Theory, relationships have been established between purely descriptive point of view and dynamical and functional one: similarities were interpreted as genetic distances between species, then genetics distances as time intervals in the evolutionary process of life on Earth.

In this chapter, we will focus mainly on these two last points. Those are historically important because during the 19th century, scientific point of view on living world changed from a static and flat representation to a dynamic, evolutionary and structured one, with dual hierarchical representation: the first of nested systems, the second of a time organized phylogeny (or genealogy) of species. Hierarchical representations being also useful to organize knowledge, a question arises: to what extent is this representation related to living systems natural organisation (is it a truthful model of living world) or is it mainly convenient to organize human knowledge on these systems (because it is efficient to answer questions, to solve problems, such as recognition ones)?

1. ORGANISATION LEVELS AND SCALES: LIVING SYSTEMS HIERARCHIES

Today, as mentioned above, we know living systems are structured in a set of successive organisation levels, from genome to large sets of interacting organisms. The most known are cells, organisms, populations, ecosystems and… biosphere. However, there are intermediate levels such as “transcriptome and proteome”, between genes and cells, or “communities”, between populations and ecosystems. This hierarchy can be represented as shown in figure 1. Some elements are easily identified. They are indeed visible elements in our environment even if instruments, such as microscopes, are required to watch them. At the opposite some elements, such as populations or communities, are not obvious structures; they are consequences of a conceptual effort. A whole organism can be observed; a whole population can usually not. That is the reason why we developed specific techniques (e.g. statistical ones) to analyse them. But even frontiers and limits are not always exactly determined, and often its size itself, a population has a physical meaning, an existence in time and space. Another difficulty has to be underlined: an organism can be multicellular, but also unicellular (e.g. micro-organisms such as bacteria). This example shows that some organisation levels can be “naturally” observed, because they correspond to visible structures, other ones are more difficult, less natural, to define. Thus before moving forward it is necessary to give some definitions. However, we have to underline that one of the fundamental question, after the description of structures and functions, is: When and how did such structures emerge in the pass? Obviously, we cannot easily answer this question; in the following section we just try to give some actual hypothesis about that.

1.1 Organisation levels

1.1.1 Sub-cellular systems

The discovery of cell components did require a great deal of efforts, both conceptual and experimental. For instance, the concept of gene was enounced before the discovery of its physical organisation (chromosomes) and support (DNA). Progress also arose with the distinction between structures and functions (manufactory walls, engines, and engines' production). To date, the principal sub-cellular entities, constituted by macromolecules, are the following:

- Genome: among the macromolecules, DNA plays a particular role (a self-reproductive molecule, it supports genetic information, codes for structures and functions). Without genome, organised set of genes, we cannot speak of living system.
- Other macromolecular systems are not really levels of organisations; but represent functional units (e.g. transcriptome, the whole set of translated genes) or cell sub-structures (e.g. nucleus, Golgi apparatus, cellular wall).

1.1.2 Cells

Cells are the first largely autonomous living systems: we can distinguish a cell from its environment: a wall isolates an internal part from an outside one; exchanges are possible through this wall. Cells are characterized by their morphology and metabolism. They can be on their own, e.g. bacteria, hence cells are organisms of their own (i.e. microorganisms, capable of reproduction) or they are specialized and organized forming organs, substructures of a multi-cellular organism.

1.1.3 Organisms

Organisms are mono-cellular or multi-cellular, autonomous and reproductive. They come to life and they die. Organisms belong to species: minimum sets of organisms sharing a maximum of characteristics or similarities (in particular morphological, biochemical and genetic ones). Organisms can exchange or bring together genetic information during the reproductive phase (non sexed or sexed organisms). For sexed species, characters are kept in the offspring from the reproductive phase. Interbreeding ability between individuals is a proof too for belonging to the same species. There might be slight differences between parents and children; the accumulation of these differences, along successive generations, might progressively lead to new species. Today, it is even the most common admitted genetic evolution mechanism. However, as stated below, in order to survive new organisms have to be adapted to their environment (ecological constraint).

1.1.4 Populations

Populations are sets of organisms belonging to same species. An "individual" designate an organism, the population elementary unit. They share further characteristics than belonging to the same species. In particular, gene exchanges -gene flows- are greater between members of a population than with other individuals from the same species. That characteristic is usually related to a geographical location: interbreeding is more frequent between neighbouring individuals than between individuals separated by a great distance. Exchanges between distant populations are still possible, consequence of seasonal migrations for example. Sets of remote population, sharing properties, in particular genetic ones, are called metapopulations. Populations are usually permanent through generations.

Although it is not always the case in many biologists' everyday language, it is important to distinguish species from populations. The concept of species derives from taxonomic and genetic classifications. The concept of population comes from functional consideration and is defined in space and time (geographical location, and time interval of effective presence at this location). Populations are also characterized by demographic parameters (e.g. generation time, average number of individuals in offspring). A species is a set of all living and non-living individuals (dead or yet to be born) belonging to old, current or future populations all around in the world and sharing characteristics, in particular the highest level of genetic similarities that specifically led to the definition of the species.

1.1.5 Communities

Populations interacting in same locations constitute communities[1]. In fact, more or less populations, sharing the same geographical space, do interact at least in regards to the space occupation itself. However there are other ecological interactions: competition, mutualism, predation, etc. Communities can be considered as the first *ecological level*.

1.1.6 Ecosystems

Ecosystems include biological (communities) and non-biological (air, water and minerals) components. They are limited in space and relatively

[1] This word has another meaning in Ecology, particularly in French: it can designate a set of organisms sharing properties, independently from their location (e.g. the community of herbivores).

biologically homogeneous; limits between ecosystems are called ecotones. Broad classes of ecosystems can be defined (e.g. terrestrial ecosystems: forests, grasslands, savannas, desert, wetlands; aquatic ecosystems: sea and ocean, coastal zones, and also continental aquatic ecosystems such as rivers[2] and lakes). The first difference between these broad classes is the living environment: air for terrestrial ecosystems, water for aquatic ecosystems. Soils and sediments can also be considered as specific milieus. Besides, ecosystems types (e.g. forests) being widely spread on the globe surface, we can distinguish bioclimatic zones leading to some specificity (e.g. intertropical and tropical forests, temperate and boreal forest); in that case the difference mainly lies in species constituting the ecosystems (e.g. oak is a temperate forests' tree, when teak grows in intertropical forests).

For a long time the tree did hide the forest: ecosystem's study was restricted to the dominant category of constitutive organisms. We know better today: in order to understand ecosystems' functioning then dynamics, we must, as much as necessary, integrate other organisms and milieus.

1.1.7 Ecosystems' associations

Ecosystem associations are often considered as upper levels of organisation. We will first with stick to the general terminology (Lamotte and Blandin, 2 French ecologists, also spoke of "ecocomplex"). Ecosystem associations are more or less heterogeneous, made of different ecosystems (e.g. a patchwork of forests and meadows, of rivers and lakes). They are: landscapes, watersheds, bio (or eco) regions. Watersheds are defined around a river basin hydrology and geomorphology. Bioregions (or ecoregions) are naturally dependent of climate (e.g. continental, Mediterranean, wet and dry intertropical) and of geomorphology (e.g. hills, mountains, plains). Landscapes, intermediate structures, are often the results of human activities. Thus at this scale, if there are natural constraints (all is not possible everywhere: a palm tree cannot grow in polar regions, … except may be in greenhouses), the human role is evident (direct or indirect effects on structures, environment planning and management). In fact, it is now always evidence at lower levels of organisation. We'll come back this point further on.

[2] Sometimes, rivers are included in terrestrial ecosystems. Indeed their functioning is strongly dependant from terrestrial components and they heavily participate to these components' dynamics. This is the case in wet intertropical forests, such as Amazonian forests.

1.2 Organisation scales and levels

Organisation levels are characterized by the appearance of new entities, sets of lower entities' units and therefore wider than these entities. For example, populations of organisms are wider than organisms. However, elementary entities' size being very diverse, populations' sizes are consequently very diverse (e.g. two dimensions: from some square micrometers or millimeters for micro-organisms populations to square kilometres for populations of biggest organisms such as elephants or whales). There still is a length interval (or area, or volume) that roughly corresponds to a specific organisation level.

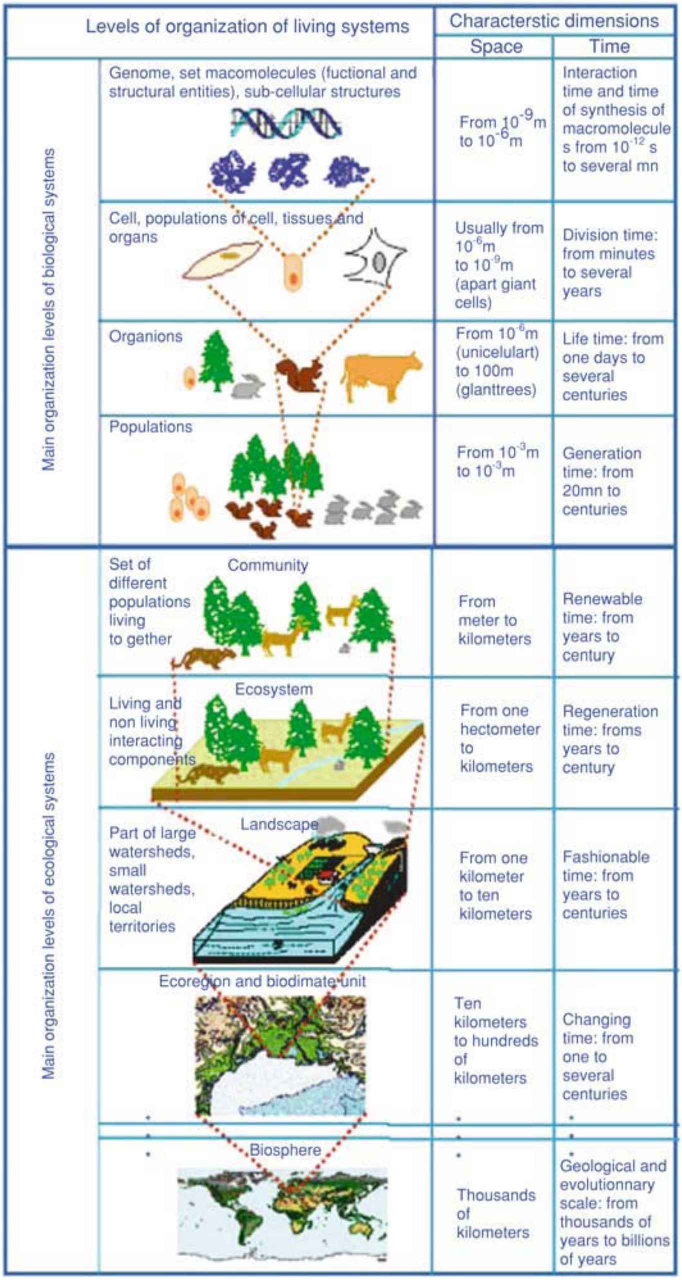
Organisation levels are characterized as well by interactions between units (e.g. between individuals of the same population). Consequently a time-lag – called *characteristic time* (figure 1) – is necessary to fulfill these interactions. The fulfilment delay will increase together with the unit size; furthermore it depends of entity size (e.g. population size), of elements' dispersion (e.g. distance between individuals) and of units living environment's "viscosity" (e.g. speed at what individuals spontaneously meet enables individual interactions). Obviously, there are other factors influencing characteristic times (e.g. attractive or repulsive chemical signals), but on average *characteristic times are positively correlated to characteristic sizes* (figure 2). These concepts of characteristic scales are well defined in a recent publication (André, Mégie and Schmidt-Lainé, 2003).

1.3 Basic interaction processes

Properties or processes, having no significance at lower levels, appear at each level. This is called *emergence mechanism*. Usually those properties or processes characterize interactions between elements. For example, mammal populations' social behaviour is not relevant for cells or macromolecules populations. Other processes, or analogous processes, can be found at nearly all levels. Thus, competition well exists between macromolecules (spatial competition at least: the area occupied by a molecule is not to be occupied by another one; furthermore competition between enzyme molecules for substrate has been demonstrated. At molecular level substrates look like resources at the ecological one). That specific interaction also exists between cells, between organisms, between populations and even between ecosystems (always and at least for distribution in space). It is not our purpose here to detail each and every process, but roughly the principal ones are:

- Synthesis (or reproductive) processes
- Spontaneous or external elements induced degradation (or death) processes: proteases induced proteins' degradation, predation of one organism by another one (in trophic networks)
- Competition (see above): roughly basic competition is either for space occupation or for access to and consumption of resources
- Association and cooperation: cells associations building organ and organism, or enzymes' cooperation in metabolic pathway, or even cooperation between organisms enabling a population survival and development
- Adaptation: for a given living system survival and development is also dependant of its environment and of its abilities to adapt itself to fluctuations

These basic interaction processes can be applied to all living systems. Real interactions look like these basic ones. However, most advanced organisms (animals) populations are structured in societies exhibiting further refined interactions through social behaviours (social dominance, breeding parades etc.). Actually, it seems that social structures are consequences of these interactions. During the emergence of social structure, we observe the reinforcement of the interactions role till an apparent stabilization of the structure occurs. All this has been demonstrated not only from observations and experiments, particularly on simple animal societies such as ant societies, but also from simulation, e.g. multi-agents systems use (Ferber, 1995).



(1) Biological systems. Intermediate structures and systems can be identified such as macromolecules populations ensuring functional roles: transcriptome, proteome, metabolome. On multicellular level, between cells and organism: cells populations constituting tissues and organs.

(2) Ecological systems *. If ecosystems' structure and limits as well as interfaces between ecosystems can usually be easily identified (e.g. ecotones, or watershed limits, or even ecoregions bioclimatic frontiers), other ecological systems bounds are not always easy to determine (e.g. landscapes).

Figure 1: Principal levels of organisation of living systems (from André *et al.*, 2003)

**Living systems are presented here on the same figure. However, there is a qualitative break between biological and ecological systems: ecological systems include different interacting populations (i.e. communities), environment's physical components (i.e. ecosystems), and more and more scientists take into account anthropogenic actions. On the other hand biological systems, from genome to population, have a genetic homogeneity, this not being the case for communities.*

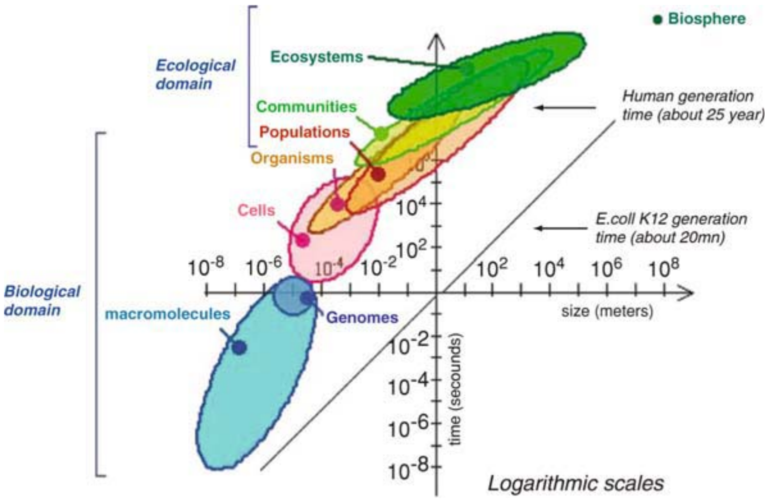


Figure 2: Approximate relationship between characteristic times of interactions between elements of successive entities and characteristic scales of space along organisations levels. Entities of a level are units of the immediate upper level (e.g. when considering the organisms-populations couple, organisms are units of populations. Now regarding the populations-communities couple, populations become units of communities (Barbault and Pavé, 2003).

2. HOW LIVING SYSTEMS ARE ORGANIZED: SELF-ORGANISATION AND HUMAN DRIVEN ORGANISATION

For a very long time living systems organisation has been dependent only on these "natural" or "spontaneous" interactions: we refer to this organisation type when using "self-organisation". Then since the emergence of human societies, this organisation becomes more and more dependent on man actions, either directly, e.g. development of agro-systems and more generally of artificial environments (from arranged ecosystems to urban zones), or indirectly, for example man productions effects, such as pollutants, on living systems. Furthermore we globally observe variations of physical parameters such as temperature and hygrometry. As a result, today, when speaking of living systems organisation we have to envisage these two points of view, as well as the way natural and human factors interact.

2.1 Biological systems' hierarchical self-organisation

2.1.1 History

Life evolution is characterized by the emergence of biological organisation, from cells to ecosystems and biosphere. This organisation spontaneously and progressively appears since about 3.8 billion of years, cadenced by critical events:

- first observed prokaryotic cells, organisation of macromolecules working together within a self-reproductive organism limited by a wall enabling exchanges of molecules with surrounding (3.8 Billion of years Before Present);
- first observed eukaryotic cells (about 1,8 By BP), presenting intracellular structures, in particular the core,
- first observed multicellular organisms (about 0.7 By BP);
- Amazing Ediacara radiation of Metazoa (about 0.650 By B.P.);
- At last Cambrian radiation (which began about 0.550 By BP).

Existing organisms come from this last radiation. Since this period, general tendency is complexification and diversification (Caroll, 2001); "variations on a same theme" the diversification was around general architectures designed during the Cambrian period.

After the Cambrian period, roughly at the end of Silurian period or the beginning of Devonian period (around – 400 My B.P.), continents colonization began and led to terrestrial organisms and ecosystems. This evolution too has been punctuated by important crisis (large scale extinctions of species, *cf.* figure 3), followed by biodiversity explosions. It is remarkable to note that, in spite of crisis, the tendency is an increase of biodiversity.

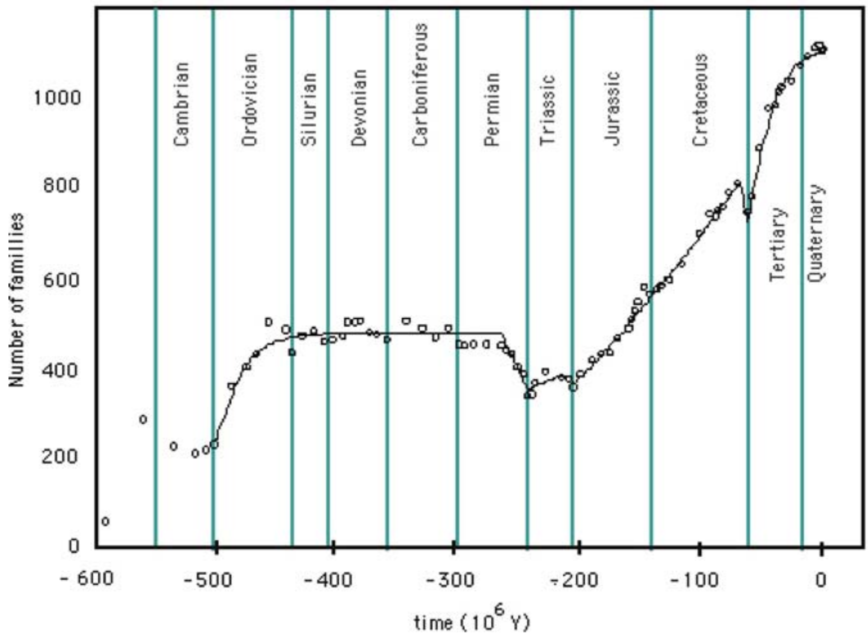


Figure 3: Biodiversity changes at the geological time scale. Data are represented by circles (from Courtillot and Gaudemer, 1996, extracted from the well-known Sepkosky's fossils data-base). The continuous line is the graphic of a chained logistic model[3], which provides a good representation of the major tendencies. Data are taken into account from the end of the Cambrian period on, i.e. after 500 MYBP (Pavé *et al.*, 2002).

[3] The logistic model can be formulated the following way: $dN/dt = r\,N\,(1-N/K)$ -where *N* is the number of families, *K* the plateau (maximum number of families) and *t*, the time-. It depends of 3 parameters: *r* (the growth rate), *K* (the plateau), that are explicit in the differential equation, and also the initial condition N_0, implicit. Courtillot and Gaudemer proposed to use this model to represent the biodiversity variations during the main periods, and for the growing parts only: from the end of the Cambrian to the end of the Carboniferous period (1), for the beginning of the Triassic period (2), for the Jurasic-Cretaceous period (3), and for the last periods, the Tertiary-Quaternary (4). For each time interval, *r*, *K* and N_0 were estimated (12 independent estimations). In our article, we have also used this logistic model for decreasing periods (e.g. Permian), but only the initial condition N_0 for $t_0 = -500$ My was estimated, and r and K for each periods, including decreasing ones. The initial condition for each interval results from the calculated value obtained at the considered time from the model of the preceding one (e.g. the initial condition for Permian period decreasing part is the value calculated by the model at the end of the Ordovician-Carboniferous period, with parameter values estimated for this period). This is a chained model. Estimates are better than the "piece by piece" model proposed by Courtillot and Gaudemer: it gives additional information concerning the decreasing periods in particular.

2.1.2 Mechanisms

One of the main questions is to understand biological evolution mechanisms (new species appearance, increase in living systems diversity and complexity, and those systems organisations). This question is the focus of many paleontologists and evolutionists' works since the beginning of the 19th century and more recently the focus of geneticists and ecologists' researches.

Darwin's and Wallace's famous works established the bases of the well known Theory of Evolution (Darwin, 1859), following Buffon's, Lamark's (who invented the term "Biology") and Geoffroy de Saint-Hilaire's transformationism hypothesis (as opposed to fixism, defended in France by Cuvier) of the History of Life (Grimoult, 1998). The theory focuses on organisms' diversification mechanisms (mainly on speciation obviously leading to new species). Although Mendel and Darwin lived at the same epoch, the genetic interpretation of their theory was only expressed beginning of the 20^{th} century (Morgan's works on drosophila genetics, and in the 1930's, population geneticists works, e.g. Fisher and Haldane). In 1964, Kimura proposed a stochastic model of this theory, centered on genetic "neutral mutation" called "neutralist theory of evolution". S. J. Gould, the well-known paleontologist, criticized the latter on the basis of his observations (instead he did propose the theory of "punctuated equilibriums"). Practically, we can roughly summarize biological diversification mechanisms by using current language:

1. Genomic level: modifications occurs at the genome level (DNA modifications: mutations, recombination, DNA segments acquisitions, transpositions, SOS enzymes transformations); these modifications are stabilized and transmitted to offspring.
2. Ecological levels: individuals from offspring are submitted to environmental constraints, only adapted ones survive and thrive (this is natural selection). The most competitive wins the "game of the life".

In fact, evolution seems to act as a race between species. Species must evolve at the same time their competitors do. It is the "Red Queen" hypothesis (this analogy comes from Lewis Caroll's book "Through the looking glass", has been proposed by Van Valen 1973. Plus check Barbault, 1992, Combes, 1995). This set of mechanisms is at the origin of the coevolution concept. Today however this competitive centered point of view is attenuated, as cooperative mechanisms seems to be also of a great importance (Michod, 2000).

As seen above, the history of Earth teaches us life is submitted to drastic events. Thus diversification appears to be a mechanism that could ensure its permanence. Species well adapted to different conditions may already exist in the wide mix of species, shall a major "accident" occur. In order to increase chance of relicts, evolution processes can be spontaneously accelerated to generate new species, where adapted ones have some chance to be found (effects of "SOS enzymes").

Competitive vs. cooperative mechanisms

For a long time, particularly since Darwin's book, competition was considered as one of the principal interaction between entities (e.g. when considering individuals at the ecological level: intra-specific competition is between individuals of same species, inter-specific one between individuals of different species). Surprisingly, cooperation was neglected. However, theoretical studies implied that if relationships were limited to competition, then ecosystems should probably converge towards simplification (i.e. decreasing of biodiversity) rather than towards diversification: most competitive species indeed wining "the game of life" and excluding others. But then how can many ecosystems' current diversity are explained? Furthermore how can species' coexistence in a same space be? At subcellular and organismic level, R. Michod has recently shown this organisation cannot be explained solely by competition, but that cooperative mechanisms are also involved (for example cooperation between macromolecules to draw metabolic pathways). Furthermore he has demonstrated this hypothesis was not in contradiction with Darwinian Theory (he spoke of "Darwinian Dynamics") and could explained organisation of life, at least from subcellular to organismic levels. We did suggest biodiversity increase, at the geological time scale, should also be explained by such mechanisms appearance at the ecological level (Pave *et al.*, *op. cit.*). These kinds of mechanisms are more and more found in wildlife. But how to explain the time needed by the scientific community to consider the importance of such mechanisms? There might be two major reasons for that delay: (1) Human societies justifying an ideology, liberalism, by stating competition is the main factor of progress in wildlife too; (2) cooperative mechanisms being difficult to identify because of their diversity and complexity. Whatever the reason is, even if cooperative mechanisms seem to be important, we should not in turn neglect competition. In fact, in the appearance of structures, both mechanisms play an important role; acting simultaneously their relative importance may vary during time and space, and depend on the background dynamics of the system. For a general overview of cooperative processes, both in biological and social systems, one can refer to the collective book edited by Harmmerstein (2003).

Be that as it may, even if exact mechanisms are still under debate, the general framework postulated by Darwin in his famous book « On the origin of species by means of natural selection. Or the preservation of favoured races in the struggle for life », is still admitted (Darwin, 1859). The title in itself is a summary of the book. However if it gives an explanation for biological diversification, it doesn't give a framework for biological organisation, at least in the common interpretation.

Today, not only evolutionists, but biologists and ecologists too pose this problem of global biological organisation "from genes to ecosystems and biosphere", and the implied mechanisms that led to the actual nested hierarchy of organized entities and trophic networks (e.g. eater-eaten relationships, and more generally the flow of matter and energy through the living world and exchanges with the non living one). This is one of the main research topics developed for the last years in "Integrated Biology". More generally the biological part of the *complex systems theory* is also under development (cf. section 4).

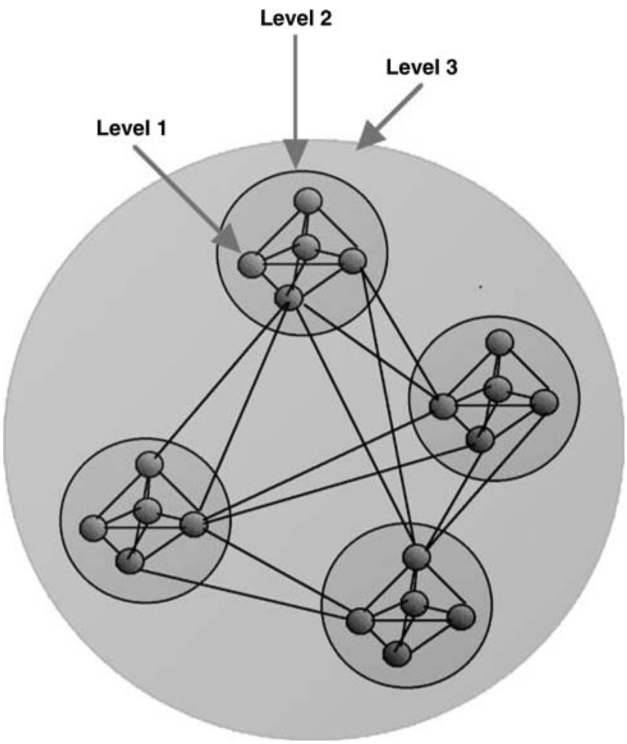


Figure 4: Organisation levels: living systems are self-organizing by assembling in networks. That leads to a hierarchy of nested levels. For example an organism (level 3) is composed of organs (level 2, even if not properly a level of organisation) and organs composed of cells (level 1); in ecology, a community (level 3) is constituted by interacting populations (level 2), themselves sets of interconnected organisms (level 1).

One of the problems concerning self-organisation was the apparent contradiction with the classical thermodynamics (developed for close equilibrium states). G. Nicolis and I. Prigogine have demonstrated that spontaneous self-organisation can occur in systems far from the equilibrium (dissipative structures, characterized by collective coherent behaviours: Nicolis and Prigogine, 1977). Living systems belonging to this category of non-equilibrium systems, the contradiction is removed. On this subject, Kauffman's key work, from Santa-Fe Institute, must be considered as fundamental (e.g. Kaufman, 1993, 1999). He is also a reference regarding both concepts of complexity and complex system (detailed in section 4.1.). A general and recent point of view concerning physical, biological and social sciences, is presented in a special issue of PNAS (Turcotte and Rundle, 2002).

We discussed assumed biological systems trends to self-organisation by considering elementary mechanisms: reproduction, individuals' mortality and balance between cooperation (following Michod's theory on Darwinian dynamics, *cf.* abov*e*) and competition in inter-individual relationships. In the recent paper already mentioned (Pavé *et al.*, *op. cit.*), based on the analysis of global biodiversity variations since the Ordovician period, we have assumed biodiversity growth after a crisis, more than an increase in ecological niches' number, corresponds at the ecological level to appearance of coexistence and cooperative mechanisms -postulated by Michod at the cellular and organismic levels-, at least since the Triassic period. Consequently, we emphasized the need for the development of an ecological paleontology and an evolutionary ecology. In this framework, it seems particularly important not only to reconstitute old age ecosystems' biodiversity, but also as far as possible to elaborate a spatial representation of these paleoecosystems.

To help us in this query apart from paleontology, biological and ecological observations and experiments, modelling plays a particular role. We will examine more closely this role in section 4.

2.1.3 Chance and necessity, and necessity of chance?[4]

Finally, the role of chance has to be examined in new species appearance, implementation and the establishment of relationships with other species in ecological context (e.g. cooperative interactions). Early in the development

[4] A specific paper on this subject: Pavé A. Chance and Necessity and Necessity of Chance: towards a general theory on biodiversity. (book to be published). See also: http://www.academie-agriculture.fr/files/publications/notes_conjoncture/20041207note5.pdf

of modern biology and ecology, stochastic processes had to be considered as good models for many natural dynamics (such as Bernoulli, Markov or Galton-Watson processes). Here we solely focus on the evolutionary point of view:

- Stochastic processes at molecular and genomic levels, during life or at the time of reproduction (Kimura's neutralist theory of evolution, local mutations, DNA recombination and DNA acquisition),
- At ecological level, trees' spatial distribution in tropical rainforests for example, or, even though it is not expressed in terms of stochasticity, the neutralist theory of biodiversity (Hubbell, 2001) and discussions about it (Chave, 2004). Establishment of ecological relationships, between individuals of a population, could, at least for a part, be stochastic (try and error, only "working" interactions are selected, and subsist those giving an evolutionary advantage).

Stochasticity participates to diversification (at genetic level), to its upholding and reinforcing (at ecological one), and probably also to hierarchies' emergence. Considering stochasticity is a term used for games and/or for solving problems on computers too, we should remain careful about analogies. In those cases however, randomness is allowed by physical device (e.g. roulette) or manipulation (e.g. card shuffling) to ensure diversity and therefore games unpredictability, or also simulated by an algorithm (e.g. problem solving by the Monte Carlo algorithm). Processes leading to stochasticity (natural roulettes or card shuffling) may have appeared and evolved in natural systems; they did generate diversity then permit permanence of life despite extinction hazards (a great number of species leads to a better chance for some of them to resist to drastic environmental variations).

Thus, complexity and hierarchical organisation result from two fundamental types of processes: deterministic processes on one hand ones, stochastic ones on the other hand.

2.2 Human activities' role

Living organisms interact more or less together through a variety of mechanism (see above). Some play a particular and important role: human organisms, particularly with many other species domestication since the beginning of agriculture and rearing (about 10000 BP). It also corresponds to the first managements of nature: fields design, irrigation, crops, … that progressively led to existing arranged and artificial landscapes in many countries. Human activities have created news levels of living systems organisation: fields, meadows, planted forests, gardens, farms, large scale crops (such as in north American Middle West)…, landscapes far from spontaneous natural systems organisation. When these systems are left to their own dynamics, they return to a quasi-natural state (e.g. fallows) very different from the human driven system (for example, mono-specific crop does not maintain itself in this state but field vegetation rapidly becomes very diversified). These are direct effects of man action, specifically on land uses, land cover and land structure (anyone flying above cultivated region, looking at the soil, can see that: a mosaic of fields covered by diverse crops). They are strongly dependents on agronomic technologies, on economy and related policies.

These human activities and other ones, mainly industries and transports, also have indirect effects: further geographical space management (roads, canals, built-up areas, …) and pollutions. Today the better-known consequence is the global modification of the greenhouse effect and related increase in the average temperature on the globe. These modifications have led to changes in natural vegetation as well as in crop and plantation compositions.

2.2.1 Land use, land cover changes, spatial organisation evolution

This question of human continental areas organisation is important. It is the reason why an international programme has been launched on the subject some years ago (LUCC: land use, land cover change, effectively launched in 1997; the current scientific program has been defined in 2001; a synthetic presentation can be found in Pavé *et al.*, 2003). Terrestrial systems types are globally given in figure 4. The main part of continental areas changes is consequences of farming strategies mostly inspired by environmental constraints (climate, soils, hydrology), technical possibilities (machinery, chemical products, crops', rearing and livestock products' varieties), economy, national and international agriculture policies. Consequences are: crop changes, deforestation. A non-desired effect is desertification, resulting from an over exploitation of lands. On shall note urban areas are still negligible at the continental scale, but we must point out that very often towns are built on rich lands.

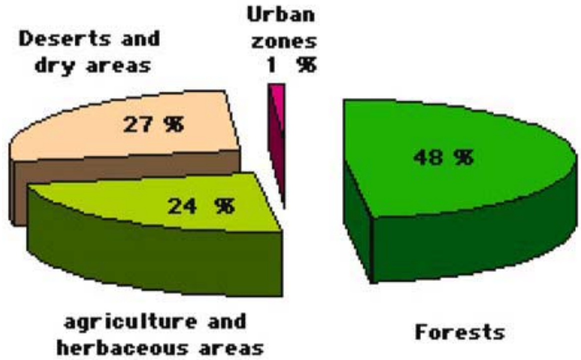


Figure 5: Global repartition of terrestrial ecosystems, except polar areas (statistics correspond to 114 million km2; data from Melillo *et al.*, 1993, and scheme from Schmidt-Lainé, 1999).

Observation of organisation and changes of land cover is not an easy task. One of the principal tools is Remote Sensing. However, pictures direct use is not sufficient; data analyses, particularly classification algorithms, have to be used (Di Gregorio and Jansen, 2000). Apart the global hierarchical representation of terrestrial systems, already known (see figure 6), Remote Sensing analysis gives a map of the distribution of corresponding objects in geographic space.

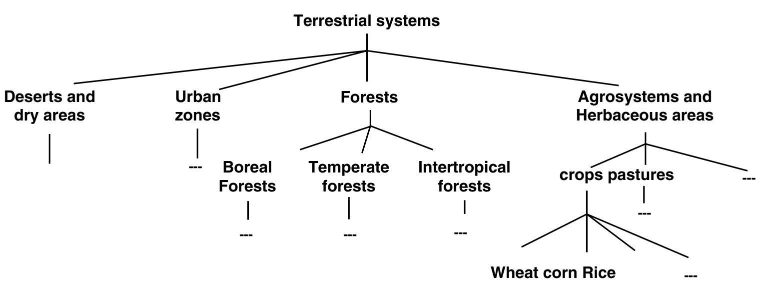


Figure 6: Example of hierarchical representation of land use and land cover. This kind of representation does not take into account the spatial distribution of objects belonging to these categories (apart for forest, but very roughly following bioclimatic zones).

2.2.2 Ecological systems engineering

Ecological systems are living systems, of higher level of organisation than populations (cf. figures 1 and 2). They are natural (i.e. mainly under the dependence of spontaneous processes) or more or less artificial. Crops and gardens, town parks are among the most artificial ones. Facing existing problems, such as effects of chemical pollutants, of new species, human selected and manipulated plants and animals introduction, we have to consider ecological manipulations effects on the whole ecosystem (e.g., introduction of a predator in order to limit a pest extension) and to introduce more and more ecological knowledge for their management: ecological components (populations and communities), hierarchical organisation, biochemical role (contribution to biogeochemical cycles and to biochemical products production), geophysical function (soil stabilisation, hydrology,…) and spread in geographical space. Exactly we have to forecast dynamics after perturbation as well as state after a period of time.

Ultimately, the dream is to envisage a desired state and manage the transition from an initial sate to that expected final one. This is the goal of what we could call an "ecological systems engineering" (Barbault and Pavé, *op. cit.*). It requires not only ecological knowledge, but also specific research to pass from basic results to operational ones, like the necessary research to pass from a chemical reaction in laboratory to an industrial process. That field of research should be call "sciences for ecological systems engineering".

3. MODELLING APPROACHES – COMPLEXITY AND HIERARCHY

Models are very efficient tools enabling a better understanding of how hierarchies emerge in biological systems. Another concept, necessary when speaking of hierarchies, is complexity. Hierarchical organisation can appear mainly in complex systems. What is complexity? What are complex systems?

3.1 Complexity and complex systems

"*Complexity, Science of 21st century*" is the title of a recent issue of the French edition of Scientific American (Pour la Science, 2003). Previously, Science published a special folder on the subject (Callangher and

Appenzeller, 1999) as well as Nature (Allen *et al.*, 2001). However Vicsek wrote in Nature: "*If a concept is not well defined, it can be abused. This is particularly true of complexity.*" (Vicsek, 2003).

In fact, the notion of complexity is not new, but the principal modern base is certainly the Weaver's paper published in American Scientist (1947). Beginning of the 90's, complexity became really operational, mainly when it was associated to the concept of system making what is called "complex system theory". The Santa-Fe Institute is one of the principal locations where the reflection has been elaborated. Shortly, a complex system is a network of elementary units that has the following characteristics: emergence of new properties (it distinguish complex systems from complicated ones); different relationships between units (short-range, non-linear, presence of feedback loops); they are open (a part cannot contain the whole); they have an history; they are nested; boundaries are difficult to determine and mostly result from the observer's choice.

3.1.1 Towards a definition of complexity

We do largely agree with Vicsek position and, without specific criticism of the above listed characteristics, have tried to give a definition of this concept from the analysis of its common uses in literature in a recent paper (Schmidt-Lainé and Pavé, to be published). First we proposed to distinguish between *structural complexity* (complex spatial architecture or interconnected sub-systems' functional organisation, see figure 7) and *behavioural or dynamical complexity* (mainly resulting from non-linear dynamics). Furthermore we pointed out that a simple structure can still have complex dynamics (e.g. the well known case of a single population displaying chaotic dynamics (May, 1976)), and a complex structure can exhibit simple dynamics (*e.g.* a multi-compartmental system with many compartments and exchanges between compartments, but where exchanges are simply governed by linear relationships).

The first part of this definition, structural complexity, refers to networks. As a matter of fact, hierarchical organisation appears when, within a network, subsets of elements are more related together than with other ones (figure 4). Thus hierarchies' emergence is first of all a way to organize complexity (spontaneous or human driven emergence), meaning somewhere to simplify it. A second step can be observed when subsets become specialized in tasks' and functions' categories, alike organs are for an organism.

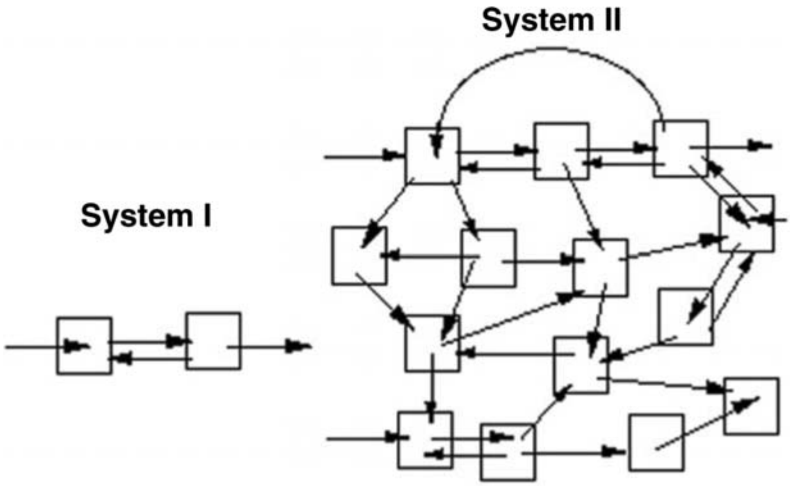


Figure 7: Examples of simple (I) and complex systems (II) from (Schmidt-Lainé and Pavé, *op.cit.*)

The dynamical complexity is related to the changes on a system over time. Sometimes theses changes are smooth and progressive, they may also seem to be erratic, and we might observe rapid transition. The discrete time logistic model of population dynamics give good examples of a great variety of qualitative behaviours (figure 8).

Usually, one of the basic questions when analysing complexity is to find an order in an apparent disorder. Zoologists and botanists' works provide a paradigm. For the last three centuries, they have shown how taxonomy leads to comprehensive classifications of animals and plants: that is to say a *hierarchical model* of life organisation. Before that the living world appeared extremely complex and not immediately understandable.

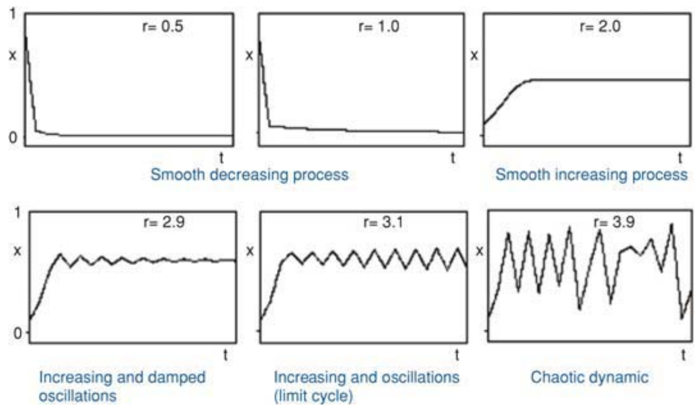


Figure 8: Example of dynamical complexity: the classical discrete time logistic model (May 1976). In spite of the simplicity of its mathematical expression: $x_{t+1} = r\, x_t\, (1 - x_t)$ complex dynamics are observed for values of parameter $r > 3.6$. This example, which is a paradigm of dynamical complexity, also shows what is called "bifurcation": variation of parameter "r" changes the qualitative behaviour of the model, from smooth to chaotic dynamics.

3.1.2 Complexity and modelling – modelling complexity

The problem often comes down to finding a good model that represents (and sometimes explains) natural systems' complexity and to detecting underlying organisations. We consider, for example, two populations displaying chaotic dynamics, with the same demographic parameters but independent dynamics when isolated (independence is ensured by differences in initial conditions; sensitivity to initial conditions is indeed a well-known property of chaotic systems). If they are interacting through a competitive relationship then dynamical structures progressively appear, strange structures at first (being in fact a "strange attractor") and progressively structures become simpler than previous ones. Finally dynamics synchronized (in "phase space" $\{x_t, y_t\}$ successive points representing these populations' simultaneous density arranged themselves along a straight line). These dynamics are drawn in figure 9. That kind of emergence can be considered as the signature of an organisation level.

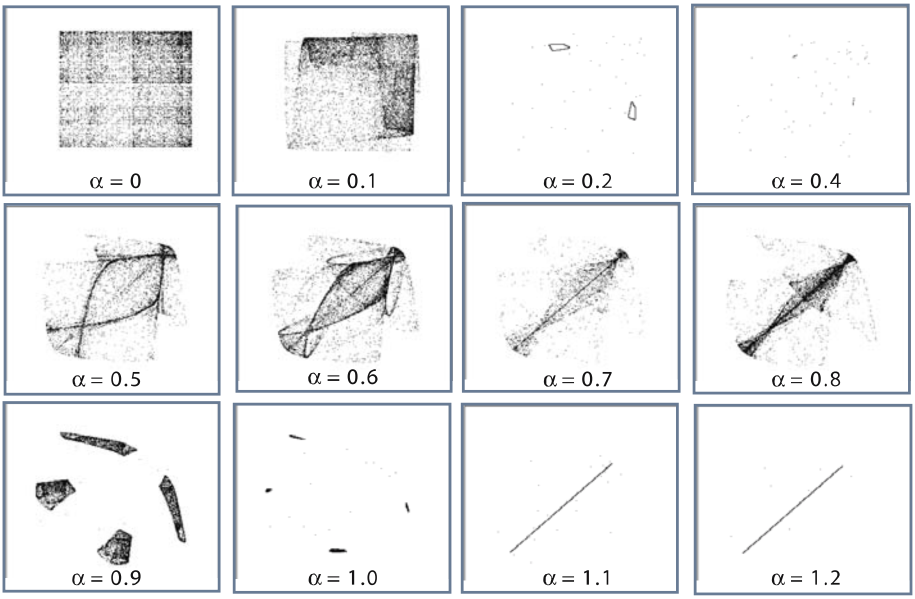


Figure 9: An organisation emergence in two chaotic and interacting populations simulated from the discrete time model:

$$x_{t+1} = r\,x_t\,(1 - x_t) - \alpha\,x_t\,y_t$$
$$y_{t+1} = r\,y_t\,(1 - y_t) - \alpha\,x_t\,y_t$$

Where $r = 3.77$, $x_0 \neq y_0$, graphs' x-axis: x_t, y-axis: y_t, $t = 0, 1, 2, \ldots, 10^4$.
Values of parameter α control the intensity of the interaction: if $\alpha = 0$, the dynamics are independent, if $\alpha \geq 1.1$ the dynamics are synchronized (straight line of slope 1). Note the strange forms of the attractor for values between 0.2 and 1.1 (from Pavé and Schmidt-Lainé, 2003).

3.2 Some different modelling approaches

Mathematical, statistical or computer-based tools are becoming increasingly efficient and more widely used. Some examples are:

- Statistical data analysis (multivariate analysis) enables detection of complex structures (Thioulouse *et al.*, 1995). There are lots of applications to ecosystem and to genome analysis. Based on linear algebra and Euclidian geometry, today non-linear extensions are also taken into account. They also enable to recognize organisation by detecting groups and clusters. For example, a simple principal components' analysis of frequencies of codons from *Escherichia coli* genes distinguishes high expressed genes from low expressed ones. Based on the same theoretical corpus, hierarchical classification and numerical taxonomy allow organizing data in trees. These techniques can be applied to a large variety of data (for example, to analyse remote sensing data, ecological data, genetic code data, etc.). That explains its success. Today new developments even enable to detect complex non-linear structures (Gershenfeld *et al.*, 1999).

- Mathematical modelling and numerical simulations, now largely developed, lead both to basic and practical results (e.g. to answer theoretical questions about stability or emergence of regularities in complex systems, or to provide guidelines for managing biological, ecological, technical or environmental systems). Some results have been obtained about the emergence of simple global laws, for example the power law summarizing many complex structures (West *et al.*, 1999; Ferrière and Cazelles, 1999). It is amazing to note that many more or less interconnected units behaviour, may have a global simple dynamics (for example, the logistic model is known, since its establishment by P.F. Verhulst (1838-1842), to give a good representation of human demographic data). Does the emergence of "simplicity" from complexity characterize the existence of an organisation level? The example presented in figure 9 suggests such hypothesis.
- Computer based modelling, quantitative and qualitative simulations (e.g. cellular automata, multi-agents systems, individual-centered modeling) are more and more developed. They enable representations of mechanisms difficult to take into account in mathematical models (e.g. agents or individuals' behaviors as well as their environment's spatial structure and dynamics). Consequently these methods are well adapted to study the emergence process (Ferber, 1995). Specific computer systems have also been developed; it is for example the case for plant's architecture (cf. figure 10).

3.2.1 Mathematical models vs computer based models

Mathematical models are based on a strong theoretical framework, developed over more than 2,000 years. Mathematics offers well-defined concepts, which can be successfully used in other fields of science. For example, the theory of dynamical systems introduces the concepts of stability, attractive (or repulsive) sets, chaotic trajectories, resilience, bifurcation, etc. Theorems lead to general results. When numerical simulations of these models are used, their coherence and consistence can be often verified. However it is difficult to represent some mechanisms or structures, such as individual or social behaviours, biological structures, such as the morphology of an organism, or even spatial heterogeneity.

Computer models, such as cellular automata or multi-agent systems can do that, but there are only few theoretical results (mainly for cellular automata). In that case, coherence and consistence of simulations are difficult to verify. Generalizations can be obtained only when multiplying simulations, and even then, we are never sure the results are "universally true": a simulation is indeed not a demonstration. Be that as it may, they are useful and illustrative tools, particularly to visualize structures' geometry (one can find many examples in the Wolfram's book devoted to this kind of study, Wolfram, 2002). Exciting results have also been obtained from "artificial life" simulations. Here we must mention Stuart Kauffman's works; he was a pioneer in the domain (see for example, Kauffman, 1995). We can also quote other important references (Hall *et al.*, 2003 and Salther, 1993)[5].

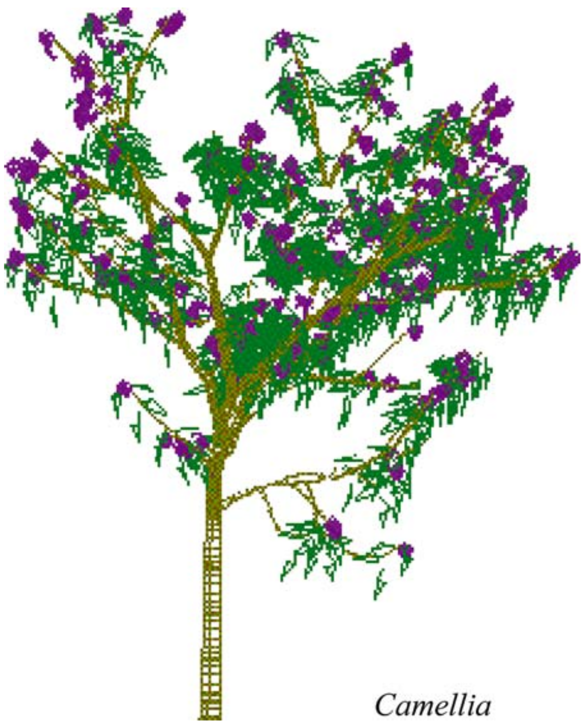


Figure 10: Example of a tree simulated by AMAP software (*Atelier de Modélisation de l'Architecture des Plantes:* Plant Architecture Modelling System, cf. de Reffie and Edelin, 1989, de Reffie *et al.*, 1997). Beside its spectacular aspect, it is also an example of a model of a basic hierarchical biological structure (we use the word "tree" and the model of its structure to represent hierarchies).

[5] Be careful, these books refer to "development", but in the context that word means development of biological organism and not development of human societies.

Today, a complementary approach is usually advised, linking mathematical and computer based modelling. For example, simulations of a well-known system of which a mathematical model exists can help us verifying the functioning of a multi-agent simulator (e.g. the simulation of a Lotka-Volterra predator-prey system). Aggregating results of simulations can also lead to simplified results, enabling then mathematical global approaches. For example, despite microscopic interactions' complexity, an organism or a population's growth is often well represented by a simple logistic model (as mentioned above for human populations).

3.2.2 Integrated modelling: models' association and hierarchies' models

Integrated modelling has two meanings: the first one is to combine together different processes' models at a given level in order to lead to a global model; the second one is to represent a hierarchical organisation (often related to different scales).

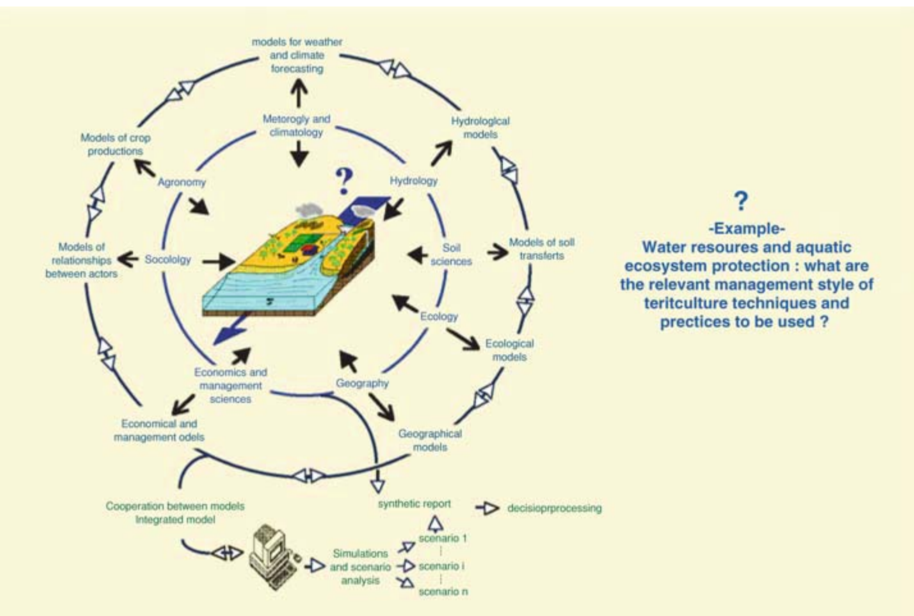


Figure 11: An example of integrated modelling approach to solve an environmental problem (Schmidt-Lainé & Pavé, 2002). The model links together elementary process-based models of the same kind (e.g. models of enzyme kinetics to simulate a metabolic pathway) or of different kinds (e.g. models of physical processes, biological processes, economical processes, etc. as showed in this example).

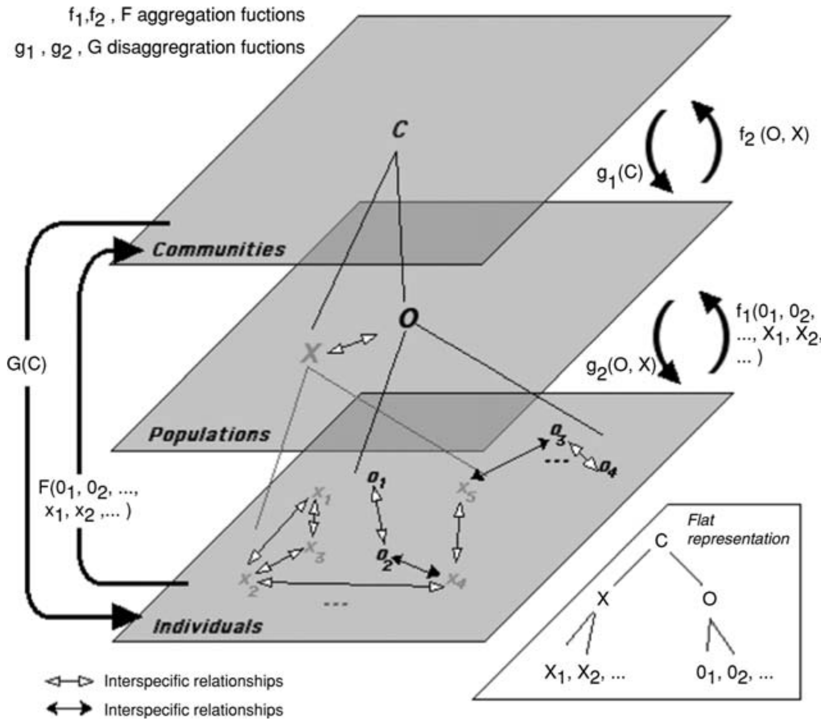


Figure 12: Modelling problems of hierarchical organisations are: (1) how individual dynamics' models can be summarized at population level in a simpler model than a combination of individual ones and also pass from populations to communities (f_1, f_2); (2) How models at population level can be disaggregated at the individual one (g1), and the same from communities to populations (g_2); (3) How to take into account the effect of an upper level on a lower one (i.e. how a community can influence population or individual dynamics) and conversely (G and F). Finally note that successive planes could represent geographical spaces where these entities are functioning. Somewhere, these problems are related to the complexity's analysis and reducing. This kind of representation looks like the one used in objects centered programming, with a hierarchical structure (e.g. in some knowledge base systems). Such elementary models' organisation may be useful for the conception of an integrated model of the hierarchical system.

References about these topics can be found in Science special issues (about models in population's biology: Levin *et al.*, 1997, and about bioinformatics: Shatz *et a.l*, 1997). We have just terminated an article about that (Pavé, to be published, 2005).

Sometimes elementary models of processes governing a system are known or can be reasonably spelled out. Linking these models together enables dynamics' simulation. However, it is essential to carefully and progressively proceed in establishing these links, or rather in writing the models of these links. If not, artefacts may be introduced. This constructive approach is increasingly used. The well-known general circulation models for climate simulation are elaborated following such a strategy. It is also the case for global biosphere or ecosystem models (e.g. Melillo *et al.*, *op. cit.*). Architecture of such a model, which assembles different processes' elementary models is displayed in figure 12.

CONCLUSION

Evidence of living systems' hierarchical organisation is no longer to be demonstrated. It is obvious. Nevertheless the following problems remain: (1) what are exactly the mechanisms leading to such an organisation? (2) What are the relative roles of deterministic and stochastic processes? Does hierarchical organisation result from the complexity of living systems (which is also evidence)? (3) What are complexity and hierarchical organisation's "advantages" for living systems (stability, resilience, permanence, diversity, etc., cf. the discussion initiated by R. May, 1973)? (4) Are complexification and organisation obligatory consequences of a systems dynamics' (living or not) "general law", as suggested by Prigogine and Nicolis (1977)? (6) What are chance and determinism's relative parts in such organisations? (7) How to take into account these characteristics into human decisions and actions on these systems?

This is a non exhaustive list of questions for experimental and theoretical research. One thing is sure: models have to play an essential role in these queries.

Moreover, one of the main practical questions we have to answer in the near future is how to deal with complex systems and their capabilities to organize themselves in structured hierarchies. It is now evident that most natural processes lead to diversity, complexity and self-organisation. This is also true for many human activities, including technological developments, social structures, and human planned and built systems, such as agrosystems, landscapes and urban systems. In the past, the tendency was to simplify, with the result that most management rules and techniques were elaborated for simplified systems (for example, for mono-specific crops in agriculture). We now know that was not always a good solution: since we are dealing with complexity, we must therefore find new rules adapted to such systems' management. Modelling and simulation can provide efficient methodologies for developing new and appropriate management tools.

REFERENCES

Allen, L., Publisher, 2001, Nature insight: Complex systems, *Nature*. **410**.

André, J.-C., Mégie, G., Schmidt-Lainé, C., 2003, Échelles et changements d'échelles, problématiques et outils, in: *Études sur l'environnement : de l'échelle du territoire à celle du continent*, P. Caseau, ed., Rapport sur la science et la technologie, Académie des sciences, Tech&Doc, Lavoisier, Paris, pp. 167-199.

Barabási, A-L., 2002, *Linked: The New Science of Networks*, Perseus Pub, New-York.

Barbault, R., 1992, *Ecologie des peuplements,* Masson, Paris.

Barbault, R., Guégan, J.-F., Hoshi, M., Mounolou, J.-C., van Baalen, M., Wake, M., Younès, T., eds., 2003, Integrative Biology, Complexity in Natural Systems. Keys to Adressing Emerging Challenges, *Biology international*.

Barbault, R., Pavé, A., 2003, Écologie des territoires et territoires de l'écologie, in: *Études sur l'environnement : de l'échelle du territoire à celle du continent*, P. Caseau, ed., Rapport sur la science et la technologie, Académie des sciences, Tech&Doc, Lavoisier, Paris, pp.1-49.

Begon, M., Harper, J.L., and Townsen, C.R., 1996, *Ecology. Individuals, Populations and Communities*, Blackwell Science, Oxford, 1068p.

Blasco, F., ed., 1997, *Tendances nouvelles en modélisation pour l'environnement*. Elsevier, Paris.

Blasco, F., Weill, A., ed., 1999, *Advances in Environmental and Ecological Modelling*. Elsevier, Paris.

Bourgeon, P.S., Humphries, H.C., Jensen, M.E., Brown, B.A., 2001, Integrated Ecological Assessements and Land-Use Planning, in: *Ecological Principles of Land-Use*, V.H. Dale, and R.A Haueber, eds, Springer-Verlag.

Bousquet, F., Cambier, C., Mullon, C., Morand, P., Quensière J., and Pavé, A., 1993, Simulating the Interaction Between a Society and a Renewable Resource, *Journal of Biological Systems*. **1/2**: 199-214.

Brown, J.H., West, G.B., 2000, *Scaling in Biology*, OUP.

Callangher, R., Appenzeller, T., 1999, Beyond Reductionism. *Science*, **284** (special issue).

Carroll, S.B., 2001, Chance and necessity: the evolution of morphological complexity and diversity, *Nature*. **409**: 1102-1109.

Caseau, P., ed., 2003, *Études sur l'environnement : de l'échelle du territoire à celle du continent*, Rapport sur la science et la technologie, Académie des sciences, Tech&Doc, Lavoisier, Paris.

Chave, J., 1999, Study of structural, successional and spatial patterns in tropical rain forests using TROLL, a spatially explicit forest model, *Ecological Modelling*.**124**: 233-254.

Chavez *et al*., 2003, From anchovies to sardines and back: Multidecadal change in the Pacific Ocean, *Science*. **299**: 217-221.

Coll., 1997, Human-Dominated Ecosytems, *Science*. **277**: 445-608.

Coll., 2003, La complexité. La science du XXIe siècle, *Pour La Science*. **314**: 28-156.

Combes, C., 1995, *Interactions durables – Écologie et évolution du parasitisme*, Masson, Paris, 524p.

Costantino, R.F., Desharnais, R.A., Cushing, J.M., Dennis, B., 1997, Chaotic Dynamics in an Insect Population, *Science*. **275**: 389-391.

Courtillot, V., Gaudemer, Y., 1996, Effects of mass extinctions on biodiversity, *Nature.* **381**: 146-148.

Darwin, C.M.A., 1859, On the origin of species by means of natural selection. Or the preservation of favoured races in the struggle for life, John Murray, Albemarle Street, London.

de Reffye, Ph., Edelin, C., Jaeger, M., 1989, La modélisation de la croissance des plantes, *La Recherche.* **207**: 159-168.

de Reffye, Ph., Houllier, F., Blaise, F., Barthélémy, D., Dauzat, J., Auclair, D., 1997, Modélisation et simulation de la croissance d'une architecture végétale : une approche morphogénétique expérimentale, in: *Tendances nouvelles en modélisation pour l'environnement*, F. Blasco, ed., Elsevier, Paris, pp. 91-112.

Di Gregorio, A., Jansen, L.J.M., 2000, *Land Cover Classification System (LCCS)*, FAO, Rome, 179 p.

Ferber, J., 1999, *Multi-agent systems. An introduction to distributed artificial intelligence.* Addison Wesley (translation of the book: Ferber, J., 1995 *Les systèmes multiagents : vers une intelligence collective*, InterEdition, Paris).

Ferrière, R., Cazelles, B., 1999, Universal power laws govern intermittent rarity in communities of interacting species, *Ecology*. **80(5)**: 1505-1521.

Gershenfeld, N., Schoner, B., Metois, E., 1999, Cluster-weighted modelling for time-series analysis, *Nature*. **397**: 329-332.

Grimoult, C., 1998, *Évolutionnisme et fixisme en France : Histoire d'un combat (1800-1882)*, CNRS Editions, Paris.

Hall, B.K., Peason, R.D., Müller, G.B., 2003, *Environment, Development and Evolution*, Bradford Book, 352 p.

Hammerstein, P., ed, 2003, *Genetic and Cultural Evolution of Cooperation.* MIT Press, 450 p.

Hubbell, S.P., 2001, *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton Univ. Press, Princeton. See also, Whithfield J., 2002, Neutrality versus the niche, *Nature*. **417**: 480-481.

Kauffman, S., 1993, *Origin of order self-organisation and selection in Evolution*. Oxford University Press, Oxford.

Kauffman, S., 1995, *At Home in the Universe*, Oxford University Press, Oxford.

Legay, J.-M., 1988, Méthodes et modèles dans l'étude des systèmes complexes, in: *Pour une agriculture diversifiée*, M. Jollivet, ed., l'Harmattan, Paris, pp. 14-24.

Levin, S.A., Grenfell, B., Hastings, A., Perelson, A.S., 1997, Mathematical and Computational Challenges in Population Biology and Ecosystems Science, *Science*. **275:** 334-342.

Lewin, R.,1993, *Complexity: Life and the Edge of Chaos*. Macmillan Publishing Co., New York, 208p.

Lions, J.L., 1997, Modélisation mathématique et environnement – Quelques remarques, in: *Tendances nouvelles en modélisation pour l'environnement*, F. Blasco, ed., Elsevier, Paris, pp. 79-90.

Mackenzie, D., 1999, New Clues to Why Size Equals Destiny, *Science*. **284**: 1607-1607.

May, R.M., 1976, Simple mathematical model with very complicated dynamics, *Nature*. **261**: 459-467.

May, R.M., 1973, *Stability and Complexity in Model Ecosystems.* Princeton University Press, Princeton.

McCann, K., Hastings, A. and Huxel, G.R., 1998, Weak trophic interactions and the balance of nature, *Nature*. **395**: 794-798.

Melillo, J.M., Mc Guire, A.D., Kicklighter, D.W., Moore, B., Voromarty, C.J., Schloss, A.L., 1993, Global climate change and terrestrial net primary production, *Nature.* **353**: 234-240.

Michod, R.E., 2000, *Darwinian Dynamics*, Princeton paperbacks, Princeton.

Nicolis, G., Prigogine, I., 1977, *Self-Organisation in Nonequilibrium Systems*, J. Wiley & Sons, New York.

Pavé, A., 1994. *Modélisation en biologie et en écologie*, Aléas, Lyon.

Pavé, A., 2000, French environmental labs may get “big science” funds (Interview), *Nature*. **403**: 822.

Pavé, A. Modelling Living Systems, their diversity and their complexity. Some Methodological and Theoretical Problems. *C.R. Acad. Sciences – Biologies* (to be published, 2005).

Pavé, A., and Schmidt-Lainé, Cl., 2003, Integrative biology: modelling and simulation of the complexity of natural systems, *Biology International*. **44**: 13-24.

Pavé, A., Hervé, J.C., Schmidt-Lainé, Cl., 2002, Mass extinctions, biodiversity explosions and ecological niches, *Biologies*. **325(7)**: 755-765.

Pavé, A., Pumain, D., Schmidt-Lainé, Cl., 2003. Dynamique des territoires : les changements de couverture et d’utilisation des terres, in: *Études sur l’environnement : de l’échelle du territoire à celle du continent*, P. Caseau, ed., Rapport sur la science et la technologie, Académie des sciences, Tech&Doc, Lavoisier, Paris, pp. 51-77.

Salthe, W., 1993, *Development and Evolution – Complexity and Change in Biology*, Bradford Books, 368p.

Schatz, B.R., Levin, S.A., Weinstein, J.N., *et al.*, 1997, Bioinformatics, *Science*. **275**: 327-349.

Schesinger, M., 1998, From chaos to complexity (a review of books), *Nature*. **394**: 538.

Schmidt-Lainé, Cl., 1999, *Échanges forêts-atmosphère. Le cas du C02,* Rapport à l’Académie des Sciences.

Schmidt-Lainé, Cl., Pavé A., 2002, Environnement : modélisation et modèles pour comprendre, agir et décider dans un contexte interdisciplinaire, *Natures, Sciences, Société.* special issue « Sciences pour l’ingénierie de l’environnement », **10 (s.1)**: 5-25. (Summary in English).

Schmidt-Lainé, Cl., Pavé, A., 2005, Bringing complexity back to life. *Biologist* (to be published).

Solbrig, O., Nicolis, G., 1991, Perspectives on Biological Complexity, *IUBS Monograph series*, **n°6**, Paris.

Thioulouse, J., Chessel, D., Champély, S., 1995, Multivariate analysis of spatial patterns: an unified approach to local and global structures. *Environmental and Ecological Statistics*. **2**: 1-14.

Turcotte, D.L., Rundle, J.B., 2002. Self organized complexity in the physical, biological, and social sciences, *Proc. Nat. Acad. Scien*., **99**: 2463-2465 and all the rest of this issue.

Vicsek, T., 2002, The bigger picture, *Nature*. **418**:131.

Weawer, W., 1947, Science and Complexity, *American Scientist*. **36**: 536-544.

West, G.B., Brown, J.H. and Enquist, B.J., 1999, A general model for the structure and allometry of plant vascular systems, *Nature*. **400**: 664-667.

Wolfram, S., 2002, *A New Kind of Science*, Wolfram media.

Chapter 3

SIZE, SCALE AND THE BOAT RACE; CONCEPTIONS, CONNECTIONS AND MISCONCEPTIONS

Geoffrey B. WEST
President
Santa Fe Institute

Considerations of size and scale play a central role across the entire spectrum of science and technology from practical problems of medicine and engineering to some of the most fundamental conceptual questions of physics and biology. Scaling laws typically reflect, and often reveal, the general principles underlying the structure of a physical problem. This idea is illustrated by considering the conceptual role of scaling in fundamental problems in physics and biology that have their origins in hierarchical fractal-like structures. Pitfalls and misconceptions concerning scaling are illustrated by examining and interpreting the result of the recent Oxford-Cambridge Boat Race.

Beginning with the seminal ideas of Galileo on the size of structures and limits to growth, scaling in its various guises has been a powerful tool in understanding many physical phenomena. Among the many important basic problems that have been usefully addressed by casting them in these terms are the design of ships, buildings and bridges, the size of pharmacological dosages, the understanding of critical exponents in phase transitions, the nature of chaos, the unification of the fundamental forces of nature and their relationship to the evolution of the universe from the big bang. General techniques for analysing such problems include classic dimensional and similarity analyses (and their extensions to modelling theory such as used in wind tunnels) (Sedov, 1959; Birkhoff, 1960), and most recently, their generalisation to the sophisticated machinery of the renormalisation group (Goldenfeld, 1993; Cheng and Li, 1984) and fractal analysis (Feder, 1998).

D. Pumain (ed.), Hierarchy in Natural and Social Sciences, 71-80.

A deceptively elementary, yet important, property central to many scaling arguments is that there exist natural scales appropriate to the specific problem at hand. At a trivial level this is reflected in the use of special units such as the fermi for measuring lengths in nuclear physics or the micron in cellular phenomena, rather than using miles or kilometers, for instance. Conversely, inappropriate scales should not appear in a solution to a problem. For example, in the quantum regime Planck's constant (h) sets the natural scale whereas in relativity it is the velocity of light (c). Consequently, in the classical Newtonian limit, where velocities are much smaller than c and distances much greater than atomic scales, neither h nor c can explicitly occur in formulae describing a physical system. Thus, for example, the classical dynamics of flight or the mammalian circulatory system are neither quantum mechanical nor relativistic so neither h nor c appear in the equations describing them. However, these constants are, of course, implicitly hidden in phenomenological macroscopic quantities such as conductivity, viscosity, and friction which encode the corresponding fundamental physics which occurs at the microscopic scale.

At a somewhat more sophisticated level is the idea that appropriately similar scales must be used when making a comparison between similar problems viewed at different resolutions. For example, consider the classic claim that ants can lift loads "hundreds of times" their own weight (Shuster and Siegel, 1938); (in actuality, this factor is rather more like a hundred than hundreds). In contrast to ants, most of us can barely lift loads comparable to our weight so, does this mean that ants are hundreds of times stronger than we are? This sort of problem was addressed some 500 years ago by Galileo who observed that "a small dog could probably carry on his back two or three dogs of his own size; but I believe that a horse could not carry even one of his own size." Galileo realised that the strength (S) of a beam holding up a building, for example, increases linearly with its cross-sectional area (A). Thus, if the building is scaled up isometrically (that is, all lengths are scaled by the same factor λ,), then $S \propto A \propto \lambda^2$, whereas the total weight of the structure, $W \propto \lambda^3$. This leads to the well-known result that $S \propto W^{2/3}$ and, consequently, that the strength to weight ratio, S/W, decreases as $W^{-1/3}$.

Applying this argument to the musculature of limbs we can ask how heavy a load, relative to his own weight could a 70Kg man lift if he were scaled down to the size of an ant (weighing 0.1g, say). From Galileo's argument the load to weight ratio of a man scaled to an ant should be approximately $(7 \times 10^5)^{1/3} \sim 90$; in other words, if a man were the size of an ant he could lift loads almost 100 times his own weight, comparable in fact to what a real ant can actually do. Ants and humans are therefore equally strong.

A powerful variant of this argument with profound consequences for fundamental physics was first made by Georgi, Quinn and Weinberg (1974). They made the extraordinary observation that three of the four basic forces of nature (electromagnetism, the weak, and the strong nuclear forces) which appear to be so vastly different at "human size" scales, have the same strength when they are all viewed at the appropriately same scale of 10^{-34}cm, the so-called "grand unification scale". This remarkable result is one of the central pillars supporting the idea that there really is a unified theory of all the fundamental interactions; furthermore, it strongly suggests that there was no distinction between these apparently different forces until the average distance between the elementary particles in the expanding universe exceeded this scale, which occurred 10^{-35} secs after the big bang when the temperature of the universe was approximately 10^{28} 0K. The argument is based on a general result derived from a renormalisation group analysis of relativistic quantum field theory (Cheng and Li, 1984; West, 1988) that, to a good approximation, physical quantities generically depend on the combination of variables $le^{1/bi\ \alpha i}$, rather than on l and α_i independently; here, l is the observational scale of resolution, α_i the strength of interaction, and *bi* a calculable constant characteristic of the force. Small space-time scales ($l \rightarrow 0$) are therefore equivalent to $\alpha_i \sim 1/bi \ln l$. Thus, d α_i /d$l \sim -1/bil(\ln l)^2$ so that the α_i either increase or decrease with resolution depending on the sign of the corresponding bi. In quantum electrodynamics (QED), α_i is the familiar fine structure constant ($\alpha \sim 1/137$ at atomic scales) and *bi* a (known) negative number. Thus, in QED α slowly *increases* as the resolution scale *decreases*. The weak force behaves similarly, but the strong interaction (quantum chromodynamics, the theory describing the interaction between quarks and gluons, whose $\alpha_i \sim 1/10$ at nuclear scales) has a positive value for *bi* and so decreases with decreasing scale. Georgi, Quinn and Weinberg (1974) discovered that the scaling of the different α_i coincide at $l \sim 10^{-35}$cm. This idea has now become so well accepted that many researchers believe that it provides compelling evidence for the existence of a new fundamental symmmetry of nature, supersymmetry, which relates fermions to bosons, because it significantly improves the coincidence of the forces. There is no generally accepted theory of quantised gravity, so the precise scaling of this fourth basic force of nature remains uncertain. Nevertheless, there is general agreement that its strength, which is extraordinarily weak at ordinary human scales (its $\alpha_i \sim 10^{-45}$), dramatically increases at very short distances to coincide with the scaling of the force that unifies the other three basic interactions at a scale of ~ 10^{-40}cm (the "Planck scale"). These scaling arguments lie at the heart of the quest to find a unified "theory of everything" and are one of the major driving forces for constructing ever larger accelerators to investigate and test such bold and grandiose theories, including those inspired by string theory.

The generic dependence of the α_i on the observational scale ($\alpha_i \sim 1/b \ln l$) has its physical origins in the uncertainty principle: as the space-time resolution of a particle becomes finer, precise knowledge of its energy and momentum becomes coarser. As such, this allows energy conservation to be violated for the short duration of the observation: the finer the resolution, the greater the violation. The excess energy accessible via this uncertainty principle mechanism permits "virtual" particles and anti-particles to be produced alongside the original "real" particle. Thus, for example, a physical electron always has a cloud of "virtual" photons and electron-positron pairs associated with it though their distribution and characteristics depend on the resolution. Consequently, the nature of any observable particle and its various strengths of interaction (the α_i) depend on the scale of observation. In the pictorial representation of these processes invented by Feynman the dominant structure of these "virtual" particle states appear as tree-like hierarchies. The renormalisation group provides an elegant technique for summing up this complicated hierarchy to give the simple mathematical structure $l\mathrm{e}^{1/b\alpha i}$. This structure reflects the invariance of physical quantities to the choice of the arbitrary normalisation scales necessary to define the physical α_i.

This renormalisation group invariance and the resulting dependence of physical quantities on the combination $l\mathrm{e}^{1/b\alpha i}$ in quantum field theory can be viewed as a generalisation of the conventional use of dimensionless combinations of variables familiar in classical physics (Sedov, 1959; Birkhoff, 1960; West, 1988). These result from the invariance of physical equations to the arbitrary choice of units used to make measurements. For example, in hydrodynamics physical quantities must typically depend upon functions of the well-known dimensionless combinations $\mathrm{R} = vl\rho/\mu$ (the Reynold's number) and $\mathrm{F} = v^2/lg$ (the Froude number), where, v is the velocity, l some characteristic length, μ the fluid viscosity, ρ its density and g the acceleration due to gravity; (see below). This structure forms the basis for classic modelling theory whereby, for example, large ships moving slowly can be simulated by small scale models moving quickly, if they have similar characteristics.

Curiously, the mathematical structure of the scaling of the fundamental forces in quantum field theory, $le^{1/b\alpha i}$, also arises in biology. It has recently been shown that time scales for basic biological processes depend on the combination of variables $M^{1/4}e^{E/kT}$, rather than on M and T independently (Gillooly *et al.*, 2001 and 2002a); here, M is the mass of the organism, T absolute temperature, k Boltzmann's constant, and E the activation energy of rate limiting processes for energy production in the molecular respiratory complex. This similarity becomes even more striking when we realise that functional biological lengths (L) typically scale as $M^{1/4}$ (Schmidt-Nielsen, 1984; Calder, 1984; West *et al.*, 1997a) so this combination can be expressed as $Le^{E/kT}$. Whether this similarity to the scaling of the fundamental forces is just a coincidence or expressive of some deeper underlying mathematical connection, possibly related to generalised scaling arguments based on the renormalisation group, is far from clear. A hint suggestive of a similar origin may lie in the observation that at all scales many biological structures exhibit hierarchical, fractal-like networks which are topologically similar to the tree-like hierarchies in the Feynman diagrams driving the scaling of the strengths of the fundamental forces.

The $M^{1/4}$ scaling behaviour is just one example of ubiquitous quarter powers that occur in biological phenomena over scales ranging from the molecular up to ecosystems (Schmidt-Nielsen, 1984; Calder, 1984). Many basic physiological quantities (including metabolic rate, growth rate, mitochondrial density, prokaryotic genome length and lifespan) scale with size as a power of body mass with an exponent which is typically close to a simple multiple of 1/4. For example, metabolic rate scales as $M^{3/4}$ over almost 27 orders of magnitude from the molecules of the respiratory complex up to the largest mammals (West *et al.*, 2002). It has been proposed that these universal quarter-powers have their origin in generic properties of the various hierarchical fractal-like branching network systems that sustain life at all scales (West *et al.*, 1997b). Examples of such networks include the circulatory, respiratory, renal and neural systems, plant vascular systems, and possible intra-cellular transport networks. The scaling with temperature is determined by the temperature dependence of the energy producing biochemical reactions within mitochondria which are governed by the classic Boltzman factor, $e^{E/kT}$. The combination $Le^{E/kT}$ therefore represents the joint effects of the scaling of the production of energy at the "microscopic" intra-mitochondrial level and transport constraints at the "macroscopic" whole-body level. Biological rates scale as the inverse of time and so behave as $M^{-1/4}e^{-E/kT}$. Thus, the vast differences in biological rates from the smallest to the largest organisms, covering over 20 orders of magnitude, can be largely accounted for by these scaling effects. In other words, if all growth rates, metabolic rates, life spans, and even rates of evolution were viewed at the same scale (that is, rescaled to the same mass and temperature) each would be an approximate invariant (Gillooly *et al.*, 2001; Gillooly *et al.*, 2002a; Gilloly *et al.*, 2002b), in the same way that the strength of an ant is approximately equal to that of a man if both are scaled to the same mass.

Size and scale play a central role not only in science and technology but also in the non-scientific arena where they are often conspicuous, either by their absence, or because of misconceptions. Quoting statistics as per capita, or having different weight classes in sporting events such as boxing or weightlifting, are attempts to recognise a crude concept of scale, albeit a linear one. This naivety is not always confined to the non-scientific; classic cases of assuming that dosages increase linearly with body mass are well-known (Schmidt-Nielsen, 1977). Its remnant can still be seen in the recommended dosage on the label of children's Tylenol bottles: double the weight of the child, doubles the dosage! On the other hand, most cooks are aware that doubling the size of a turkey does not require doubling the cooking time, although they would probably be hard-pressed to say why.

Sporting events provide interesting examples that illustrate the concept of appropriate scale and, at the same time, highlight the difference between popular and scientific standards for what constitutes the "best" or the "strongest" and therefore the "winner". Lietzke (1956), many years ago plotted the world records for the total load lifted by weightlifters (their "strength" S) versus their body weight (W) on a log-log plot. He obtained a straight line whose slope was very close to 2/3, as predicted by the Galilean argument ($S \propto W^{2/3}$). Thus, when properly normalised to the same weight, champion weightlifters in different classes (including presumably champion ants) have approximately the same strength. However, Lietzke noted that the heavyweight (who weighed 198lbs and lifted 1023lbs weight) actually lay slightly below the 2/3 line, whereas the middleweight (who weighed 148lbs and lifted 853lbs weight) lay slightly above. Since weightlifters in the other classes lay on the 2/3 line he concluded that the middleweight, Kostilev, was the greatest champion weightlifter at that time, whereas ironically the heavyweight was the weakest. Indeed, if Kostilev were scaled up to 198lbs he would have been able to lift 1036lbs, a respectable 13lbs more than the heavyweight! It is deviations from the idealised constancy of the ratio $S \propto W^{2/3}$ that is the relevant quantity that determines relative strength. In an analogous fashion it could be argued that it is deviations from idealised quarter power scaling that reflect the interesting differences between biological organisms!

A more subtle example is provided by the recent Oxford − Cambridge boat race[1]. This race is a major international spectacle with almost 250,000 spectators lining the banks of the Thames and a TV audience estimated at almost 400 million. Typically, the race is effectively decided after one crew opens up clear water between the boats, which usually happens relatively early on in the race. This year, however, the two boats stayed in contact over the entire 4 1/4 mile course with Cambridge in a slight lead for most of the race following a poor start. In an exciting finish Oxford came through in the last 20 strokes to win by almost 2/3 of a length. The race was hailed as one of the most enthralling in its long history and the winning crew heartily praised in the press for its extraordinary performance, particularly because they overtook Cambridge on the outside of the last of the three hairpin bends in the river, a feat rarely accomplished in almost 150 years of competition. It was indeed a wonderful performance. The race, however, wasn't entirely without incident: for, with a mile or so remaining and ahead by over half a length, the final bend of the river in their favour (worth approximately 2/3 of a length), one of the Cambridge crew "broke" and effectively stopped rowing. So, for the last mile the race was between a crew of eight oarsmen versus a crew of seven. In winning, the eight (Oxford) effectively gained almost two lengths on the seven (Cambridge) over this last mile. Was this margin of victory to be expected or was this an extraordinary feat, as intimated by many in the press? To put it slightly differently: if the crews were scaled to the same size who would have won and by how much?

The framework for answering this question was developed by McMahon (1971) who showed that the speed of a boat, when all n oarsmen are pulling, scales as $v \propto n^{1/9}$. The derivation is based almost entirely on general scaling arguments which are independent of detailed dynamics. It is elegant yet elementary, and the result agrees very well with data. Now, when one oarsman is not pulling, two major effects come into play: the total power output of the crew decreases whereas its relative weight increases. At a constant speed the power delivered by the crew is exactly balanced by the power, P, dissipated in the viscous drag of the water. As intimated above, general dimensional scaling arguments (Birkhoff, 1960) require $P = \rho v^3 A f(R, F)$, where A is the wetted surface area of the boat and f some function of the Reynolds and Froude numbers. For the case considered here, where the characteristics of both boats and their crews are approximately the same, ρ,R, F and A remain constant, independent of the number of oarsmen, nr, actually rowing. If the power supplied by each oarsman, P_0, is also

[1] http://www.theboatrace.org

approximately constant, then $P = n_r P_0$, so that $v \propto n_r^{1/3}$. This differs from McMahon's scaling law because, even though the number of rowers changes as in his case, the weight of the crew remains the same so the wetted surface area (or equivalently, the depth the boat sits in the water) remains constant. Thus the relative change in velocity when one oarsman stops rowing is given by $\Delta v/v \approx 1/(3n) \approx 1/24$, since $n = 8$; i.e., the velocity decreases by approximately 4%. Over a distance d, this translates into a relative gain by the intact crew of $\Delta d/d \approx 1/24$. If $d \approx 1$ mile, then $\Delta d \approx 220$ feet $\approx 3^{1/2}$ lengths. Since Cambridge were leading by approximately 1/2 length when they were reduced to seven men, and gained a further 2/3 length advantage from the final bend, Oxford should have gone into the lead with 8 over 3/4 mile remaining and have won by approximately 2 1/3 lengths. The victory, however, was gained only in the last few hundred feet with the margin closer to 2/3 length. We therefore conclude that it was actually Cambridge that performed significantly beyond expectation over the last mile! A straightforward calculation shows that their relative pulling power, $\Delta P_0/P_0$, was a remarkable 5% higher than Oxford's. Had they not lost one oarsman the calculation indicates that they would have won by approximately 2 2/3 lengths and completed the course in 17 secs less time than they actually did.

These estimates do not take into account various corrections of secondary importance such as the extra drag on the Cambridge boat because it was unbalanced, or how much the disabled oarsman either contributed to, or hindered, Cambridge's progress. In addition, there are imponderable "psychological" factors which go to the very nature of sport. Although Oxford were behind for most of the race their indefatigable tenacity and continuous challenge throughout was undoubtedly a major contributing factor that led to the loss of one of the Cambridge crew and ultimately set the stage for their victory. The point here, however, is not that, when appropriately scaled, Cambridge "should" have won, but rather that their extraordinary performance over the last mile of the race can only be really appreciated by a quantitative scaling argument analogous to that which showed that Kostilev was the greatest weightlifter in 1956.

To summarise: understanding the nature and origin of scaling laws has been enormously productive in gaining deeper insights into problems ranging across the entire spectrum of science and technology from the mundane to the profound. The observation of relatively simple phenomenological scaling laws typically reflects underlying generic features and physical principles that are independent of detailed dynamics or specific characteristics of particular models. This is in part because scaling laws often span a wide range of parameter space, effectively averaging over details thereby leaving a residue that is primarily sensitive to the basic underlying principles. Chaos, phase transitions, unification of forces, and the discovery of quarks are but a few of the more significant examples where scaling has illuminated important universal underlying principles. Other examples, such as Zipf's laws for the distribution of words in languages or the size of cities (Zipf, 1949), suggest the existence of as yet to be formulated basic laws in these areas. Biology is surely one of the most compelling of these: the existence of a very large number of simple quarter-power scaling laws which span the entire range of life from molecules to ecosystems is undoubtedly telling us something important about the generic principles governing life's structure, function and organisation. It is surely neither an accident nor some diabolic coincidence that the lengths of aortas, tree trunks and prokaryotic genomes all scale in the same way. Galileo's seemingly innocent question regarding how and why things scale has indeed led both to some remarkable insights as well as to some mundane observations, such as how extraordinarily well the Cambridge crew rowed in the last mile of this year's boat race, in spite of losing.

REFERENCES

Birkhoff, G., 1960, *Hydrodynamics: A Study in Logic, Fact and Similitude*, Princeton University Press, Princeton, NJ.

Calder III, W. A., 1984, *Size, Function and Life History*, Harvard University Press, Cambridge, MA.

Cheng, T.-P. and Li, L.-F., 1984, *Gauge Theory of Elementary Particle Physics*, Oxford University Press, Oxford.

Feder, J., 1998, *Fractals*, Plenum Press, New York.

Galileo, Galilei, 1914, *Dialogues Concerning the Two New Sciences*, MacMillan, London.

Georgi, H., Quinn H., and Weinberg, S., 1974, Hierarchy of interactions in unified gauge theories, *Phys. Rev. Lett.* **33**: 451.

Gillooly, J. F. *et al.*, 2001, Effects of Size and Temperature on Metabolic Rate, *Science*. **293**: 2248.

Gillooly, J. F. *et al.*, 2002a, Effects of Size and Temperature on Developmental Time, *Nature*. **417**: 70.

Gillooly, J. F. *et al.*, 2002b, Effects of Body Size and Temperature on DNA and the Molecular Clock, *Nature*. Submitted for publication.

Goldenfeld, N., 1993, *Lectures on Phase Transitions and the Renormalisation Group*, Addison-Wesley, Reading, MA.

Lietzke, M. H., 1956, Relation Between Weightlifting Totals and Body weight, *Science.* **124**: 486.

McMahon, T., 1971, Rowing: a Similarity Analysis, *Science.* **173**: 349.

Schmidt-Nielsen, K., 1977, Problems of Scaling: Locomotion and Physiological Correlates, in: *Scale Effects in Animal Locomotion*, T. J., Pedley, ed., Academic Press, New York.

Schmidt-Nielsen, K., 1984, *Scaling; Why is Animal Size So Important*, Cambridge University Press, Cambridge.

Sedov, L. I., 1959, *Similarity and Dimensional Methods in Mechanics*, Academic Press, New York, NY.

Shuster, J. and Siegel, J., 1938, Superman, *Action Comics*, **1**.

West, G. B., 1988, Scale and Dimension from Animals to Quarks, in: *Particle Physics*, N. G. Cooper and G. B West, eds., Cambridge University Press, Cambridge.

West, G. B., Brown, J. H. and Enquist, B. J., 1997a, The Fourth Dimension of Life: Fractal Geometry and Allometric Scaling of Organisms, *Science*. **276**: 122.

West, G. B., Brown, J. H. and Enquist, B. J., 1997b, A General Model for the Origin of Allometric Scaling Laws in Biology, *Science.* **276:** 122.

West, G. B., Brown, J. H. and Woodruff, W. H., 2002, Allometric Scaling of Metabolism from Molecules and Mitochondria to Cells and Mammal, *Proc. Nat. Acad. Sc.* **99**: 2473.

Zipf, G. K., 1949, *Human Behavior and The Principle of Least Effort*, Addison-Wesley, Cambridge, MA.

Chapter 4

HIERARCHY, COMPLEXITY, SOCIETY

David LANE[1]
Professor of Economics
University of Modena and Reggio Emilia
External faculty, Santa Fe Institute, USA

> "It is a commonplace observation that nature loves hierarchies. Most of the complex systems that occur in nature find their place in one or more of four intertwined hierarchic sequences."
>
> Herbert A. Simon

> "No power has been given by nature the right to decide on the relative importance and the respective hierarchy of the entities that compose, at any given moment, the common world."
>
> Bruno Latour

Over the last several decades, increasingly many researchers – physicists, biologists, social scientists – invoke the word "complexity" to describe their orientation to the problems on which they work. Obviously, there is no consensus about just what "complexity" means, but there is a cluster of

[1] I would like to acknowledge support in the preparation of this chapter from grants from the European Commission, through the Future and Emerging Technologies program in the Information Sciente Technology directorate, and the Italian Ministry of Instruction, Universities and Research. The charter (and I) benefitted from comments on earlier drafts from various friends and colleagues, especially Denise Pumain, Roberto Serra, Andrea Ginzburg and Melissa Lane, none of whom have any responsibility for the errors and ambiguities that remain.

D. Pumain (ed.), Hierarchy in Natural and Social Sciences, 81-119.

associated ideas whose co-occurrence establishes some common intellectual terrain among these researchers: "system", "interactions", "emergence", "self- organisation", "learning and adaptation", "evolution and coevolution", "positive feedbacks", "networks", "distributed control". Some prominent researchers have maintained that "hierarchy" should occupy a central position in this cluster of concepts around complexity. At first sight, it might seem strange to put hierarchy together with concepts like bottom-up emergence, networks or distributed control.[2] However, like complexity, hierarchy partakes of several meanings, and the relation among these meanings – as well as the relation between them and the concepts associated with complexity – may yield productive and deep connections that make an apparent contradiction seem trivial by comparison.

This essay investigates some of the relationships between hierarchy and complexity. After a brief look in Section 1 at some of the possible meanings of "hierarchy", we turn in Sections 2-4 to a discussion of seminal essays by three scientists who thought deeply about hierarchy and complexity: Herbert Simon, Phil Anderson and John Holland. As we shall see, both Anderson and Simon argued in the late 1960's and early 1970's that hierarchy was fundamental for understanding complexity, but their views of what forges the link between these two concepts are very different, even contradictory. In work over four decades, Holland advocated ideas that bear a superficial resemblance to Simon's about why hierarchy is inextricably linked to the organisation of complex systems, but in reality Holland's perspective is very different from Simon's: he focuses on *process* while Simon's approach is primarily *structural*; Holland sees hierarchy as emerging from the *bottom up*, while Simon concentrates on its capacity to order a system, from the *top down*. From the differences in the perspectives of these three scientists, the concept of "level hierarchy" takes on an ambiguity that makes its relationship with "complexity" seem much more problematic than any of the them taken separately would lead us to expect.

In the last two sections, we focus on level hierarchies "over our heads" – that is, in which the levels consist of human aggregations, like those of which societies, cultures, states, or economies are composed. Here, the concept seems particularly ambiguous. For example, Simon, Anderson, Holland and most other proponents of ontological hierarchies have no trouble identifying the levels that comprise physico-chemical or biological hierarchies ("elementary particles", "nuclei", "atoms", "molecules";

[2] After all, what could be less bottom-up than a military or ecclesiastical hierarchy? And in transaction costs economics, the fundamental contrast is between the completely distributed "market" and the centralized "hierarchy"; for the new institutionalists in economic sociology, networks go "beyond markets and hierarchy."

"organelles", "cells", "organs", "multicellular individuals", "populations," "species", "ecosystems"), but unanimity tends to break down about the "natural" level structure of human aggregations. Indeed, some thoughtful social scientists deny the value, even the existence, of the concept of "level hierarchy" applied to social organisation and historical processes. In Section 5, we summarize the attitude toward level hierarchy expressed by two of these scholars, the anthropologist-philosopher Bruno Latour and the historian Carlo Ginzburg. For both these writers, it is the very complexity of social and historical processes that undercuts attempts to define them in terms of levels separated by distinct characteristic spatiotemporal scales. Section 6 concludes the paper with a few modest proposals about the relationship between hierarchy, complexity and society.

1. WHICH HIERARCHY?

"Hierarchy", like other words that name deep and powerful concepts, is polysemous. In this section, I distinguish four different kinds of "hierarchy": order hierarchy, inclusion hierarchy, control hierarchy and level hierarchy. As one might expect from the surname they all share, these kinds in part overlap; in part, though, they are contradictory.

1. **Order hierarchy**: In various technical contexts, hierarchy is sometimes taken to be equivalent to an ordering induced by the values of a variable defined on some set of elements. Consider, for example, the definition offered by Batty in his chapter in this volume: "a hierarchy is a natural ordering that is initially based on size…"[3] This definition has the

[3] This definition seems to be standard among the geographers and economists who study "urban systems". Now an urban system consists of cities that engage in many kinds of interactions and relationships among themselves; that is why it is legitimate to speak of them as a "system", rather than just a "set" of cities. But according to Batty's definition, none of these interactions and relationships is relevant to the existence of an "urban hierarchy." Indeed, various models – from Gibrat (1931) to Gabaix (1999) – attempt to "explain" urban hierarchy via models that deny any role for systemic effects in generating the hierarchy: that is, they derive a distribution for city size on the assumption that the dynamics of population growth for any one city is independent from that of any other! Certainly, as Batty's chapter well illustrates, any serious urban geographer or historian understands that urban systems are systems, and the values of many variables associated to the cities in such a system, including but by no means merely restricted to their population sizes, are tied to the values of corresponding variables to other cities in the system, through networks of political, cultural, economic and social interaction among the cities themselves and the entities of which they are composed. Moreover, these systems have important "hierarchical" aspects, in the stronger senses of the word described in the texts.

advantages of clarity[4] and abstraction, but it also weakens the richness and depth of the concept by reducing hierarchy to an ordered set.[5] In contrast to the other senses of hierarchy discussed below, order hierarchy does not even refer to relationships and interactions among the entities that comprise the hierarchy, much less give any role to hierarchy in conditioning entity relationships and interaction structures. As far as the relation between hierarchy and complexity is concerned, order is essential to hierarchy – but order alone is not what makes hierarchy important for complex systems, nor is hierarchy the source of the order in complex systems.

2. **Inclusion hierarchy**: Sometimes, hierarchy is used to refer to a recursive organisation of entities. For example, in the essays discussed in section 2 below, Herbert Simon defines hierarchy with his famous "Chinese boxes" image: "In application to the architecture of complex systems, 'hierarchy' simply means a set of Chinese boxes of a particular kind. A set of Chinese boxes usually consists of a box enclosing a second box, which, in turn, encloses a third – the recursion continuing as long as the patience of the craftsman holds out. The Chinese boxes called 'hierarchies' are a variant of that pattern. Opening any given box in a hierarchy discloses not just one new box within, but a whole small set of boxes; and opening any one of these component boxes discloses a new set in turn."[6] Of course, this notion is just a special case of an order hierarchy, where the ordering variable is the number of boxes one opens before arriving at the particular box of interest. Clearly, however, it is more than this: it makes an ontological claim. An entity is a container, and what it contains are other entities; this is the structure of reality (at least that part of reality contained in an inclusion hierarchy) "all the way down." The Chinese box image actually suggests more than this: not only is an entity a container, but *there is nothing else* in it but other entities. Note that an "urban hierarchy" as order hierarchy is just a set of cities ordered by, say, population size, and this is not an inclusion hierarchy; but if we think of an urban system as a collection of cities, which in turn contain, say, firms, households, institutions and organisations, which in

[4] At least if we set aside the puzzling words "natural" and "initially."

[5] Or "partially ordered," as Simon and others insist, since in most applications of order hierarchy, the ordering variable is often discretized, with ties permitted. In principle, "size" in an order hierarchy could follow any distribution; in the literature, a set is described as an "order hierarchy" only when the associated variable has a long-tail, particularly power-law, distribution. This is undoubtedly tied to the theme of our paper, the linking of hierarchy and complexity: the "soft" literature on complexity abounds now with references to such distributions as "the signature of complexity."

[6] "The organization of complex systems," pp. 4 and 5.

turn contain people, then an urban system so regarded is an inclusion hierarchy.

3. **Control hierarchy:** Probably the most commonly used sense of "hierarchy", particularly in reference to social organisations, has to do with who gives orders to whom. In this context, hierarchy refers to a control system in which every entity has an assigned rank, and all power is concentrated in the (usually single) entity with the highest rank. Entities with a specified rank are entitled to give orders to entities with lower ranks, while they must obey orders received from elements with a higher rank. Orders flow rank-downwards in a control hierarchy; information and requests rank-upwards. In this sense, a church, political party, or army can[7] be organized as a control hierarchy; but so can non-human systems, like the set of devices that comprise a building or factory automation system. Note that the entities that comprise a control hierarchy do not form an inclusion hierarchy, even though an army, for example, may be described in either way, depending on the units[8] that correspond to "entities."
4. **Level hierarchy:** Many authors use "hierarchy" to describe a particular kind of ontological organisation, in which entities are posited to exist at different "levels". Each level is characterized by a particular spatio-temporal scale for its associated entities and for the processes through which the entities at this level interact with one another: the higher the level, the more extended the associated spatio-temporal scale. Entities at a given level may, through their interactions, construct and maintain entities at higher levels, and higher-level entities may be, at least in part, composed of lower-level entities: these are often described by the term *upward causation*. Through upward causations, level hierarchies may form inclusion hierarchies. However, level hierarchies are also characterized by *downward causation*: incorporation into a higher-level entity can change the properties and interaction modalities of lower-level entities. Moreover, even if entities in a level hierarchy may be included in entities of a higher level, they need not be: entities at any level in a level hierarchy are usually regarded in some sense as "autonomous." Standard examples of level hierarchies, some of which are also inclusion hierarchies, are: physico-chemical (e.g. elementary particles – atoms – molecules); biological (e.g. cells – organs – individuals – species); economic (individuals – working groups or departments – firms –

[7] At least in principle. The social work necessary to maintain such a control hierarchy is immense, and it cannot be carried out by downward order and upward information flows alone.

[8] In this case, the soldiers, ordered by rank, form a control hierarchy, while the command units (platoon, company, regiment, corps etc.) comprise an inclusion hierarchy.

national economies); linguistic (e.g. letters – words – phrases – sentences – paragraphs – texts).

For the rest of this paper, we will be particularly interested in level hierarchies, which are those most implicated in discussions of complexity. However, we will find that hierarchy's polysemy is not accidental or superficial: we cannot escape issues of control, ordering and inclusion as we follow the discussion over the relation of level-hierarchical ontology and complexity in natural and social systems.

2. HERBERT SIMON: CHINESE BOXES, WATCHMAKERS AND LOOSE COUPLING

For Simon, hierarchy was *the* key concept for understanding complex systems. Simon presented his ideas on hierarchy and complexity in two influential papers, "The architecture of complexity" (1962, reprinted in the 1969 book *The Sciences of the Artificial*) and "The organization of complex systems" (1973). In these papers, Simon laid out a general ontological framework for complex systems based on hierarchy, and argued why both natural and man-made complex systems should conform to this framework. In this section, I describe Simon's framework, arguments and examples, followed by a discussion of some puzzling features of his ideas.

According to Simon, complex systems are composed of entities organized in inclusion hierarchies, "Chinese boxes." These entities interact, and the dynamics induced by their interactions are *nearly-decomposable*: that is, they are characterized by loose coupling both *vertically* and *horizontally*. Loose *vertical* coupling means that entities interact mainly with other entities at the *same* level of the inclusion hierarchy. Moreover, processes at different levels have quite different temporal scales. As a consequence, it is generally possible to study processes at each level in isolation: processes at higher levels are sufficiently slow that they may be regarded as "constants" with respect to focal level processes, while lower level processes are so fast in comparison that they "reach equilibrium" and so average out with respect to the time scale of the processes at the focal level. Loose *horizontal* coupling implies that entities at a particular level tend to cluster into weakly interacting sub-systems, whose detailed dynamics are nearly independent of one another. Indeed, the sub-systems interact with one another on an "input-output" basis, so their internal structure and dynamics can change, without inducing system-wide cascades of change, as long as they remain *functionally equivalent* – that is, able to produce the same outputs from the same set of inputs.

Simon offered a variety of examples of systems that he claimed conform to his hierarchical ontological framework. In the 1973 essay, he presents four "intertwined" examples, in which "most of the complex systems that occur in nature find their place." The first starts with "observable chemical substances", in which one finds "sets of component molecules," then atoms, then elementary particles. The second is biological: "living organisms to tissues and organs to cells, to macromolecules, to organic compounds, to a junction with the molecules of the first hierarchy." The third is meant to describe genetics, but the biggest "box" is hard (at least for me) to identify ontologically: "the statistics of inheritance;" then come "genes and chromosomes, to DNA, and all that."[9] Finally, the fourth refers to human social organisation, with a peculiar cognitive twist: it "leads from human societies to organizations, to small groups, to individual human beings, to cognitive programs in the central nervous system, to elementary information processes – where the junctions with the tissues and organs of neurobiology largely remains to be discovered." As with the third Simonian hierarchy, it is not clear to me what some of these fourth hierarchy "boxes", especially the last two, are supposed to represent; Simon attempts to identify them with analogues from the "artificial" world of digital computing: programs and machine-specific instructions. But do these correspond to "real" entities in human cognition? Of course, Simon's pioneering work in cognitive science was premised on the assumption of just such a correspondence; but a working assumption in a developing research program hardly constitutes a convincing claim for the structure of complex systems "that occur in nature."

Even if we assume that these hierarchies describe "real" entities, the fundamental question from the point of view of the argument that Simon tries to develop is whether or not they are organized as a nearly decomposable inclusion hierarchy. Simon argues that his first example satisfies "near decomposability" because the forces that govern interactions

[9] "Statistics of inheritance" is an aggregated property of "genes and chromosomes", not a separate ontological level, as the "Chinese box" metaphor suggests. Moreover, it is unclear what Simon wants to indicate by "statistics of inheritance", since population genetics theory depends on natural selection, which operates at the level of interactors and their phenotypes, not just on "genes and chromosomes." It would seem to be more consistent with the proposed ontology to label this level "biological individuals", with a higher level of "species" (or perhaps "populations"). But then the tight structure-function coupling that underlies the idea of near-decomposability would be lost, since it is would have been implausible, even around 1970, to suppose that the genotype-phenotype relationship was characterized by sparse connectivity and input-output modularity.

between entities at each level have energies and operative spatial and temporal scales that differ sharply from level to level. Rather surprisingly, given that his own research concentrated on his fourth hierarchy, Simon fails to present any argument that explains why the Chinese box metaphor applies to this hierarchy and why the hierarchy satisfies near decomposability. It seems to me that both these assumptions are highly questionable. The Chinese box metaphor suggests that each box at level n is included in only one box at level n+1. Clearly, this is not the case for the fourth hierarchy: individual human beings generally participate in many small groups and even organisations, while the members of even the most tightly-bound small groups often belong to different higher-level organisations (like firms, political polities, churches and so on). Through crosscutting networks of individuals, the small group level of organisation may be sparsely connected but is very unlikely to be loosely coupled horizontally, as the spate of recent research on the small-world property of most social networks shows. Moreover, the idea of functional equivalence – that is, classes of entities fungible with respect to inputs and outputs in their interactions with other entities at the same level – hardly corresponds to the complex network of relationships among entities at Simon's " organisations" level. As for vertical loose coupling, the "heterogeneous engineering" of social organisation, as we will see in the examples from Bruno Latour's work presented in section 4 below, brings into interaction entities from very different levels of Simon's social inclusion hierarchy – and *nonhuman* entities as well, from various levels of the other three Simonian hierarchies. To say the least, Simon's fourth example no longer seems to offer convincing phenomenological validation for the pervasiveness of his proposed "architecture of complexity."

Simon did not intend that his claim for the pervasiveness of hierarchy in the organisation of complex systems should rest just on the plausibility of his examples. He develops a theoretical argument that inclusion hierarchy is overwhelmingly the most probable form for a complex system that arises through an "evolutionary process". The probabilistic calculation is simple and straightforward, though its relation with the mechanisms of any particular "evolutionary process" are not at all obvious; rather, the force of the argument is conveyed rhetorically through the famous fable of the two watchmakers. According to this fable, the two watchmakers both make watches from 10,000 parts. One of them first builds 100 subassemblies, "stable arrangements of 100 parts," and then puts these subassemblies together to make a watch. The other has no such two-stage production procedure. Now comes the nub of the argument: both watchmakers are interrupted, on average in the time it takes to put together 150 elements ("primary" parts or subassemblies). Interruptions make any incomplete construction (not a watch or finished subassembly) fall apart. Calculation: the first watchmaker finishes a watch after about eleven interruptions, the second watchmaker almost never finishes one. Conclusion: "hierarchies will evolve much more rapidly from elementary constituents than will non-hierarchic systems containing the same number of elements. Hence, almost all the very large systems will have hierarchic organization. And this is what we do, in fact, observe in nature."

I have seen this argument cited many times over the years,[10] and I have read it in Simon's papers many times. I must confess, though, that I have never understood how the three sentences quoted at the end of the last paragraph have any relation to the point of the fable (and the more general probabilistic argument it is intended to illustrate). There are four principal difficulties with the argument. First, the conclusion talks about "evolving", while the successful watchmaker *designed* his subassembly scheme, knowing that his aim was to produce a watch and presumably having analyzed the component functionalities that make a watch work.[11] Second, the *innovation* in the *conclusion* is the generation of new kinds of higher-level entities from stable lower-level components, while the *innovation* in the *story* is the introduction of a new mid-level class of entities, with the low-level and highest-level entities (the elementary components and the

[10] Usually favorably, but not always. Agre (2004) shares some of the same perplexity I describe above and provides an historical context for Simon's reasoning in an effort to understand it. Sabel (2004) reads the fable as an argument in favor of the robustness of hierarchical organisation – and points out that the story only works because the watchmakers always make the same watch, in the same way; otherwise, the commitment to subassemblies might hamper rather than augment effective watchmaking.

[11] Designed systems and evolved systems have very different properties, despite attempts to various authors to adapt Darwinian evolutionary theory to processes of technological change. One of these differences has to do with the role of emergence: evolved systems "rely" on emergent structure to generate new system functionality, while designers generally work hard to suppress emergent features. It is interesting to note that recent large-scale distributed systems, like the Internet, which are hybrids between designed and evolved, are beginning to exhibit emergent structure. It is not clear whether the emergent structure is adding desirable or exploitable system functionality. Another difference between designed and evolved systems, which is relevant to Simon's arguments is the different means used to achieve functional robustness. Edelman and Gally (2000) argue that designed systems rely on redundancy, while biological (evolved) systems are characterized by degeneracy of the structure-function mapping (that is, biological structures typically have more than one functionality, and biological functionality is achieved by more than one structure, in very different ways – contradictoring Simon's assumption of functional equivalence, according to which different structures that deliver the same functionality have similar input-output descriptions).

watch respectively) fixed.[12] Thus, the relevant "speed of evolution" (in the sense of how "innovative" or evolvable is a system) in the fable is about how long it took the watchmaker to develop his subassembly system, not about how robustly that system performs in production, which is what is calculated. Third, even in its abstract version, the argument presumes that generating new kinds of higher-level entities from lower-level components is simply an exercise in combinatorics. The logic of combination, explored more deeply in John Holland's work, is a powerful transformation engine for entities at a particular hierarchical level, but it does not account for the transformations that construct new levels – at least the kind of levels that Simon describes in his example hierarchies. For example: a multicellular individual is not just a combinatory assemblage of cells! As Leo Buss argues in his splendid book *The Evolution of Individuality*, to "evolve" multicellular individuals from unicellular ones required the solution of an entire suite of control problems[13], in which variation and selection processes at the cellular level were either repressed or channelled into development mechanisms at the individual level. It is hard to see the relevance to such a process of a model that starts with a constant rate of combination of components at a given level. Fourth, the argument begs the question of why complexity, and hence (according to the argument) hierarchy, exist at all. This is because the form of the argument is: IF there are complex systems to be observed in nature, THEN they must have the (hierarchical) form of a recursive assemblage of lower-level components. But why should Simon's *deus ex machina*, his mysterious "evolution", produce complex assemblages at all? Is Simon slipping in teleology here, positing an "inexorable" and inherent tendency of evolution to "move" in the direction of complexity? Or is it an application of the anthropic principle that, given we – surely complex systems – are here to observe that nature has created complex systems, it is acceptable that this question be begged, and a tendency to aggregate into assemblages be assumed *a priori*, as the probabilistic argument in fact does, merely making some assumptions about the rate at which these assemblages form (and disaggregate, if a stable configuration hasn't yet formed)? Neither of these possibilities seems a satisfactory basis for a theory of hierarchy and

[12] It is worth noting that the evolution of hierarchical systems often proceeds through the emergence of a new "intermediate" level, just as in Simon's design example. For example: the level of the multicellular individual surely evolved BEFORE the level of specialized organs within these individuals. As pointed out in the text, Simon's combinatorial argument says nothing about why we should expect to encounter this kind of evolving hierarchy.

[13] Illustrating the point that the kinds of hierarchy are difficult to separate – even if many of the control mechanisms employed to repress or channel lower-level processes inimical to the higher-level entity are anything but examples of hierarchical control. Some, though, seem to be…

its relation to complexity, but I don't find any hints in Simon's work to an alternative.

If both the arguments and the phenomenological evidence favoring Simon's vision of the hierarchical organisation of complexity are substantially weaker than he intends them to be, as I have tried to suggest, what can we say about his contribution to linking the two concepts of hierarchy and complexity? Certainly, through his work in the 1960's and early 1970's, he was an important voice in raising the issue: encouraging the idea from general systems theory and cybernetics that "complex systems" formed a class with many common properties and mechanisms worth studying as a class, and throwing "hierarchy" into the stew of properties and mechanisms that characterized the class. But there were serious limitations to the way Simon thought about complexity that prevented him from seeing as deeply as, say, Anderson or Holland into the relation between hierarchy and complex systems. The Chinese box metaphor reveals some of these limitations, by what it *doesn't* incorporate about level hierarchy in complex systems:

- A box is a static structure: take it out of a larger box and it doesn't change at all. This is emphatically NOT the case with complex systems that are hierarchically organized: for example, the *in vivo* metabolic rate of cells scales (linearly in the log-log scale) with the mass of the animal "housing" the cell, while the metabolic rate of cells in culture is independent of the animal from which the cells originate![14] Thus, the phenomenon of downward causation is missing from the metaphor, but it is surely as important as upward causation in complex systems (and the interplay between them, giving rise to the feedback loops of reciprocal causation, is more important than either considered separately).
- The Chinese box metaphor is purely structural, while functionality (and, again, structure-function relationships) is fundamental to the role of hierarchical organisation in complex systems.
- Boxes at different levels are always boxes; they differ only in scale. This points in two misleading directions: first, it hints (and Simon maintains explicitly) that entities (and processes) at different levels necessarily differ with respect to their characteristic spatio(temporal) scales; second, that hierarchies in complex systems necessarily display self-similarity across scales. Both of these are valid for some complex systems, but I do

[14] I learned this "well-known" biological fact from Geoff West, who explains the scaling by an argument that couples hierarchy (in this case, defined in terms of a space-filling, optimized branching network) and complexity in a way not addressed in this paper.

not think that either is characteristic across the "class" of complex systems.[15]

- Chinese box hierarchies are never "tangled". Each box belongs to a well-defined level of the hierarchy (count the number of boxes you open to reach it – although this doesn't imply that every box at level n has the same properties as every other!); and each level n box is inside one and only one level n-1 box (and doesn't appear when you open any other box at any level). I suspect that such tangles are generic in complex system hierarchies, especially when levels higher than the human individual are present.

3. PHIL ANDERSON: MORE IS DIFFERENT

Phil Anderson's short 1972 *Science* article, "More is different: Broken symmetry and the hierarchical structure of science," has become one of the classics in the complexity literature. The article reflects the frustration of doing solid-state physics in the 1950's and 1960's, during the heyday of experimental and theoretical breakthroughs in the study of elementary particles. The elementary particle physicists were wont to describe their own work as "fundamental physics", and at least some of them liked to use the pun "squalid state physics" to describe what their underprivileged solid-state colleagues were doing. "More is different" is a manifesto in opposition to these attributions.[16] Its primary aim is to redefine what we mean by the "fundamental laws" of science. In order to do this, Anderson opened up the issue about hierarchy and complexity that Simon had begged: why and how do new ontological levels emerge?

The paper begins by asserting that no serious scientist denies that "reductionist hypothesis": that is, that "the same set of fundamental laws" hold throughout all domains of matter, living and non-living. The question, though, is whether this reductionist hypothesis implies the "constructionist hypothesis:" that it is possible to start from these laws and "reconstruct the universe." It is this hypothesis that Anderson wants to refute: "the more the

[15] At the very least, it is arguable in both cases – and since there are theorists who are ready to assume both as generally true, it would have be better to avoid a metaphor that incorporates these assumptions.

[16] The article certainly helped turn the situation around, as did the path-breaking research of Anderson and others on problems like superconductivity. Now Anderson's discipline has a more respectable name, condensed matter physics, and its role in providing the foundations for nanotechnology will probably confer it in the near future with something like the public reverence that the atomic bomb provided for elementary particle physics in the postwar decades.

elementary particle physicists tell us about the nature of the fundamental laws, the less relevance they seem to have to the rest of science, much less to those of society." The problem, says Anderson, is that the constructionist hypothesis "breaks down when confronted with the twin difficulties of scale and complexity."

It is of course important to understand just what Anderson means here by "scale" and "complexity." Scale seems to refer just to number of entities, the "more" of his title; complexity seems to refer to the *organisation* of these entities, which is what gives rise to the "different." To be more precise about what "organisation" means and the sense in which it introduces "difference," we need to follow his argument a bit further. He claims that at each new "level of complexity", new kinds of properties arise, and research into these properties is "as fundamental as any other." That is, the "fundamental laws" of elementary particle physics apply to all matter, but particular configurations of matter require additional laws that are not deductive consequences of the "fundamental laws". To explain what he means by a "level of complexity," Anderson refers to an ordered sequence – in his terms, a hierarchy – of what he calls "sciences:" "elementary particle physics, solid state or many-body physics, chemistry, molecular biology, cell biology,…, physiology, psychology, social sciences." Substituting the name of a "science" for the entities the science primarily studies, this list bears a startling resemblance to a concatenation of three of Simon's four hierarchies – with even the "…" in the same place that Simon placed it. Indeed, I think it is correct to say that Anderson isn't making an epistemological claim here: he means the list to represent an ontological level hierarchy. He argues that the "elementary entities" in each level of this hierarchy "obey the laws of" the science preceding it in the list. "But this hierarchy does not imply that Science [n + 1] is 'just applied [n].' At each stage, entirely new laws, concepts and generalizations are necessary… Psychology is not applied biology, nor is biology applied chemistry."

Before addressing the key issue for us, which is where these new laws – and the new entities to which they refer – come from, I think it is worth pointing out here that Anderson the physicist endows much more autonomy than Simon the social scientist to the entities that inhabit successively higher levels of their shared ontological hierarchy. Anderson claims that at each level new laws "are necessary" to account for the phenomena to which the new entities through their interactions give rise. Simon claimed much less: he believed that the "laws" for phenomena at higher levels of the hierarchy were merely a short-hand and efficient coding that approximated deductions from the "fundamental laws" at the lowest levels of the hierarchy; these approximations generally worked well enough because of the loose vertical coupling that, according to him, characterized the ontological hierarchy. Hence, the behavior of lower-level entities could effectively be regarded as irrelevant to higher-level phenomena, because they would have reached "equilibrium", while the state of higher-level entities could be regarded as constant in describing lower-level dynamics. So Simon believed in the "constructivist hypothesis" Anderson set out to refute; he just didn't consider it very practical, an understandable point of view for the inventor of bounded rationality.

So why is "more" "different"? This becomes the central question for Anderson's version of a hierarchy-based theory of complexity. He doesn't pretend to have an answer, but he thinks he may be well positioned to suggest some promising directions to explore, because in his field, the first step up from elementary particles in his hierarchy of the sciences, substantial progress has been made in understanding how the "shift from quantitative to qualitative differentiation takes place:" the theory of broken symmetry. A stationary system must exhibit the same symmetries as the laws of motion that govern it, but this result need not apply to processes that occur when the system is not in its stationary state. In such cases, symmetry may be broken – not "violated." The key distinction is that the internal structure of a piece of matter need not be symmetrical even if the total state is: a distinction that Anderson illustrates through the pyramidal structure of an ammonia molecule, important chemically because of the electric dipole moment it provides, which because of quantum mechanical tunnelling and the consequent inversion of the pyramid, doesn't violate the symmetric stationary state, an equal superposition of the pyramid and its inverse. Inversion is very fast for ammonia, very slow for big molecules; the inversion time gives space for symmetry breaking and hence new kinds of phenomena at a "higher" (in this case molecular or chemical) level, due to the number and organisation of lower-level entities involved. Many other examples illustrate the same kind of phenomena, in which a "structure" lacks the symmetries of a "stationary state." The theory of broken symmetry states that the "state" of a very big system need not have all the symmetries of the laws that govern its constituents, and in general has less. In many-body physics, such "states" can be studied by letting the size of the macroscopic system go to infinity, where they undergo phase transitions to states that violate microscopic symmetries. In real systems, "there is, of course, no question of the systems really violating, as opposed to breaking, the symmetry of space and time, but because its parts find it energetically more favourable to maintain certain fixed relationships with each other, the symmetry allows only the body as a whole to respond to external forces."

Because of this "rigidity", "the whole becomes not only more than but very different from the sum of its parts."

Something like this, Anderson says, is how it works at the level of the emergence of new kinds of entities at the macroscopic level of inert matter. Is there a more "complete" destruction of fundamental symmetries that may give rise to emergence of entities at even higher levels of organisation? Anderson speculates that this may be so, with the introduction of "information," in the form of heterogeneous lower-level entities varying from "cell" to "cell" in some spatially or temporally regular structure: like DNA, for example, or spoken or written language. On the other hand, perhaps there are no universal mechanisms as we ascend the complexity hierarchy; in any case, Anderson suspects that all along the way, the synthesis of the whole from its parts may be impossible, while analysis may be always fruitful – level by level, *not* all the way down the hierarchy.

In this paper, Anderson, like Simon, places hierarchy squarely in the center of the phenomenology of complex systems. To Simon, the *structure* of hierarchy explained how complex systems worked. To Anderson, hierarchy is a complex phenomenon to be explained, while the mechanisms of emergence provide the explanations. Through these mechanisms, quantitative change ("more entities") becomes qualitative, in the form of new organisations that present "rigidities", which in turn induce collective action on the part of their constituent entities. These organisations are then entities on which the mechanisms of emergence may act, when enough of them interact.

Anderson's professional expertise is of course at the lower end of his hierarchy of sciences, so it is probably not surprising that at the high end of his hierarchy we find only one category, with a plural name: "social sciences." It is interesting that these "sciences" do not follow the implicit rule behind Anderson's hierarchy of sciences: that is, that the level of the sciences increases with the ontological level of the entities it studies. Instead, the social sciences are organized functionally: economics (production and exchange), political science (power), sociology (social organisation), anthropology (culture), geography (place). Each of these sciences concerns entities that differ by several orders of magnitude in their size, complexity, spatiotemporal scales. Yet there does not seem to be any widely shared taxonomy that describes a hierarchy of levels of social organisation; certainly Simon's "human societies – organisations – small groups – individual human beings" seems a feeble attempt at carving society at its joints, compared to his physical hierarchy of "molecules – atoms – nuclei and electrons – elementary particles" or the biological hierarchy of "living organisms – tissues and organs – cells – macromolecules – organic compounds."

Why is this so? One possible explanation is that societies do not conform to a level hierarchical organisation, a possibility to which we return in section 5 below. Another is that social level hierarchies are much harder to detect and categorize than physical or biological ones, because they are not inclusion hierarchies. Or perhaps Anderson and Simon are right when they assert that hierarchy is a universal feature of the organisation of complex systems, and the difficulty in classifying the levels of social organisation is simply an artifact of our own particular location in nature's level hierarchy: it may be easier for us to look down from our own level, the biological individual, and find "our" components, than to look up at the levels of which each of us *are* merely components.[17]

4. JOHN HOLLAND: BUILDING BLOCKS AND PERSISTENT PATTERNS

In his stimulating 1998 book *Emergence: From Chaos to Order*, John Holland outlines a framework for thinking about emergence. Holland identifies hierarchical organisation as one of the four key "landmarks" to the "terrain of emergent phenomena."[18] In his treatment of hierarchy, Holland favourably cites Simon's work, especially the watchmaker fable, several times. However, his approach is very different from Simon's:

- Simon regarded hierarchy as the key to understanding "the organisation of complexity." In contrast to this structuralist perspective, for Holland the *process* of emergence is primary, and hierarchical organisation is regarded as a *consequence* of this process.
- Simon's approach, in particular his central image of the Chinese box, is oriented from the top-level down.[19] In contrast, Holland proceeds from the bottom up: his principal image is the building block.

[17] Ahl and Allen (1996) provide a lucid and stimulating introduction to an epistemological hierarchy theory, in which the observer's role in determining levels is fundamental. It ought to be possible to address hypotheses of this sort from the point of view advocated in their book.

[18] Along with "mechanisms (building blocks, generators, agents)", "perpetual novelty (very large numbers of generated configurations," and "regularities (persistent, recurring structures or patterns in generated configurations)" (Emergence, p. 9).

[19] You have to open the biggest box to see what is inside. Moreover, as we have seen, when Simon lists the levels in an inclusion hierarchy, he always starts with the highest level and moves down, in contrast to the practice of Anderson and of Holland.

- For Simon, level hierarchies are nearly always inclusion hierarchies. In contrast, Holland points to various kinds of tangles, especially in his discussion of conceptual default hierarchies.[20] In addition, he describes many higher-level "entities" as patterns of organisation rather than stable aggregates of lower-level entities.[21]
- In all of Simon's four basic hierarchies, described above, levels are absolute and well ordered. In contrast, except for a brief introductory invocation of a "hierarchy of sciences",[22] Holland eschews a nested sequence of well-defined levels, instead invoking "level" only as a comparative term (x is on a "higher" level than y, or a "new" level emerges).[23] Indeed, the concept of hierarchical organisation that emerges in *Emergence* is not even an order hierarchy.

Holland's treatment of hierarchy in *Emergence* may be interpreted as an attempt to implement the program I described in my discussion of Anderson's "More is Different": to discover common causal principles operating across many levels of the "hierarchy of sciences" (perhaps generalizing Anderson's broken symmetry), which act to construct that hierarchy through the emergence of new levels of organisation. Holland's approach is to provide a setting in which emergence may be defined and a methodology in which its causal principles may be identified. In the next paragraphs, I describe briefly some features of this setting and methodology, though I cannot hope to convey the richness and subtlety of Holland's extended treatment, nor match his lucidity of expression. The section

[20] Default hierarchies describe the organisation of conceptual categories, as well as "if-then" behavioral rules. Categories may be subcategories of more than one kind of higher-level concept: elephants are mammals, as well as large objects. Contradictory behavioral rules – one more general ("when a vehicle approaches the sidewalk where you are standing, flee"), the other specific ("when a bus approaches the sidewalk at your bus top, move towards it") – can support one another with advantage to the behavioral system. Default hiearchies represent a very different cognitive organisation than Simon's logically coherent decomposition of computer programs (and even his deterministic production systems). Holland discusses default hierarchies in detail in two of his other books, *Induction* and *Hidden Order*.

[21] For example: a glider in Conway's Game of Life, or New York City. Holland points out – with evident glee – that the human body turns over all its molecules in two years, which not only contradicts a strict entity inclusion relation between levels, but makes for a very porous Chinese box!

[22] *Emergence*, page 8. Holland's version of the hierarchy goes from nuclear physics to ecology – no human sciences appear on the list.

[23] See for example the discussion of the level concept in Chapter 10 of *Emergence*, in which he defines a "new level" in terms of "the possibility of combining mechanisms to make a more complex mechanism." This definition does not lead to a well-ordered set of levels.

concludes with a discussion of some loose ends in Holland's derivation of level hierarchical organisation from his concept of emergence.

Holland's theory of emergence is premised on a particular and original concept of science, which he develops in *Emergence* from a kind of origin myth. The myth starts with an ancestral capacity for metaphor. Like other thoughtful scientists, for example Hofstadter (whom Holland cites frequently) and Lakoff (whom he doesn't), Holland regards metaphor as the fundamental mode of thought of our species, underlying all our intellectual attainments from poetry to physics. For Holland, metaphor starts with a "source system, with an established aura of facts and regularities" (p. 210), which is coupled with a "target system" by means of the identification of some elements in the source with corresponding elements in the target – and then many aspects of the "aura" are transferred as well, leading to new ways of viewing and understanding what are now seen to be corresponding "facts and regularities," and relations among them, in the target.

Holland credits two pre-historical inventions for facilitating the move from our ancestral capacity for metaphor to (proto-) science. The first is *number*: abstracting away from a collection of objects all properties except how many of them there are. Once numbers were invented, our ancestors begin "to recognize their organizing power…from the counting of herds, to a basis for trade, to Pythagorean and Archimedean theories of the world that replaced myths, to current practice that puts number at the center of the human scientific endeavor." (p. 202). The second seminal invention is board games, which date back to the early Egyptian dynasties. Board games may have originated from a metaphoric transfer from human warfare to the more convivial and safer world of a board and stone or wood pieces. By experimenting with the rules to discover some that led to "games worth playing,"[24] humans attained their first complexity laboratories, where they could observe the generation of persistent patterns that could be interpreted to provide information about future outcomes of the game, like pawn formations in chess. Once this lesson was assimilated, it became possible to develop a metaphor in the other direction, from board games to our natural or social worlds: perhaps in these worlds too there are phenomena driven by simple rules, which were obscured by the complexity of the configurations they are able to generate. From this idea, according to Holland's origin myth, logic, mathematics and eventually science emerged.

[24] This term is due to Eric Leifer, who in *Actors as Observers* distinguishes "games worth playing" from those on which game theorists tend to concentrate, for which optimal or equilibrium strategies can be calculated *a priori*, with no need actually to play the game to find out what happens.

Holland concludes from his origin myth that science is essentially a kind of reductionism: the search for simple rules, by means of which it is possible to uncover previously hidden regularities in phenomena that present themselves as an ever-changing sequence of complex configurations. Note that this is very different from the sort of reductionism that claims that everything is derivable from a few fundamental laws that describe interactions among elementary entities: for any given phenomenon, it is essential to find the right level of detail and the right simple rules.[25] Unlike Simon's ontological commitment to "real" entities embedded via an inclusion hierarchy in other "real" entities, Holland maintains an ontologically agnostic position: "unlike the tiled automaton, where the possibility of reduction is guaranteed, the efforts of basic research go forward without such guarantees but encouraged by past successes" (p. 128). Indeed, Holland tends to avoid ontological terms like "entities," preferring process terms like "mechanism" or "generator."[26]

For Holland, *emergence* is the inverse of reductionism: persistent aggregate- or global-level regularities displayed by "interesting" systems generated by simple rules.[27] One consequence of this idea is the obliteration of what many regard as a fundamental distinction between deduction and induction. For Holland, Newton's laws and Maxwell's equations posit a set of simple rules, the consequences of which we are still discovering, just as we are still discovering new principles in chess. It is a secondary consideration that in the first case those consequences may be obtained deductively (that is, they can be derived "logically" or mathematically from the posited relations), while in the second case, the significance of the patterns that underlie the new principles have to be inferred from observation. Holland describes both new theorems in mechanics or electro-magnetism and new ways of evaluating pawn positions as emergent properties.

[25] Thus, Holland's concept of reductionism is consistent with Anderson's idea that fundamental laws arise at each new "level" in the scientific hierarchy. Holland's formulation seems to go beyond Anderson's, since it doesn't appear to rely upon a neat nesting of levels, although Holland does argue that rules for "higher-level" systems cannot contradict whatever rules the scientist (or the scientific community?) acknowledges as valid for "lower-level" systems.

[26] The exception is "building block," although as will be clear in a few paragraphs, I find his use of this term ambiguous.

[27] Here Holland is taking a position in strong contrast with many self-proclaimed followers of complexity theory, who regard reductionism as the enemy. For Holland, it is merely the other side of complexity theory's principal coin!

In Holland's conception of science, models and metaphors occupy the central role in science that more traditional accounts of the "scientific method" assign to cycles of data gathering, hypothesis generation and experimental tests. Mathematical models are the scientific analogue of the rules of a board game. As such, following the epistemology of the origin myth, modelling is the lynchpin of the scientific enterprise. The *scientific* value of models depends on a metaphor, in which the model is the source and some phenomenological domain is the target; the scientist transfers to that domain understanding, predictions and control strategies calculated via mathematical analysis or computer simulation from the model. Moreover, metaphors play a double role in Holland's account: models are often derived through metaphors in which a well-understood phenomenological domain serves as source and a less-understood domain of interest is target; a model that served to enhance understanding of the source domain is then transferred, perhaps with modification, to the target domain. Holland claims that Maxwell posited his equations in exactly this way, on the basis of a metaphor from fluid mechanics to electromagnetism.

With this background, it is easy to understand that Holland's main project in *Emergence* is to define a class of models in which simple rules generate both perpetual novelty and persistent patterns, the latter of which will count as emergent phenomena. He calls these models *constrained generating procedures* (*cpg*s). *Cpg*s perform the same two seemingly opposed but in fact fundamentally intertwined roles as board games, whose boards provide space to *generate* a huge number of possible configurations, which are interestingly *constrained* by the game's rules. Essentially, a *cpg* consists of a set of functions, called "mechanisms," each of which maps a set of input variables to an output state. The model is specified via a "wiring diagram" among mechanisms, whether the output state of one mechanism determines the value of a particular input variable to another mechanism. Once a dynamic for the "free" input variables and initial states for each mechanism are specified, the model generates a sequence of states, the vector of states of the component mechanisms. Holland shows how many rich dynamic models can be represented as cpgs, from seminal research projects by himself and by Art Samuels in the early 1950's on neural nets with cycles and checker-player programs, through cellular automata and billiard-ball "perfect gas" models. Like board games, some *cpg*s are "interesting" and some are not: the interesting ones are those characterized by *perpetual novelty*,[28] whose state vectors tend to remain inside relatively small subsets of their history spaces. These subsets correspond to emergent phenomena for *cpg*s.

[28] Or more accurately, very long cycle times.

Scientists can then use *cpg*s to study complex natural or social phenomena. The *art* of this science consists in choosing the right set of abstractions from the phenomenological domain of interest to posit as building blocks and interaction rules. Once these choices are made, the scientist then constructs a *cpg* whose mechanisms and wiring diagrams induce state transformations consistent with interactions among the selected building blocks. We can then call a persistent pattern defined in terms of building block interactions through time in the phenomenological domain "emergent", if we can link it a corresponding emergent phenomenon in the *cpg*. The "simple rules" described by the mechanisms and wiring diagrams in the *cpg* then provide an explanation of the observed persistent pattern in the phenomenological domain. This strategy provides a basis for a constructive approach to complexity science, although Holland warns us that we still have a long way to go, both in formulating interesting *cpg*s and in developing techniques (beyond inspired hunches) for identifying emergent phenomena to which they give rise, never mind providing proofs rather than merely simulated evidence for their existence and properties. The ultimate goal for Holland would be to derive sufficient conditions for emergence to occur in a *cpg*.

We now turn to Holland's argument for his claim that hierarchical organisation is one of the key "landmarks" of systems supporting emergence. The argument depends on his interpretation of two additional concepts, *macrolaw* and *level*, which he develops in Chapter 10 of *Emergence*. A macrolaw is a description of the behavior of an emergent phenomenon that makes no reference to the mechanisms and connection structure (the microlaws) of the *cpg* in which it arises. Holland's favorite simple example of a macrolaw is the glider formation in Conway's Game of Life.[29] One can recognize the shape and trajectory of a glider as it moves diagonally down the screen (in absence of interference from other formations), without any knowledge of the rules of the Game. The glider macrolaw turns out to be important, because it provides a means of communication between distant sites in the lattice of which the Game takes place, thus playing a key role in the derivation of the universal computational capacity of the Game. That the glider pattern is not robust to interference from other formations is, according to Holland, typical of macrolaws: they hold only under certain conditions on the state of the *cpg*,

[29] His favorite complex example are the laws that describe the properties of chemical bonds, which permit the development of chemistry without constant reference to the underlying quantum mechanics. Since quantum mechanical derivation of chemical transformations would be not only tedious but beyond our capabilities, this example shows the considerable advantage of macrolaw descriptions.

the violation of which forces a reversion to description in terms of the state itself.

For Holland, "level" is a dangerous idea: "if we read 'level' wrongly, we end up with notions of emergence that are so trivial as to destroy the usefulness of the concept ('the levels in organisation charts show that organisations are emergent'). Or, we can end up at the other extreme, treating emergence as something holistic that cannot be reduced to anything more basic ('consciousness is distinct from the activity of the central nervous system'). Neither extreme will help us in our quest." Precision is the way out. Holland defines "level" in terms of *cgp*s: a *cgp*, which as we recall consists of a set of interconnected mechanisms, is itself a mechanism. Thus, it can be combined with other mechanisms to generate "still more complicated" *cgp*s. "More generally, we can produce hierarchical definitions of *cgp*s, using *cgp*s defined early in the process as building blocks for later, more complicated *cgp*s. We gain, thereby, a precise notion of level."[30]

In the final chapter of *Emergence*, Holland summarizes in eight points some general conclusions about emergence derived from the *cpg*s investigated in his book, though of course he concedes that at present these findings are only suggestive. Four of these points provide the outline for his claim that hierarchical organisation is one of the key "landmarks" of systems supporting emergence:

[30] What we gain certainly is a criterion that allows us to conclude that mechanism x is at a lower level than mechanism y, if we can find a copy of mechanism x inside mechanism y. Whether the nesting property provides a "precise notion of level" is more arguable. For example, consider a single mechanism x, an "adder": x's state s is an integer, it takes another integer i as input, and it returns the integer s + i. Take two copies of x and bind the state of the first to the input of the second. Then we have a new, presumably level 2, mechanism, y, which has state (s_1,s_2), takes i as input and returns s_1+s_2+i. Now take another copy of x and a copy of y, and bind the state of the first to the input of the second. We have a new mechanism z. Presumably z is a level 3 mechanism, but its two components at two different levels, 1 and 2. Thus, we are no longer in Simon's world, where entities at each level interact primarily (or only?) with entities at the same level, with a characteristic spatio-temporal scale of interaction. Clearly, Holland's definition does not privilege interactions restricted to mechanisms at a particolar level. Worse, consider the mechanism p that wires together three x copies, binding the state of the first to the input of the second and the state of the second to the input of the third. Presumably, p is a level 2 mechanism, since all its inputs are level 1. But p and z are functionally identical: both have states in Z^3 (if we identify Z^2xZ with Z^3, as is usual), with a single free input, which they sum to the coordinates of their state vectors. From this example, it seems as though a level hierarchy deriving from this definition could qualify as an inclusion hierarchy, but not as an order hierarchy!

Point 3. "Emergent phenomena in generated systems are, typically, persistent patterns with changing components."

Point 6. "Persistent patterns often satisfy macrolaws."

Point 7. "Differential persistence is a typical consequence of the laws that generate emergent phenomena." In particular, different macrolaws have different typical time scales.

Point 8. "Higher-level generating procedures can result from enhanced persistence."

Points 3, 6 and 7, though all qualified by "typically" or "often", establish the claim that emergence may give rise to some relatively long-lasting macrolaws. Point 8 implies that some of the persistent patterns associated with long-lasting macrolaws may generate new composite building blocks. Since this last point is far from self-evident, Holland provides an example from biological evolution: the origin of the mammalian eye. The idea is that already existing elements, which were persistent for the most part because they were adapted to other functions, combined to provide a new functionality that itself proved adaptive and hence attained an "enhanced persistence." For example, pre-existing crystalline proteins could focus light energy on simple light-sensitive compounds, which could trigger neurons to fire. "What was extremely unlikely on inspection of the generating procedure based on the interaction of atoms to form molecules, becomes likely – almost inevitable – once we take into account the formation of a higher-level generating procedure" (p. 230). Thus we are back to Simon's "probability-of-evolution" argument, though in a considerably more evolutionarily sophisticated and hence plausible form.

Holland's treatment of emergence and its relation to level hierarchical organisation is intriguing, but I do not find it entirely convincing, for two reasons.[31] The first is essentially technological. With good reason, Holland switches frequently in the book between at least three levels[32] of rigor: *cgp*s; models that are not in *cgp* form – but could be, albeit at some cost for the detection difficulty of emergent phenomena; and verbal theories or descriptions of natural phenomena. *All* of the multi-level building blocks that Holland talks about occur in the third of these categories. In the second, the

[31] For that matter, I doubt that Holland does either. As he says, much remains to be done.

[32] I apologize for further abusing a word that has already too much work to do in this essay. But polysemy is often the key to latent meanings, and I believe this may be a fruitful example of this phenomenon.

best we get are gliders in the Game of Life and cell assemblies in Holland's neural net with cycles. These are both legitimate candidates for second-level building blocks, as Holland explains (verbally), but in the models themselves we don't see them building anything. Many researchers have tried to create models in which higher-level building blocks emerge, but I have yet to see any convincing success stories.[33] Taking Holland's constructive approach to complexity science at face value, this lack of success in generating higher-level generating procedures suggests that Holland's notion of emergence as persistent patterns in a *cgp* history space may not be nearly enough to explain the origin of level hierarchical organisation. Of course, we are in the early days of concentrated research on *cgp*-like models, so perhaps this technological objection will be overcome in the near future.

The second problem with Holland's argument is more serious, because it calls into question the adequacy of some of the concepts underlying the *cpg* framework. The problem is ontological. Look back at the first three summary points quoted above: they all refer to "persistent patterns." Now look at the fourth: here, we have jumped a level, and Holland refers to "generating procedures." The passage from "persistent pattern" – the form emergent phenomena take in Holland's theory and in his *cpg*s – to "building block" is just a matter of time. And yet in the "precise" definition of level, Holland is not talking about persistent patterns at all, rather about the inclusion of one mechanism as a component in another. I have no doubt that Holland is right when he says, in effect, that emergent persistent patterns are critical materials in the processes that assemble higher-level building blocks, but at least in the socioeconomic situations that I study (see section 6), "persistent patterns" do not simply transmute into "building blocks." Here, issues of control – that inevitable component of hierarchy, left lurking on the sidelines by all three of our complexity theorists in their discussions of hierarchy and complexity – occupy center stage. Holland's *cpg* framework may be useful for generating understanding of persistent patterns, but I don't think it (yet) has much to say about how they become transmuted into building blocks.[34] At the end of *Emergence*, we are not much farther along

[33] Walter Fontana's Artificial Chemistry model perhaps arrived to a second-level building block, and some recent work by Jim Crutchfield likely does that as well and may do even better. As Holland warns us, "level" can be a slippery notion, and he is surely right that a precise idea of what we mean by it would help.

[34] In fact, this observation is supported by the very examples Holland uses to illustrate his "precise definition" of levels in Chapter 10. To show how mechanisms can be incorporated into "higher-level" mechanisms, he tiles a cellular automaton, converting a 3x3 block of sites in the Game of Life into a *cpg* that can be treated as a site in a new cellular automaton. The new *cpg* is big enough to contain a glider, but it loses the capacity to track its gliding. In fact, the new *cpg* has too many states and inputs to be much use for

than we were with Simon and Anderson towards an understanding of the phenomenological observation they all share: complex systems tend to manifest a level hierarchical organisation. But we are also left with the unsettling feeling that their interpretations of what how that organisation might best be described, and why (and under what circumstances) it might be expected to emerge, are very different and far from precise.

5. DISSENTING VOICES: COLLECTIVES, CONSTITUTIONS AND THE CONJECTURAL PARADIGM

Many other complexity researchers besides Simon, Anderson and Holland have investigated the complicated relation between hierarchy and complexity, without having definitively resolved the ambiguities and difficulties that I highlighted in the previous three sections.[35] Indeed, some scholars believe that these problems may be irresolvable or even irrelevant. In this section, we see why two eminent scholars believe that the very complexity of society renders level hierarchy useless as a conceptual tool in reconstructing historical experience and analyzing social, economic or cultural change.

Warning (stylistic) to the readers of this section: Neither of scholars I discuss here identifies himself with the complex systems science community, as was the case for the three scientists discussed in the previous three sections. Indeed, the writings reviewed here have a completely different style than that which the complex systems science community usually admits as appropriate to acceptable scientific discourse. As a result, I change my own style of presentation in this section. My aim here is just to let our two authors raise issues about the place of hierarchy in understanding complex phenomena, not to criticize their work or to subject it to standards of rigor they themselves do not share. My own attempt to integrate what I draw from them into the complex system science discourse about the relations between hierarchy and complexity is incorporated into the proposals I present in Section 6.

---

anything – except as an example of a definition. In contrast, while Holland shows a glider can be represented by means of a cpg, the representation that he uses is incompatible with his "moving up a level" definition based on embedding mechanisms.

[35] In particular, see the works cited by Heylighen, Allen, Salthe and Lemke, and references therein.

Bruno Latour: collectives and constitutions I have enjoyed following Bruno Latour's intellectual journeys ever since I first read *Science in Action* around fifteen years ago, although the difficulty in translating what he writes into a language I can understand seems to increase with every new work. Fortunately, for the purposes of this paper I do not have to analyze Latour's thought in any detail; it will be sufficient just to understand why he totally rejects an ontology based upon level hierarchical organisation.

Let's start with the opening lines of *We Have Never Been Modern*:

> "on page four of my daily newspaper, I learn that the measurements taken above the Antarctic are not good this year: the hole in the ozone layer is growing ominously larger. Reading on, I turn from upper-atmosphere chemists to Chief Executive Officers of Atochem and Monsanto, companies that are modifying their assembly lines in order to replace the innocent chlorofluorocarbons, accused of crimes against the ecosphere. A few paragraphs later, I come across heads of state of major industrialized countries who are getting involved with chemistry, refrigerators, aerosols and inert gases. But at the end of the article, I discover that the meteorologists don't agree with the chemists; they're talking about cyclical fluctuations unrelated to human activity. So now the industrialists don't know what to do. The heads of state are also holding back. Should we wait? Is it already too late? Toward the bottom of the page, Third World countries and ecologists add their grain of salt and talk about international treaties, moratoriums, the rights of future generations, and the right to development."

The message is clear: try as we may to induce order on "nature" and "society"[36] by assigning each kind of entity that supposedly composes them into its place in a hierarchical ordering relative to all the others, as soon as we follow any process in which facts or values are contested, the networks of actors that have to be mobilized to settle the contest involve humans and nonhumans grappling with one another, forming alliances, employing all the resources of science, politics and discourse they can muster. As Latour summarizes the situation described in the previous paragraph,

> "The same article mixes together chemical reactions and political reactions. A single thread links the most esoteric sciences and the most sordid politics, the most distant sky and some factory in the Lyon suburbs, dangers on a global scale and the impending local elections or

[36] The quotation marks, as any Latour reader will readily appreciate, indicate that his major intellectual task over the past several years is to show that neither "nature" or "society" exists, except as conceptual weapons of domination.

> the next board meeting. The horizons, the stakes, the time frames, the actors – none of these is commensurable, yet there they are, caught up in the same story."

Can complexity theory help us to understand the processes through which his heterogeneous networks form? Not in Latour's opinion. The actors who create these networks do not

> "proceed thanks to a complex Science whose model and means would moreover entirely escape poor thinking, searching humanity… [They] *do not know* what does or does not constitute a system … what is connected to what… Neither cybernetics nor hierarchies make it possible to understand the chaotic, Darwinian, sometimes local and sometimes global, sometimes rapid and sometimes slow agents that [they] bring to light through a multitude of original experimental arrangements, all of which taken together fortunately do not constitute a secure Science."[37]

The process of assembling these networks, which Latour calls "collectives," is the heart of social change, and he is too committed now to the political challenge of forming a new "constitution" that will apportion appropriate "power and ability to speak, mandate, and will" to the things and people that compose collectives to spend much effort in providing detailed refutations of recent developments in the complex systems sciences. Still, I think we have already heard enough to understand why level hierarchical organisation could not possibly provide a foundation for understanding social change.

Carlo Ginzburg: the conjectural paradigm The relevance of Ginzburg's work to a criticism of level hierarchy is much less direct than Latour's, and we must approach it with more patience and subtlety. Its roots lie in an epistemological commitment that Ginzburg shares with other historians. Our problem will be to understand the ontological assumptions that justify this commitment. We begin by trying to describe what that commitment entails.

Many years ago I was a student in an American history seminar directed by Oscar Handlin. Handlin had the disconcerting habit of appearing to sleep during student presentations, occasionally rousing himself (usually with eyes still closed) to make an observation that, on sufficient reflection, undercut everything the speaker had been trying to maintain. On one such occasion, apropos of what I cannot recall, Handlin interrupted the speaker to declare "anything you want to know about how people make history, you can find out by digging deeply enough in Des Moines, Iowa, in the 1880's."

[37] Politics of Nature, pp. 21-2.

Handlin's remark reverberated in my mind for quite some time. I had the uncomfortable sensation that it was both significant and incomprehensible.

Years later, Handin's aphorism came to mind again when, in a discussion about the possibility of modelling innovation processes, I cited the familiar adage[38] "God is in the details," to which my collaborator Bob Maxfield – an electrical engineer by training and an entrepreneur by profession – replied that in his world, people always said "the devil is in the details." Again, the difference between these two variants struck me as significant but elusive: it was clear, though, that Oscar Handlin would have been on God's side. As for the devil's, I recall another conversation, this time with the physicist Geoff West, who was describing his experience in collaborating with biologists. "If Galileo had been a biologist," Geoff said, "he would have written a three volume treatise describing the differences in the trajectories and landing times observed when you dropped objects ranging from feathers to horses from the tower of Pisa." But of course Galileo was a physicist, and he realized that all these differences were incidental and epiphenomenal: the important point was that, stripping away all the irrelevant details and caveats, a falling object's velocity is proportional to how much time it has been falling.

Simon, Anderson and Holland are all on Galileo's side. For all of them, level hierarchy and its associated macrolaws make "more" (as in "more is different") a *relevant* detail, allowing the Galilean program to be extended to study complex phenomena, one level at a time, each with its own tools[39] that allow observers to hone in with the appropriate spatiotemporal resolution to measure the essential quantitative variables and filter out the myriad of *irrelevant* details. Indeed, could there be any alternative for understanding complex phenomena? Handlin's comment suggests that, at least for phenomena in which "people make history," he believes the there may be an alternative, somewhere deep down in Des Moines. In a brilliant and provocative essay, "Clues: Roots of an Evidential Paradigm,"[40] to which we now turn, the historian Carlo Ginzburg offers some hints about what such an alternative might be.

[38] Commonly attributed to Flaubert, Aby Warburg and Mies van der Rohe, but probably predating any of them.

[39] In Galileo's day, from microscope to telescopes.

[40] Originally published in Italian in 1979. The first English translation appeared in 1980; my account is based on the translation by John and Anne Tedeschi in the 1989 volume of essays by Ginzburg published by Johns Hopkins Press under the title *Clues, Myths, and the Historical Method.*

Like Holland, Ginzburg starts with an origin myth, even older than numbers or board games. He asks us to imagine an ancestral hunter, who "learned to reconstruct the shapes and movements of his invisible prey from tracks on the ground, broken branches, excrement, tufts of hair, entangled feathers, stagnating odors. He learned to sniff out, record, interpret, and classify such infinitesimal traces as trails of spittle" (p. 102). If we were to associate a trope with this capacity to generate coherent scenes extended in both space and time from seemingly insignificant details, it would be metonymy, not Holland's metaphor. The hunter's capacity derives neither from abstraction nor rules, but from immersion in a particular world of experience, for which every part gives evidence about the whole to which it belongs.[41] For millennia, "reading" signs or clues to interpret context has been a fundamental human skill, acknowledged and celebrated in folk tales as well as such arts as divination, astrology and medicine. Far from partitioning experience into distinct levels separated by spatiotemporal scale, metonymy makes meaning by linking the smallest of observable scales to the largest in a coherent unity, which cannot itself be directly observed.[42]

For Ginzburg, the invention of phonetic writing was a watershed in the history of the uses of signs for meaning making. The abstraction of the concept of phoneme and the relation between phoneme and sign, and the abstraction that *these* abstractions allowed in written expression (unthinkable in a purely oral culture), provided the impetus to an alternative, deductive path to knowledge, which the Greeks were quick to pursue. The older "conjectural paradigm" remained the mainstay of "physicians, historians, politicians, potters, carpenters, sailors, hunters, fishermen, women" (p. 105), but it was subordinated to the "prestigious (and socially higher) model of knowledge developed by Plato." In the next two millennia, this newer model continued to develop. Ginzburg locates a decisive turning point in this development, where what we now call "science" can be said to have well and truly begun, with the emergence of the Galilean paradigm of physics, founded in number, mathematical models and experiment.[43]

Not all disciplines dedicated to generating knowledge followed Galileo's lead. According to Ginzburg, those that did not were primarily interested in knowledge that referred to individual cases, "precisely *because they are*

[41] To use a modern image, the hunter's picture of the world is a hologram, not a photograph.

[42] As Ginzburg reminds us, "The Hippocratic school maintained that only by attentively observing and recording all symptoms in great detail could one develop precise 'histories' of individual diseases; disease, in itself, was out of reach" (p. 105). Anthony Grafton explores a similar idea in his stimulating exploration of the mental world of Girolamo Cardano, the brilliant 16th century physician, mathematician and astrologer.

[43] Or alternatively and more succinctly, following Holland, in abstraction and metaphor.

individual" [italics his], like natural history, artistic connoisseurship or clinical medicine.[44] But the spread of the Galilean paradigm brought such knowledge increasingly under suspicion, and scholars and disciplines felt under increasing pressure to sacrifice "knowledge of the individual element for generalizations (more or less scientific, more or less capable of being formulated in mathematical terms)." Since "the tendency to obliterate the individual traits of an object is directly proportional to the emotional distance of the observer" (p. 112), it is not surprising that the disciplines that concerned themselves with human beings were the slowest to follow Galileo. From the 17th century on, even here there were notable attempts to introduce the mathematical method, from the work of the political arithmeticians who attained rigor by reducing their field of study to the biological fundamentals, births and deaths; to the work of Bernoulli and later Laplace on the mathematical theory of probability as a foundation for understanding human decision-making at the individual and social levels; to the 19th century creation of statistics, as a set of methods to count and categorize populations at the level of the national state and as a rigorous approach to characterizing lower-level entities in terms of an "ideal" or "normal" representative, with a measurable range of allowable variation; to the creation of positivist social sciences, especially economics, whose practitioners proudly considered themselves to be the most Galilean of all.[45]

Of course, the ancient conjectural paradigm continued to endure, especially outside the disciplinary and professional milieu. It remained an invaluable component of daily life:

> "The ability to identify a defective horse by the condition of hocks, an impending storm by sudden changes in the wind, a hostile intention in a sudden change of expression, was certainly not to be learned from a farrier's manual or meteorological or psychological treatises. Knowledge of this sort…was richer than any written codification… These insights were bound by a subtle relationship: they had all originated in concrete experience. The force behind this knowledge resided in this concreteness, but so did its limitation – the inability to make use of the powerful and terrible weapon of abstraction." (pp. 114-5)

Not that the new scientists didn't try to extend their hold over even this domain of knowledge: the *Encyclopedie* is only the most ambitious and well known of a substantial body of work in the late 18th and early 19th centuries attempting to codify and appropriate this kind of knowledge.

44 Ginzburg rightly reminds us that "individual" may mean "a social group or an entire society" (p. 106) – that is, the issue is not ontological level but uniqueness.

45 At least after the marginalist revolution.

Then, around the middle of the 19th century, the conjectural paradigm suddenly gained a new respectability, raising it above the level of popular or folk knowledge to the status of an alternative scientific paradigm. The reason, as usual, was social necessity, in this case manifesting itself first through the development of new methods to identify individuals for the purposes of social control, followed on by the articulation of new methodologies in several "humane sciences" that specifically aimed to infer identity from seemingly meaningless clues. The key example of putting the conjectural paradigm to work for social control is fingerprinting, first used by the colonial administrator Sir William Herschel to distinguish among his Bengali subjects, and then put on a firm scientific footing by the great statistician Francis Galton. Ginzburg offers two particularly notable examples of "scientific" developments that infer identity from the seemingly irrelevant and meaningless:

- Freudian psychology, with its systematic deployment of such otherwise meaningless phenomena as dreams or slips of the tongue to lift the veil covering the "unconscious", the "real" locus of the identity of the individual; and
- Morelli's rejection of "schools" and "styles" as the key to authentication in art history, substituting the careful cataloguing and matching of "irrelevant" details, such as the shape of ears an artists draws when he portrays a face. Style, according to Morelli, is an abstraction, and anyone who masters its rules (and has sufficient manual dexterity and familiarity with technique) can produce a convincing copy of any style. We should seek instead "the most trivial details that have been influenced least by the mannerisms of the artist's school: earlobes, fingernails, shapes of fingers and of toes" (p. 97). Artists paint such details automatically, and copyists or forgers automatically produce *their own* versions of them, not that of the artist they are attempting to imitate.[46]

Morelli's work is perhaps less known today than Freud's, but it too generated a lot of attention in its day, both because Morelli used it to expose several works in major museums as inauthentic, and because his method undermined the art-historical basis of authentication, challenging the legitimacy of the knowledge claims of a lot of art historians. Not surprisingly, the experts dismissed Morelli's approach as crude positivism. Ginzburg claims that many art historians still use it, surreptitiously, today.

[46] Of course, as so often in strategic and self-referential situations, once Morelli pointed this out, ears could be stylized, and so the connoisseur who wants to use Morelli's method has to find some new "automatic" element to separate the true from the fake.

In the century and a half since the high culture surfacing of the conjectural paradigm in the work of Morelli and Freud, it has gained a solid foothold in the humane sciences:

> "Minute palaeographical details have been adopted as traits permitting the reconstruction of cultural exchanges and transformations… The depiction of flowing vestments in Florentine Quattrocento painters, the neologisms of Rabelais, the cure of scrofula patients by the kings of France and England, are only a few examples[47] of how slender clues have been adopted from time to time as indications of more general phenomena: the world view of a social class, a single writer, or an entire society".[48] (pp. 123-4)

Examples like these of applications of the conjectural paradigm can easily be found in the literatures of other humane sciences, including ethnography, archaeology, linguistics and other social scientific disciplines, as well as clinical medicine and paleontology.

What does all this have to do with hierarchy? For the Galilean paradigm, level hierarchy is essential: whether a scientist uses a linear accelerator, an electron microscope, or a telescope depends on the spatiotemporal scales of the processes he studies and the ontological level of the entities that compose these processes. In any case, whether he subscribes to near decomposability or credits with "more is different" emergent macrolaws, whichever level he chooses as focal, he may rest assured that his work can be productive and even "fundamental" without requiring him to spend his time peering into the logbooks, papers and instruments of scientists busy investigating phenomena involving entities on levels different from his. The situation of Ginzburg's humane scientists following the conjectural paradigm is different. They start from an ontological assumption of coherence, the source and nature of which

[47] The references are to classic historical studies by Traube, Warburg, Spitzer and Bloch.

[48] Another beautiful illustration of the method is Ginzburg's own study of the mental world of the miller Menocchio in his *The Cheese and the Worms*. In this work, Menocchio stands in metonymically for all the others in his era who, for the first time, had access to texts with which they could interact, without intermediaries imposing their own interpretations. With the help of the details Ginzburg assembles of what Menocchio read and what meanings he drew from these texts, we can recover a sense not only of the strange world Menocchio created, but of the strange world he inhabited, in which texts were for the first time "actors", not yet domesticated by the institutions that have since taken on the role of determining the order in which we encounter them and the ways in which we extract meaning from them. It is hard for me to imagine how the Galilean paradigm could address a question like "how can new communication technology affect patterns of cognition and social organization?"; while it seems to me Ginzburg's book gives an excellent example of a setting and a methodology with which metonymy and the conjectural paradigm can provide an answer.

are unfortunately unobservable, so that entities like "the world view of a social class, a single write, or an entire society" are inextricable entangled with their lower-level manifestations or origins, in the thoughts and actions of individual human beings. Where to look, and what instruments to use, to find clues that might reveal the assumed coherence is problematic. As Ginzburg puts it,

> "Though pretensions to systematic knowledge may appear more and more far-fetched, the idea of totality does not necessarily need to be abandoned. On the contrary, the existence of a deeply rooted relationship that explains superficial phenomena is confirmed the very moment it is stated that direct knowledge of such a connection is not possible.[49] Though reality may seem to be opaque, there are privileged zones – signs, clues – which allow us to penetrate it."

What gives us the capacity to interpret these clues is experience – our own immersion in a coherent world to which we have become exquisitely attuned, so that we have learned how to recognize scenes from traces. Indeed, if we immerse ourselves sufficiently in the details of another world – like Des Moines in the 1880's – we can come to recognize some of its hidden unity from the clues it has left behind. And underlying all such worlds is the coherence, the hidden unity, of people making history. It is because of this unity that metonymy is a feasible basis for understanding, and we can indeed learn everything it is possible to know about how people make history by digging deeply enough in Des Moines in the 1880's.

We can characterize and understand complexity, from this point of view, only after a thorough exploration of all the details, however irrelevant they might appear *a priori*, about all the entities at all the ontological levels of an individual complex system. As Ginzburg reminds us, though, "everything it is possible to know" is very far from everything.

Is all this talk of an ontological commitment to an underlying coherence and unity compatible with what most of us mean by science? Before I address this question, let me remind you that Simon's work is premised on some strong ontological commitments as well, in his case to a nearly-decomposable partition of entities segregated by spatiotemporal interaction scales; and Holland, although he personally maintained an ontologically agnostic position, agreed that his vision of science really makes sense only if underlying the complex phenomena we observe, there is some small set of

[49] I admit to finding to this sentence mysterious, but I quote it because it asserts the reality of these "deeply rooted relationships" that I think are the ontological core of the world Ginzburg is describing.

building blocks with simple generative interaction rules. So inherently unknowable ontological commitments are not a monopoly of the conjectural paradigm. Moreover, many social scientists – for example most cultural anthropologists, at least before the post-modern turn, many sociologists, and all economists who like to talk about invisible hands – share a form of the ontological commitment described in the previous paragraph, if not all the conclusions that Ginzburg and Handlin seem to draw from it.

Ginzburg himself concludes his essay with his answer to the "But is it science?" question:

> "The quantitative and antianthropocentric orientation of natural sciences from Galileo on forced an unpleasant dilemma on the human sciences: either assume a lax scientific system in order to attain noteworthy results, or assume a meticulous, scientific one to achieve results of scant significance… The question arises, however, whether exactness of this type is attainable or even desirable for forms of knowledge most linked to daily experience – or, more precisely, to all those situations in which the unique and indispensable nature of the data is decisive to the persons involved… In such situations, the flexible rigor (pardon the oxymoron) of the conjectural paradigm seems impossible to suppress. These are essentially mute forms of knowledge in the sense that their precepts do not lend themselves to being either formalized or spoken. No one learns to be a connoisseur or diagnostician by restricting himself to practicing only pre-existent rules." (pp. 124-125)

It is curious that both Holland and Ginzburg seem to agree that whether one thinks that level hierarchy is an essential ingredient of a complexity perspective to social phenomena depends on whether he believes that human and social cognitive systems are "really" rule-based. Holland does (or like the good Peircean that he is, is willing to act as though he does), Ginzburg doesn't.

6. A CONCLUDING QUESTION – AND SOME MODEST PROPOSALS

None of the authors discussed in this chapter, nor I, would dispute Phil Anderson's conclusion that "more is different," in the sense that interaction structures involving many entities of a particular kind may display structural and functional regularities that, without violating any "laws"[50] to which the interacting entities themselves are bound, can be described in a vocabulary that makes no reference to these entities or their laws. Furthermore, the new vocabulary may even identify some of these regularities as new entities and some macrolaws that these new entities obey. Thus, we can speak of "levels" of entities in a loose, comparative sense, to indicate for example that human individuals are a different level than firms or political parties or religions or states or cultures; or, less obviously, that firms are on a different level than national economies, even though many firms are multinational.

What is under discussion is the concept of "level hierarchy", its relation to "complexity," and its utility in helping us to understand complex social phenomena. As defined in the first section, in a level hierarchy we should expect to see a strict ordering on levels, with each entity associated with a particular level, interacting directly with entities at the same level, and indirectly, through processes of upward and downward causation, with the levels immediately above and below its own. Moreover, we would also expect that the interaction processes at each level would be characterized by particular space and time scales, the magnitudes of which increase as we ascend to higher levels. This corresponds with what Simon calls "the architecture of complex systems" and with Anderson's description of the hierarchy of the sciences. While Holland seems to put less weight on the idea that levels are strictly ordered or that each entity need belong to only one level, mainly because he focuses much more on processes than entities, he subscribes to the rest of the characteristics of level hierarchy and argues that they form a distinguishing landmark of systems that display emergence, the *sine qua non* of complexity.

But Latour and Ginzburg do not agree. The world we live in is not like this, Latour argues. Not only do the processes he studies refuse to keep "natural" and "social" actors separate, but he finds that the stories he wants to tell have casts of characters that vary in scale by many orders of magnitude, all interacting with each other, while the effects of their interactions might last for microseconds or millennia – or sometimes, both

[50] I have put quotation marks around "laws" as a reminder that not all of our authors would interpret this concept in the same way.

and everything in between. Ginzburg frames the issue differently: what we really want to know is about a totality that hides itself, but which can be glimpsed and interpreted on the basis of clues that might involve entities at any level, just as a Sherlock Holmes might use cigar ash (or now fragments of DNA) to solve a mystery involving, say, an attempt to steal a vital piece of information that, in the wrong hands, might initiate a world war. What kind of level hierarchy could introduce order into stories like these?

Why the difference between our two groups of scientists? One possibility is that Latour and Ginzburg expect something different from science than the others; in that case, level hierarchy may be essential for what the latter want to do, but not helpful for what Latour and Ginzburg want. But what do they want? Given their propensity to excavate the past by rummaging through archives or conducting repeated open-ended interviews with key actors,[51] let us take them as representatives of Ginzburg's "humane sciences" and suppose that they aspire to what Ginzburg calls "a scientific knowledge of the individual"[52] – recalling, of course, that the "individual" in question might be a person, a social class, an entire society – or, in Latour's process-oriented work, a particular instance of science or technology "in action". Note that this does *not* mean that their science is meant to yield insights only about the individuals whose detailed descriptions they recount; rather, whatever general understanding or "totality" they seek is to found only *through* the detailed exploration of some set of specific cases.

We need to consider what role theory might play in "a scientific knowledge of the individual". I think the answer must be that theory provides a *framework* for recounting the details that cohere into the "individual" in question. In particular, if the "individual" is an instance of a process, as in Latour's work, then the theory is about the "kind" of process, and the framework describes what we might call an ontology: the *kinds of entities* that may be instantiated in any instance, the possible *modes of interaction* among these entities and the transformations in entity properties that result from each of them, and a *dynamic* that specifies the ordering through time of entity interactions. Such a framework provides a minimal vocabulary for recounting particular instances, and the value of the theory depends on how causally convincing are the instances so recounted. If the "individual" is a structure, as for example the cognitive structures or belief systems that are the subjects of much of Ginzburg's work, the theory would specify the organisation (properties; parts and connections) of the relevant structural type (in particular, the relation between the observable parts and

[51] See Latour's Aramis, or the Love of Technology.
[52] "Clues," p. 112.

properties with those that are unobservable), and the value of the theory would depend on how convincing were the resulting reconstruction of an entire instance from the observed “clues.”

In this context, I can frame the question with which I would like to conclude this paper: for theories of this kind, about *which* sorts of complex social organisations or processes are we obliged – by the requirement to produce causally convincing instances – to postulate an ontology characterized by *which* kind of level hierarchical organisation?

Unfortunately, I can’t answer the question. I conclude with three modest proposals about what a full answer might reveal:

- **Tangled hierarchies** Latour is right: the entities that populate his collectives don’t stay stably spatiotemporally segregated, despite the intentions of some of the human actors to force them to do so. That makes it difficult to imagine *any* social processes whose ontology would feature strictly ordered levels, each populated by entities that interact only with others at the same level and characterized by distinct spatial and temporal scales that increase with increasing levels. The generic ontology would allow tangles, in the form of entities that might inhabit more than one level, be components of more than one “higher-level” entity, and engage in interactions with entities at a variety of different levels.
- **Sandwiched emergence** In general, in social processes new levels don’t emerge “bottom-up”, but intermediate between already existing levels. That is, there is a “whole” before its “parts” emerge, just as in biology temporary functional cell aggregations preceded the emergence of multi-cellular individuals, and multi-cellular individuals preceded the emergence of differentiated organs within the individual. In this sense, Ginzburg is right: some form of coherence or totality is ontologically prior to social level hierarchy.
- **Triadic hierarchies:** Holland is right. Persistent patterns arising from agent interaction are the key to emergence, which is the key to the formation of new levels of organisation. For social organisation, this leads to the idea of a triadic hierarchy, which I think is a fundamental ingredient of any social ontology. Start with a set of interacting agents, who constitute the micro-level of the triadic hierarchy. The meso-level is constituted from the history of interactions among sets of agents. If we represent interactions as binary, then the meso-level can be represented as a set of networks, with micro-level agents as nodes and histories of particular kinds of interactions as ties. The meso-level does more than

describe the past of the process: it provides the pathway whereby agents move into the future, since social agents interact with agents with whom they have already interacted, or whom they can reach via interactions with agents with whom they have already interacted.[53] Moreover, agents can endow agency upon recurring patterns of interaction, thereby inducing the emergence of new agents and even new intermediate levels of agents. But the meso-level is important for another reason: it is the locus of process functionality. Social agents are intentional: they act "for" something. Many recurring patterns of interaction among agents lead to attributions of functionality from some of the participating or observing agents: they "achieve" something, and if that something is valued by some of the agents, the pattern is reinforced. In this way, the meso-level carries the functionality of social processes. But interactions are transient, and patterns can be disrupted by any number of perturbations. The patterns of interaction that deliver valued functionality need to be nurtured, maintained and sometimes reconfigured if the functionality they deliver is to be sustained. Social organisations – sometimes agents, sometimes configurations of interactions among agents and artefacts – emerge to carry out this functionality-preserving functionality. Lane and Maxfield (2005) call such organisations scaffolding structures; they constitute the macro-level of the triadic hierarchy.

REFERENCES

Agre, P., 2003, Hierarchy and history in Simon's "Architecture of complexity," *Journal of Learning Sciences*. **12**: 413-426.

Ahl, V. and Allen, T., 1996, *Hierarchy Theory: A Vision, Vocabulary, and Epistemology* Columbia University Press, NY.

Allen, T. and Starr, T., 1982, *Hierarchy: Perspectives for Ecological Complexity*, Chicago University Press, Chicago.

Anderson, P., 1972, More is different: Broken symmetry and the hierarchical nature of science, *Science*. **177**: 393-396.

Batty, M., 2005, Hierarchies in cities and city systems, *this volume*.

Buss, L., 1989, *The Evolution of Individuality*, Princeton University Press, Princeton.

Crutchfield, J. and Gornerup O., 2004, Objects that make objects: The population dynamics of structural complexity, *Santa Fe Institute Working Paper*.

Edelman, G. and Gally J., 2001, Degeneracy and complexity in biological systems, *PNAS*. **98**: 13763-13768.

Fontana, W., 1991, Algorithmic chemistry, in: *Artificial Life II,* C. Langton *et al.*, eds., Addison Wesley, Redwood City, pp. 159-209.

[53] This is an instance of "action as re-enactment"; see the discussion of a narrative theory of action in Lane and Maxfield (2005).

Gabaix, X., 1999, Zipf's law for cities: An explanation, *Quarterly Journal of Economics*. **114**: 739-767.

Gibrat, R., 1931, *Les Inégalités Économiques*, Sirey, Paris.

Ginzburg, C., 1980, *The Cheese and the Worms: The Cosmos of a Sixteenth-Century Miller*, translated by J. and A. Tedeschi, Johns Hopkins Press, Baltimore.

Ginzburg, C., 1989, Clues: Roots of an evidential paradigm, in: *Clues, Myths, and the Historical Method*, translated by J. and A. Tedeschi, Johns Hopkins Press, Baltimore, pp. 96-125.

Grafton, A., 1999, *Cardano's Cosmos: The Worlds and Works of a Renaissance Astrologer* Harvard University Press, Cambridge.

Heylighen, F., 2000, Evolutionary transitions: How do levels of complexity emerge, *Complexity*. **6**: 53-57.

Holland, J., 1995, *Hidden Order: How Adaptation Builds Complexity*, Addison-Wesley, Reading.

Holland, J., 1998, *Emergence: From Chaos to Order* , Addison-Wesley, Reading.

Holland, J., Holyoak, K., Nisbet, R., and Thagard, P., 1986, *Induction: Processes of Inference, Learning, and Discovery*, MIT Press, Cambridge.

Lane, D., and Maxfield R., 2005, Ontological uncertainty and innovation, *Journal of Evolutionary Economics*. **15**: 3-50.

Latour, B., 1987, *Science in Action: How to Follow Scientists and Engineers Through Society* Harvard University Press, Cambridge.

Latour, B., 1991, *We Have Never Been Modern*, translated by C. Porter, Harvard University Press, Cambridge.

Latour, B., 1996, *Aramis or the Love of Technology*, translated by C. Porter, Harvard University Press, Cambridge.

Latour, B., 2004, *Politics of Nature: How to Bring the Sciences into Democracy*, translated by C. Porter, Harvard University Press, Cambridge.

Leifer, E., 1991, *Actors as Observers: A Theory of Skill in Social Relationships*, Garland, New York.

Lemke, J., 2000, Across the scales of time: Artifacts, activities, and meanings in ecosocial systems, *Mind, Culture, and Activity*. **7**: 273-290.

Sabel, C., 2005, Theory of a real-time revolution, forthcoming in *Organization Studies*.

Salthe, S., 1985, *Evolving Hierarchical Systems: Their Structure and Representations*, Columbia University Press, NY.

Simon, H., 1962, The architecture of complexity: Hierarchic systems, *Proceedings of the American Philosophical Society.* **106**: 467-482.

Simon, H., 1973, The organization of complex systems, in: *Hierarchy Theory: The Challenge of Complex Systems*, H. Pattee, ed., George Braziller, NY.

Chapter 5

HIERARCHY IN LEXICAL ORGANISATION OF NATURAL LANGUAGES

Bruno GAUME*, Fabienne VENANT**, Bernard VICTORRI***
** Researcher*
Laboratoire IRIT-CNRS, Université de Toulouse

*** PhD Student*
Laboratoire Lattice-CNRS, ENS, Paris

**** Senior Scientist*
Laboratoire Lattice-CNRS, ENS, Paris

Words have often been compared to living organisms. One of the first authors to develop this idea was the French linguist Arsène Darmesteter who wrote a book entitled *La vie des mots étudiée dans leurs significations* (Darmesteter, 1887). He described the evolution of the meanings of words as a 'struggle for life' (*concurrence vitale* in his own words). To stay alive, words need to occupy as much 'semantic ground' as possible, particularly by taking over 'territories' in new semantic domains, creating thus the important phenomenon of polysemy (see also Bréal, 1897). Some words enjoy real 'success stories', expanding their meanings in many different directions, in a rather monopolistic way, while others decline and eventually die.

Even though Darmesteter was more concerned by the analogy with biology rather than with social sciences, it is interesting to notice that his conception of interactions between words as a 'struggle for life' could also be applied to many social, political and economical human interactions. As a matter of fact, words are organized in the lexicon as a complex network of evolving semantic relations. It is not surprising that such a system shares many important properties with complex systems of social relations as well as complex systems of biological relations.

D. Pumain (ed.), Hierarchy in Natural and Social Sciences, 121-142.

In this paper, we focus on the comparison between lexical systems and social structures. As we shall see below, it has been discovered very recently that several graphs of semantic relationships between words belonged to the class of what is called ‘small world’ graphs, i.e. they first characterized graphs of social relationships. This result opens new perspectives in lexical semantics. It suggests that lexical graphs contain a rich amount of information concerning the semantic structure of the lexicon. In particular, we can expect that analyzing these graphs will enable a better understanding of its hierarchical organisation.

We present here a mathematical model in which each word is associated with a region in a global semantic space. In this representation, polysemy is taken into account by the size of the regions: words with many different meanings are represented by very large regions, while words with unique precise meaning are represented by very small (point-like) regions. If the regions associated with two words intersect, these two words share one or several meanings. Said differently, overlaps of regions in the semantic space correspond to (partial) synonymy between words.

Thus this model brings an interesting light to the similarity between graphs of lexical and social relationships. The semantic space plays the same role for words that the geographical space does for humans. Words meet in the semantic space like people meet in the world. Each meeting between two words means that there is a place in the semantic space, i.e. a precise meaning, that is common to both of them, exactly as a meeting between two persons shows that there is a place which belongs to the geographical fields of activity of both people.

As we know, fields of activity are not homogenously distributed on the map. There is a scaling structure from big cities to small villages corresponding to a scaling distribution of the density of fields of activity on the geographical map. Small world properties of many social networks are clearly related to the scaling structure of the underlying geographical space. Since lexical graphs are also small world graphs, we assume that the same holds for their underlying space. Meanings must have a scaling distribution in the semantic space, from places of high density (covered by many words) to the equivalent of villages, i.e. meanings that are poorly covered by the lexicon.

To test this hypothesis, we designed several methods to build a semantic space from a graph of synonymy. We present here these methods, illustrating them with the example of the French verb lexicon. As we will see, the different methods lead to rather similar results, showing that they reveal intrinsic properties of the semantic structure of the lexicon.

1. SMALL WORLD GRAPHS

Watts and Strogatz (1998) defined small world graphs as graphs combining two features: a high 'clustering coefficient' and a short 'characteristic path length'.

The clustering coefficient is a measure of how tightly the neighbors of a node in the graph are connected to each other. Numerically, it is defined as the proportion of pairs of nodes linked with one another among all the neighbors of a node[1]. In social terms, it measures how many of one's acquaintances know each other. So, it is not surprising that social networks have a high clustering coefficient (most of my friends are friends of each other).

The characteristic path length is a measure of how far two nodes are situated one from the other in the graph. The distance between two nodes is defined as the minimum number of edges traversed to get from one of them to the other. The characteristic path length is the average of the distance over all pairs of nodes. In the social context, a short characteristic path length means that there is generally a small number of go-betweens in the smallest chain which connects two people. This is the popular notion of "6-degrees of separation" (Guare, 1990) resulting from the famous experiments devised by Stanley Milgram who introduced the term of "small world" (Milgram, 1967).

A third property of small world graphs was put forward after Watts and Strogatz's work. It concerns the distribution of the number of edges among the nodes. It was discovered that the degree of a randomly selected node (the number of its neighbors) follows a power-law distribution[2]. The power-law was first verified on the Web network, which is also a small world graph (Barabási *et al.*, 2000, Huberman & Adamic, 1999, Kleinberg *et al.*, 1999), but it also holds for social networks (Newman, 2001, Barabási *et al.*, 2002). An important consequence is that small world graphs have a "scale-free" topology. Roughly speaking, it means that the ratio of very connected nodes to the number of nodes in the rest of the network remains constant as the network changes in size.

As shown by Ravasz & Barabási (2003), the two features, high clustering coefficient and scale-free topology, determine an original combination of modularity and hierarchical organisation. As the authors put it, "we should

[1] More precisely, it is computed as follows. Let p be a node, k its degree (number of its neighbors) and n the number of edges among them. The clustering coefficient at node p is $c(p) = 2n/k(k-1)$. It is easy to check that c(p) lies between 0 and 1. It equals 0 if there is no edge linking any pair of neighbors of p, and 1 if all neighbors are connected with one another. Then the clustering coefficient C of the graph is the average of c(p) over all nodes.

[2] The probability P(k) that a randomly selected node has k links follows the law $P(k) \sim k^{-\lambda}$ where λ is a constant for the given graph.

not think of modularity as the coexistence of relatively independent groups of nodes. Instead, we have many small clusters, which are densely interconnected. These combine to form larger, but less cohesive groups, which combine again to form even larger and even less interconnected clusters. This self-similar nesting of different groups or modules into each other forces a strict fine structure on real networks".

So, hierarchy appears as an emergent feature of the network. It is not a simple pyramidal organisation. No node can be viewed as dominating other nodes. The hierarchy is made of groups of nodes, with small clusters at the bottom and very large groups at the top. Moreover, groups of nodes may overlap at any level. A group (or a part of a group) of the lower level can be included in more than one group at the higher level, since it can belong to several different groupings having approximately the same clustering coefficient.

As far as social networks are concerned, such a hierarchical structure can, in many cases, be related with the underlying geographical space. For instance, acquaintance relationship is highly correlated with geographical proximity. So we can expect a duality relation between the hierarchical organisation of a graph of acquaintance and the hierarchical structure of the geographical distribution of humans. Each person (node of the graph) is associated with her spatial zone of activity, which may be a very large area for some individuals. Then, small clusters of strongly interconnected people correspond to relatively small areas where few people often meet, such as villages and districts in cities (notice that the same individual can belong to several different clusters, corresponding for instance to his home and his workplace). As we climb up the hierarchy on the graph by considering larger and larger groups (less and less interconnected), we obtain a smaller number of more densely occupied places. At the top level, the largest groups correspond to the centres of the largest cities.

Now if we consider other types of small world graphs, we can assume that there is always an underlying space with a dual hierarchical structure, even though most of the time the nature of this space is more abstract than a geographical map. This is the main idea that we will develop here to study the semantic structure of the lexicon. But before focusing on lexical graphs, we have to remark that the approach could be applied to any 'semantic' graph. For instance, let's consider, in the Internet universe, the small world graph whose nodes are all the web pages and whose edges indicate the presence of a hypertext link. Clearly the geographical factor is not relevant. But if we build an abstract semantic space whose dimensions are the different topics that a website may deal with, every website can be conceived as occupying a region of the space. Generalist sites will be represented by rather large areas, whereas more specialized ones will occupy smaller regions. We can expect that some places of the space will play the role of big cities in being densely covered by many sites, and others the role of countryside in being rarely broached on the web. Studying the hierarchical organisation of the semantic space and its evolution could bring interesting insights of what is going on on the web: what are the hottest topics, which ones are growing up and which ones are declining. Of course, the two most important problems with this approach is first to design the abstract semantic space (how to choose the relevant dimensions and the relevant metric on the space), and second to compute automatically the region associated with each website. The methods we present here provide the beginnings of a solution to both problems since they allow to derive the whole geometrical representation from computations on the initial graph, which is (relatively) easy to obtain.

2. LEXICAL GRAPHS

Lexical graphs have been a more and more important topic for the last few years, following the tremendous development of electronic linguistic resources (dictionaries and large corpora). The most famous example is WordNet, a very rich lexical database for English (cf. Fellbaum, 1998) comprising more than 150 000 words and many different relations between them. There are different types of lexical graphs, depending on the semantic relation used to build the graph. This relation can be a paradigmatic one such as synonymy, hyperonymy or translation (when more than one language is involved). It also can be a syntagmatic one, when, for instance, two words are linked if they appear in a same sentence in a given corpus. It can also be a more general semantic proximity relation, mixing syntagmatic and paradigmatic dimensions, as is the case when two words are linked if one appears in the definition of the other in a given general dictionary (cf. Gaume *et al.*, 2002).

The structures of many lexical graphs of all sorts have been studied (see, among others, Ferrer & Solé, 2001, Sigman & Cecchi, 2002, Ravasz & Barabási, 2003, Gaume, 2003, Gaume *et al.*, 2001). All the studies lead to the same conclusion. It seems that every lexical graphs have a small world structure, whatever the nature of the semantic relation involved. This result is important, since it shows that what is at stake is an intrinsic property of the semantic organisation of the lexicon in natural languages. It sustains the idea of an underlying semantic space whose hierarchical topological organisation could explain why different semantic relations share a same small world graph structure.

The graph we worked on for the present study, Synoverbe, is typical of these lexical graphs. It is a synonymy graph of French verbs which has been extracted from a general dictionary of French synonyms[3] by one of us (Bruno Gaume). Synoverbe has roughly 9000 nodes and 50,000 links. It has the three characteristic features of small world graphs. Its characteristic path length is small, around 4, which is the order of magnitude that can be expected from a random graph with the same number of nodes and links. Its clustering coefficient is very large, around 0.3, five hundred times higher than a random graph[4]. As shown on figure 1, the distribution of the degrees (number of links per node) follows a power-law distribution.

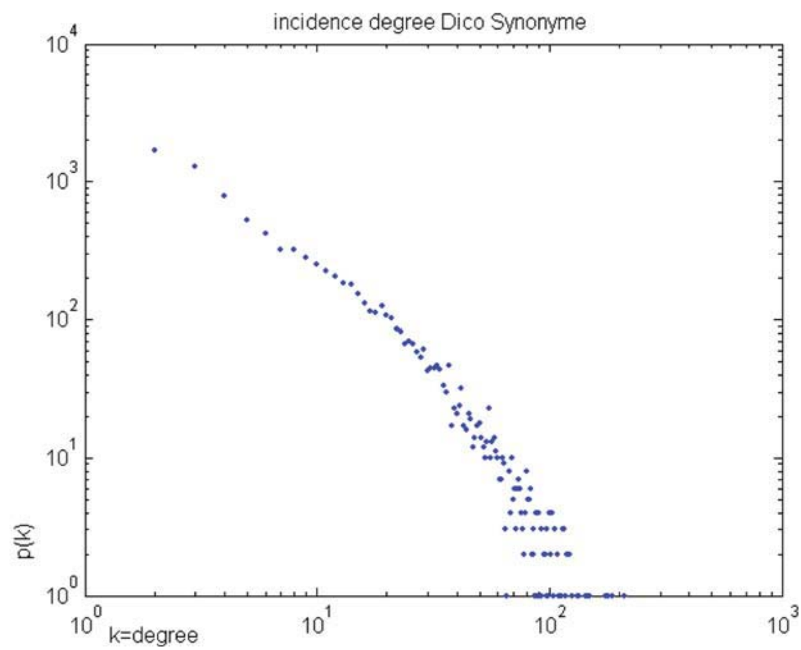


Figure 1: Synoverbe: log-log plot of the distribution of the degrees.

A more detailed description would give a more concrete idea of what this distribution means. While the average degree is less than 12 among the 9000 nodes, about 1000 nodes have more than 30 links and about 100 more than

[3] The general dictionary of French synonyms is managed by J.L. Manguin at the CRISCO research laboratory in linguistics, at the University of Caen. It is available on the Web (http://www.crisco.unicaen.fr/).

[4] For a random graph of n nodes and p links, the characteristic path length is $L = \log(n)/(\log(p)-\log(n))$ on average, and the clustering coefficient is $C = p/n^2$ on average. In our case ($n = 9000$, $p = 50{,}000$), the computation gives $L = 5.31$ and $C = 0.0006$. The precise figures for Synoverbe are $L = 4.17$ and $C = 0.318$.

80 links. Furthermore, nearly 90% of the nodes are directly linked to at least one of the 1000 most connected nodes, and nearly 50% are directly linked to at least one of the 100 most connected ones. In other words, among the nearly 10,000 French verbs, we can extract a subset of 1000 verbs which covers virtually all the meanings covered by the entire set, since nearly all the French verbs are synonyms of verbs of the subset. Moreover, a subset of only a hundred verbs covers half of the verb meanings. These verbs are of course the most highly polysemic ones, since each of them has a hundred or so synonyms. The most connected ones, like *faire* (translations: make, do …) and *prendre* (translations: take, get …) have even more than 200 synonyms. They have two other interesting properties: (1) they are the most frequently used by French speakers, and (2) they are the first to be acquired by children. No doubt that they are the winners in the 'struggle of life' evoked by Darmesteter! It is also worth noticing that they are rather tightly interconnected. In fact, the subgraph composed by these 100 verbs has basically the same properties as the whole graph: an average of 6 links by node, a characteristic path length of the same order of magnitude as the one of a random graph of the same size and connectivity, and a clustering coefficient markedly higher than the one of a random graph.

Thus we can then describe the structure of the French verb lexicon as a hierarchical structure with three levels:

- at the top, a first subset of 100 verbs, each with several general meanings. It represents the basic vocabulary for the verb semantic domain. With these 100 verbs, one can express most actions and events, but vaguely and without accuracy.
- at the second level, a subset of 1000 verbs presenting a rather important degree of polysemy (more than 30 synonyms each). It covers all verb semantics, quite sufficient to describe any action or event in everyday life. In fact, the verb lexicon used by most people in production is only a part of this subset.
- at the third level, the entire set of nearly 10,000 verbs, permitting very precise descriptions, subtle uses of qualifications, and different styles and levels of language (formal, technical, poetic, slang, etc.).

Even though it gives a first idea of the structure of the French verb lexicon, the above description is neither accurate nor satisfying enough. The problem comes from the arbitrary nature of the choice of our levels. Why three levels, rather than four or five? As a matter of fact, the hierarchy is not discrete, with well identified intrinsic levels: it is a continuous scaling. Therefore, we need mathematical tools suited to continuous representations in order to model the lexical hierarchy in a more appropriate way. Here is the main reason why we turned to geometrical tools and quantitative measures such as the notion of proxemy, that we introduce in the following section.

3. PROXEMY: A MEASURE OF SEMANTIC NEARNESS

Bruno Gaume defined a new measure of the nearness of nodes in a graph that takes into account the density of the graph along the different paths linking them. This measure, that he called *proxemy* (Gaume, 2002, 2003, 2004), is well suited for small world graphs because it relies on the structural properties of the graph. We know that somehow two nodes are never far from one another in a small world graph, since the characteristic path length is small. But for the same minimal path length, two nodes may be very loosely related by only one path linking two separate dense regions clearly apart, or they may belong to a same dense region with many different paths of minimal length connecting them. Obviously, the nodes must be qualified as "closer" in the latter case than in the former one: this is exactly what the measure of proxemy does.

A good idea of the notion of proxemy can be given by considering a particle wandering randomly on the graph, going from one node to any of its neighbors with equal probability. Let the particle be at node A at the beginning of the process. After the first time step, the only nodes that can be reached by the particle are the direct neighbors of A[5], each with a probability of 1/n, where n is the degree of A. After k time steps, any node B located at a distance of k links or less can be reached, the probability of this event depending on the number of paths between A and B, and the structure of the graph around the intermediary nodes along those paths. The more interconnections between these nodes, the higher the probability of reaching B from A will be. In other words the probability for a random particle to go from A to B is a good candidate for the measure we were looking for: we call it the k-proxemy of B with respect to A.

More generally, we define the *k-proxemy of a node with respect to a given subset of nodes* as the probability for a particle to reach it after k time steps if the particle were at time 0 on one of the nodes of the subset (if the subset contains p nodes, each of them is endowed with a probability of 1/p to be the starting point of the particle)[6]. When the subset includes all the nodes of the graph, we will speak of *global k-proxemy*.

[5] Including A itself: for technical reasons (property of ergodicity, see note 7 below), it is preferable to consider that the graph is reflexive, i.e. that each node is its own neighbor.

[6] From a mathematical point of view, it is easy to show that the random process we described is a markovian process. If we call $A = (a_{ij})$ the matrix of adjacency of the graph ($a_{ij} = 1$ if nodes i and j are connected, else $a_{ij} = 0$), the markovian matrix $M = (m_{ij})$ associated with the random walk of the particle is given by $m_{ij} = a_{ij}/s_i$ where s_i is the sum of the row i of the matrix A. The k-proxemy of a node n with respect to a subset S can be computed as follows. Let $U=(u_i)$ be the vector associated with the uniform probability density over S ($u_i=1/s$ if the node i belongs to S, else $u_i=0$, s being the number of elements of S). Then the

It must be emphasized that the value of k plays a crucial role in the definition of proxemy. For very small k, the k-proxemy fails to catch the structural properties of the graph because it is too local: the proxemy of most nodes of the graph is zero with respect to any given node. On the other hand, if k is too large, the k-proxemy of a given node with respect to any subset does not depend on the subset any longer: it tends towards a value that only depends on the degree of the given node[7]. Thus interesting values of k lie between the two extremes. Empirically, it seems that the best results are obtained with values belonging to the interval (L, 2L) where L is the characteristic path length of the graph. For instance, in the case of Synoverbe (L = 4.17), the value k = 6 proved to be the best one. From now on, we drop the “k” prefix in the term k-proxemy, assuming a choice of k in the right interval (and a value of 6 for the examples from Synoverbe).

Using proxemy, a geometrical representation of the graph can be built, which preserves its structural properties (Gaume, 2004). To each node A of the graph is associated its *proxemic representation*, a vector whose n^{th} component is the proxemy of the n^{th} node of the graph with respect to the node A. In other words, the proxemic representation of a node gives the probability distribution over the whole graph for the random walk of a particle originating from this node[8]. This means that the proxemic representation takes into account the relations of a node with all the others: it is characteristic of the structural position of the node in the whole graph.

When dealing with a lexical graph, the proxemic representation could be qualified as ‘Saussurian’, since it fits exactly Saussure’s structuralist theory according to which the semantic value of a lexical unit cannot be defined in absolute terms, but only by its relative position in the entire system. As a matter of fact, Karine Duvignau and Bruno Gaume have shown that proxemy is also relevant for psycholinguistic considerations, in particular in studying lexical acquisition and children production (Duvignau, 2002, Duvignau & Gaume, 2003, 2004) as well as in modeling disambiguation processing.

---

k-proxemy of n with respect to S is given by the n^{th} component of the vector V obtained by applying k times the transformation M to the vector U (in matrix notation: $V = U.M^k$ where U and V are row vectors).

[7] When the graph is reflexive, it can be shown (Gaume 2004) that the markovian process is ergodic. Then, a corollary of the theorem of Perron Froebenius implies that there is a unique stationary probability and that the process converges towards this stationary probability for any initial conditions (see for instance Semata 1981 and Bermann & Plemons 1994). In our case, it is easy to verify that the stationary probability is the vector $W = (d_i/2p)$ where d_i where d_i is the degree of the node i and p is the number of links in the whole graph.

[8] Computationally speaking, the proxemic representation of the node i is the i^{th} row of the matrix M^k, where M is the Markovian matrix defined above (see note 6).

Here we will focus on the use of proxemy for visualizing the hierarchical organisation of the lexicon. To begin with, we must notice that the global proxemy of a node (its proxemy with respect to the entire set of nodes) is a better indicator of its semantic extent than its degree. Whereas the degree of a node only indicates the number of its synonyms, its global proxemy gives more precise information about the more or less central role played by the node in the whole graph. Moreover, the nodes can be located in the same geometrical space thanks to their proxemic representation. Of course, the geometrical space cannot be faithfully visualized because of its high dimensionality, but the use of a classical method of dimension reduction (principal component analysis) allows to obtain a two or three dimensional representation preserving the main geometrical relations between the nodes we choose to visualize. We show on figure 2 the proxemic representation of the 200 highest-ranked French verbs according to their global proxemy (computed from Synoverbe)[9]. Each verb is represented by a sphere. As can be seen on the figure, the most general French verbs (the top of the hierarchy) are organized along four semantic axes structuring the whole lexicon. As can be seen on the figure, the most general French verbs (the top of the hierarchy) are organized along four semantic axes structuring the whole lexicon as a sort of conceptual tetrahedron. Around the first vertex (labeled A on the figure) can be found verbs expressing escaping and rejecting actions (*partir*, *fuir*, *disparaître*, *abandonner*, *sortir*… Interestingly, *quitter* is located between *disparaître* and *abandonner*). The zone around vertex B is composed by verbs expressing productive and enhancing actions like *exciter, enflammer, exalter, animer, soulever, transporter, soulever, provoquer, agiter, augmenter* (and *entraîner* between *attirer* and *provoquer*). The third vertex C is characterized by the ideas of connecting and communicating (*assembler, joindre, accorder, fixer, établir, indiquer, montrer, exposer, marquer, dire, composer…*, *réunir* between *attacher* and *joindre,* and *révéler* between *montrer* and *indiquer*). At last, vertex D corresponds to destructive and damaging actions such as *briser, détruire, anéantir, abattre, affaiblir, ruiner, épuiser, écraser, casser, dégrader*… The verb *tuer* is located there, between *altérer*, *dégrader* and *supprimer*.

It must be noticed that we can observe gradual semantic changes as we move from one vertex to another. For instance, moving from A to B we find successively *s'enfuir, fuir, partir, sortir, passer, courir, venir, marcher, aller, suivre, avancer, revenir, introduire, faire.* From A to D the gradation involves *s'enfuir, fuir, disparaître, quitter, abandonner, mourir, cesser, perdre, diminuer, supprimer, casser, anéantir, détruire.* Between B and D

[9] Proxemic representations of different lexical graphs are available on the Web: http://dilan.irit.fr/.

can be found the series *exciter, enflammer, agiter, tourmenter, troubler, ennuyer, bouleverser, fatiguer, ruiner, détruire, anéantir, briser*, whereas one passes from B to C through e*xciter, exalter, animer, soulever, provoquer, entraîner, augmenter, élever, conduire, déterminer, produire, former, dire, établir, exposer, indiquer, montrer, révéler*. Last example, here is the series from C to D: *fixer, assembler, joindre, réunir, arranger, attacher, retenir, serrer, fermer, arrêter, cesser, rompre, séparer, couper, étouffer, supprimer, diminuer, casser, affaiblir, abattre, anéantir, briser*.

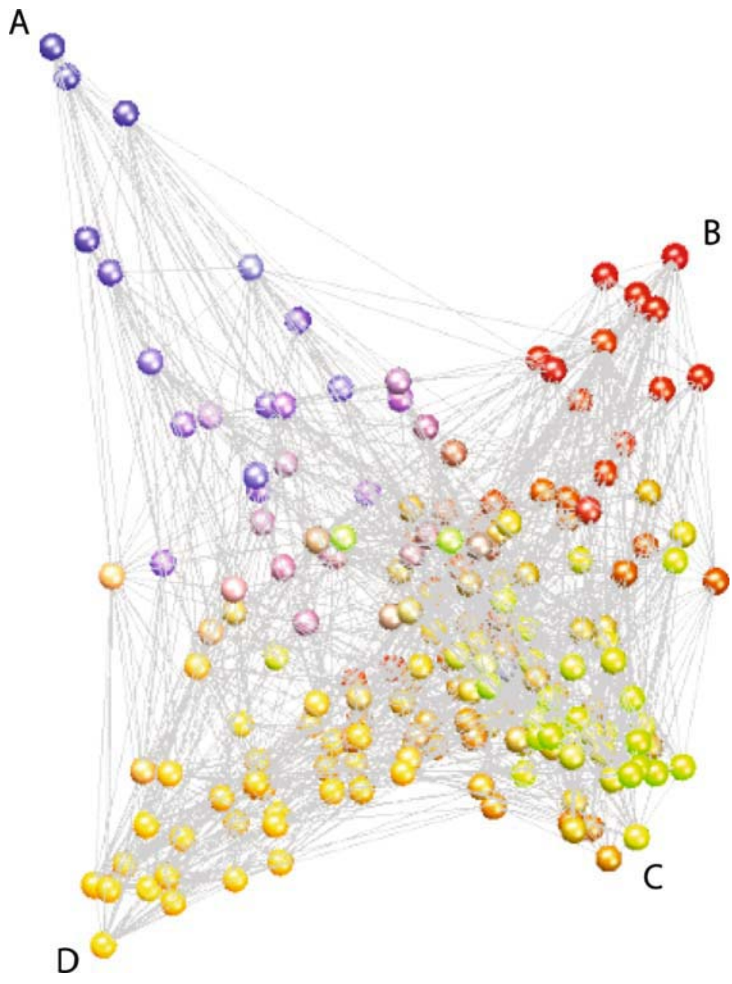


Figure 2: Representation of the first 200 French verbs with highest global proxemy.

Three important comments are worth emphazising:

- First, as shown by these examples, the geometrical distance between spheres presents a very good correlation with the semantic distance between lexical units: close verbs on the figure are also close by their meanings. This proves that proxemic representation actually catches semantic properties of the lexical units.
- Second, the global structure, sort of tetrahedron with its four vertices, is relatively independent of the precise number of top-ranked verbs used to build it: a very similar form is obtained with the first 100 or 300 verbs instead of the first 200. This means that the method is well suited to the continuous aspect of the hierarchical structure of the lexicon. Thanks to our geometrical representation, we do not need to define any 'levels' of hierarchy. The choice of the number of verbs taken into account is not a crucial decision, but a question of convenience: taking more verbs leads to a more accurate representation, but at the same time a less readable figure.
- The third remark is also a consequence of the continuous aspect of the geometrical tools. Once the representation has been built with a small number of top-ranked node, we can represent any of the remaining nodes in the same figure. In other words, the geometrical representation is a global referential frame in which we can locate all the nodes of the graph. As regards our example of Synoverbe, it follows that any French verb can be characterized by its location in the tetrahedron. For instance, if we add the verb *accabler*, which is not among the 200 top-ranked nodes, to the representation, we find that it is located in the D region, between *écraser*, *fatiguer* and *bouleverser*, as could be expected from its meaning. Far from being restricted to the 200 verbs used to build them, the four vertices correspond to four semantic dimensions whose relevance is general all over the French verb lexicon[10].

We can also use the proxemic representation of the nodes to visualize more local parts of the graph. Instead of using global proxemy to choose the nodes to be represented, we can choose to study any subset of nodes of particular interest by representing the verbs having the highest proxemy with respect to the selected subset. We will call such a representation a *proxemic zoom onto* the given subset. Actually, since all the nodes can theoretically be represented in the same high dimensional space, we can consider that we really zoom into a part of this representation when we select some nodes to visualize the relative positions of these nodes in the high dimensional space. Of course, we practically need to use principal component analysis to reduce

[10] An important question is whether these dimensions are *universal*, i.e. shared by all human languages. This is one of the issues that we intend to explore in the near future.

the dimensionality of the space, exactly as we proceeded when we visualized the global structure.

Figures 3 to 5 show such proxemic zooms. In each case, we have chosen a couple of antonyms as subsets defining the proxemy: {*monter*, *descendre*}, {*commencer*, *finir*} and {*aimer*, *haïr*} (respectively go up/go down, begin/end and love/hate). Each time, one of the two verbs is on the 'positive' axis and the other on the 'negative' one. It is interesting to see how the antonyms are connected by relatively short paths through their semantic domain, with semantically very relevant intermediary verbs: *sauter* (to jump) between *monter* and *descendre*, *partir* (to depart) between *commencer* and *finir*, *envier* (to envy) between *aimer* et *haïr*.

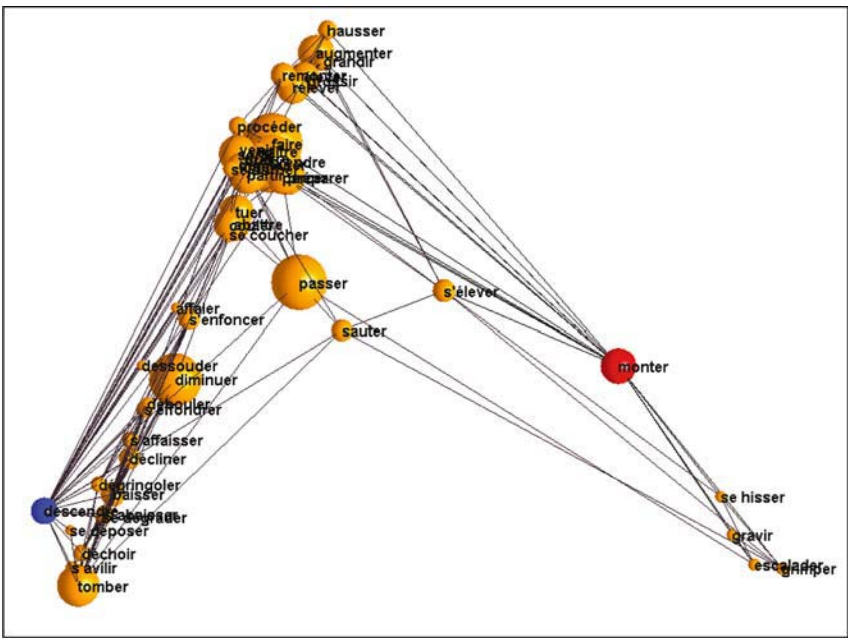


Figure 3: Proxemic zoom onto {*monter, descendre*}.

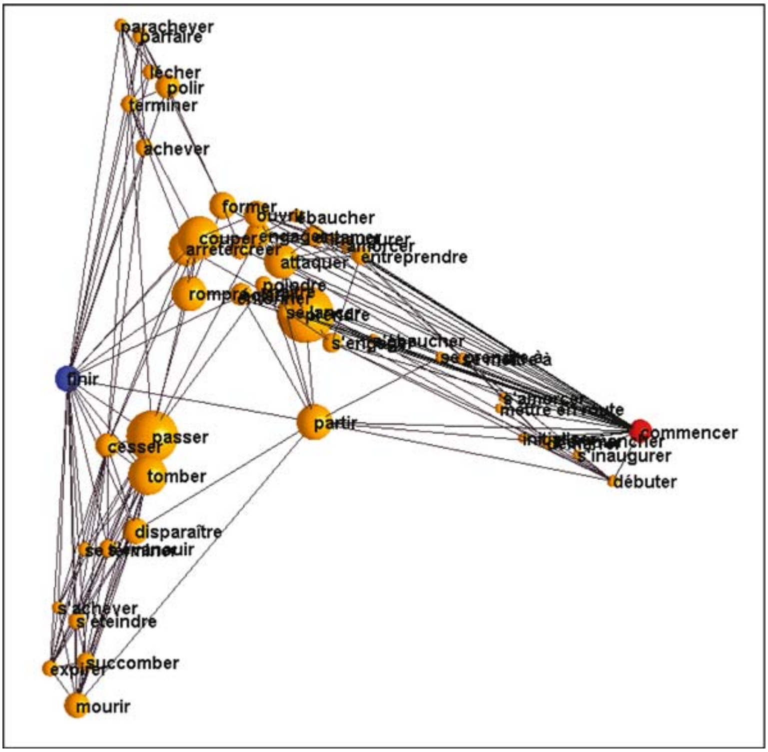


Figure 4: Proxemic zoom onto {*commencer, finir*}.

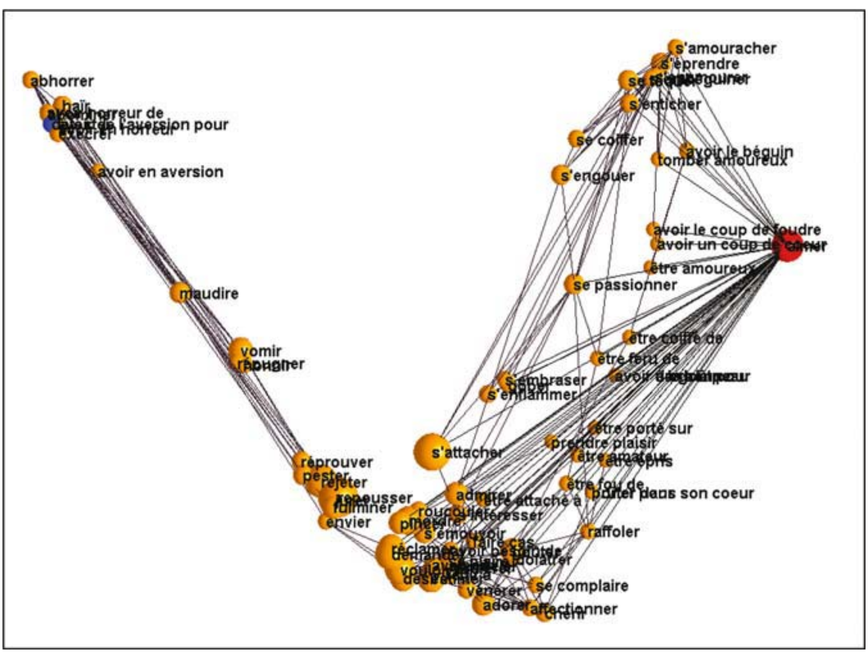


Figure 5: Proxemic zoom onto {*aimer, haïr*}.

Proxemic zooms can also be used to study the polysemic organisation of a single verb. We just have to zoom with respect to the subset reduced to this verb. For instance, figure 6 shows the representation around the verb *jouer* (to play). As can be seen on the figure, four specific meanings of the verb emerge: *s'amuser* (have fun), *risquer* (*jouer de l'argent*: to gamble), *tromper* (*se jouer de quelqu'un*: to deceive somebody), and *imiter* (*jouer les victimes*: to play the victim), the center of the representation corresponding to the more general meaning *pratiquer* (*jouer aux cartes, au tennis, du piano*: to play cards, tennis, piano).

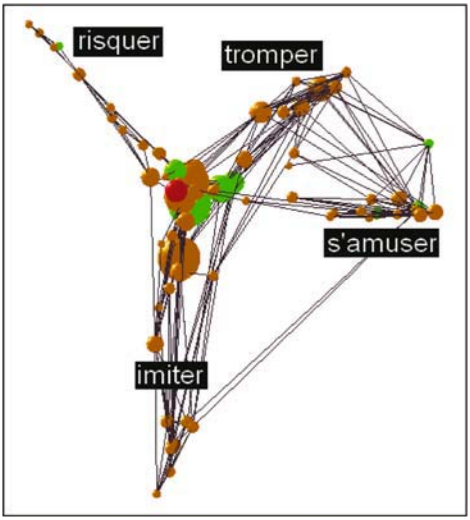


Figure 6: Proxemic zoom onto *jouer*.

Thus, proxemy offers an interesting method to study a lexical graph at different scales, from the most global structure to the most detailed meanings of a word.

4. SEMANTIC SPACES

Can the geometrical figure obtained by the proxemic method be considered as the abstract semantic space we were looking for? As we said at the beginning of the paper, lexical units must be represented by regions rather than points in the semantic space, if we want to take into account their polysemy and the overlap of meanings characterizing partial synonymy between several units. In order to maintain coherence in our model, we must consider that the different meanings of a lexical unit are scattered over an area surrounding the proxemic vector representing the unit. As a matter of fact, this is exactly what we just did when we looked at the proxemic zoom onto the verb *jouer*. In the figure 6, the sphere labeled by *jouer* is located in the middle of the representation, where we found the more general meaning *pratiquer*, but all the other meanings are spread and situated relatively far from the sphere of *jouer*, which plays a role similar to a center of gravity. This can be better visualized in the figure 7, where *jouer* and its proxemic neighbors are represented in the global conceptual tetrahedron of the French verbs.

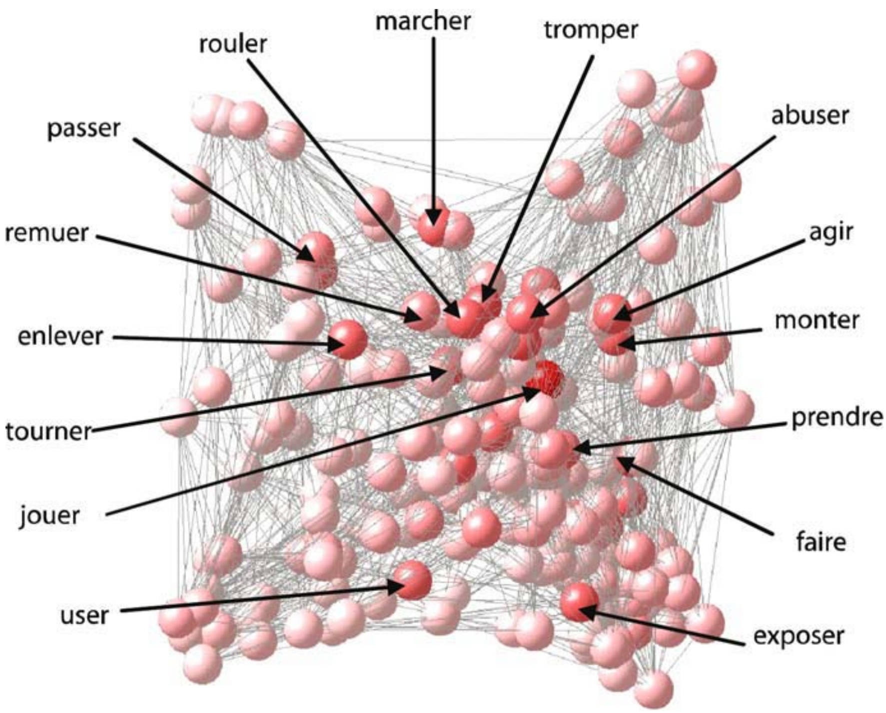


Figure 7: Localisation of *jouer* and its proxemic neighbors in the global representation of French verbs (hue indicates the proxemy with respect to *jouer*).

Therefore, we will define the semantic area associated with a given unit as the region containing all the units having a high proxemy with respect to it. With this definition, we can assume that proxemic representation gives a good approximation of the semantic space needed in our model.

In order to check this assumption, we used a completely different method of construction of a global semantic space. This method has been used for several years by one of us, Bernard Victorri, to build local semantic spaces associated with polysemic lexical units (Ploux & Victorri,1998). Fabienne Venant (2004) and Nabil Abdellaoui (2004) have extended very recently this method so as to apply it to the building of global semantic spaces.

The main idea of the method consists in associating points of the semantic space with the *cliques* of the lexical graph. The cliques of a graph are its maximal completely interconnected subsets of nodes, i.e., in our case, maximal sets of lexical units that are all synonyms for one another. The cliques define very precise meanings that can be considered as the intersection of the meanings of all the units belonging to the clique[11]. It is worth noticing that the "synsets" of WordNet (Fellbaum,1998) are analogous to the cliques in that they are also sets of words designed to represent the different meanings of a lexical unit.

Let us take an example to illustrate this point. As we saw, the French verb *jouer* displays a rather extended polysemy, with a large number of synonyms (precisely 94). Of course, most of its synonyms are far from being synonyms for one another. For instance, if we look at the synonyms that we presented above to characterize the different parts of the proxemic zoom of figure 6 (*s'amuser*, *risquer*, *tromper*, *imiter*, *pratiquer*), they convey very different meanings. On the opposite, the cliques containing *jouer* evoke a unique nuance of meaning of *jouer.* For instance, we find, among others, the three following cliques:

{*jouer, aventurer, compromettre, exposer, hasarder, risquer*}
{*jouer, miser, boursicoter*}
{*jouer, miser, parier, ponter*}

All of them can be considered as instances of the '*risquer*' meaning of *jouer*, but each of them enhances a precise determination of this meaning (the first one evokes venturing and hazarding, the second speculating, and the third gambling and betting). It is then sensible to assume that each clique has to be represented by a point in the semantic space. As the number of cliques containing *jouer* is also rather large (precisely 98), we have enough points to design what we called the semantic space associated to *jouer*

In order to build the semantic space of a given unit, we compute a distance between the cliques containing the unit. We use the chi-square distance[12], a metric which is well known in statistical analysis, intensely used to compute correspondences between subsets of individuals and subsets of qualitative characteristics. As usual, principal component analysis is applied

[11] The algorithm used to compute the cliques can be found in Reingold *et al.*, 1977. For a similar approach using also a graph of synonymy, see Warnesson, 1985.

[12] More precisely, let $u_1, u_2, \ldots, u_n$ be the synonyms of the given unit, $c_1, c_2, \ldots, c_p$ the cliques containing the unit, and x_{ki} the coordinates of the cliques over the synonyms: $x_{ki} = 1$ if $u_i \in c_k$ and $x_{ki} = 0$ si $u_i \notin c_k$. Then the distance $d(c_k, c_l)$ between two cliques is given by the following formula:

$$d^2(c_k, c_l) = \sum_{i=1}^{n} \frac{x}{x_{\bullet i}} \left(\frac{x_{ki}}{x_{k\bullet}} - \frac{x_{li}}{x_{l\bullet}} \right)^2 \text{ where } x_{\bullet i} = \sum_{j=1}^{p} x_{ji} \text{ , } x_{k\bullet} = \sum_{i=1}^{n} x_{ki} \text{ , and } x = \sum_{i=1}^{n} \sum_{j=1}^{p} x_{ji}$$

to reduce the dimensionality of the space (for all the technical details and a thorough discussion of the model, see Ploux & Victorri, 1998).

As can be seen in figure 8, the semantic space of *jouer* obtained by this method is strikingly similar to the proxemic zoom we presented above, as far as structure is concerned: four branches for the same specific meanings (*s'amuser*, *risquer*, *tromper, imiter*), with the general '*pratiquer*' meaning in the center. It must be emphasized that a synonym can appear in different regions of the semantic space. For instance, one can see figure 8 that the verb *rouler*, a highly polysemic French verb, is present in two regions: in the center, with the meaning 'to swing', 'to oscillate', and in the *tromper* branch, with the meaning 'to deceive', 'to trick'. This is one of the main qualities of the model: as expected (cf. introduction), words meet in the semantic space at different places, each place corresponding to a precise meaning.

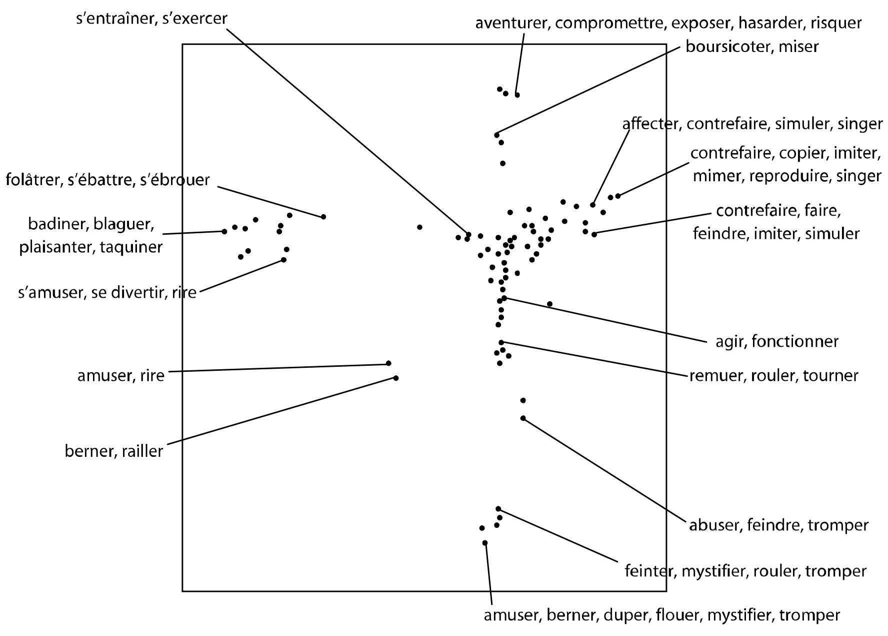


Figure 8: The semantic space associated with jouer.

Thus, it is interesting to see if this method also gives the same general semantic dimensions as the proxemy method, when applied at the global level. We can obtain all the cliques of the complete graph Synoverbe (there are more than 25,000 cliques for less than 10,000 nodes) and compute the distance between any couple of them. Obviously, we cannot visualize a so large semantic space with the simple technique we used for local semantic spaces.

To solve this problem, we first build all the balls of a given radius centered on a clique, and we select the first hundred ones that have the highest density. In other words, we select the largest groups of strongly interconnected units, analogically corresponding to the centers of the largest cities (see the first part of this paper). Each high-density ball is assimilated to its center, which is associated with the whole set of synonyms corresponding to the union of the cliques included in the ball. Principal component analysis is then applied to the centers.

The results can be visualized on figure 9 and figure 10. Once again, the similarity with the results obtained with global proxemy is striking. We observe the same four main semantic axes structuring the meanings of French verbs. Figure 9 shows the projection of the global semantic space onto the first two dimensions. Three semantic zones appear, corresponding to three semantic axes that can be called 'positive' (*exciter, provoquer, produire*), 'negative' (*détruire, enlever, affaiblir*), and 'expressive' (*dire, montrer*). Figure 10 is a three-dimensional representation of the same space. It reveals the fourth semantic axis, which we called 'repulsive' (*disparaître, quitter, partir, sortir*).

On each figure, some information is given for a few representative balls, namely the content of the clique which is at the center of the ball, and the number of cliques and synonyms belonging to the ball. It can be observed that high density of cliques is not necessarily correlated with high density of synonyms. It means that this method actually brings out semantic zones where lexical units are highly interconnected. Moreover, as we already saw for the local semantic space of *jouer*, highly polysemic units cover very large regions of the global space. For instance the verb *sortir* extends over a large part of the 'expressive' zone and a large part of the 'repulsive' zone. Thus, this model gives a method to classify polysemic units, depending on the size of the associated region in the global representation: the most highly polysemic unit is not the most connected one (i.e. the node of highest degree in the graph), but rather the most extended one in the global semantic space.

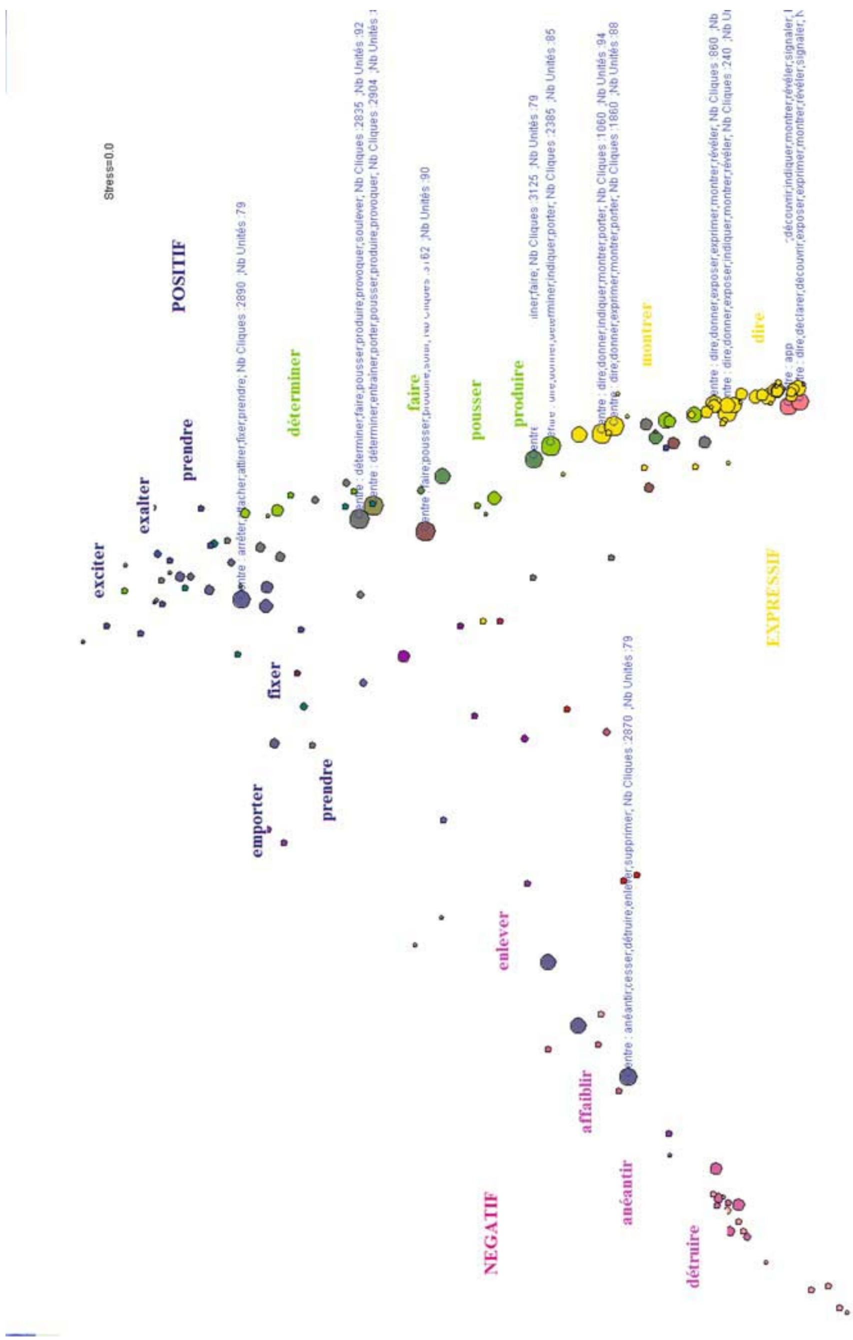


Figure 9: Two-dimensional representation of the global semantic space of Synoverbe.

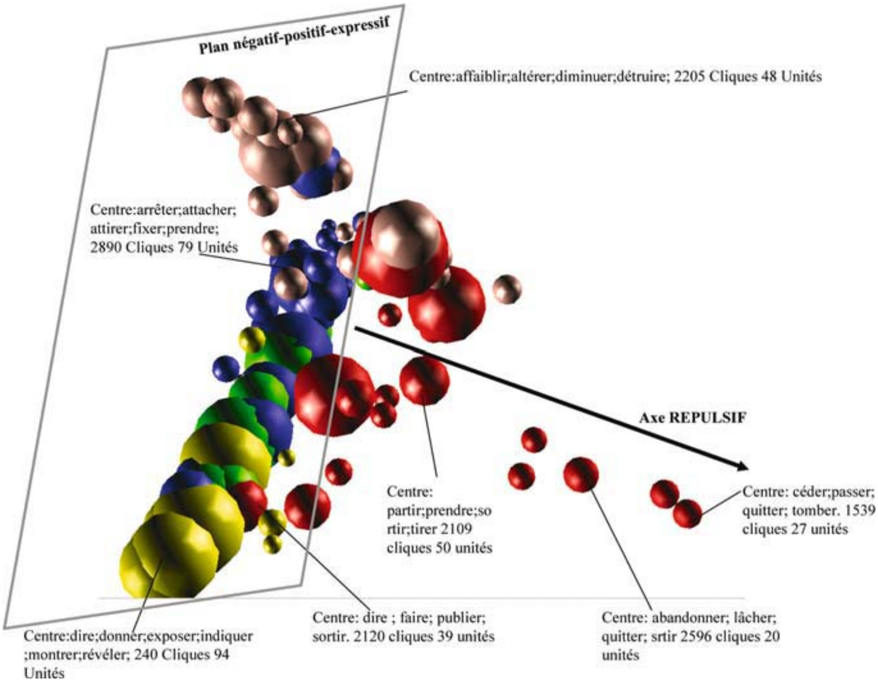


Figure 10: Three-dimensional representation of the global semantic space of Synoverbe.

The convergence of two independent methods seems to prove that the features which are revealed by both of them are really intrinsic to the structural properties of the French verb lexicon. We have then at our disposal two tools to explore the hierarchical structure of small world graphs. As we said, the construction of 'semantic spaces' is not only interesting for lexical systems: it can prove very valuable for other 'semantic' graphs like the Web, as well as social graphs where the involved relationship depends more on conceptual factors than on geographical ones.

REFERENCES

Abdellaoui, N., 2004, Géométrisation et exploration du sens, *Mémoire de DEA,* EHESS, Paris.

Barabási, A. L., Albert, R., Jeong, H., and Bianconi, G., 2000, Power-Law Distribution of the World Wide Web, *Science*. **287/2**115a.

Barabási, A. L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., and Vicsek, T., 2002, Evolution of the social network of scientific collaboration, *Physica A*. **311(3-4)**: 590-614.

Bermann, A., Plemons, R. J., 1994, *Nonnegative Matrices in the Mathematical Sciences*, Siam, Philadelphia.

Bréal, M, 1897, *Essai de sémantique science des significations*, Hachette, Paris, rééd. 1983.

Darmesteter, A., 1887, *La vie des mots étudiés dans leurs significations*, Éditions Champ Libre, Paris, rééd. 1979.

Duvignau, K., 2002, *La métaphore, berceau et enfant de la langue. La métaphore verbale comme approximation sémantique par analogie dans les textes scientifiques et les productions enfantines (2-4 ans)*. Université Toulouse Le-Mirail, thèse Sciences du Langage.

Duvignau, K., Gaume, B., 2003, Linguistic, Psycholinguistic and Computational Approaches to the Lexicon: For Early Verb-Learning, *Cognitive Systems*. **6 (1)**.

Duvignau, K., Gaume, B. 2004, First Words and Small Worlds: Flexibility and proximity in normal development. *Proceedings of the interdisciplinary conference* "*Architectures and Mechanisms for Language Processing*", Aix en Provence.

Fellbaum, C., ed., 1998, *WordNet, an Electronic Lexical Database*, MIT Press.

Ferrer, R., Solé, R. V., 2001, The small world of human language, *Proceedings of The Royal Society of London. Series B, Biological Sciences*. **268(1482)**: 2261-2265.

Gaume, B., Duvignau, K., Gasquet, O., Gineste, M-D., 2002, Forms of Meaning, Meaning of Forms, *Journal of Experimental and Theoretical Artificial Intelligence*. **14(1)**: 61-74.

Gaume, B., 2003, Analogie et Proxémie dans les réseaux petits mondes, Regards Croisés sur l'Analogie, *Revue d'intelligence Artificielle*. **n° 5-6**, Hermès.

Gaume, B., 2004, *Balades aléatoires dans les petits mondes lexicaux,* Cepadues, http://dilan.irit.fr/.

Guare, J., 1990, *Six degrees of separation: A play*, Vintage Books, New York.

Huberman, B. A., and Adamic, L.A., 1999, Growth dynamics of the world-wide web, *Nature*. **401/131**.

Kleinberg, J. M., Kumar, R., Raghavan, P., Rajagopalan, S., Tomkins, A. S., 1999, The Web as a Graph: Measurements, Models and Methods, *Lecture Notes in Computer Science*, **1627**: 1-17.

Milgram, S., 1967, The small world problem, *Psychol. Today*. **2**: 60-67.

Newman, M. E. J., 2001, The structure of scientific collaboration networks, *Proc. Natl. Acad. Sci. USA*. **98**: 404-409.

Ploux, S., Victorri, B., 1998, Construction d'espaces sémantiques à l'aide de dictionnaires de synonymes, *Traitement automatique des langues*. **39(1)**: 161-182.

Ravasz, E., Barabási, A. L., 2003, Hierarchical Organization in Complex Networks, *Phys. Rev. E*. **67/026112**.

Reingold, E. M., Nievergelt, J., Deo, N., 1977, *Combinatorial Algorithms, Theory and Practice*, Prentice-Hall, Englewood Cliffs.

Senata, E., 1981, *Nonnegative Matrices and Markov Chains*, Springer, New York, 2nd ed.

Sigman, M., Cecchi, G. A., 2002, Global organization of the Wordnet lexicon, *Proc. Natl. Acad. Sci.* **99(3)**: 1742-7.

Venant, F., 2004, Géométriser le sens, *RECITAL Conference*, Fès, Morocco.

Warnesson, I., 1985, Applied Linguistics: Optimization of Semantic Relations by Data Aggregation Techniques, *Journal of Applied Stochastic Models and Data Analysis*. **vol.1/2**: 121-143.

Watts, D. J., 1988, Strogatz S. H., Collective dynamics of 'small-world' networks. *Nature*. **393**: 440-442.

Chapter 6

HIERARCHY IN CITIES AND CITY SYSTEMS

Michael BATTY
Bartlett Professor of Planning
University College London
Centre for Advanced Spatial Analysis (CASA)

> "To a Platonic mind, everything in the world is connected to everything else and perhaps it is. Everything is connected but some things are more connected then others. The world is a large matrix of interactions in which most of the entries are close to zero, and which by ordering those entries by their orders of magnitude, a distinct hierarchic structure can be discerned." (page 258)
>
> Herbert A. Simon (1977)

Hierarchy is implicit in the very term city. Cities grow from hamlets and villages into small towns and thence into larger forms such as 'metropolis', 'megalopolis' and world cities which are 'gigalopolis'. In one sense, all urban agglomerations are referred to generically as cities but this sequence of city size from the smallest identifiable urban units to the largest contains an implicit hierarchy in which there are many more smaller cities than larger ones. This organisation approximately scales in a regular but simple manner, city sizes following a rank-size rule whose explanation is both mysterious and obvious. In this chapter, we begin with a simple but well-known model of urban growth where growth is randomly proportionate to city size and where it is increasingly unlikely that a small city becomes very big. It is easy to show that this process generates a hierarchy which is statistically self-similar, hence fractal but this does not contain any economic interactions that we know must be present in the way cities grow and compete. We thus modify the model adding mild diffusion and then note how these ideas can be fashioned using network models which generate outcomes consistent with

D. Pumain (ed.), Hierarchy in Natural and Social Sciences, 143-168.

these kinds of order and scaling. We then turn this argument on its head and describe how the same sorts of morphology can be explained using ideas from central place theory. These notions are intrinsic to the way cities evolve and we conclude by noting how city design must take account of natural hierarchies which grow organically, rather than being established using top-down, centralized planning.

1. THE URBAN SOUP

Conventional wisdom concerning the origins of life on earth are now largely fashioned around the notion that in the beginning, life began through some chance spark setting off a reaction in a sea of undifferentiated chemical soup, leading to the formation of the various nucleotides that constitute the buildings blocks of life – RNA and DNA. In the same way, we can speculate that societies and cities began with household units randomly located across a landscape where the spacing of individuals was determined by food available from hunting and gathering. These units of course made contact in their quest for survival and although the dominant mode was one in which households competed with one another for territory which was synonymous with survival, there was a dawning realisation that cooperation rather than competition could ensure greater prospects for survival. Hamlets and villages were formed initially to ensure strength in numbers, for protection, but in time, the social contact which resulted, reinforced a division of labour leading to increased prosperity.

The simplest possible model is one in which some individuals in this undifferentiated urban soup grow more than others simply due to the fact that they continually get ahead, while others fall behind, often disappearing. Eventually clusters that are differentiated by size which we call cities, appear in this landscape, and it is these that give structure to the urban soup. Hierarchy is an intimate part of this structure but before we show how such hierarchies emerge as a natural part of the growth process, we will take one step back and show how cities in this artificial world first organise themselves according to size.

A hierarchy is a natural ordering that is initially based on size but size can be measured in many different ways. In cities, size is typically based on the number of individuals or households or workers – on populations – but it may also be based on the area over which such location occurs or energy is used, or the field of influence over which individuals in the hierarchy have control. Let us begin with the simplest of possibilities: places of the same size are randomly scattered over a uniform plane. In such a world, if clusters exist, then these are random occurrences. We will assume a process in which a place grows randomly but this growth rate is applied proportionately to the size of that place. So if a place i at time t has size P_{it} and the growth rate ε_{it} is chosen randomly, the place grows (or declines) as

$$P_{it+1} = (1+\varepsilon_{it})P_{it} = P_{it} + \varepsilon_{it}P_{it} \quad (1)$$

The consequences of this process are surprising at first until one pauses to reflect. In a system of many places, the distribution of growth rates will be uniform at any time t over a range from small to large, which might also be from negative to positive. However the chances of any particular place getting a series of very high growth rates allocated to it one after another, and thus growing very big is increasingly small. Equally the same goes for a place getting increasingly small and of course in this model if a place gets too small, it disappears so there is some asymmetry within the process. It is very easy to work out what happens if we apply this growth process to a small number of objects, with random growth rates chosen from a given range, and then apply these using equation (1) over and over again. An increasingly small number of the objects grow big, most remain small, quite a lot disappear but the crucial issue is 'does the resulting size distribution show any kind of order'. In a sense we have anticipated that it does: there are far fewer bigger objects than smaller but let us take a worked example, which although somewhat artificial, graphically demonstrates the point.

Our example is based on a grid of objects of dimension 21 x 21 giving 441 objects or spaces where the initial populations are uniformly distributed with $P_{i0} = 1, \forall i, t = 0$. The rates of growth ε_{it} are chosen randomly in the range $-0.1 < \varepsilon_{it} < 0.1$. The proportionate growth model in equation (1) quickly sorts out the objects into a size distribution and by time $t = 100$, the frequency distribution shows every sign of being lognormal. In fact, we have run the model for 1000 iterations which we refer to somewhat euphemistically as 'years' and during this simulation there is much movement between those objects in terms of their relative size. The popular way of organising such frequency distributions is by ordering them hierarchically in terms of size and then plotting them in this order which is against their rank. The so-called rank-size distribution or Zipf plot is in fact the counter-cumulative; that is, the rank is the number of objects above a certain size threshold, so the highest ranked population in the hierarchy at rank $r = 1$ is the population $P_{it}(1)$, the second highest population $P_{it}(r)$ is at rank $r = 2$ and so on down the size distribution to $P_{it}(r)$. This rank-size distribution is plotted for $t = 1000$ in Figure 1, and it is immediately apparent that the signature is that of a lognormal distribution where the plot is visualised as a log transformation of population size against rank.

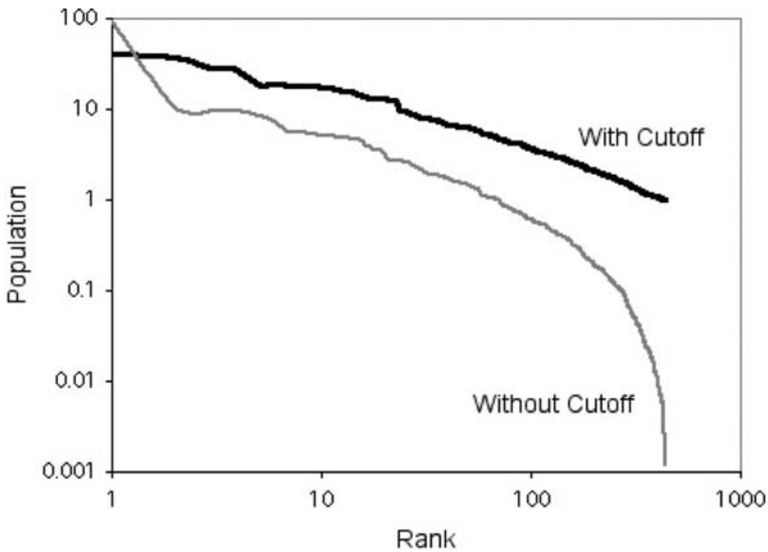


Figure 1: Generating a Lognormal Distribution Using Proportionate Effect and Power Law Scaling from Proportionate Effect with a Minimum Size Threshold

In fact, many researchers have shown that proportionate random growth of the kind we have described leads to lognormal size distributions (Pumain, 2000). Gibrat (1931) produced the first comprehensive argument for cities and income distributions but the English statisticians Fisher and Yule knew of the model and its consequences a generation earlier. If we were to continue the simulation beyond $t = 1000$, then more and more populations would converge to zero and ultimately, we hypothesise that in discrete systems of this kind, all activity would be attracted to a single cell. In fact, in city systems, such a simulation is bounded from below by indivisibilities and it thus makes sense to modify our model by introducing a size threshold below which populations cannot fall. Then whenever a population cell i falls below this number, it is restored to that number, this mechanism acting as a safety net or subsidy of sorts. This can also be viewed as a way of killing off a city and introducing a new one at the same time, thus incorporating a perfectly balancing birth and death process; formally

$$if \quad P_{it} < \Psi \quad then \quad P_{it} = \Psi \qquad (2)$$

Examining the lognormal distribution in Figure 1 reveals two regimes – the long tail which is almost linear and the short tail which accounts for the order of the smaller settlements. It is tempting to think that the long tail could be approximated by a linear or scaling relation and in effect, if we use the cut-off mechanism just postulated then in a purely phenomenological sense, this is effectively cutting off this short tail. We have run the simulation again with equation (2) now operative and indeed the almost straight line distribution in Figure 1 is that which is generated by time $t = 1000$. The cut-off in fact works and what we end up with is a distribution which is no longer log-normal. In fact it is scaling as it can be approximated by a power function where the population size which we will now call $P_{it}(r)$ varies inversely with the rank r as $P_{it}(r) \sim r^{-\alpha}$ with α the so-called scaling parameter of the distribution. As cities are moving up and down this hierarchy, it is tempting to think of the long tail as a 'steady state' to which cities are 'attracted' and indeed, theorists such as Gabaix (1999) demonstrate that this is indeed the case for the Gibrat process which can converge to the a pure scaling law with the parameter $\alpha = 1$.

2. RANK SIZE AND THE LAW OF PROPORTIONATE EFFECT

What we have just demonstrated has been known as an empirical fact about cities and many other distributions for over one hundred years. The most popular exposition of rank order which conforms to a scaling relation is provided by Zipf (1949) in his remarkable book which examined many such distributions from word frequencies to cities. Zipf argued that these distributions were not only scaling, conforming to the power law, but also that many such distributions – indeed the implication that all such distributions – were such that the power law was a pure inverse. This means that the rank of the population in continents, countries, and counties, at any level or scale, would conform to $P_{it}(r) \sim r^{-1} = 1/r$. This is the strong form of Zipf's Law. Zipf's Law implies that city size distributions are fractal in that if one examines the relationship at any scale, then the distribution is the same. This is self-similarity in its pure form which in terms of a power law means that if the distribution is rescaled, then this is simply a scaling up of the original distribution. Imagine that the rank size is rescaled by rank to another order s, then the rank size scales as $sP_{it}(r) \sim (sr)^{-1} = s^{-1}r^{-1} \sim r^{-1} \sim P_{it}(r)$ which implies that the scaling is the same over any order of magnitude (Batty and Shiode, 2003).

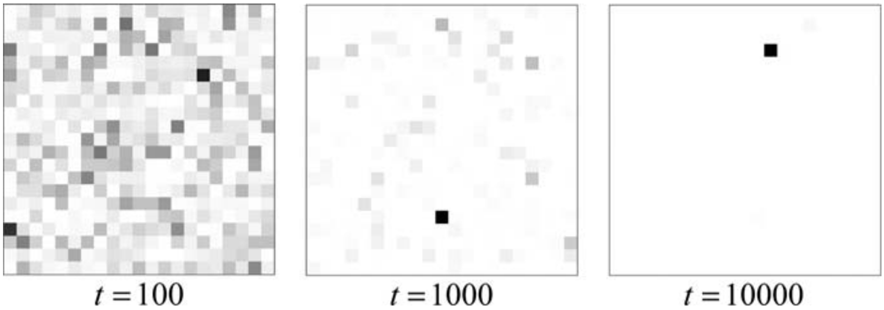


Figure 2: Emergence of the Rank-size Distribution Using Proportionate Effect with Cut-Off

The law of proportionate effect with a lower bound is akin to a random walk with a reflecting barrier (Sornette, 2000). The model is simplistic, perhaps nihilistic in that it does not include any form of competition or interaction between the objects. This is extremely odd as cities compete and interact and many models of their formation emphasise such interactions. Our model of proportionate effect with the lower bound clearly generates distributions which appear to be scaling and follow Zipf's Law, but in many ways this model is unstable. The time over which such distributions emerge and the volatility of the top ranked cells or places is sufficient to suggest that the model does not have enough inertia to mirror real places. The fact that it produces size distributions which concur with reality is not sufficient to mean that this is a good model. For example, as we move through the time periods, then the distributions which are generated change not in their scaling but in their shape. We can see this in two ways. In Figure 2, we show the pattern of distribution after 100, 1000 and then 10000 iterations ('years') which reveals the convergence to extreme distributions as the simulation continues. In Figure 2, the distributions for small number of interactions are much flatter and gentle than those for larger numbers.

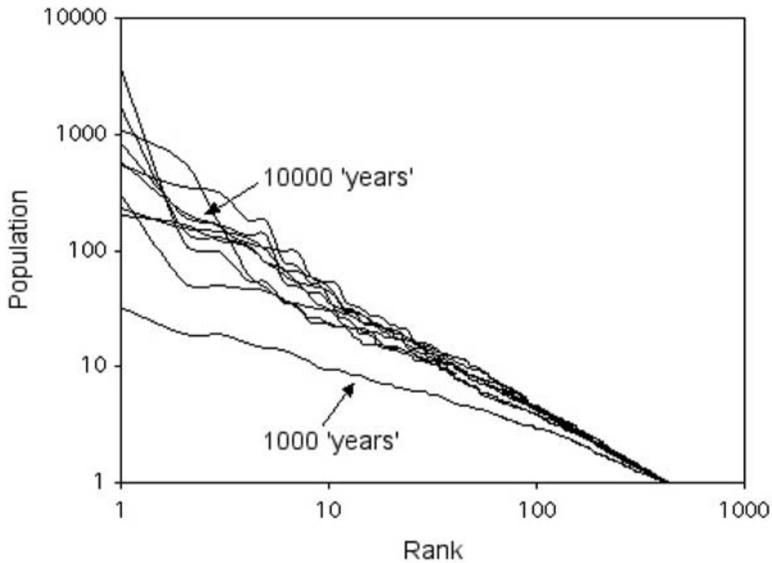


Figure 3: Power Law Scaling as the Population Distribution Emerges

Figure 3 however shows the distributions for $t = 1000, t = 2000, \ldots t = 10000$ which clearly get steeper – implying the parameter α gets larger as the population grows. In fact it appears that the parameter is converging on the pure Zipf case of unity although from these results this is inconclusive. There is a variety of theoretical evidence that suggests this is the case for growth by random walk with a reflecting barrier as Gabaix (1999) and Blank and Solomon (2000), for example, show. In Figure 3, the parameter α rises from $t = 1000$ to $t = 10000$ as 0.668, 0.862, 0.907, 0.977, 1.008, 0.984, 0.980, 0.978, 1.053, to 0.962 which shows the final value hovering around 1 with the straight line logarithmic fits all explaining more than 99 percent of the variance for each time slice. This is a fairly remarkable result. What we have shown is that an almost nihilistic model with no spatial competition can generate highly ordered simple hierarchies which in fact mirror the empirical evidence that has been compiled for many cities in many places during the last 50 years. Figure 4 shows the rank size of incorporated places in the United States from 1970 (some 7000 places) to the year 2000 (some 25000). In Figure 4(a), the entire distributions are shown and these are clearly lognormal. When we cut-off the short tails, the remaining long tails are quite straight implying power laws as in Figure 4(b). In fact Figure 4 is an empirical equivalent of Figure 1, and for the four time slices from 1970 to 2000, the α parameter varies from 0.986 to 0.982 to 0.995 to 1.014 with the variance explained a little lower than the theoretical model as 0.98, 0.97, 0.97, and 0.97. The same kind of dynamic analysis has been done for France by Guerin-Pace (1995) and a thorough review is presented by Pumain (2005) in a complementary chapter within this book.

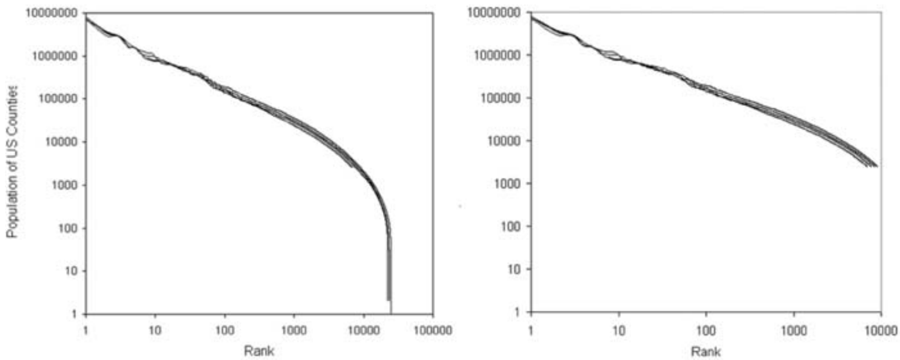


Figure 4: Lognormal and Power Law Scaling of the US Population Based on 'Incorporated Places' 1970 to 2000

Although the model produces aggregate distributions uncannily close to those we observe in most places, what is quite clear is that when we unpack the simulations, there are many inconsistencies that imply this model is nothing like as good as these results suggest. Of particular concern is the lack of apparent structural consistency as the simulation proceeds and the rankings of cells change. During the 10000 'year' simulation, the number of different cells at the top of the rank order is 18. We have only sampled the rankings at every 50 time periods and thus it is likely that there are many more than 18 cells which appear at the top of the ranks during the simulation. To give an idea of the volatility of these ranks, we show how the top ranked cells 1, 6, 12 and 18 from these top orders change over the 10000 year history in Figure 5. These cells appear at different times as we indicate but what is quite clear is that the length of time they occupy the top position is small, thus implying that there is no inertia in the model. This is manifestly the case and as the simulation time periods have no relationship to real times, then it is unclear what the 10000 model 'years' actually mean in terms of the evolution of actual urban systems such as the US system shown in Figure 4. Our guess is that the volatility of the distribution through time is much greater than any real system where change is slower and inertia greater.

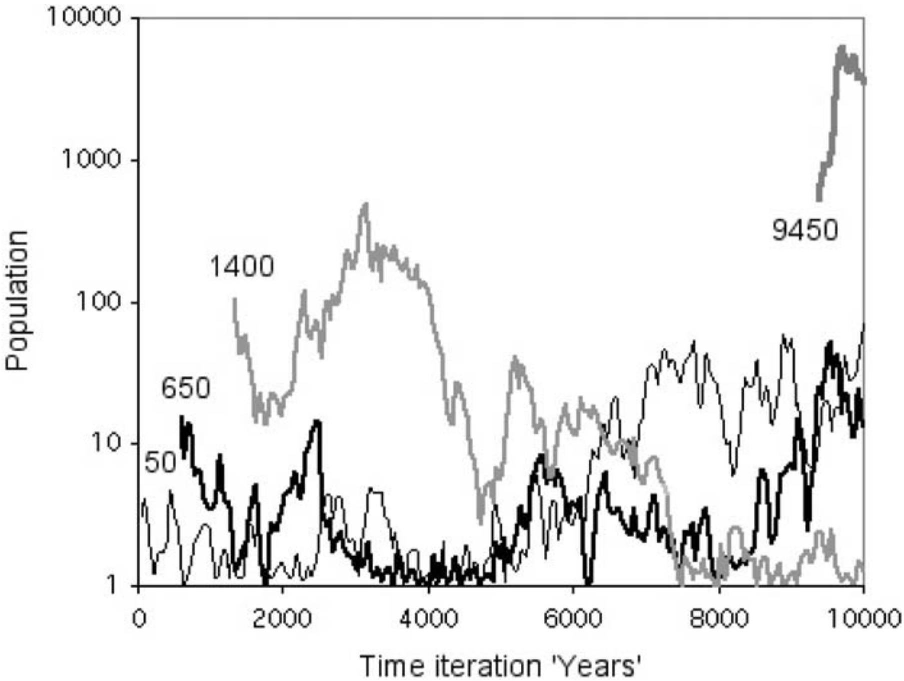


Figure 5: The 1st, 6th, 12th and 18th Top Ranked Population Cells and Their Progress Through the Simulation

The last thing that we want to show before we try to improve the model and generate hierarchies which imply spatial interaction and competition, is the effect of changing the geographic dimensions of the space within which the simulation takes place. We have changed the grid from 21 x 21 to 51 x 51 and then to 101 x 101 and run the simulation with the cut-off for 10000 'years'. We show the three rank-size distributions in Figure 6 where it is quite clear that the slopes are similar implying that the model does indeed hold up as we scale the system in geographic size. This might be expected as there is no interaction between the parts but what is of interest is the increased size of the populations as the systems scales. This is a bit of a mystery but it probably occurs because there are more and more opportunities for extreme growth as the spatial system gets larger, yet it requires further investigation as do many other aspects of these simulations. It does not however detract from the main result that the model scales spatially. It is surprising that so simple a model which has had so much effort devoted to it in terms of simulations and mathematical analysis is still far from being thoroughly understood.

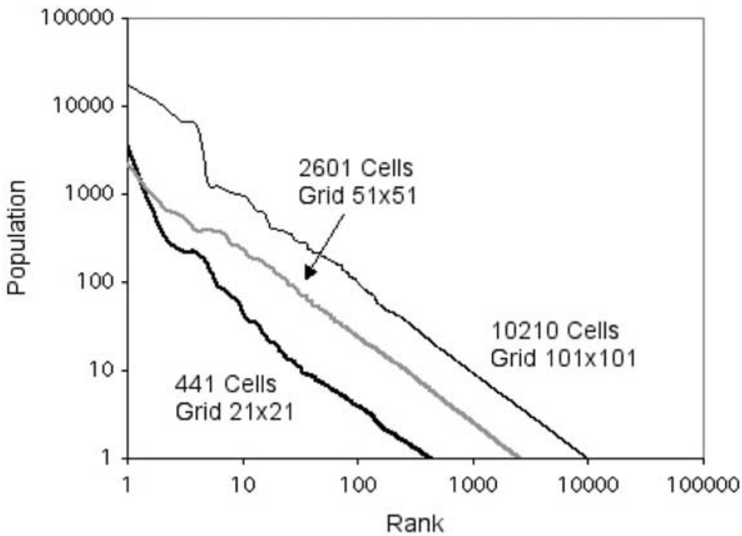


Figure 6: Consistent Scaling Behaviour for Different Sizes of Lattice

3. GENERATING HIERARCHY: COMPETITION, INTERACTION AND SPATIAL DIFFUSION

The hierarchy generated by the model of proportionate effect is the simplest possible – a simple rank order or unidirectional hierarchy where the order of objects is simply one of size and where each object is independent of any other. This cannot be a good model for the growth of cities because is does not admit competition or interaction of any kind. Cities are completely disconnected from one another. Simon's opening quote is largely irrelevant to this definition of hierarchy for nowhere in such a model are there clusters of connected activity which provide the kind of connectivity from which hierarchic structure can be derived. What we require is some form of interaction between cities or places, between the points on the lattice and to explore this, we will add some simple diffusion to adjacent grid cells at each stage of the model simulation. In short, at each time step, a fixed proportion λ of the population in each cell k diffuses to its nearest neighbours in the von Neumann neighbourhood comprising the cells which are north, south, east, and west of the cell in question. Thus for cell i, the population at time $t+1$, P_{it+1}, is now computed as

$$P_{it+1} = (1+\varepsilon_{it})P_{it} + \lambda \frac{\sum_{k\in\Omega_i} P_{kt}}{4} \quad (3)$$

where the neighbourhood for diffusion is defined as $\Omega_i = k_{north}, k_{south}, k_{east}, k_{west}$. In this model, minimal action at a distance is admitted and within a short time interval which is proportional to one dimension of the lattice, every cell influences every other cell. This kind of diffusion is still somewhat nihilistic in that it is only based on the notion that a proportion of people move to be with their neighbours, without specifying any particular reason, other than the implication that such movement is social and/or economic.

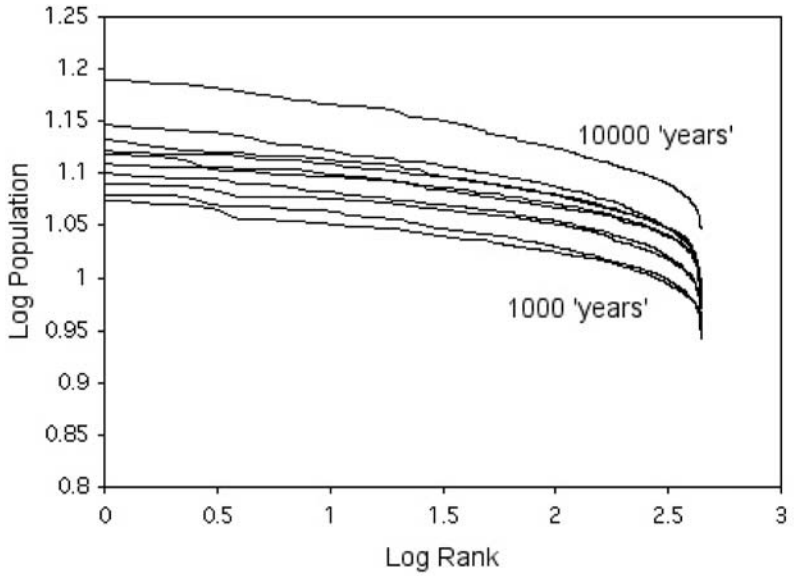


Figure 7: Lognormal Distributions Generated from Proportionate Effect With Diffusion

We have run the model in equation (3) retaining the cut-off in equation (2) for $t = 1000, 2000, ..., 10000$ time periods, and this generates the rank-size distributions shown as Zipf plots in Figure 7. The level of diffusion used involves setting the parameter $\lambda = 0.3$, implying 30 percent of the population in each cell gets redistributed into adjacent cells in each time period. What this leads to are not scaling distributions as the model without diffusion does but lognormal distributions. The cut-off is in fact discounted by the diffusion. This is a little different from the similar model posed by Manrubia and Zanette (1998) which is based on the same process but without a cut-off and with only positive growth which produces a scaling law. Yet in a sense, whether these kinds of model produce scaling or lognormal distributions is of less concern because at a phenomenological level, all multiplicative processes such as these variants belong to the same class of model (Sornette and Cont, 1997).

The diffusion in this model is so intensive in each time period – all cells are affected, and also extensive due to the fact that the number of time periods is far greater than the size of the system which in this example is based on the 21 x 21 lattice, that it is impossible to track all interactions which accumulate between all pairs of cells. Action-at-distance occurs through the medium of adjacent cells and the number of combinations of diffusion paths is thus enormous. What effectively this diffusion leads to are densities which fall around the cells with the largest populations just as a city core attracts and diffuses activity around it. We take an impressionist view of the hierarchy formed where we simply plot the hierarchy by associating cells with their higher order centre (based on population size), deciding whether or not they are connected simply through adjacency. In Figure 8, we first illustrate the patterns of growth for the model at $t = 100$, $t = 1000$, and $t = 10000$ and it is clear that the type of pattern produced occurs within 100 time periods and simply repeats itself – in different locations of course – through time. In Figure 8, we also show simplifications of these pictures by first identifying the top ranked cell, the next 3 followed by the next 8, then the next 24, and finally the next 64 around these cores. This provides a crude picture of population density which we can represent as a hierarchy. We do this for the pattern at $t = 10000$ where we simply associate each cell at each level with the cells above it if they are connected directly or indirectly to that level through cells of similar value.

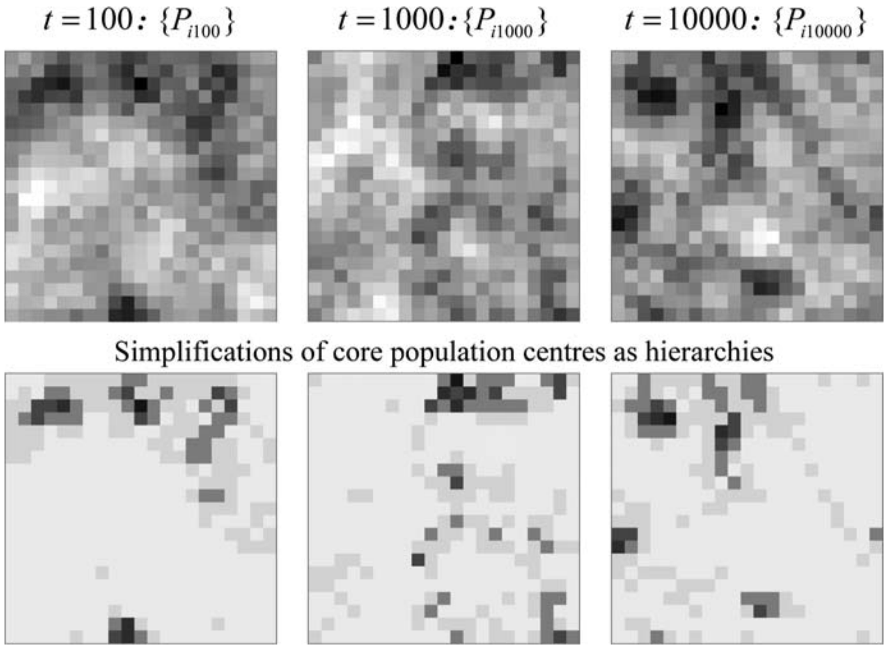


Figure 8: Patterns of Diffusion

The hierarchy produced is plotted as a semi-lattice in Figure 9. It is not possible to uniquely associate every cell with a single cell at the next level of hierarchy due to the fact that we do not have network links between cells that we can cut to define separate regions. In fact this representation of hierarchy is much more realistic and supports the long standing notion of overlapping fields of influence which was articulated rather well by Alexander (1966) almost 40 years ago in his article "A City is Not a Tree". There is considerable structure in this hierarchy which is introduced through the diffusion process but this model appears just as volatile as the pure Gibrat process. Over the simulation period of 10000 'years', of the 441 distinct cities or cells, all these cells occur at the top of the hierarchy at some point while the pattern of these top ranked cells would appear quite random. We show this pattern in Figure 10 where it is quite clear that there are no particular clusters of cells or individual cells that predominate over any others. It is quite clear from these simulations that there is too little inertia in the system to mirror our experience of real city systems for it is most unlikely that for the simulation times used here, all cells would at some point dominate. Other models are thus required.

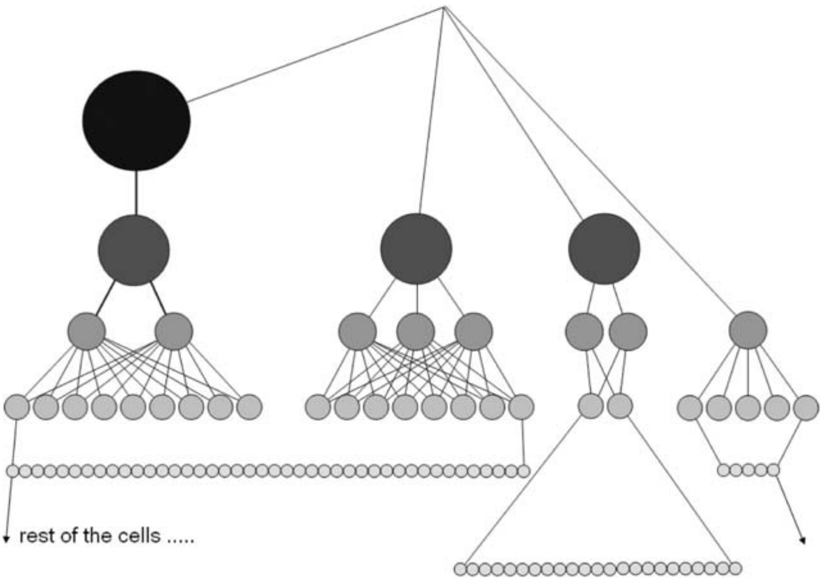


Figure 9: A Hierarchy for the Pattern at $t = 10000$

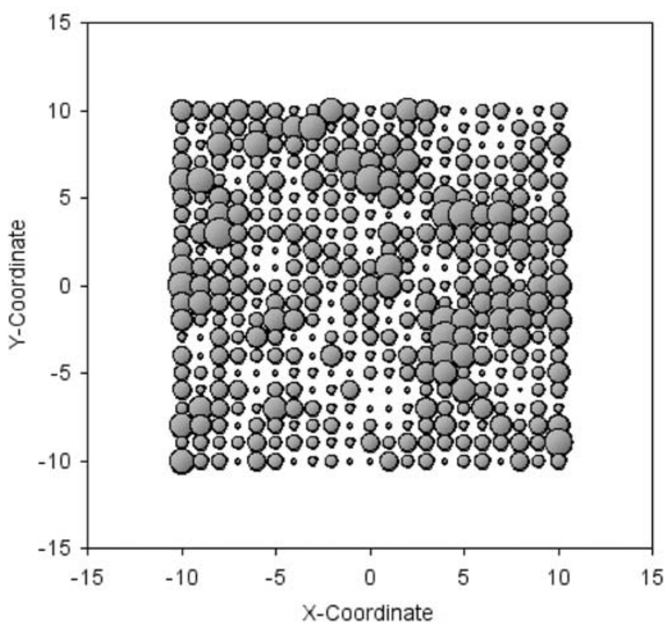


Figure 10: Top Ranked Cells During the 10000 'Year' Simulation (the size of the bubbles range from 1 to 83 time periods in which the relevant cells dominate with an average size of 23 time periods)

The distributions generated from the Gibrat process with diffusion are somewhat flat and as the level of diffusion is reduced, the hierarchical structure begins to disappear. We have attempted to inject more structure into the model by departing from the Gibrat process and introducing agglomeration economies into the model adding a term reflecting current city size. Our model thus becomes

$$P_{it+1} = (1+\varepsilon_{it})P_{it} + \lambda \frac{\sum_{k\in\Omega_i} P_{kt}}{4} + \phi P_{it}^{\eta} \tag{4}$$

where ϕ and η reflect the proportionality and the scaling imposed by agglomeration economies. We have set $\phi = 0.2$ and $\eta = 1.08$ and with these parameters we do indeed succeed in sharpening the distribution of city sizes but the lognormality of these distributions remains as we show in Figure 11. There do not appear to be any real qualitative differences produced by this model. To introduce a different form of hierarchy into the urban soup, we require much more explicit networks of interaction exploiting results from the burgeoning science of networks (Watts, 2003) which we will now present.

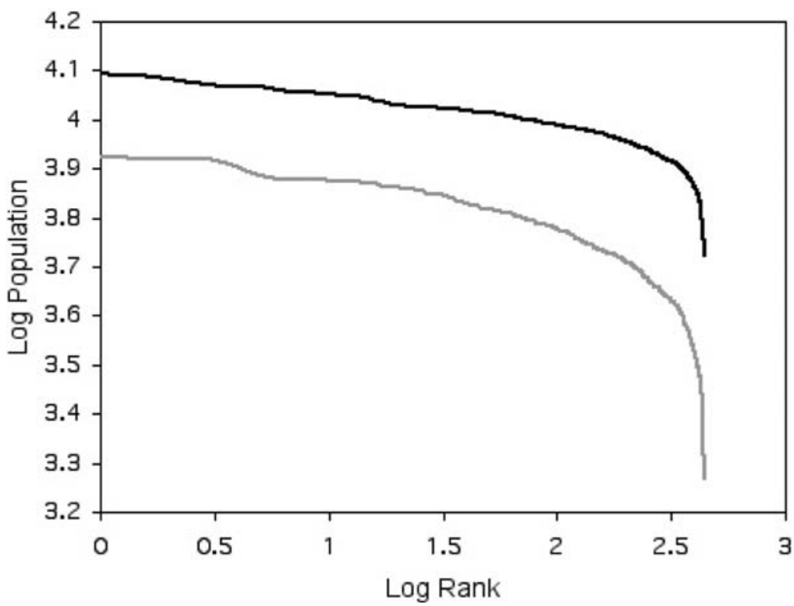


Figure 11: City Size Distributions for the Agglomeration Model at $t = 1000$ and $t = 10000$

4. NETWORK HIERARCHIES: THE GIBRAT INTERACTION MODEL

So far our models of hierarchy have focused on their evolution whereas Simon's (1977) definition tends to assume that such hierarchies are already developed. To detect them, we thus need to observe the interconnections between the system's parts in order to define the clusters of highly connected subsystems that form the whole. Our model needs to be extended to make explicit these interconnections and this requires us to generalise Gibrat's model to networks. We do this by adding both cells and their links randomly, one in each time period. The mechanics of the model are contained in the following equations. In each time period, for a node which is already established and linked to other nodes, we consider the random addition of a link volume as $\delta_{ikt+1} = 1$ where P_{ikt} is the total number of links from node i to j. The total links associated with i, the new population size of i, is P_{it+1}. The equation for total links is thus

$$P_{it+1} = \sum_{j} P_{ijt} + \delta_{ikt+1} \tag{5}$$

where the number of links is updated in each time period as

$$P_{ijt+1} = P_{ijt} + \delta_{ijt+1} \tag{6}$$

Whether a link is added or not depends on both the size of the node and its distance to other nodes which is reflected in an exponentially weighted gravitational function of the form

$$\delta_{ijt+1} = \begin{cases} 1 & if \quad rnd(\varepsilon_{ijt+1}) = KP_{ijt} \exp(-\beta d_{ij}) \\ 0 & \end{cases} \tag{7}$$

The term $rnd(\varepsilon_{ijt+1})$ determines a random choice based on size of the potential interaction where d_{ij} is the distance from node *i* to node *j* and the parameter β reflects the frictional effects of this distance. Essentially this process is one of preferential attachment in that links are added in proportion to the size of existing links and the population that the node has already attracted. It has been very widely exploited recently by Barabasi (2002) and his colleagues who have shown that the model does indeed lead to what that call 'scale free' networks (Barabasi and Albert, 1999).

This process does not show how nodes are established in the first place and thus we must add a mechanism for the birth of new nodes, akin to that added by Simon (1955) in his classic model of the rank-size process. A new node is added if a random variable $rnd(\varepsilon_{it+1})$ is greater than a predetermined threshold z which is given as

$$\delta_{it+1} = \begin{cases} 1 & if \quad rnd(\varepsilon_{it+1}) > z \\ 0 \end{cases} \tag{8}$$

where the value of z is small compared to the probability for the addition of new links as reflected in equations (5) to (7) above. For the 21 x 21 lattice, we choose the threshold for the addition of new nodes as $z = 0.1$ and this implies that at the beginning of the process, there is a 1 in 10 chance that a new node is added. Of course as the process continues, this chance falls for if a node is chosen that is already established, this is abandoned. In terms of the generation of links to established nodes, then a node is first chosen randomly but in proportion to its size P_{it}, and then a link to another node j from i is chosen in proportion to its inverse distance function as defined in equation (7). In this way, the network builds up through preferential attachment to existing nodes. The overall dimension of the system is 300 x 300 *x-y* coordinate units for each grid square and thus we have set the deterrence parameter β in equation (7) as 0.001 which implies an average distance of around 1000 units.

In Figure 12, we illustrate the final distribution of population by node $\{P_{i1000}\}$ and alongside this, the distribution of link volumes between nodes for all links greater than 1, those greater than 2 and finally those greater that 4. A hierarchical pattern is revealed by these figures and it would be possible to cut the link volumes at points where the cluster density falls below various thresholds, thus uniquely partitioning the space into different areas and then orders of hierarchy. We do not do this for our concern is not hierarchy *per se* but ways of generating this. In Figure 13, we plot the size of population per node against their rank as a Zipf plot which is the logarithmic transform. We have not connected the points in this plot because of the comparative low volume levels generated in this example but the plot is mildly lognormal. Fitting a straight line to this gives a scaling parameter of 1.05 with 90 percent of the variation in this plot explained. This is remarkably close to the pure Zipf scaling where $\alpha = 1$ and it is confirmation that this model of preferential attachment based on Gibrat (1931) does indeed generate the same profile as in the simpler non-network cases which we discussed above.

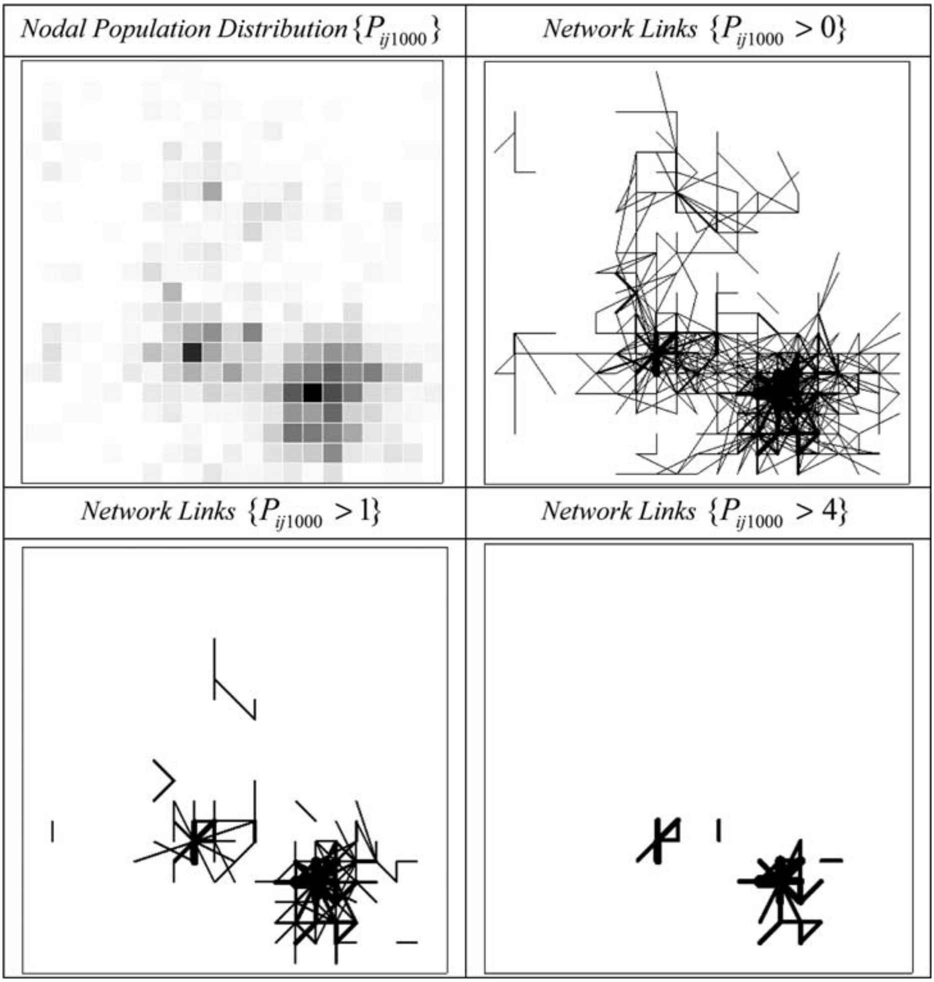


Figure 12: Patterns of Network Connectivity

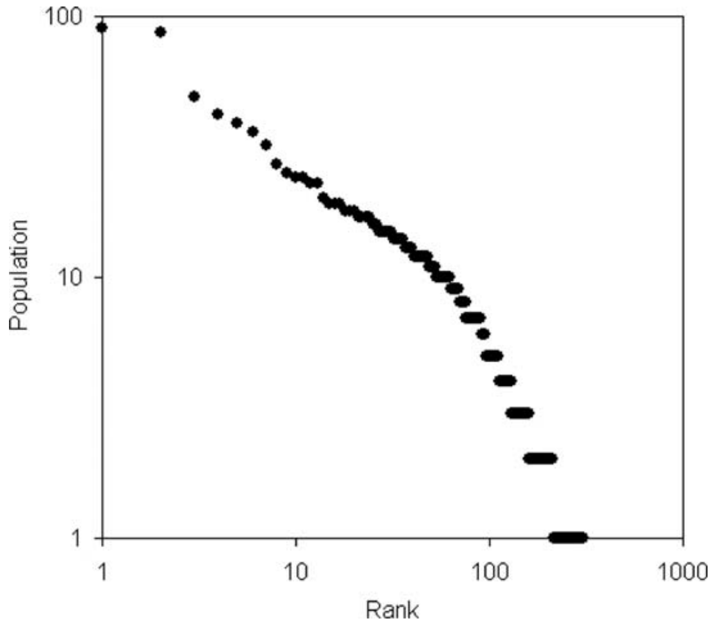


Figure 13: Rank-size Distribution of the Nodal Network Distribution

Before we introduce our final model where we will turn back to theory and generate the rank-size rule for hierarchies based on central place theory, we will illustrate an example of hierarchy for retailing within London. We have developed an index of retail intensity which is a linear weighted sum of some 42 separate indicators, each suitably normalised, and tagged to the postcode geography which at its finest scale represents retailing at an average resolution of some 50 metres (Thurstain Goodwin and Batty, 2002). We have interpolated a surface from this data and have then sliced it at some 5 different levels which provides a picture of the retail hierarchy which we show in Figure 14. This is an implicit hierarchy similar to those which can be derived from the population distributions illustrated earlier in Figures 8 and 12. What we do not have from this analysis is the detailed interaction pattern that links consumers to the retailing activity through their movements to purchase retail goods at different points or centres on this surface. But the pattern is consistent with all that we have seen previously and the distribution of retailing activity is rank size. This is only one way of implying that a hierarchy exists and in a sense, this is less explicit than the more formal approaches rooted in location theory to which we will now turn.

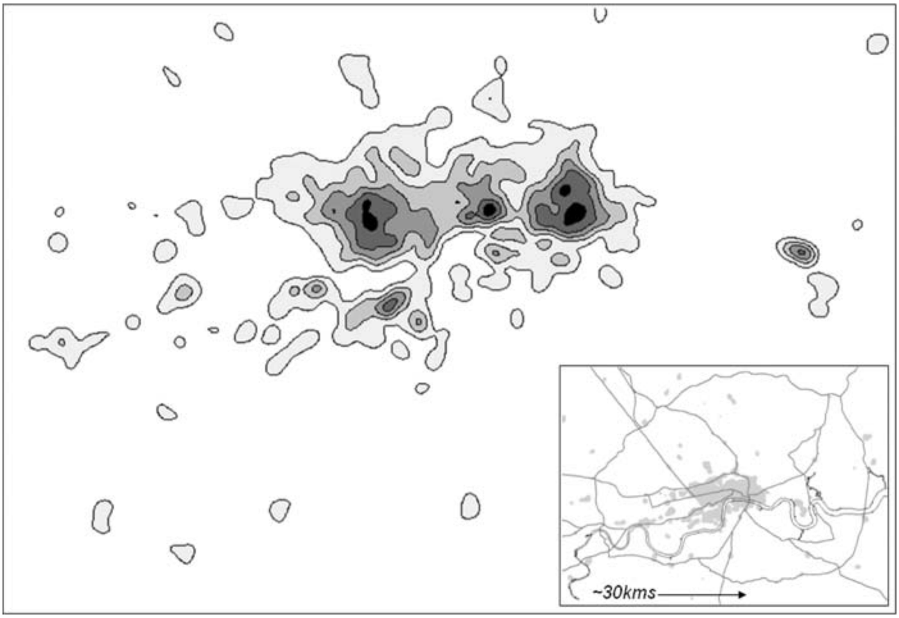


Figure 14: The Implicit Retail Hierarchy in Central London

5. CENTRAL PLACES: RANK SIZE FROM GEOGRAPHICAL DEPENDENCE

Although we have illustrated models which produce quite distinct hierarchies, we have only introduced space as action-at-a-distance from distinct nodes. In so far as competition has entered the argument, this has been either through intersecting and overlapping diffusion or through an implicit ordering where larger places get preferential treatment relative to smaller places, as in the network model of the previous section. One of the first expositions of how geographical areas based on spheres of influence around towns and cities are consistent with the rank-size rule was developed by Beckmann (1958) and his argument is so clear that we will repeat it here, thus providing some sense of closure on our more general discussion of hierarchy through rank-size scaling. Beckmann (1958) defined two key elements in the way cities are organised with respect to their functional and spatial dependence. He first assumed that a city or rather a small seed which sparked off the growth of a city was proportional in size to the population on which it depended in its surrounding hinterland or sphere of influence. He then noted that each city had a 'span of control', which related to the number of lower order hinterlands which could be said to depend spatially and economically on the centre city or seed at its core. This second kind of dependence leads directly to a series of hinterlands at different orders, increasing in number and decreasing in geographical area as they descend the hierarchy and it is from this that the rank of any city can be established.

Formally, the initial dependence ξ of the city seed p_n on its wider population P_n for any order of city *n* is

$$p_n = \xi P_n \tag{9}$$

where the order *n* is from the largest city which we call *N* to the smallest which is defined by the index 1. The second spatial dependence involves the fact that the population of the higher order level P_n is a sum of populations at the next lower order P_{n-1} defined as

$$\begin{aligned} P_n &= p_n + s P_{n-1} \\ &= \frac{s}{1-\xi} P_{n-1} \end{aligned} \tag{10}$$

Recurrence on equation (10) leads to

$$P_n = \left(\frac{s}{1-\xi} \right)^m P_{n-m} \tag{11}$$

and at the bottom of the hierarchy where the population is at the lowest level $P = P_1$, then the exponential dependence within the hierarchy is clear

$$P_n = \left(\frac{s}{1-\xi} \right)^{n-1} P \tag{12}$$

If we assume that the seed city is small or even zero, then $\xi = 0$, and equation (12) simplifies to $P_n = s^{n-1} P$.

Using the reverse order which is from 1 to *N*, the total number of cities at each level is s^m and the total number up to *m* is given as

$$\sigma(m) = 1 + s + s^2 + s^3 + \quad \ldots \quad + s^m \tag{13}$$

where this is a diverging but geometric series whose sum is $(s^m - 1)/(s-1)$. Thus the rank of the first city at level m is $\{[(s^m - 1)/(s-1)] + 1\}$ and the rank of the city which is midway through this order – the average rank for this order – is

$$r(m) = \frac{s^m - 1}{s-1} + \frac{s^m}{2} \tag{14}$$

Examining components of this sum in equation (14), we can assume that $1/s-1$ is small relative to other terms and thus equation (14) can be simplified to

$$r(m) = s^m \left(\frac{1}{s-1} + \frac{1}{2} \right) \tag{15}$$

The rank-size relation is based on population size and rank and if we multiply the relevant equation for population in equation (12) (which we convert from order n to m as $n = N - m + 1$) with the rank in equation (15), we get

$$\begin{aligned} P_{N-m+1}\, r(m) &= \left(\frac{s}{1-\xi} \right)^{N-m} P\, s^m \left(\frac{1}{s-1} + \frac{1}{2} \right) \\ &= \Phi\, (1-\xi)^m \end{aligned} \tag{16}$$

This equation is a constant if $\xi = 0$ and thus the entire argument hinges on this. To an extent this is arbitrary although it is easy to assume that the hinterland population dominates and the core or seed is near zero. If this holds, we can simplify equation (16) as $P_{N-m+1}\, r(m) = \Phi$. Writing this in a more familiar way where we suppress the order indices and define population at a rank *r* as P_r, we get $P_r\, r = \Phi$ or

$$P_r \sim \frac{1}{r} \tag{17}$$

which is the pure Zipf case. Many assumptions have been made to get this far and of course we have not tried to generalise Beckmann's (1958) discrete case to a continuous one. Nevertheless, it would appear that this kind of geographic, indeed geometric reasoning which assumes that space is nested hierarchically through its economic dependence does lead to rank-size distributions of activities such as population. At one level of course, this is all too obvious in that we have assumed hierarchical order and simply shown that geometric series which can describe such order can be manipulated to produce a rank size. This is in fact an indirect argument reflecting scaling.

Surprisingly the Beckmann model has not been widely exploited and no one (as far as I know) has developed a stochastic version of it. Nevertheless, it does serve to remind us that there is a deep underlying rationale for the existence of rank-size distributions which is essentially a spatial or geometric ordering in the geographical sense (Beckmann, 1968).

6. HIERARCHY IN THE DESIGN OF CITIES

The models we have used in this chapter to generate spatial hierarchies whose signature is the scaling of population are essentially stochastic and dynamic although the Beckmann (1958) model of the last section took a more deductive approach but whose dynamics was implicit at best. Yet there are other ways of generating spatial hierarchies. It is possible, for example, to generate such distributions as the outcome of various optimisation procedures, taking either a top-down static approach or even a quasi-dynamic one. By way of conclusion, and in our quest to square the circle and show how hierarchical systems in cities should feature in their design, it is worth noting that there is a long tradition in spatial interaction modelling in which scaling distributions of population and trip/traffic distribution can be derived using optimisation theory – from maximising utility-like or entropy/accessibility functions subject to constraints on the dispersion of such activities through their cost structures. Berry (1964) was one of the first to illustrate such an approach in his derivation of population distributions which conformed to rank size using entropy maximising techniques; and this approach was widely used by Wilson, Coelho, Macgill and Williams (1981) in their quest to embed behavioural land use-transportation models into contexts in which behaviour was considered as optimising either at the individual or collective plan-making level.

In terms of these different approaches, the rank-size distribution provides a sharp illustration of the problems we face in explaining the evolution of complex systems. Nothing could be more different than the generation of a distribution from a stochastic process where all the constituent elements are independent from one another and where the growth is from the bottom up – our first model – and the kind of top down optimisation process in which accessibility is maximised subject to some constraints on cost or energy expended. Yet the outcomes in terms of the distribution of the elements being optimised are the same. In a way, all this shows is that how we approach city systems conditions the techniques we use to generate the outcomes we expect. In terms of the design of cities which although linked to optimisation, originates from very different intellectual mindsets and professional concerns, approaches using ideas from hierarchy theory are also well established. Although many of these use hierarchy in terms of the structure of the problem-solving process where problems are partitioned into a hierarchy of sub-problems, the notion that we need equivalent simplifications to those we have sought here is instructive.

We anticipated much of this in an earlier section where we quoted Alexander (1966) who argued that the notion of strict hierarchy was far too simplistic an organising concept for design. He amongst many others drawing on ideas from organically evolving systems which latterly have been exploited in neo Darwinism by writers such as Dennett and Dawkins, argued for a paradigm in which interaction rather than hierarchy was a required design construct. Overlapping hierarchies – semi-lattices as we illustrated in Figure 9, are much more appropriate vehicles for the organisation of cities into spaces at different levels of geographical scale. In essence, this argument suggests that strict hierarchical subdivision is too simplistic a concept for the design of neighbourhoods and town spaces although it has been widely used by architects operating in top down fashion. Overlapping hierarchies although simplifying interaction capture, the diversity of behaviour and are much more suitable pictorial vehicles for progressing good urban design. In a sense, this argument has also been anticipated in urban systems science; Christaller's central place hierarchies were overlapping while the whole point of spatial interaction modelling and its link to retail centre definition has been to relax the notion of hierarchy, letting it remain implicit in the space of flows.

There are other more direct reasons for thinking of cities as overlapping hierarchies or lattices and this simply emerges from the fact that there are many such hierarchies. We have only examined the simplest here – that based on how the population in an aggregate sense arranges itself but once one disaggregates populations into the multiplicity of categories that define them, and once one adds other kinds of activities which arrange themselves hierarchically in space such as transport and other network systems, land uses, styles of buildings, social friendship nets, and so on, then the idea of overlap becomes the rule not the exception. In fact it is hard to escape from the fact that the best analysis should tackle this notion directly. The fact that most of our analysis tends to simplify the system beyond this obvious reality poses a dilemma. What we require are good, simple and plausible models that show us how different kinds of hierarchies interlock.

In terms of city size distributions, then the challenge seems to be to build on the network characterisations of Gibrat's model, possibly interlocking the network model we illustrated earlier in this chapter with some sort of dual but countervailing network based on friendship patterns rather than the economics of travel which were implicit in the model demonstrated. Interlocking networks which lead to interlocking but consistent and simple scaling of aggregate activities would seem to be the quest. We know that most distributions that we see in cities are scaling or near scaling, and the goal would be to show how these might be unpacked and linked at the network level where we are able to grapple with the diversity that characterises cities. In this way, our understanding of cities would be enriched and this would suggest ways in which we might be able to design our patterns of interaction and location more effectively.

REFERENCES

Alexander, C., 1966, A City is not a Tree, *Architectural Forum*. **122(1)**: 58-61, and **122(2)**: 58-62.

Barabási, A-L., 2002, *Linked : The New Science of Networks*, Perseus Pub, New-York.

Barabási, A-L., and Albert, R., 1999, Emergence of Scaling in Random Networks, *Science*. **286**: 509-512.

Batty, M. and Shiode, N., 2003, Population Growth Dynamics in Cities, Countries and Communication Systems, in: *Advanced Spatial Analysis,* P. A. Longley and M. Batty, eds., ESRI Press, Redlands, CA, pp. 327 –343.

Beckmann, M. J., 1958, City Hierarchies and the Distribution of City Size, *Economic Development and Cultural Change*. **6**: 243-248.

Beckmann, M., 1968, *Location Theory*, Random House, New York.

Berry, B. J. L., 1964, Cities as Systems within Systems of Cities, *Papers and Proceedings of the Regional Science Association*. **13**: 147-163.

Blank, A., and Solomon, S., 2000, Power Laws in Cities Population, Financial Markets and Internet Sites: Scaling and Systems with a Variable Number of Components, *Physica A*. **287**: 279-288.

Gabaix, X., 1999, Zipf's Law for Cities: An Explanation, *Quarterly Journal of Economics*. **114**: 739-767.

Gibrat, R., 1931, *Les Inégalités Économiques*, Sirey, Paris.

Guerin-Pace, F., 1995, Rank-size Distribution and the Process of Urban Growth, *Urban Studies*. **32**: 551-562.

Manrubia, S. C., and Zanette, D. H., 1998, Intermittency Model for Urban Development, *Physical Review E*. **58**: 295-302.

Pumain, D., 2005, Alternative Explanations of Hierarchical Differentiation in Urban Systems, in D. Pumain, ed., *Hierarchy in Natural and Social Sciences*, Kluwer Academic Publishers, Dordrecht, The Netherlands, forthcoming.

Pumain, D., 2000, Settlement Systems in the Evolution, *Geografiksa Annaler*. **82B**: 73-87.

Simon, H. A., 1955, On a Class of Skew Distribution Functions, *Biometrika*, **42**: 425-440.

Simon, H. A., 1977, *Models of Discovery and Other Topics in the Methods of Science*, D. Reidel Publishing Company, Dordrecht Holland, Boston, MA.

Sornette, D., 2000, *Critical Phenomena In Natural Sciences: Chaos, Fractals, Self-Organization and Disorder*, Springer-Verlag, Berlin.

Sornette, D., and Cont, R., 1997, Convergent Multiplicative Processes Repelled from Zero: Power Laws and Truncated Power Laws, *Journal de Physique I (France)*. **7**: 431-444.

Thurstain-Goodwin, M. and Batty, M., 2001, The Sustainable Town Centre, in: *Planning for a Sustainable Future,* A. Layard, S. Davoudi, and S. Batty, eds., Routledge/Spon, London, pp. 253-268.

Watts, D. J., 2003, *Six Degrees: The Science of a Connected Age*, W. W. Norton and Company, New York.

Wilson, A.G., Coelho, J. D., Macgill. S. M., and Williams, H. C. W. L., 1981, *Optimisation in Locational and Transport Analysis*, John Wiley and Sons Ltd, Chichester, UK.

Zipf, G. K., 1949, *Human Behavior and The Principle of Least Effort*, Addison-Wesley, Cambridge, MA.

Chapter 7

ALTERNATIVE EXPLANATIONS OF HIERARCHICAL DIFFERENTIATION IN URBAN SYSTEMS

Denise PUMAIN
Professor,
Institut Universitaire de France, University Paris I

Cities[1] 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

[1] The word « city » is used throughout in the general acceptation of town or city (urban area including its suburbs).

D. Pumain (ed.), Hierarchy in Natural and Social Sciences, 169-222.

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 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 1a and b). 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.

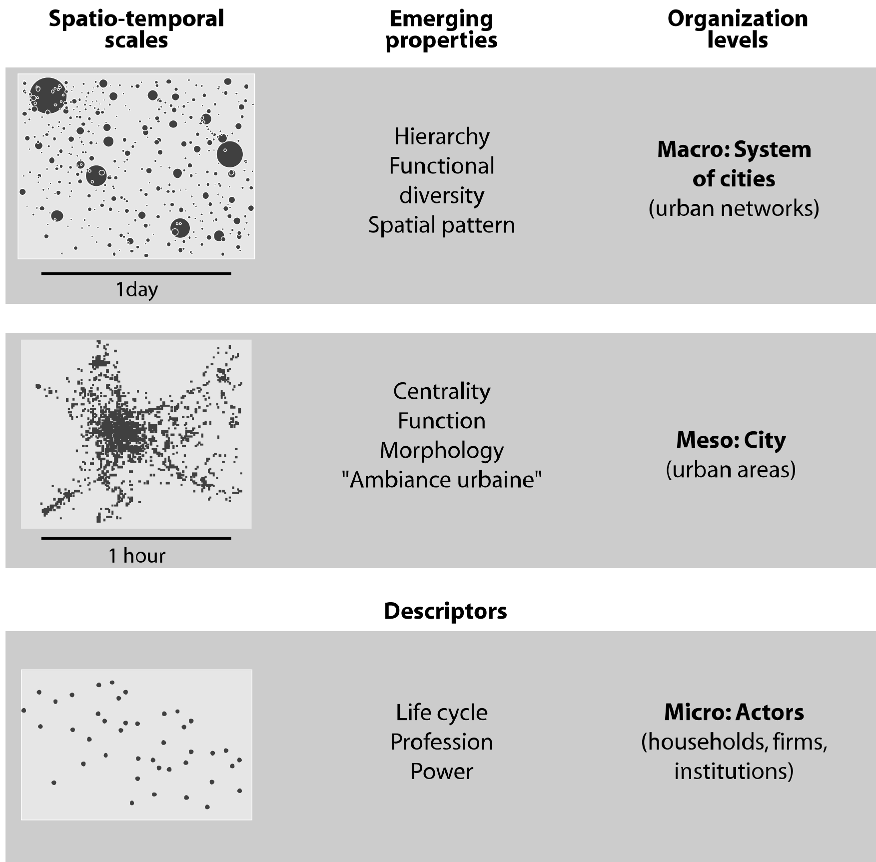


Figure 1a: Scale and urban systems - Emerging structural properties

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).

Figure 1 : Scale and urban systems

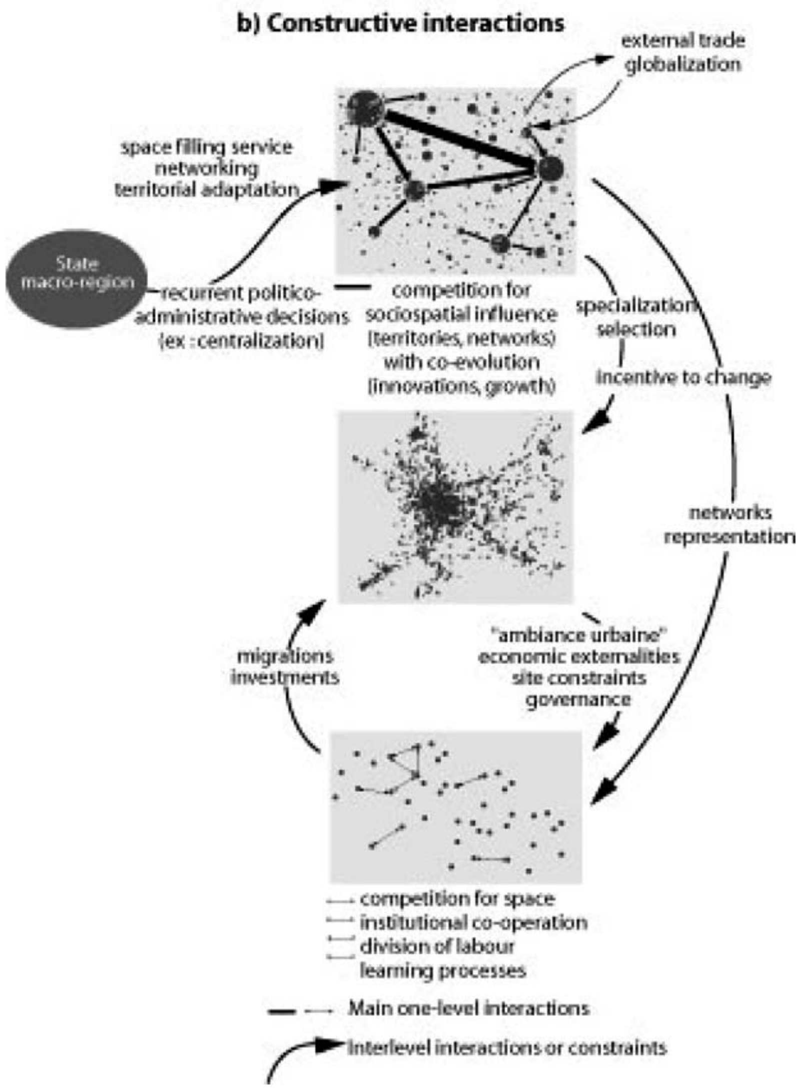
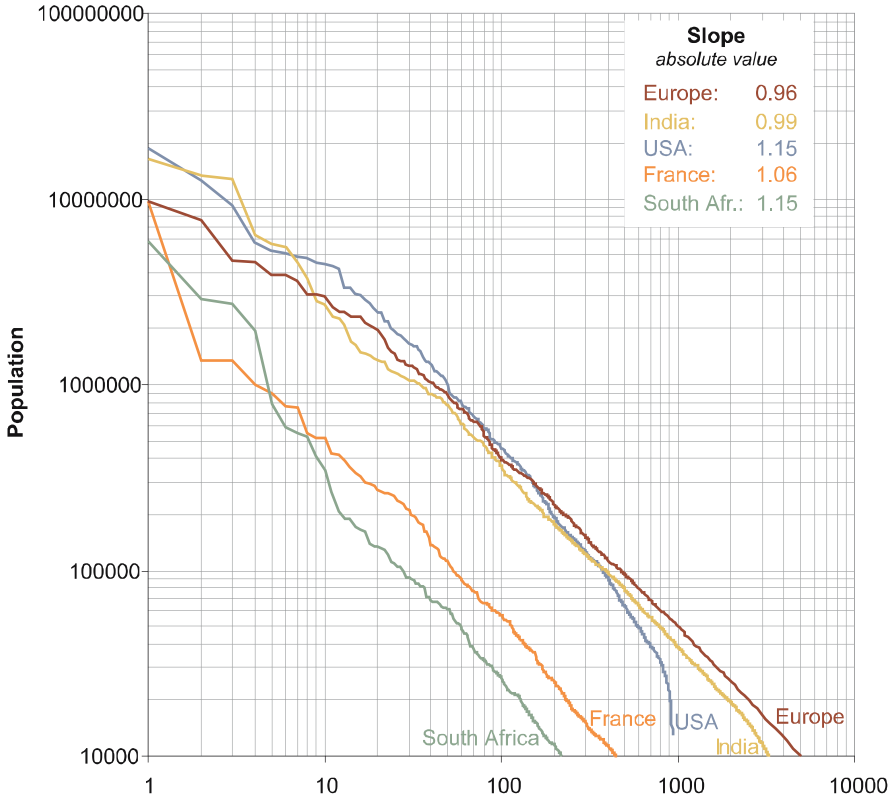


Figure 1b: Scale and urban systems - Constructive interactions

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.



Sources : Europe : Moriconi-Ebrard F., 1994, GEOPOLIS / India : Census of India 2001 / USA : United States Census 2000 / France : INSEE, Recensement de la Population 1999 / South Africa : Statistics South Africa, Census 2001, Base CVM

Figure 2: Hierarchical differentiation in city sizes

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 or 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 21th 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.

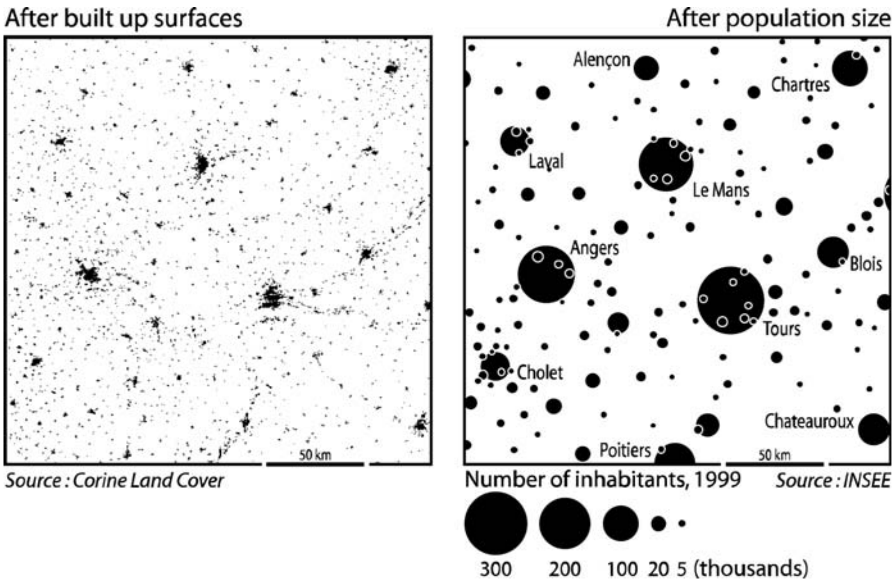


Figure 3: Spatial patterns of urban hierarchies

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 acceptation 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.

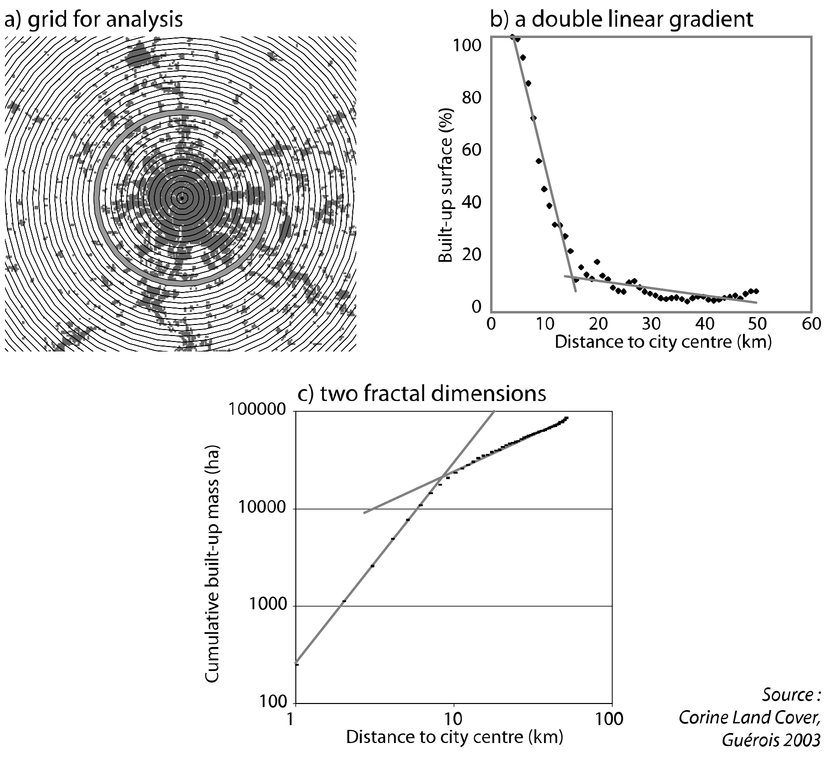


Figure 4: The urban field and the delimitation of urban entities

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).

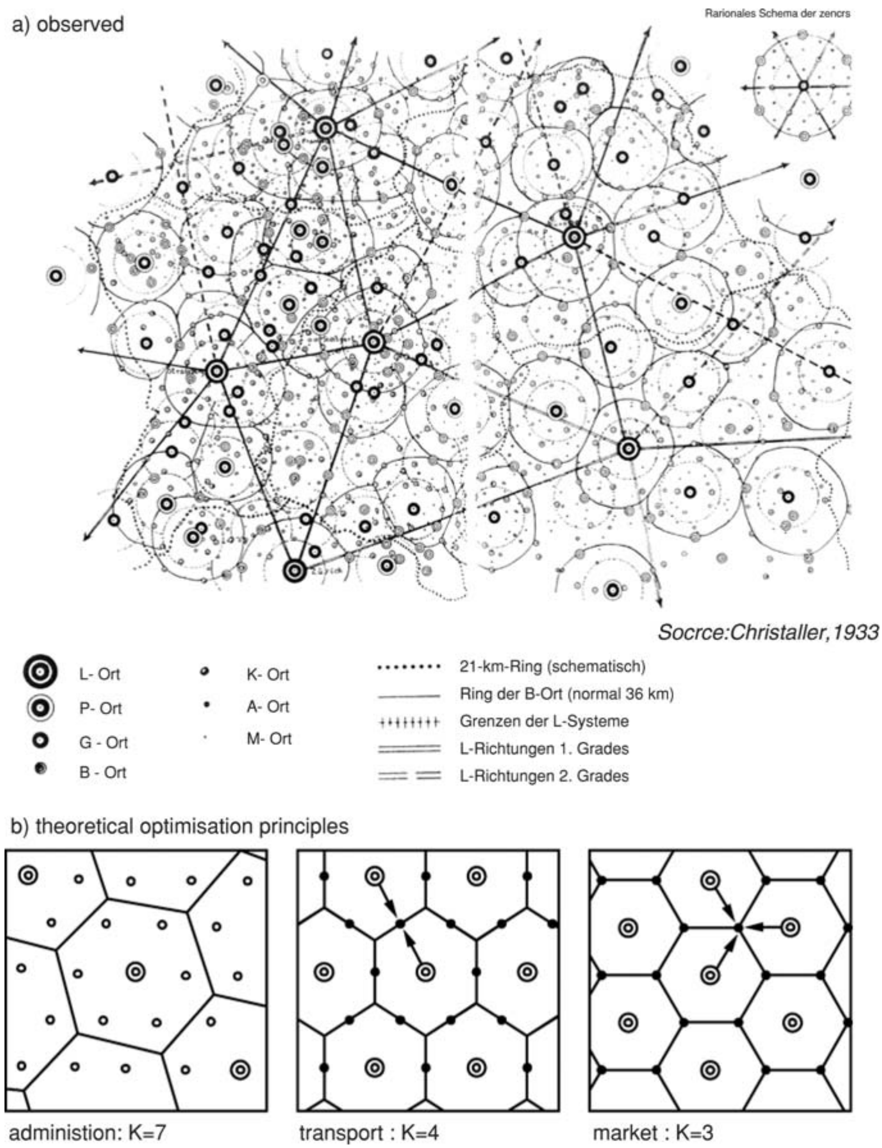


Figure 5: Urban hierarchies after W. Christaller

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.

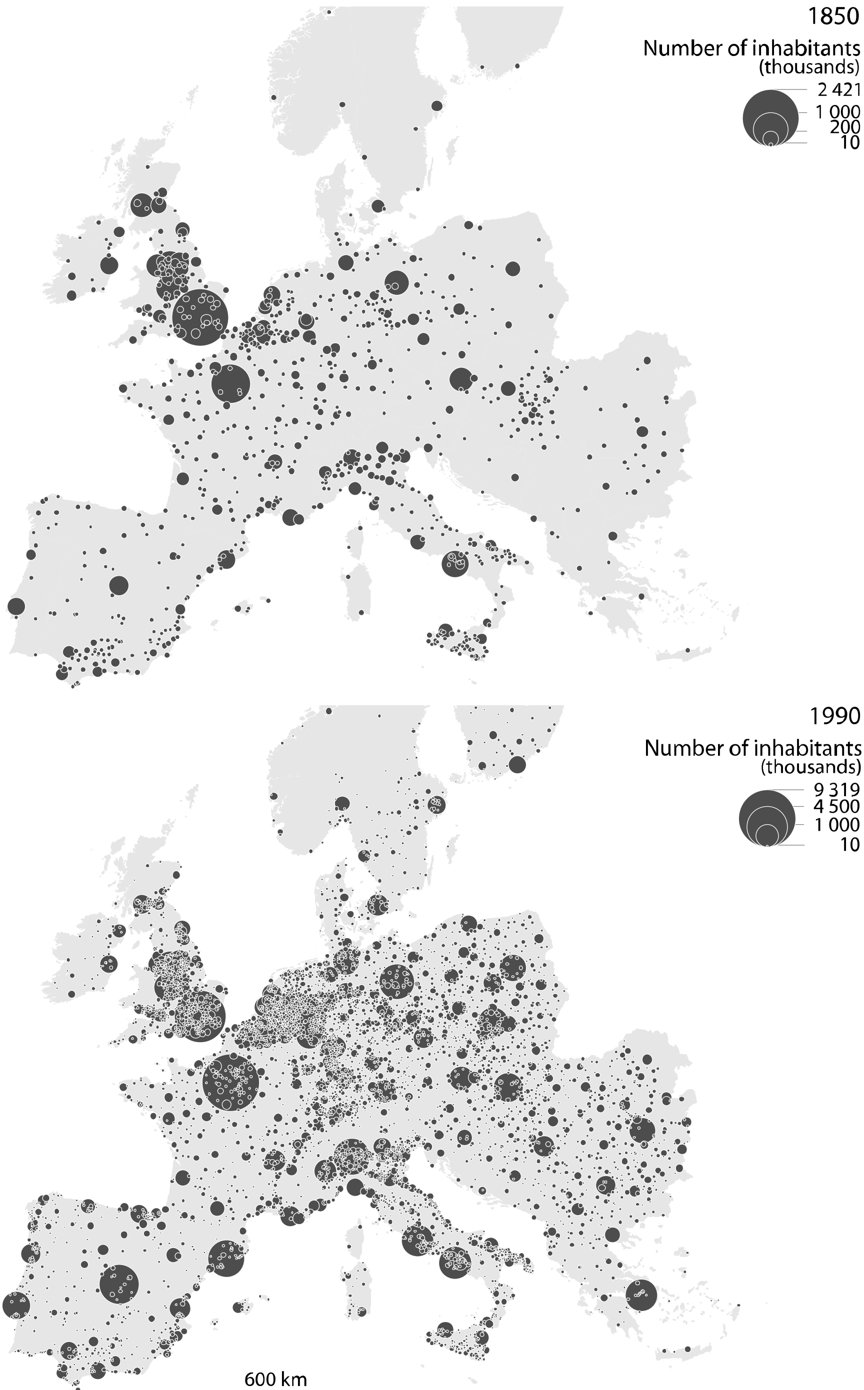


Figure 6: Distributed Urban Growth in the European Urban System

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.

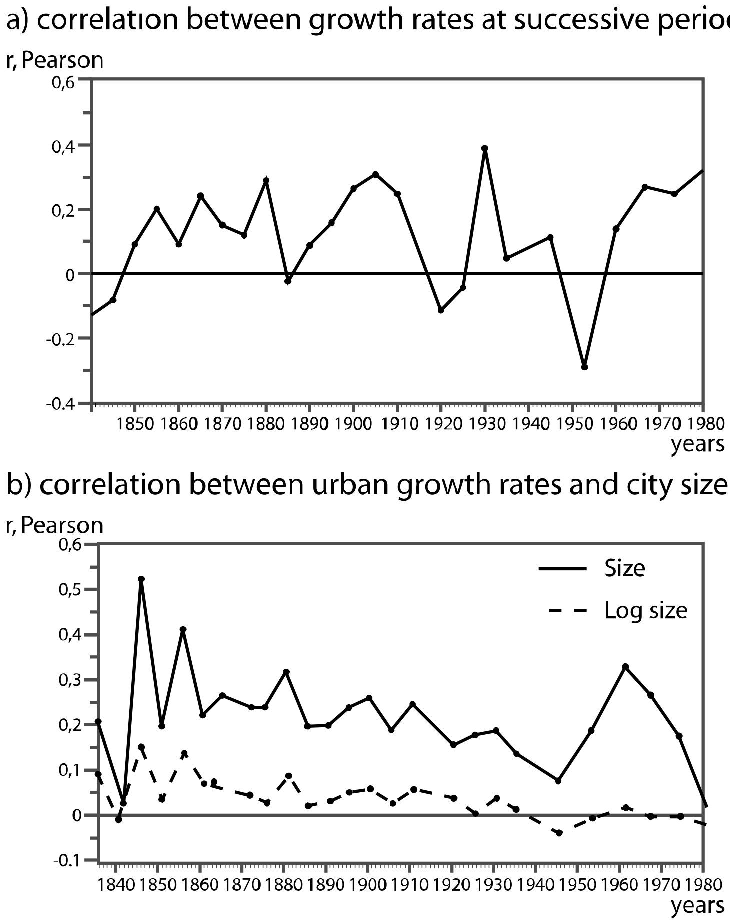
In this process, the variance of urban growth rates, as well as the almost random redistribution of growth among cities at each period of time, require explanation. The fluctuations observed in the growth rates have been related to the process of adaptation to urban change, which is not automatic or 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).



Source : Guérin-Pace, Pumain, 1990,
"150 ans de croissance urbaine en France", Economie et statistique

Figure 7: Two deviations from Gibrat's model

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.

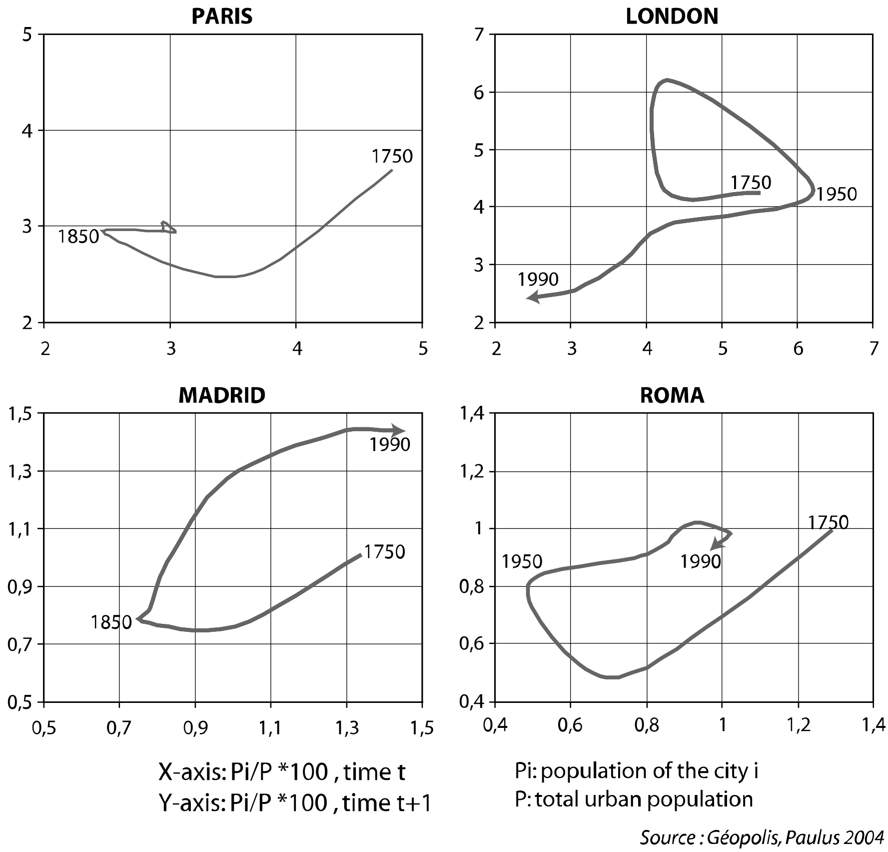


Figure 8: Trajectories of cities in the European urban system

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 eighteen 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.

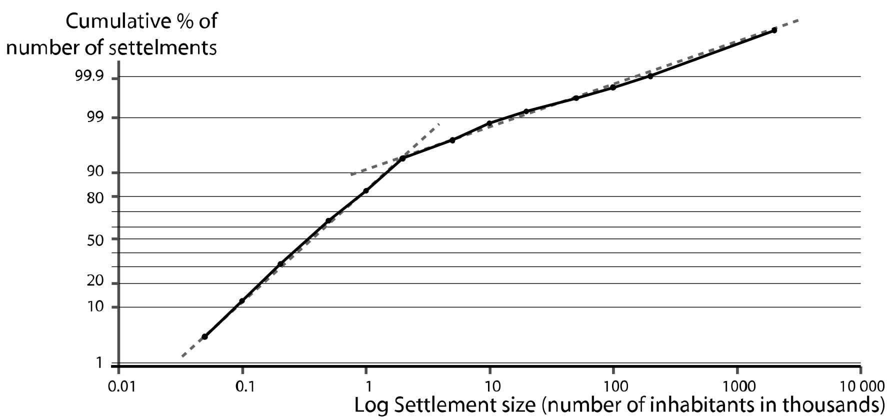


Figure 9: Adjustment of human settlement sizes by lognormal distributions

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 mico-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 experimented 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.

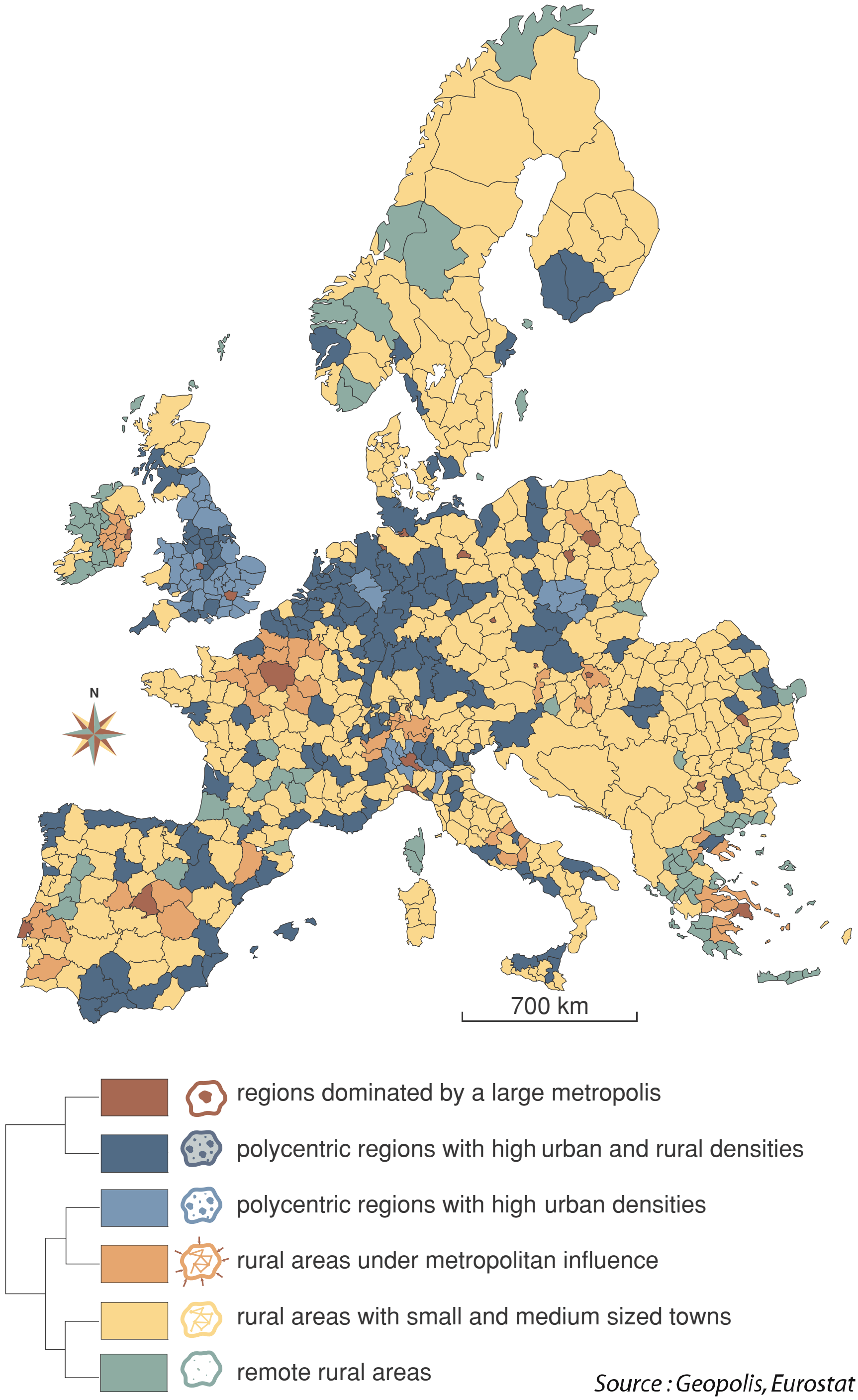


Figure 10: Settlement styles in Europe

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 attractivity, 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 Zipf'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 Sceince 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.

de Vries, J., 1984, *European urbanization,* 1500-1800, Methuen, London.

Dendrinos, D., and Mullaly, H., 1985, *Urban evolution*. Oxford University Press, Oxford.

Derycke, P.-H., Huriot, 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.

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.

Jefferson, M., 1939, The law of primate city, *Geographical Review*. **29**: 226-232.

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).

Lösch, A., 1940, *Die Räumliche Ordnung der Wirtschaft*, Fischer, Iena.

Lotka, A.J., 1941, The law of urban concentration, *Science.* **94**: 164.

Lotka, A.J., 1924, *Elements of physical biology*. Baltimore.

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, M., Parisel, C., Pumain, D., and Sanders, L., 2001, Knowledge-based simulation of settlement systems, *Computers, Environment and Urban Systems*. **25** (2): 167-193.

Page, S.E., 1998, On the emergence of cities. *Santa Fe Institute, Research in Economics*, 98-08-075a.

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 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 Moriconi-Ebrard, F., 1997, City Size distributions and metropolisation, *Geojournal*. **43** (4): 307-314.

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., and Winiwarter, P., 1985, Aggregation of independent Paretian random variables, *Advances in Applied Probability.* **17**: 465-469.

Roehner, B.M., 1991, The long-term trend toward increased dispersion in the distribution of city sizes, *Environment and Planning*. **23**: 1725-1740.

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 synergétique,* Anthropos, Paris.

Sassen, S., 1991, *The Global city: New York, London, Tokyo*. Princeton University Press.

Simon, H., 1955, On a class of Skew distributions, *Biometrika*. **42**: 425-440.

Singer, H.W., 1936, The "courbe des populations": a parallel to Pareto's law, *Economic Journal*. **46**: 254-263.

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*. Hofnur, 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.

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.).

CONCLUSION

Denise PUMAIN
Professor,
Institut Universitaire de France, University Paris I

This panorama of hierarchical organisation in social and natural sciences was intended to improve our understanding of universally emerging hierarchical organisations in nature and society. We expected new insights from the analysis of the scope for circulating concepts and methods between various disciplines. We were searching for a possible general explanation for hierarchical structures. We have reviewed a whole set of quantitative and qualitative approaches, including measures and analytical tools, which provide deeper knowledge about general and specific processes generating or maintaining hierarchies. But this theoretical and methodological investigation has also led us to reverse the question, through the discovery of a possible explanatory power of hierarchical structures themselves, as a necessary part of the architecture of complex systems. Looking for hierarchical organisation would thus become an essential methodological step in the description and understanding of complex systems.

HIERARCHY AS A CONCEPT RELATED TO LEGITIMACY

A first general explanation for the universal appearance of hierarchies in the social world is their connection with power and legitimacy. A review of the usage of the word hierarchy in dictionaries and scientific literature up to the end of 19th century by Nicolas Verdier demonstrates the long-standing strong link of this usage with the religious institutions. Reference to a "natural" order of social standing established upon divine right is common to several cultures. According to Max Weber, there were three sources for the legitimacy of political power in the history of societies, firstly reference to the sacred, secondly charismatic influence, and thirdly legal and rational standards of efficiency. Social hierarchy expresses a social order by reference to a collective system of beliefs that legitimises these sources of power. Hierarchy is still today a recognised procedural method for establishing decisions that are considered as right and legitimate. For

D. Pumain (ed.), Hierarchy in Natural and Social Sciences, 223-239.

instance, this methodological device is encountered in the pyramidal organisation of legitimate sources in Muslim law, from *Coran*, then *sunna*, to *ijmsa* (or consensus) and *qiyas* (analogy). In the Roman legal system, there is also a hierarchy of standards, each rule at the lower level having to be compatible with the upper levels (for instance, from a constitution to laws, rules and decrees in French, Spanish, Greek or Italian legal systems today). The consistency between the general order and the particular cases is governed by a set of nested rules subordinated by their succession in a hierarchical order. (There are however other legal systems, like "common law", where the hierarchical principle is not present to the same extent).

Since it is conventional in upholding the legitimacy of social rules, hierarchy could appear as a mere matter of social representation. Some archaeologists claim that the process of « verticalisation » of the mental image of the world, whereby the human species thinks of itself as dominant, occurred before the Neolithic revolution, at the time of animal domestication (around 10 000 BC), in the Middle East and the Euphrates valley. The invention of domestication could denote an alteration in mindset, placing humankind above the other components of nature, with which it previously entertained horizontal relationships. This new consciousness of domination, in the form of a vertical hierarchy, is sometimes considered as a cognitive revolution that could have preceded technological change, and could have been a necessary condition for its emergence. However, observers of animal behaviour have noticed that hierarchies, demonstrating strict or partial order, exist within a number of animal societies, in a broad variety of forms. Hierarchy is sometimes thought to confer an adaptive advantage, as it provides a non-violent solution to conflicting situations, even if in most cases it is established after a fight.

Nicolas Verdier analyses the transition between the theological acceptation of a strict ranking order and the more multidimensional social usage of the word hierarchy, which only appeared during the 19th century, after the disappearance of the Ancient Regime. Quoting Voltaire: "different hierarchical ranks are strictly incommensurable", he also emphasises the increasing neutralisation of the word accompanying the growing social rejection of hierarchical structures (for instance with Tocqueville assimilating hierarchy to despotism). This evolution is parallel to the transition of societies from agrarian economies towards more complex modes of production that cannot be represented by the simple distinction between peasants, priests and soldiers. Both the recognition that social orderings are many and various, and the general suspicion attached to the hierarchical social structures in democratic systems, could explain why contemporary sociologists do not often refer to social hierarchies (for instance social hierarchy is merely a matter of stratification for Parsons), even if the principle remains very often as an implicit or commonplace fact in their analyses (see below). This "historical-genetic derivation" (according to the terminology for the "styles of scientific thinking" developed by A.C. Crombie, 1994) of the successive meanings of the word hierarchy by Nicolas Verdier shows us that great caution is required when looking at the social historical significance of the word. We cannot assume that the word retains the same connotations over time, even if the early meaning still probably contaminates its contemporary usage.

FUNCTIONAL SOCIAL EXPLANATIONS OF HIERARCHY

Even if it is open to question that hierarchy mainly reflects the organisation of our minds, it is true that it does emerge in a large variety of social situations. We have not devoted a chapter to the presentation of social hierarchies. Another book would be probably necessary for this purpose, since the question of social hierarchy is embedded in the various possible definitions of social status, social practices and social institutions to such extent that the work of sociology overall would have to be envisaged to describe how the concept of hierarchy operates in social contexts. New theories of the hierarchical organisation of society, disconnected from the previous legitimate order based on divine right, were conceived at the time when the democratic regimes and the political power of lay society were emerging in the western world. Among these theories that are intended to explain social order in industrialised societies three main types can be recalled: 1) the notion of social class, perhaps invented by Turgot, and based by Marx upon the appropriation of production means, as reflecting the domination by capitalists over proletarians, although the class of landowners was not easy to position in a strict order in relation to these two main classes; 2) a functional model was conceived by Kingsley Davis and Wilbert Moore to explain social hierarchies by the functional importance of jobs in the division of labour, according to the duration of job training and labour availability in the relevant employments; 3) the market model, as suggested by A. Smith, based upon the same principles, but it postulates, instead of a necessary regulation process reproducing the division of labour across generations, an equilibrium between supply and demand in each type of job, according to multiple criteria. W.Pareto's theory of the circulation of elites belongs to the same liberal model.

In a systemic view, social hierarchies are the product of social interaction, which also contributes to maintaining them. They rely upon a two-way circulation of information. The social order conveying authority, power and control follows a top-down line, while admiration, respect and obedience move from bottom up. However, especially in modern societies where the division of labour has become more and more complex social hierarchies never follow a simple ordering, they use multiple dimensions which are often highly correlated but not fully redundant, like prestige, power, income, material and symbolic gratifications (capital, cultural capital, social capital in contemporary terms). Even in traditional social orders, for which G. Dumézil was able to develop a three-function theory (the priests, the soldiers and the peasants, with uncertainty about the ranking of merchants) from a comparison of all Indo-European societies, it was noticed that the social hierarchy could be more or less strict, as Louis Dumont recognised by contrasting the marked holistic hierarchy of castes in Indian society (as described in his book *Homo hierarchicus)*, with the *homo equalis* of western societies where individualism is preponderant. After considering that society is made up of different groups separated by blurred demarcation lines and organised into a hierarchical order, contemporary sociology thinks in terms of social networks. Social relatioships between actors always involve asymmetrical shares of power, but they also require a reciprocal transfer of resources and implication on the part of the actors. Instead of being embedded in a nested order, the networks that these relationships generate have many intersections, and they are not all of the hierarchical type, but display various models of possible spatial interaction. As hierarchy is a particular kind of network, we can recall briefly the methods that are provided by network analysis.

HIERARCHY AND NETWORKS

The graph theory was for a long time the most widely used tool for the analysis of networks. A graph is a simplified representation of a network in which the nodes are "vertices" and the links are "edges". It is analogous to an interaction matrix. A graph without cycles is called a tree, and can represent any hierarchical organisation such as a river system or a pyramid of levels of responsibility in an army or a firm. This pure form is rarely observed in the social world. Graphs of social relationships exhibit many cycles, which are transitive relations from node a to b then c and back to a. It is these closed loops, which may involve more than three nodes, that are called cycles. The global connectivity of the graph is measured by different ratios comparing the number of nodes to the number of possible edges, and connectivity increases with the number of cycles in the graph. In classic applications of graph theory, various indices were also used to identify relative positions of nodes in the graph, in terms of relative accessibility to the other nodes. These measures of centrality are influenced by the number of direct paths that join one node to others. B. Gaume, F. Venant and B.

Victorri (chapter 5) suggest more effective methods for analysing relative positions of this type, applied to a linguistic space defined by relationships of proximity between words (French verbs in their case) in terms of meaning. They resort to new means of analysis that were developed for social networks. Social networks are more complex than simple tree-like networks, because they include many cycles, but they nevertheless very often exhibit certain hierarchical features, since accessibility is not equally distributed within the network, in contrast to regular lattices. Nor is it distributed randomly around a mean, as in the classic Erdös-Renyi model, but on the contrary has a very large variance, a few nodes having many connections while many others are poorly connected. When the number of connections per node is distributed according to a Pareto law, the network is called "scale free". This denomination refers to the absence of any significant average in the distribution, and thus absence of any characteristic scale. In social terms, this means that the centrality (or "betweenness") is very unequally distributed within the network. Scale-free networks can reveal a "spontaneous" hierarchical ordering in society. For example, western societies and their tentacles in the era of globalisation have recently made a wide-scale experiment in creating a large network with the diffusion of Internet. Many dreamed of an equal, ubiquitous access to this new medium for circulation of information, often presented as the most democratic tool ever invented. However, analytical studies of the structure of the network have revealed that it is very unequal in terms of numbers of connections available to each node: the structure is hierarchical and similar to the model for "scale-free networks" constructed by Barabási and Albert (1998).

B. Gaume, F. Venant and B. Victorri (chapter 5) demonstrate that semantics in natural language is organised in the same way as social networks. Some words have many close synonyms, while other more precise terms are isolated. Each word can be characterised by its number of connections with others and this distribution is highly hierarchical. In fact the structure of the graph of semantic relationships is similar to networks that were identified as "small worlds" (Watts and Strogatz, 1998) because, besides their hierarchical structure (where the distribution of the degree of the nodes follows a power law, according to a scale-free topology), they also present marked clustering (many cycles in the graph) which is paradoxically combined with a rather low value for the diameter of the graph (the diameter is the length of the maximum topological separation between any pair of nodes). The diameter of a graph gives an idea of the efficiency of communication between its parts. For a network of N nodes, the diameter is about log N in the case of a small world, (that is, 6 for one million of nodes) while without that modular organisation the diameter is around the square root of N (that is, 1000 for one million nodes). Marked clustering and hierarchical organisation thus facilitate the circulation of information within this type of network, which is frequently observed in social contexts. Efficiency in conveying information has long been the main rationale for explaining the prevalence of hierarchical organisation in human activities.

Instead of restricting themselves to enumerating the degree of a node as a measure of its centrality, or hierarchical position, within the network, B. Gaume, F. Venant and B. Victorri suggest two complementary tools: the "global k proxemy" which measures the centrality of a node after its position within regions that are more or less densely linked in the graph (in the neighbourhood of dimension k), and a measure based on the number of "cliques" (or cycles comprising various numbers of nodes) which provides similar results. These methods could be applied to other types of networks, as they identify a complex hierarchy among the nodes by classifying them at different levels, according to local measurements of centrality.

Concepts from graph theory or analytical tools linked with the notion of small worlds and scale-free networks are however not entirely satisfactory for describing social networks, since they miss the fact that they are very often oriented networks, with an asymmetrical signification of the relationship. Dynamic modelling of oriented graph structures is most often provided through simulation tools, which have become more and more powerful in the last decades. This is a shift from social explanation to the statistical simulation of hierarchies.

HIERARCHICAL ORGANISATION OR HIERARCHICAL DIFFERENTIATION

We are now entering a domain where a more formal description of systems, involving measurement and enumeration, will enable the application of methods that are transversal to a large number of disciplines. We therefore require common definitions. What is a hierarchical system? Before discussing statistical approaches to hierarchies, a distinction needs to be made between two possible acceptations of what hierarchical form in a system is. A hierarchy can be conceived as an ordered succession of distinct levels, that are more or less clearly separated, but that can be considered separately since the processes that are involved in the construction of each level are very often different from one level to the next. We suggest using the term *hierarchical organisation,* when different and more or less autonomous entities can be observed at different levels of observation, and when each level needs to be described with different attributes because new properties emerge at each level of organisation. Hierarchical organisation can concern *inclusive hierarchies*, as in the biological domain, where each level is embedded in the next, proteins in cells, cells in organs, organs in organisms, and so on. However, even in biology for the definition of species or ecological systems, and *a fortiori* in the social world, most hierarchical organisations are *heterarchies*, where the levels are less easy to identify and to separate: defining and delimitating a group, a social network, a class, a culture or a civilisation is neither simple nor obvious, even if different levels can be recognised and hierarchised according to the more or less broad generality, magnitude or scope within which they are operating.

The second acceptation is the view that hierarchy in a system can also be described as a continuum of differences in size, where the elements in the system are strongly differentiated, even if they retain the same appellation and the same collection of attributes according to their ontological definition. Many examples are to be found in astronomy (size of stars or galaxies) or in the social world, as for example in the case of urbanism: we still use the same word "cities" to refer to groups of resident population that are different by several orders of magnitude, since they range from a few thousand to tens of millions of inhabitants, and the weight of their economies ranges from the turnover of a very small artisan firm to gross urban products that are equivalent to those of powerful nation states (to give a few examples: it is estimated that the economy of the Tokyo urban area is equivalent in size to the gross product of France, the economy of New York weighs as much as China's economy overall, while Paris produces as much as the whole of the Netherlands). Another example of a continuous distribution of sizes including very large differences in terms of economic power and scope of activity is observed in the case of firms, when they are ranked according to numbers of employees or turnover. We shall use the expression *hierarchical differentiation* when referring to this type of hierarchical feature.

When applied to cities or firms, the expression *hierarchy of size* does not mean that there is any relationship of subordination, direct or indirect, between the smallest and the largest elements in the series, but that their capacity for action or their weight in a social universe of competence, decision and consequence, are of very different magnitude (Pumain, 2003). Because of these inequalities in potentialities and power, the former connotation of domination is however never very far, and the expression also often refers in an indirect way to the first acceptation of social hierarchies. Whatever these social connotations, this kind of hierarchical differentiation in the size of subsystems is always reflected in a specific type of statistical distribution, known as "highly skewed" or "long tail" distribution. Many models have been suggested for describing hierarchies of size, from the various "types" of distribution that Pareto applied to the distribution of income, to the applications of lognormal distribution to firms and cities by Kapteyn and Gibrat, as well as the so-called "Zipf's law". This is very similar to a Pareto distribution, but usually figured in a simplified way on a log-log plot of size of a subsystem against its rank (the rank is equivalent to the cumulated number of subsystems that are of larger size). What is of interest in relation to these statistical models is not so much to determine which is best suited to specific observations (since the exact measurement of size in a social universe is a delicate exercise, always including significant margins of error or uncertainties in delimitation of the subsystems) but to understand how these types of distributions are generated and maintained over time. The statistical model is not in itself an explanation, but gives incentive to search for plausible and meaningful processes that may be behind its general emergence. There is also a further plausible hypothesis which is that the generative processes, from the bottom to the top of the hierarchy, may be the same in situations that are found in natural as well as in social sciences.

The analysis of hierarchical differentiation in terms of power laws is based on a simplified formal representation of complex systems, as being observable at three levels of organisation at least: the Pareto distribution of subsystem sizes is a characteristic property of the system at the macro level; This system can be broken down into subsystems (forming the meso-level) whose size is measured according to the number of elements (forming the micro-level). The identification and specification of subsystems is however a very delicate task in the realm of social sciences at least. Can we consider that a social network, or a firm, or a city, belong to the same kind of entity irrespective of size? Is it right to consider them as comparable entities whose differences can be summarised by their inequalities in size? The question remains open. On one hand, if this postulate is accepted, it is possible to gain insight into the generative process of hierarchical differentiation of this type by reviewing all the dynamic models that have been suggested for explaining highly skewed distributions, and to suggest a simple unifying statistical theory covering a large variety of systems. On the other hand, we shall see below that different rationales can be envisaged to explain hierarchical differentiation, by looking at the qualitative interaction between the elements that compose the subsystems. This will lead us to consider the possibility of using scaling laws to unify perspectives on hierarchical differentiation and hierarchical organisation in complex systems (see below).

Statistical explanation of Zipf's Law using random growth processes

Since Zipf, who suggested that his "rank-size" rule as applied to city sizes reflected an equilibrium between two forces, one of spatial concentration (of economic activities around markets) and one of dispersion (of raw material and resources), progress has been made by shifting from these static and intuitive explanations towards dynamic models that explain both the shape of the statistical distribution and its persistency over time. Even before Zipf, simple statistical growth processes were suggested as being able to generate "long tail" distributions of this type (for instance by Yule in 1924). The model proposed by the French statistician and economist Gibrat in 1931 to explain inequalities in economics establishes a clear connection, derived from the "law of large numbers" or central limit theorem, between a stochastic process of growth and the resulting lognormal distribution of sizes. The advantage of such a model is that its hypotheses can be tested onempirical examples. The main hypotheses are that in a short time interval, additional growth is proportional to the initial size (which is equivalent to saying that growth rates are on average the same), and that the fluctuations around the mean growth rate are independent of size, and independent from one period of time to the next. These hypotheses were tested and roughly validated on long-term series of urban populations by geographers like B. Robson (1973), or Pumain (1982) and Guérin-Pace (1995). Gibrat's model was recently rediscovered as a possible explanatory tool for Zipf or Pareto distributions by economists like Gabaix and Ioannides (2002). Its hypotheses have been used in a variety of simulation models.

However, even if simulations based upon random processes are able to reproduce hierarchical structures that have similar properties to the observed properties, it is very unlikely that such an explanation can be considered as final. First, it would mean accepting a purely statistical description, similar to the "empirical models" that R. Franck (2002) criticises for their lack of generality. Second, even when they are reified as "theoretical models", because they are formal, very general and transposable, the stochastic models that generate hierarchies through random growth processes are mainly operational models. They allow some predictions through projections. They provide a dynamic interpretation of the statistical shape of the hierarchies. But they neglect an essential feature of the emergence and maintenance of hierarchies in social systems, which is social interaction. Where the task is to summarize thousands of interactions using the "law of large numbers", the many independent causes that lead to differential growth, in social entities like firms or geographical entities like cities and territories, social sciences (and those who use results to understand and take action) need to know what relationships, asymmetries, or regulations underpin what could appear at a global level to be produced randomly.

CELLULAR AUTOMATA FOR SPATIAL SIMULATION

Cellular automata are regular lattices that are used in simulation models where various types of neighbouring effects or spatial interaction can be introduced, in the form of rules that interfere with the evolution of each cell (cells representing localised objects or agents). M. Batty (chapter 6) applies a deductive method for a bottom-up construction of urban systems, using cellular automata for the generation of the spatial distribution of more or less densely populated settlements. He proposes a stepwise construction of more and more realistic distributions, starting with a simple stochastic model of population growth, such as Gibrat's model generating lognormal distributions of population sizes. He demonstrates that if a growth process following a random walk with a reflecting barrier can generate a lognormal or Zipf distribution, this model produces inconsistencies when compared to observed trajectories of individual cities: changes in population size are slower and there is more inertia in the real world. Unlike the economic model suggested by Gabaix, which uses the agglomeration economies principle to justify an application of Gibrats' model, M. Batty proves that introducing spatial interaction is a necessary condition for a plausible generative process for urban settlements: "there is a deep underlying rationale for the existence of rank-size distribution which is essentially a spatial or geometric ordering in the geographical sense". A positive spatial auto-correlation in growth rates (representing a local diffusion model), that links the growth in a cell with that of its neighbour cells, provides better results in the simulation. This method is analogous with the principle of "preferential attachment" that is suggested by Barabási for generating "scale-free networks" (Barabási and Albert, 1998). It is also allied to former explanations of urban hierarchies that interpreted them in terms of "central" urban functions, linking the size of a city's population to the number and diversity of services it provides for a surrounding area of varying magnitude. Far from the rigid geometries first imagined by W. Christaller, the "inventor" of central place theory, M. Batty suggests that overlapping hierarchies, corresponding to various types of interlocking networks connecting different urban activities, could give a better model for a comprehensive explanation of urban systems. He also wonders how the knowledge about possible generative processes of real-world hierarchies could be used in design and planning of more efficient or sustainable ways of locating activities and their possible interactions.

INCLUSIVE HIERARCHIES IN BIOLOGY

Possible rationales for embedded or inclusive hierarchical organisations can be suggested by analysing biological entities. Alain Pavé (chapter 2) follows two main lines of inquiry, in an attempt to explain our representation of hierarchically embedded systems and of a time-organised phylogeny of species. He sets out the definition of a variety of hierarchical levels that can be distinguished, from sub-cellular systems to ecosystems, identifying characteristic times and scales for each of them, as well as interaction processes that are essential in their generation. He emphasises that co-operative mechanisms should be considered to be as important as competitive mechanisms to explain "the global biological organisation from genes to ecosystems and biosphere" as well as "the actual embedded hierarchy of organised entities and trophic networks". There is also a mix of stochastic and deterministic processes in the explanation of the hierarchical organisation. Reviewing a series of methods that can help in understanding these complex systems, he shows that computer-based models may be better suited for representing interactions and visualising structures than mathematical models, but to date there are too few theoretical results to validate those approaches. He therefore suggests an integrative modelling method for linking models of elementary processes involving well-known mathematical relationships to hierarchical levels of higher complexity. Thus better management tools adapted to deal with complexity could be derived through modelling and simulation.

Thus far, we are left with our initial question regarding the necessary existence of hierarchical organisations. In the biological world as well as in social situations, there is a general mystery: independent decisions unknowingly construct systems that could have being designed by engineers. Hierarchical organisations seem to appear as almost inevitable emergent phenomena in many natural and social contexts, but we do not know if they are occurring merely by chance, being configurations that have a high probability of occurring, (i.e. as a stable attractor in a very general or commonplace dynamic model), or if such configurations everywhere correspond to the optimisation of some principle – and in any event, the constraining principles need to be identified. Can we get deeper insight into this conception of self-organised criticality, which interprets hierarchies as organisations that are able to maintain their configuration within changing environments?

SCALING LAWS IN PHYSICS AND BIOLOGY

A promising tool for a better understanding of the general emergence of hierarchies in the natural and social worlds is the construction of scaling laws. Scaling laws establish invariant relationships, which are in general non-linear, between quantitative measurements of various attributes over a wide range of sizes of individual entities. G.B. West (chapter 3) suggests that they could offer much better tools for comparison than the usual ratios (which assume mostly linear effects) that we spontaneously apply when comparing the attributes of objects that are very different in size. For instance, the "social indicators" that are used for comparing countries, cities, or firms of different sizes are simple ratios to size, for instance gross product per capita, or demographic rates. G.B. West suggests that more appropriate scaling methods provide a better judgment in such comparisons, as he demonstrates in the cases of evaluating the relative strength of ants and men, or sportsmen belonging to different weight categories. But another advantage of scaling laws is that they "typically reflect underlying generic features and physical principles that are independent of detailed dynamics or specific characteristics of particular models". After underlining the similarities between the mathematical structures of the scaling of the fundamental forces in the quantum field theory and in biology, G.B. West recalls important results that have been obtained in biology, where scaling laws exhibiting universal quarter-powers are interpreted according to the "generic properties of the various hierarchical fractal-like branching network systems that sustain life at all scales" (for instance circulatory, respiratory, renal and neural systems and vascular systems in plants). Thus the connection between scaling laws and hierarchical organisation is very strong in physics and biology, and there is a "hint suggestive of a similar origin" to be found "in the observation that at all scales many biological structures exhibit hierarchical, fractal-like networks which are topologically similar to the tree-like hierarchies in the Feynman diagrams driving the scaling of the strengths of the fundamental forces".

According to G. West, a human being survives with an energy consumption equivalent to one hundred watts, whereas if the cells that compose his or her organs were isolated *in vitro*, 10 000 watts would be required. Inclusive hierarchies that characterize the organisation of the living world are therefore very efficient in terms of energy consumption. There are indeed economies of scale in the use of energy that allowed the emergence of very large organisms in the course of biological evolution. This brings evidence that, rather than the metaphor suggested by H. Simon and discussed by D. Lane (chapter 4), inclusive hierarchies in biology are not merely collections of independent "Chinese boxes", and that interactions between levels shape the whole hierarchical organisation. The scaling laws observed between metabolic rates and the size of organisms have been explained by the specific nature of the branching systems that channel the energy flows in these living bodies. According to G.B. West, the mathematical demonstration of the linkage between the universal scaling parameters (three quarters exponents) and the minimisation of energy loss seems to be a very general result that allows further predictions.

SCALING LAWS IN SOCIETY

Do scaling laws apply to social organisations? There is an interesting difference which can be noticed from the above experiments: in social systems, energy consumption, instead of decreasing, increases with development. At national (state), region or city level, energy consumption scales supralinearly with population size (which means that the scaling parameter is above one, while it is below one in the biological instance). Human concentration seems to create added value of a different type from the simple ability to survive under the constraints of the physical environment. Social purposes are obviously of a different nature and the historical trend towards an increase in the number and range of human activities seems to be supported as much as it is constrained by the extra energy expenditure incurred for their achievement. The corresponding expression from Anderson "more is different" needs to be viewed with a different meaning in the biological and social contexts (see the discussion by D. Lane, chapter 4). Societies obviously have to deal with energy and information as constraining factors that hamper their unlimited development, but they can also create and innovate to overcome these limitations.

Many other differences should be underlined, since social evolution is not driven by natural selection but by an intentional process of innovation. This can explain why the evolution rate of social structures is much more rapid and time scales very short when compared with those of natural evolution. However, the observed evolution in the social world rarely reflects any individual, or even collective intentions as they are expressed in the dominant values, beliefs, general representations or expectations, or even according to the informational resources that are available at a given period to a given society. The resulting structures are not always unwanted, undesirable or even "perverse" effects, but most often they are unpredictable and unexpected. Among the surprising emerging properties of many social structures, and especially in social networks, is the existence of hierarchies and scaling laws.

The most promising line of research for using scaling laws and detecting their related constraining principles in the dynamic exploration of social hierarchies is therefore to measure social attributes that support growth and innovation, and enable the development of larger and larger social entities and more and more complex societies. In this process, David Lane recognizes the essential role of the "scaffolding structures" that govern the shift from one structure to the next while facilitating the introduction of innovations.

ARE HIERARCHIES A RESULT OF RANDOMNESS OR OPTIMISATION ?

The idea that hierarchical organisations are so frequent because they optimise some function or principle reappears with recent work by physicists and biologists. But the explanation is no longer static: the organisation is not only the result of an intentional design following a top-down plan, it is also produced through bottom-up evolutionary mechanisms. Hierarchies cannot be considered as happening purely at random. In the living world, it seems highly probable that physical constraints, without imposing complete determinism, have widely influenced the selection of hierarchical organisations that minimise losses of energy while optimising the circulation of flows, or facilitate co-operation during the building of new organisms at a higher level during evolution. This also could be the case in the organisation of natural languages, for which the action of economy principles in the efficiency of communication seems to be recognisable. In the case of cities, the "tyranny of distance" remains the major explanatory principle of hierarchical organisations in geographical space, but the role of this effect has been continuously revisited by social institutions, which have used it in a conscious way. Similarly, the effects of competitive growth are so many and diverse that they can alternatively be considered as totally constraining, or absolutely random !

From the comparative review of work on urban hierarchies all over the world conducted by D. Pumain, (chapter 7) a first result is the rejection of randomness as a satisfactory interpretation. The explanation of the statistical distribution of city size by a static entropy maximisation process, suggested by L. Curry and supported by B. Berry in 1964, is incomplete, because it neglects the action of a constraint in producing the model. The stochastic model of distributed urban growth proposed by Gibrat, although it is dynamic and fits most observations rather well, is not a sufficient explanation either, since it does not pinpoint the interactions that are behind the dynamics it describes. In contrast to the hypotheses that are defended by authors from the "new economic geography" trend (like P. Krugman or Fujita), the existing configuration of urban settlements cannot however be considered as being the direct expression of any optimum or static equilibrium. A functional system of cities equilibrating their supply and demand in economic activities has no incentive for evolution. An evolving urban system can only be explained by evolutionary, historical dynamics, where the progressive diversification and complexification of human activities through innovation have played a large part in building and consolidating the urban hierarchies. This is in favour of the fact that, if we were to distribute the population over the surface of the earth today, it would not be relocated where it is at present: the configuration of the system at a given time does not correspond to what would be the optimal distribution for that time. What could be considered in a given economic context as the optimal size for a city is never actually achieved. Path dependency is an essential process in urban dynamics, since most cities proceed from villages, and large cities from small towns. During this evolution, constraints, especially through transportation speed set against space-filling trends, act on the system in a more or less continuous way, but the system is never in equilibrium. Moreover, it is its own dynamics (consisting in general expansion and competition between cities) that provide the most convincing explanation for the emergence and maintenance of its hierarchical structure (see below).

HIERARCHICAL ORGANISATION AND COMPLEX SYSTEMS

For David Lane (chapter 4), as opposed to the inclusive hierarchy in biology (the “Chinese boxes” that H. Simon refers to), society can be formalised in terms of an “artefact-agent space” which is organised into levels that are not exactly nested but tangled to varying degrees, producing what he calls “heterarchies”. It is in this context that emergence can occur between levels of organisation: human organisation emerges more often between existing levels. These levels are sometimes separated, segregated into spatio-temporal scales, but this is not always the case. Lane identifies three main levels:

- at macro level, there are scaffolding structures that control the evolution of the system
- at meso level, there are competence networks that deliver the system functionality
- at micro level are the individual actors, the “interactors”

This perspective is different from the interpretation of market systems, which considers the macro level of the market as the result of interacting micro-individual agents only. The meso-level is essential, it is a fluid level, which represents historical flows of interactions, or traces of past exchanges, which are called network structures. The complexity of this organisation stems from the circulation of information between levels: the notion of “methodological individualism” is not suited to organisation of this sort, since all levels are needed to understand how each operates, because of the efficacy of interaction between levels. However, while in social organisation level hierarchies need not be strict inclusion hierarchies, it is not by chance that there are connotations of order, inclusion and control attached to this use of the term. There are probable similarities linking this kind of organisation and the emergence of the properties of resilience and adaptability that are characteristics of complex systems, both in social and natural worlds.

These hypotheses are supported by the theoretical framework suggested by Denise Pumain (chapter 7) for hierarchies in urban systems. In this framework, urban hierarchy becomes an essential feature in defining the ability of complex urban systems, not only to adapt to, but also to generate almost continuous and sometimes accelerated socio-economic change. The evolutionary theory for urban systems unifies the earlier central place theory, the conception of cities as nodes in global networks, the theory of innovation cycles and urban specialisation as well as the theory of hierarchical diffusion of innovation, within a conception of complex systems in which urban systems are viewed as adaptive tools for social innovation. Conscious political and economic processes as well as physical territorial and temporal constraints are integrated into the dynamics. The theory establishes conceptual links between scaling laws in the description of urban hierarchies, space-filling processes and social (technical, economic and cultural) innovation. A few ideas are put forward for a possible transcription of these qualitative propositions into mathematical models. Demonstration, as suggested by Alain Pavé for hierarchical systems in biology, has so far been provided using simulation models (multi-agent systems) such as the SIMPOP model. In its generic version, a model of this sort is able to reproduce the main structural and evolutionary properties of urban systems, as well as a variety of urban configurations as observed in different parts of the world.

While deepening our knowledge about the possible explanations for the universality of hierarchical organisations, we have improved our understanding of what makes systems robust and resilient. The secret of their persistent organisation lies in the complex networks that mediate social, biological and physical interactions on different spatial and temporal scales. Thus, two directions for future research seem promising: The first concerns improving the definition of abstract entities and measures for a better appraisal of the scaling laws that shape the hierarchical features in complex networks. Scaling laws, when appropriately designed, can reveal the processes that ensure the maintenance and evolution of the structure of a system. Second, instead of focusing research into emerging properties in complex systems on two-level modelling, where the macro-structures are assumed to be created by interactions at the micro-level, as in many agent-based or game theory models, greater attention should be paid to multi-level structures and the circulation of information between the levels. In a social sciences perspective, it has been frequently observed that many systems are much more resilient than the dynamic statistical models describing them would have predicted. This underlines the importance of research into the processes of social reproduction. Until the 1980s, such processes were relatively easy to document and formalise, by studying the mobilisation of information resources and their asymmetries in social networks. With the emergence of the so-called "information society", including the domination of the economy by stock markets, the merchandising of information, the proliferation of large networks and the atomisation of property, without forgetting the new modes of governance that enhance participative instead of representative systems, it has become more difficult to identify the various institutions that maintain social order at meso-level by ensuring the circulation of information between levels. Although deliberately dissimulated or difficult to detect, the hierarchical patterns that operate in new networks are being brought into the open by economists or political scientists interested in regulation processes or institutions, in an evolutionary perspective.

We have thus reversed our initial challenge: in endeavouring to find methods to gain a better understanding of the emergence, universality and durability of hierarchical organisations, we can now suggest using these hierarchies as methodological tools, as significant markers, or detectors, of the operation and evolution of complex systems. This perspective will perhaps help us in solving the mystery of hierarchical organisation, which can still be considered as a commonplace or unexpected feature, but which in the light of this new perspective becomes highly meaningful. It invites further collaboration between sciences of natural and social complex systems, and this research could perhaps be considered as a first step in building a science of complex systems.

REFERENCES

Barabási, A.L., Albert, R., 1998, Emergence and scaling in random networks, *Science*. **286**: 509.

Crombie, A.C., 1994, *Styles of scientific thinking in the European tradition: the history of argument and explanation especially in the mathematical and biological sciences and arts*, Duckworth, London.

Gabaix, X., Ioannides, Y., 2002, The evolution of city sizes distributions, in: *Handbook of Urban and Regional Economics*, vol. 4, V. Henderson, and J.F. Thisse, eds, North Holland, chapter 53, pp. 2341-2378.

Guérin-Pace, F., 1995, Rank-size distribution and the process of urban growth, *Urban Studies*. **32**: 551-562.

Pumain, D., 1982, *La dynamique des villes*, Economica, Paris.

Pumain, D., 2003, L'émergence des villes, in : *Pour la Science*, décembre, numéro spécial « La complexité, la science du XXIe siècle », pp. 134-138.

Robson, B. T., 1973, *Urban growth, an approach*, Methuen, London.

Watts D.J. and Strogatz, S.H., 1998, Collective dynamics of small world networks, *Nature*. **393**: 440-442.

Author Index

Subject Index