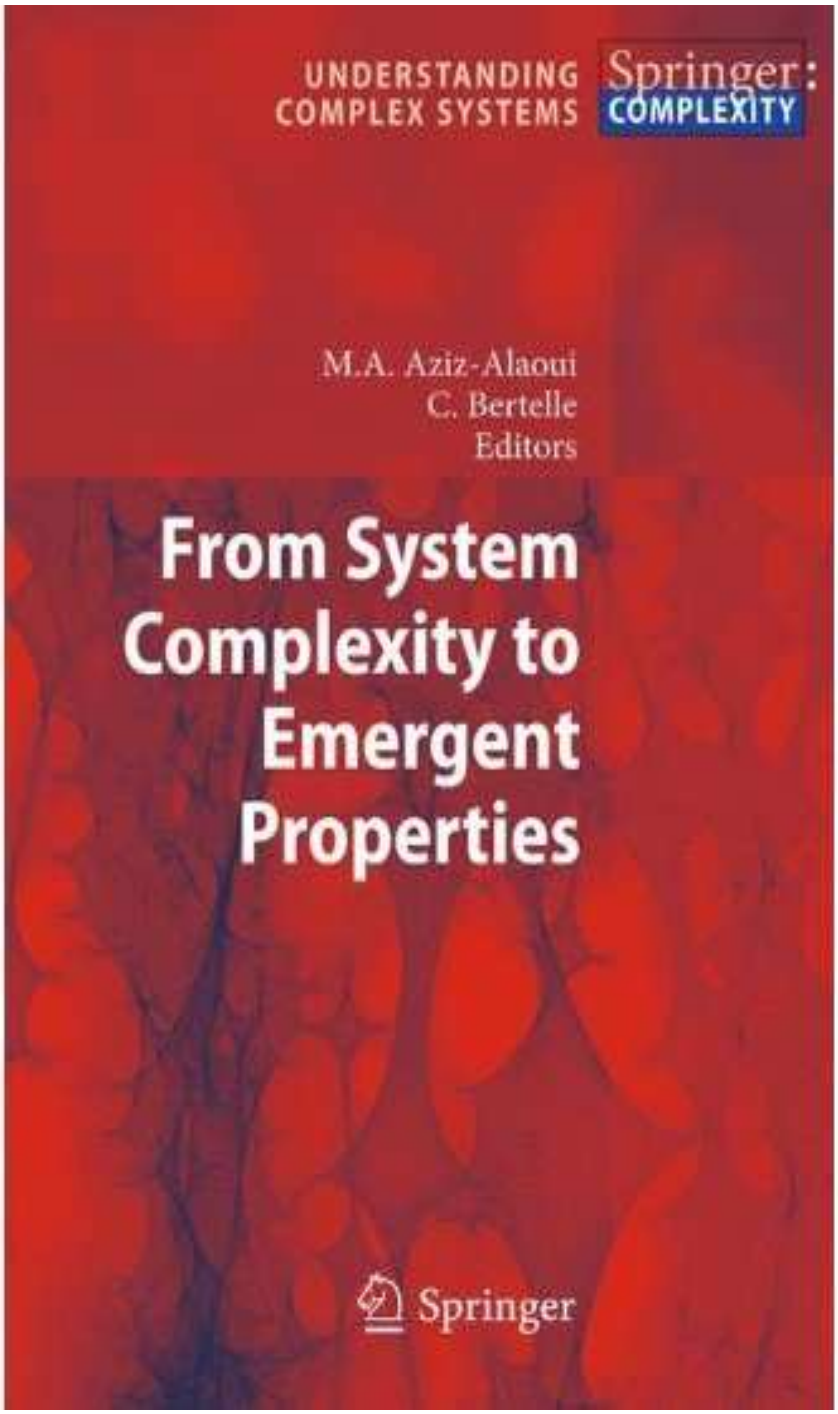

UNDERSTANDING
COMPLEX SYSTEMS
Springer:
COMPLEXITY
M.A. Aziz-Alaoui
C. Bertelle
Editors
From System Complexity to Emergent Properties
Springer

-

Springer Complexity

Springer Complexity is an interdisciplinary program publishing the best research and academic-level teaching on both fundamental and applied aspects of complex systems - cutting across all traditional disciplines of the natural and life sciences, engineering, economics, medicine, neuroscience, social and computer science.

Complex Systems are systems that comprise many interacting parts with the ability to generate a new quality of macroscopic collective behavior the manifestations of which are the spontaneous formation of distinctive temporal, spatial or functional structures. Models of such systems can be successfully mapped onto quite diverse “real-life" situations like the climate, the coherent emission of light from lasers, chemical reaction-diffusion systems, biological cellular networks, the dynamics of stock markets and of the internet, earthquake statistics and prediction, freeway traffic, the human brain, or the formation of opinions in social systems, to name just some of the popular applications.

Although their scope and methodologies overlap somewhat, one can distinguish the following main concepts and tools: self-organization, nonlinear dynamics, synergetics, turbulence, dynamical systems, catastrophes, instabilities, stochastic processes, chaos, graphs and networks, cellular automata, adaptive systems, genetic algorithms and computational intelligence.

The two major book publication platforms of the Springer Complexity program are the monograph series “Understanding Complex Systems" focusing on the various applications of complexity, and the “Springer Series in Synergetics", which is devoted to the quantitative theoretical and methodological foundations. In addition to the books in these two core series, the program also incorporates individual titles ranging from textbooks to major reference works.

Editorial and Programme Advisory Board

Understanding Complex Systems

Founding Editor: J.A. Scott Kelso

Future scientific and technological developments in many fields will necessarily depend upon coming to grips with complex systems. Such systems are complex in both their composition - typically many different kinds of components interacting simultaneously and nonlinearly with each other and their environments on multiple levels - and in the rich diversity of behavior of which they are capable.

The Springer Series in Understanding Complex Systems series (UCS) promotes new strategies and paradigms for understanding and realizing applications of complex systems research in a wide variety of fields and endeavors. UCS is explicitly transdisciplinary. It has three main goals: First, to elaborate the concepts, methods and tools of complex systems at all levels of description and in all scientific fields, especially newly emerging areas within the life, social, behavioral, economic, neuroand cognitive sciences (and derivatives thereof); second, to encourage novel applications of these ideas in various fields of engineering and computation such as robotics, nano-technology and informatics; third, to provide a single forum within which commonalities and differences in the workings of complex systems may be discerned, hence leading to deeper insight and understanding.

UCS will publish monographs, lecture notes and selected edited contributions aimed at communicating new findings to a large multidisciplinary audience.

M.A. Aziz-Alaoui · C. Bertelle (Eds.)

From System Complexity to Emergent Properties

Springer

Editors

Prof. M.A. Aziz-Alaoui
LMAH
University of Le Havre
25 Rue Ph. Lebon
BP 540, Le Havre Cedex
France
E-mail: aziz.alaoui@univ-lehavre.fr

Prof. C. Bertelle
LITIS
University of Le Havre
25 Rue Ph. Lebon
BP 540, Le Havre Cedex
France
E-mail: cyrille.bertelle@univ-lehavre.fr

ISBN 978-3-642-02198-5 e-ISBN 978-3-642-02199-2

DOI 10.1007/978-3-642-02199-2

Understanding Complex Systems ISSN 1860-0832

Library of Congress Control Number: Applied for

Typeset & Cover Design: Scientific Publishing Services Pvt. Ltd., Chennai, India.

Printed in acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

M.A. Aziz-Alaoui and Cyrille Bertelle thank “Région Haute-Normandie” for its sustainable support for research development which contributes to high level scientific collaborations. This book illustrates this purpose.

M.A. Aziz-Alaoui et Cyrille Bertelle remercient la “Région Haute-Normandie” qui, par son soutien durable à la Recherche, aide au développement de collaborations scientifiques de haut niveau. Ce livre en est l’illustration.

Preface

Emergence and complexity refer to the appearance of higher-level properties and behaviours of a system that obviously comes from the collective dynamics of that system's components. These properties are not directly deductable from the lower-level motion of that system. Emergent properties are properties of the "whole" that are not possessed by any of the individual parts making up that whole.

Such phenomena exist in various domains and can be described, using complexity concepts and thematic knowledges. Natural systems in biology and environmental science exhibit wide range of interactions' systems (food chain or neuronal systems, for example) through multi-scale phenomena where each level reproduces similar organizational emergence. Social systems in human or economical sciences exhibit similar kinds of emergent organizations, due to individual behaviour interactions. The dynamics of the components lead the system to organizational evolutions crossed by temporary critical equilibrium, for example bifurcation phenomena.

In this book, we will highlight complexity modelling through dynamical or behavioral systems. We will develop wide range of links between models and various applicative area in geography, urban systems, trafic management, biological systems, ...

Complexity science exhibits original feature by filling the communication gap between thematicians (domain experts) who hold specific knowledges of reality and phenomena and modelling designers who hold specific knowledges of some formal operational descriptors, relevent for these reality and phenomena. Complexity leads to efficient ways for thematicians to analyze practical phenomena. The thematicians' knowledges together with complex systems' concepts, can lead to emphasize innovative properties which can be generalized and formalized through a wide range of domains (resiliance, for example ...).

The first chapters of this book focus on complexity modelling concepts. In the first part, historical point of views and conceptual descriptions are given, leading to a better understanding of complexity through epistemology

(F. Varenne) or to a formalization proposal for multi-level emergent behaviours (C.-C. Chen et al.). Fundamental questions are asked: "When are things complex?" (R. Sitte) and "What makes a system c omplex?" (M. Cotsaftis).

In the second part, complexity modelling is proposed for geographical systems. Deep links of these systems with emergent properties (A. Dauphiné) are presented. Methodological approaches for risk and catastrophe analysis and management are described (E. Daudé et al., D. Provitolo). Innovative swarm intelligence algorithm is proposed for spatial self-organization simulations (R. Ghnemat et al.).

In the third part, dynamics on complex networks are studied. The emergence of chaos in networks describing adaptive systems are investigated (A. Gecow). Synchronization phenomena in neural networks are shown and lead to a power law characterizing self-organized systems (N. Corson and M.A. Aziz-Alaoui). An original validation process is proposed to study the dynamics of distributed architectures, using formal methods (I. Oliver).

In the fourth part, complexity engineering for transportation is studied. Adaptive self-organization processes for transportation on demand in urban systems are described (C. Bertelle et al.). Modelling for network intermodal transport is also given (A. Caris et al.).

In the fifth part, different aspects of engineering processes for decision making are suggested: complex systems' modelling for inventory management systems (K. Ramaekers and G.K. Janssens), medical diagnosis based on cooperation between physicians and artificial agents (B.L. Iantovics), timetable agent-based software (E. Babkin et al.), emotion modelling for problem solving applied to learning (K. Mahboub et al.).

The pluridisciplinary purposes of this book's concern are enable to design links between a wide-range of fundamental and applicative Sciences. Developping such links - instead of focusing on specific and narrow researches - is characteristic of the Science of Complexity that we try to promote by this contribution.

Le Havre, France *M.A. Aziz-Alaoui*
February 2009 *Cyrille Bertelle*

Contents

Models and Simulations in the Historical Emergence of the Science of Complexity

Franck Varenne

University of Rouen, Department of Philosophy
76821 Mont-Saint-Aignan, France
&
GEMAS, CNRS UMR 8598
Paris Sorbonne, France
fvarenne@wanadoo.fr

Summary. As brightly shown by Mainzer [24], the science of complexity has many distinct origins in many disciplines. Those various origins has led to "an interdisciplinary methodology to explain the emergence of certain macroscopic phenomena via the nonlinear interactions of microscopic elements" (ibid.). This paper suggests that the parallel and strong expansion of modeling and simulation - especially after the Second World War and the subsequent development of computers - is a rationale which also can be counted as an explanation of this emergence. With the benefit of hindsight, one can find three periods in the methodologies of modeling in the empirical sciences: 1st the simple modeling of the simple, 2nd the simple modeling of the complex, 3rd the complex modeling and simulation of the complex. Our main thesis is that the current spreading (since the 90's) of complex computer simulations of systems of models (where a simulation is no more a step by step calculus of a unique logico-mathematical model) is another promising dimension of the science of complexity. Following this claim, we propose to distinguish three different types of computer simulations in the context of complex systems' modeling. Finally, we show that these types of simulations lead to three different types of weak emergence, too.

Keywords: complexity, model, computer simulation, numerical simulation, algorithmic simulation, software-based simulation, weak emergence

1 Introduction

As brightly shown by Mainzer [24], the science of complexity has many origins in many disciplines. Whereas the definitions of complexity are numerous and did not lead to any general agreement, it is most of the time accepted that systems are complex when they are at least composed of a certain amount

of entities (elements or agents) interacting together. But it is not sufficient. Another granted condition is that the complexity of the system specifically comes from the 'kind' of interaction at stake and/or from the 'kind' of its results. Here begins a typical problem of vagueness in specification for such a 'kind'. Does the complexity come from the algorithmic incompressibility of interactions? Or from the emergent phenomena resulting from these interactions? But according to which definition of emergence? The interaction at stake seems to be of the 'kind' which precisely cannot be captured by a single and simple concept otherwise complex interactions could be generically expressed, even perhaps formalized and unambiguously computed.

My aim here is not to give the last word on this topic. But it is to show that - parallel to evolutions in mathematical approaches - recent evolutions in methodologies of modeling and simulation of complex systems have allowed the asking of such seminal questions from a different and somewhat more distinguishing point of view. My particular thesis is that the current spreading of complex computer simulations of systems of models (where a simulation is no more a step by step calculus of a single logico-mathematical model) is another promising dimension - of course not the only one - , of the science of complexity. Although it has during years been condemned as self-contradictory or sterile, the development of complex modeling and complex simulation of complex systems seems to become a real interesting challenge in many areas of science. Why? What has changed in the applied epistemologies of the modelers of complex systems? And for which reason?

To understand this shift in the recent applied epistemologies, it is necessary to characterize the kinds of relations models and simulations had with complex systems in the past. In a first part, after many authors, I briefly recall how and why simple modeling of simple systems can be seen as one of the major factors that made the success of modern science. In a second one, I suggest that the science of complexity became a foremost subject of concern essentially when it became reasonable to hope for some mathematical and (relatively) simple modelings of complex systems. In a third part, I suggest that the complex modeling and simulation of complex systems were no more seen as pitfalls in the 90's because they began to be used to formulate, replicate and disentangle simultaneously many different 'kinds' of interactions and causal intrications in complex systems. In this context, I propose to distinguish three types of computer simulations. In a fourth part, I use this distinction to differentiate *three types of weak emergence* in the case of the modeling and simulation of complex systems.

Finally, I propose that we recognize the real variety of computer simulations, and that we particularly distinguish the one I call *pluriformalized simulations* from *algorithmic simulations* and from *numerical simulations*. Those distinctions help to further explain some recent contributions of simulations

to the science of complexity. And, through that, they help to clarify a bit the debated notion of emergence.

2 Simple Models of Simple Systems

Of course, nothing is absolutely simple in itself. And all the meanings of the terms 'simple' or 'complex' in this paper have to be thought of with a connotation of relativity. As already pointed out by Mainzer [24], in the ancient Greece, celestial movements were believed to follow simple geometric laws (i.e. geometric figures constructed with compass and ruler), whereas sublunar phenomena were thought to be very complex. Although we cannot much further define the term 'complex' (which is anachronic in this context), it remains possible for us to say that *relatively* to sublunar phenomena, supralunar ones were seen to be much less attainable to simple representation or modeling. Note that this was a coherent claim in the doctrine of Aristotle, for instance, because sublunar phenomena were subject to generation and corruption. And, because of this specificity, contrary to the celestial substances, their essences could not be grasped in an everlasting, simple and perfect (i.e. achieved) geometrical law.

Not only the notion of 'complexity' but also the notion of 'model' is questionable when referring to this historical context. Nevertheless, it seems rather acceptable too when we recall the brilliant study of Pierre Duhem entitled after a famous word of Simplicius (6th century AD)"To save the phenomena" [12]. In this book, Duhem shows that many philosophers of the late antiquity had newly insisted upon the distinction between the work of an astronomer and the one of a physicist. According to Posidonius, Simplicius or Proclus, for instance, the physicist is concerned with the very essence of the celestial entities whereas the astronomer has only to "save the phenomena", i.e. to speak about the apparent figures, sizes and distances of celestial entities. Being limited to these considerations, it is understandable that his objects of study really match the discourses of arithmetic and geometry.

Put in another way: those late antique philosophers reverse the traditional arguments of the ancient philosophers to explain the same thing, i.e. the relative good matching between celestial movements and geometrical discourses or figures. For the latter, this was due to the capture of the perfect essence of celestial beings through mathematics. For the former, on the contrary, this matching was due to the fact that the astronomer, unlike the physicist, limits his work to the salvation of the pure appearances. In 1908, Duhem is keen on exhuming those troubling arguments as he is convinced that such a phenomenological epistemology for the physical theory is the good one and that it must prevail: generally speaking, scientists shall not try, nor hope, to represent what surpasses the human being. This is the reason why, according

to Duhem, a good law ought to remain mathematical and ought not to be represented through a physical model as Maxwell or Boltzmann erroneously tried to do. In the rest of his book, Duhem shows that modern science was born not in the neoplatonist works of Kepler or Galileo but rather in those earlier times of the 14th century when nominalist philosophers decided that the sublunar world itself was as complex (hence only mathematically salvable) as the late antique philosophers finally said the celestial world was. According to Nicholas of Cusa (1404-1464), for instance, the sublunar world was no more seen as "unachieved" (i.e. "imperfect" for the ancient Greeks) but, more positively, as infinite because it was seen to inherit a kind of infinity and complexity from its divine creator. So physicists, as astronomers, had to renounce to seek something more than a salvation of phenomena. They were told to limit themselves to fictitious essences and hypothetical causes (ibid.: 68).

Through this historical and epistemological contribution published in 1908, Duhem fights against the then traditional essentialist interpretation of the laws of the mechanics. Doing this, he lays a foundation for a neo-positivist reinterpretation of all the physics which will be influent, particularly through the contemporary works of the philosopher of science Quine. The traditional essentialist interpretation says, on the contrary, that the laws of mechanics are simple mathematical laws because the phenomena are *simple in themselves*. In fact, *pace* Duhem, the predominance of such a thesis from 1687 (the year of publication of the masterwork of Newton) to the end of the 19th century was confirmed at length, for instance, by the more recent studies of the historian of science Alistair Crombie [8]. In fact, Crombie recalls that, during this period, most of the physicists admitted that, with his laws, Newton had provided both an explanation and a prediction of the gravitation. They admitted that these laws had captured in a simple way the simple essence of the mechanical phenomena. But Duhem had a strong religious conviction which obliged him to recognize the infinite complexity of the creation, be it sub- or supra-lunar. Simultaneously, he felt that he had to put the stress on a pure mathematical presentation of laws and that, consequently, he had to fight against the British physicists of his time who used physical models to assess the laws of electromagnetism.

It is noteworthy that most of the historical accounts of the science of complexity stress on the traditional essentialist interpretation. By quoting Duhem here, my aim is not to contest this point. What I want to suggest is that when I say that in the times of modern science "we modeled simply the simple", it can be put in this more correct form: we made mathematico-physical theories to explain simply what was generally thought to be simple in itself. Contrasting with this view, Duhem had in fact a modelist and positivist interpretation of physical theories, although he rejected the physical modeling

for the mathematical laws. So, let's say that the dynamical 'model' of Newton was 'simple' and had a 'simple' target system.

On the one hand, the simplicity of the model came from the fact that Newton (along with Leibniz) had constructed well-fitted and well-designed mathematics (techniques of notation and techniques of computation) at the same time. The simplicity of the model resides in the efficiency of its notation and in the ease of the symbolic manipulation, combination and resolution it permits. On the other hand, the simplicity of the target system came from three main ontological hypotheses: supposedly, there was only one framework of space and time in nature; all the phenomena in this framework were thought to be of the same mechanical kind (one type of underlying causes); any dynamical systems were (or could be) isolated from any other. The simplicity of the target system (TS) resides in a representation of the world which entails ubiquity, genericity and separability.

3 Simple Models of Complex Systems

The science of complexity has known multiple births during the 20th century. It has known many historical reconstructions too. But it is only in the 1980's that it began to be a well recognized academic domain. At that time, interdisciplinary research programs became more systematic. There are many reasons for this long maturation. According to me, one of the main factors is quite prosaic. But simultaneously, as we will see, this factor reveals a relative continuity in epistemological positions, contrasting with what is often said about the paradigm shift from simplicity to complexity. This factor is the brutal and large diffusion of PC's in all types of labs all over the world in those years.

Let's recall that, in 1892, Poincaré showed that easily writable non-linear hamiltonian equations could lead to chaotic behavior. Doing this, he did not show that the world is complex neither that the solar system really is chaotic[1]. But he introduced a decisive split between the *attributes* of the different properties of the Hamiltonian formalism. Such a formalism possesses at least three distinct properties: (1) it is a notation; (2) it enables symbolic manipulation and combination; (3) it leads to formalized solutions. Poincaré shows that a particular mathematical formalism that possesses the two properties to be a notation and a working symbolic system (which can each be attributed a kind of simplicity), does not necessarily possess the (simple) property to give simple result of its combinations.

So as to facilitate the comparative exposition, let's define now a 3 dimensional vector called 'F' (for 'Formalism' or 'Properties of the formalism'). Its three

[1] This will be demonstrated only in 1989 by J. Laskar, with the aid of a computer and the "pursuit lemma".

dimensions represent the *degrees of simplicity attributed* to each of the three main properties of the formalism F, namely: notation, combination, solution. Let's assume that these degrees can only be of the two extreme kinds: S (for simple) and C (for complex). Beware that this notation has no pretention to be neither exact nor comprehensive. It is only an *illustrative notation* which will be useful for the clarity of some of my arguments.

Hence, the claim about the formalism used by Poincaré can be represented by the complexity vector:

$$F(S, S, C)$$

Roughly speaking, the possibility to lead always with ease to a simple solution is denied. I.e. the Hamiltonian formalism reveals a kind of complexity in *itself*. But it affects only one of its properties. As it presents two other properties to which simplicity still can be attributed, such a formalism can still be put in the set of the 'simple' ones. Following this major advance of Poincaré, researches in the mathematics of dynamic systems may focus on the internal relations between those apparently contradictory *attributes of properties* of some family of formalisms. In this context, one of the key questions is this one: how is it possible for a same formalism to be at the same time simple and complex?

Another distinct (and more recent) question is this one: what are the relations between such a complexity of the formalisms and the complexity that can be detected or measured in experimental works in physics or chemistry?[2] To what extent are those complexities of properties of different entities (a formalism and a real system) of the 'same' nature? Contrary to the first one, these last two questions are not only on complexity but on the *external validity* of the model of complexity for real systems (such as chaotic models of turbulent fluids).

When studying the spreading of models of deterministic chaos in theoretical ecology, the historian of science Deléage [9] has shown that the most impressive dimensions of the work of Robert May in the 70's was the fact that he gave a simple and "unified formalism" to explain both the stochastic variations and the cyclic oscillations of populations. Note that, in this particular case, because of the remaining simplicity of their notation (and not of their resolution), the non-linear equations of the theoretical model of chaos could serve as a theoretical argument against any further complexification in the notation of ecological models. Why complexifying the notation when a simple notation seems to give a quite realistic (complex) appearance to its numerically computed results? The complexity of real systems seemed to be sufficiently captured and/or represented by the internal relation between the attributes of the different properties of the model. So, the epistemological

[2] For this question, see [3]

position of Robert May was rather conservative from this standpoint. More generally, we can say that, in a first period, the shift to chaos models in many disciplines remains in fact in the continuity of the traditional hope to capture and reproduce in a simple way (at least at the level of the notation) what is complex in reality. So, these approaches mostly consist of confronting "simple models (at least at the level of notation) to complex reality".

Indeed, the schools working on dissipative systems or on systems far from equilibrium (Nicholis & Prigogine [27], Kauffman [22]) and on synergetics (Haken [16]) are fully aware of the non validity of some hypotheses of modern sciences concerning the real systems. They consider that most of the systems are open, not closed nor separable. But, similarly, they are still in search of the simplest and most generic mathematical or symbolic notation for a wide range of complex systems.

See, for instance, Kauffman:

> "If all properties of living systems depend on every detail of their structure and logic, if organisms are arbitrary widgets inside arbitrary contraptions all the way down, then the epistemological problems that confronts us in attempting to understand the wonder of the biosphere will be vast. If, instead, core phenomena of the deepest importance do not depend on all the details, then we can hope to find beautiful and deep theories." [22].

In another context, see Jensen on self-organized criticality (SOC):

> "The paper by Bak, Tang, and Wiesenfeld (1987) [Phys. Rev. Lett., 59: 381] contained the hypothesis that, indeed, systems consisting of many interacting constituents may exhibit some general characteristic behavior. The seductive claim was that, under very general conditions, dynamical systems organize themselves into a state with a complex but rather general structure. The systems are complex in the sense that no single characteristic event size exists: there is not just one time and one length scale that controls the temporal evolution of these systems. Although the dynamical response of the systems is complex, the simplifying aspect is that the statistical properties are described by simple power laws." [21].

Jensen again, in his chapter 5 entitled 'The Search for a Formalism':

> "One wants a model with sufficient structure to contain nonobvious behavior, but the model should not be so complicated that analytic approach cannot be carried out." [21]

The search for "general laws" or for "simplifying aspects" is then often *a search for an homogenization of the degrees of simplicity within the formalization* once it has be shown to have some realistic acceptable results. That

is: physicists or theoretical biologists most of the time want to build a formalism $F2(S,S,S)$ from a formalism $F1(S,S,C)$.

We could object that the works on Cellular Automata [34, 33, 37] remains very empirical. It is true. Their theoretical "empiricity" comes from the format of the complexity vector of their formalism which is not $F(S,S,C)$ but $F(S,C,C)$. I.e. not only the resolution is not simple but the combination is not too, and both have to be assisted by computers. In the case of the SOC, Jensen proposed to build an approximate $F2(S,S,S)$ (with the hypotheses of a Mean Field Theory) using the work on cellular automata made by Bak and his colleagues and which was first published in a $F1(S,C,C)$ form.

Specifically, researches on CAs were innovating at their beginning in that they contrasted with the classical numerical simulations of manually intractable analytical models of the form $F(S,S,C)$. But, the *trend to model simply the complex* goes back in this field too, when some research programs finally - and very interestingly - seek for some generic CAs, i.e. CAs simulating back analytical results. Wolfram [37], for instance, is seeking a move $F1(S,C,C) \rightarrow F2(S,S,S)$. Similarly, people working on Artificial Life (Santa Fe) often look for some similitude with cases of physical phase transitions so as to make possible the applications of subsuming models which would be simpler from a combinatorial point of view. This would be a move of this kind: $F1(S,C,C) \rightarrow F2(S,S,C)$ or $F1(C,C,C) \rightarrow F2(S,S,C)$. Note that this second situation happens when we are in front of rich formal agents with 'complex' - i.e. non-trivial - notations instead of simple CAs.

For all those reasons, the emergence of the science of complexity has given a central role to formal modeling. And, in conformity either with a positivist and ontologically parsimonious epistemology (devoted to Ockham's razor) or with an essentialist one seeking great transdisciplinary, theoretical and simple principles, the quest for a remaining dimension of simplicity was not abandoned. Hence, most of the first approaches in the science of complexity have been guided by the quest of simple modeling, i.e. of *modeling with at least simple notations*. From this epistemological viewpoint (be it positivist or essentialist), the complexity of the model is accepted. But it has to reside in its treatment, not in its notation.

This is the very reason why most of the important roots of the science of complexity can be found in the pre-computer era. But this is the reason why too the entrance of this science in the computer era - and particularly in the PCs' era - has led to rapid and new convergences and - globally - to a noteworthy academic reinforcement. The famous tale about Edward Lorenz (1963) gives us a key to understand this point. With the aid of the computer, he rediscovers quasi-empirically the sensitivity to initial conditions of non-linear equations (with an additional important factor of dissipativity). But by doing this, he

was entering a kingdom of a new and fertile ambiguity. He first thought he was ordering the computer to treat numerically some supposed analytical model of the kind $F(S, S, C)$. But, by visualizing the results, he couldn't be very sure of the epistemic status of what he was doing: wasn't he doing a *virtual experiment* on a stylization of the target system $TS(NL, D)$, i.e. a Target System with Non-Linear causal relations and with Dissipativity? What was in view: the complexity of the model or the complexity of the stylized fact?

With this example, we understand that before the use of the computer as a plain simulator, the exploration of the internal relations between simple properties and complex properties of formalisms of the type $F(S, S, C)$ could easily be dissociated from the question of the empirical matching between the complexity of a $TS(NL, D, SO, SOC...)$ (where $NL =$ Non-linearity, $D =$ Dissipativity, $SO =$ Self-Organization, $SOC =$ Self-Organized Criticality...) and the kind of complexity which appeared by making use of such a $F(S, S, C)$. But, the more recent development of computer-aided simulations - in parallel with more classic computer-aided resolutions of mathematical models - has led to a perplexity. On the one hand, it has led to a significant facilitation for direct applications of the concepts of the science of complexity (such as in CAs or in Multi-Agent Systems). But it has caused a reconfiguration of the problem of the relations between the simple and the complex. These relations were not only thought *within* the properties of formalisms but within the different uses of the formalisms (be they computer-aided or not). Consequently, the till then constant idea that simplicity must remain a *minima* an inner attribute of the model seen as a system of notation directly speaking to our mind (so that we "understand" the laws of nature) begins to be questioned [6, 7].

4 Models and Complex Simulations of Complex Systems

But what is a model?

According to a characterization due to Minsky,

> "To an observer B, an object $A*$ is a model of an object A to the extent that B can use $A*$ to answer questions that interest him about A." [25]

This minimal characterization is interesting because it adopts a pragmatic approach. In particular, it does not make use of the notion of representation. So, it lets open the question whether a useful model is a true or approximate representation of the target system.

The main characteristics of a model are (1) its objectivity, i.e. its ontological independency compared to analogy; (2) the relativity (to an observer) of its property to model the TS (target system); (3) the property of the modeling

relation to be interpreted in terms of a simplicity in the asking and answering some questions about the TS (compared to a direct questioning of the TS).

Let's focus on this simplicity in the questioning and answering. As the diversity of the practices of modeling shows, it can have many interpretations. Among them are the following:

1. Simplicity in the reproduction of an observable behavior: models of prediction, operational models...
2. Simplicity of an experimenting on the model instead of an experimenting on the TS (due to time, space, technical, material or deontological limits)
3. Simplicity in the answering a question about some main causal factors: explanative models.
4. Simplicity in the providing of an easy global understanding (for the human mind): theoretical functional models. According to many theoretical biologists, such a simplicity necessarily entails beauty, deepness and ... truth. See Kauffman, *supra*.
5. Simplicity in the testing of the coherence or realizability of a formal theory: models in logic, mathematics or theoretical physics.

In particular, a formal model should be characterized as a formal construct possessing a kind of simplicity either in its notational power (simplicity of symbols, unity, formal homogeneity) or in its combinatorial power or in its ability to lead to a solution. This simplicity is chosen so as to satisfy a specific request: prediction, explanation, decision ...

Note that the characterization of Minsky entails that such simplicity, whatever its nature, remains relative to the observer. But the question is: where are the limits of the observer and the limits of the observed object? Where is their mutual frontier? Hence, to whom or to what this simplicity has to be relative? Is this simplicity relative to the system 'human programmer and observer + computer' or to the system 'human mathematician and observer'? We have to be aware that, at this level, different answers lead to different epistemological choices. A formalism appearing as a $F(S, S, S)$ to the former can appear as a $F(S, S, C)$ to the latter.

A similar problem is at stake in a rather popular question among the epistemologists of computer simulations: is a computer simulation an "extended observation" (Humphreys [20]) or only a "conceptual argument" (Stöckler [32])? The main argument of Humphreys lies on the fact that computers are kinds of instruments which are very similar to material ones. Whereas, according to Stöckler, any use of a computer is founded on systematic and deterministic uses of pure symbols (at least at the level of the machine language). It implies that a fictitious human mind, with only more time, attention and patience than ours, could always deduce by itself the result of any computer

simulation. Whereas it is at the material level that a pure difference of degree prevails for Humphreys, it is at the symbolic level that such a difference prevails for Stöckler. So, let's have a more precise look at computer simulations.

What is a computer simulation?

It is often said that "a simulation is a model in time". Hence, there seems to be no computer simulation without a unique underlying model. In 1996, Stephan Hartmann follows this widely accepted characterization:

> "Simulations are closely related to dynamic models [i.e. models with assumptions about the time-evolution of the system] ... More concretely, a simulation results when the equations of the underlying dynamic model are solved. This model is designed to imitate the time evolution of a real system. To put it another way, a simulation imitates a process by another process." [17].

In this view, the imitating process is always supposed to be a delegation to the computer of the complicated combination of symbols of some model of the type $F(S, S, C)$ or $F(S, C, C)$ which would lead to its resolution. But it has been shown [35, 36] that this dynamic imitation is not fundamental in every Computer Simulation (CS). Some CSs are imitative (compared to an eventual formal calculus of the model or to the dynamic of the TS itself) in their results but not in their dynamic. If, for instance, the CS uses a numerical trick - such as fictitious finite elements - to solve an analytical model, each phase of the step-by-step computation hasn't to be realistic or imitating in itself. Similarly, Varenne [35] has shown that there are simulations of the growth of botanical plants which are imitating the process of the real plants at the end of each computation, in its result, but not during the computation itself.

That is the reason why I propose to characterize CS in another way. In particular, it can be useful to avoid the term "model" because it is no more obvious that every CS is based on a unique model. Today, we can observe that many CSs are rather based on *systems of models*, and sometimes on *complex systems of models*. If we do not confound the algorithm and the models at stake in a computer program dedicated to a CS, it is necessary to avow the spreading of CSs using, at the same time, a *multiplicity of models*. As it is visible through some contributions to this book, some CSs of complex systems tend to be more *software-based* (such as agent-based CSs) than directly *model-based* as in the classical fashion (mathematical model + computer-aided numerical simulation). Note that we can hardly say that our times see the shift from simple models to complex ones. The current doubtless move toward a complexification of formalisms impinges more directly on CSs than on models. It is sometimes said that "computational models" become more and more complex. But it is not a very different claim from mine. Because

what is called "computational models" are in fact *software based systems of models* which are built to enable complex CSs.

Of course, software-based CSs are no panaceas. But they are chosen in some scientific fields (especially in biology and social sciences) because they enable to insert a certain degree of *heterogeneity* in the various models scientists aim to implement. I call *pluriformalization* those CSs that enable the coexistence and co-calculation of a multiplicity of formalisms [35]. Pluriformalized complex CSs enable to build multiaspectual and/or multiscale systems of formalisms. The notation's properties of their formalisms are much less constraining than the ones of traditional mathematical or physico-mathematical models. Beware that the system of notation can become complex in itself. But it is its ability to denote directly (without intermediate fictions) some trivial or well-known (in a scientific domain) or commonly-recognized (in the opinion, in the collective representations ...) things that is facilitated. This compromise is obvious in the definitions and the uses of multi-agent systems in social sciences. Consequently, as far as a characterization of CSs is concerned, I recommend focusing more directly 1st on the relation between symbols inside a CS and 2nd between the symbols of the CS and the symbols of the associated model(s).

So, let's say that a CS is minimally characterized by a *general strategy of symbolization* taking the form of at least one step by step treatment. This step by step treatment possesses at least two *major phases*:

- 1st phase: a certain amount of *operations running on symbolic entities (taken as such) which are supposed to denote* either real or fictional entities, reified rules, global phenomena, etc.;
- 2nd phase: any observation or any measure or mathematical or computational *re-use* (e.g.: the simulated "data" are taken as real data for a model or for another simulation, etc.) of the result of this amount of operations *taken as given* through a visualizing display or a statistical treatment or any kind of external or internal evaluations.

CSs are not always supposed to be iconic modeling, i.e. similar representations of systems. For instance, in the case of a *numerical* CS of a mathematical model, the symbolic entities denote fictitious entities. They really are taken as denoting entities, i.e. as real symbols. But they are particular denoting entities as they are denoting nothing really existing in the TS. As shown by eminent philosophers of language and science (Russell among others), these cases are not invalidating in ordinary language. For instance, it is meaningful to say that "Santa Claus does not exist". By doing this, we use a symbolic notation ('Santa Claus') which denotes nothing really existing. But it does not prevent the whole sentence from having a meaning and, even, from being true. Similarly, in a numerical CS, what is important is the phase of aggregation

and of measure (2nd phase) of the effects of this aggregation: the symbols are reified and treated as *relative subsymbols*.

It is in the context of the second AI that Smolensky [31] coined the term *subsymbol*. By this term, he denoted the entities processing in a connectionist network at a lower level and which aggregation can be called a symbol at an upper level. Subsymbols are *constituents of symbols: "they participate in numerical - not symbolic - computation"* [31]. Berkeley [5] has shown that Smolensky's notion can be interpreted from a relativistic point of view. This is this *relativity of the symbolic power of symbols in CSs* I want to express through my relativistic connotation of the term. In the second phase of any CS, aggregations of symbols finally are taken themselves as symbols of some features of the TS. And from this new standpoint, the elementary and aggregated symbols are taken as *subsymbols*.

A *numerical CS* depends heavily on a model and can be seen as a direct subsymbolization of this model. For instance, a continuous model is treated by a strategy of discretization. Hence discretization is a subsymbolization of the previous explicit model. On the contrary, the kind of CS I propose to call *algorithmic* [35] is directly based on iterative and uniform rules. It can be for instance simple CAs, formal grammars, rewriting formal systems, L-systems ... Their denotational power is often emphasized because such CSs use symbols that tend to denote directly those rules that are supposed to really exist in the TS. For instance, CSs of biological morphogenesis which are based on L-systems are seen as algorithmic ones (see [28]). Note that, contrary to some computer scientists, many botanists think that such rules are fictitious entities too. They claim that they are more similar to computational tricks than to symbols really denoting something in the TS. We see that the denotational power of the symbols at stake is relative. It depends on the ontological commitment of the scientific field.

Finally, *Software-based CSs* are different from *numerical* ones and from *algorithmic* ones. Numerical CSs are most of the time designed to treat a previous intractable model of the type $F(S, S, C)$. The discretization of the notation of the model enables to smooth the difficulty of the resolution through its transformation into an iterative combination. The numerical CS has this kind of effect: $F1(S, S, C) \rightarrow F2(S, C, S)$. With the play on the non-denotational symbols, the difficulty of resolution is transformed in a difficulty (in terms of amount) of combinations and iterations. *Algorithmic or rule based CSs* tend to treat directly formalisms of the type $F(S, C, C)$. Most of the time (but there are exceptions: especially those CAs that are built to simulate other formalisms), their performance can be more easily interpreted as an iconic modeling and as a quasi-real experimenting on the TS because the denotational power of the symbols plays a major role.

Software-based CSs operate on systems of models. According to my notation, a system of models can be presented as a vector of formalisms (or a matrix of properties) having different attributes of simplicity for each of their properties. For instance, as in [11], a multileveled CS in population ecology can intricate (1) some solvable differential equations working at each step at the level of the population $F1(S, S, S)$, (2) some models of real space or social space in a graph or network formalism $F2(S, S, C)$ and (3) some models of agents possessing a complexity of combination (in terms of diversity and variability) because they are cognitive agents or relatively complex reactive ones $F3(S, C, C)$. It can be represented as follows:

$$[F1(S, S, S), F2(S, S, C), F3(S, C, C), ...]$$

The relations between the ease of notation and the difficulty to iterate are more difficult to understand in this case. The main result of these analyzes is that such a CS is no trivial numerical treatment of a model. This vector of formalism (or matrix of properties) cannot be easily reduced to or transformed into a huge formalism having its own homogeneity, simplicity or complexity properties. So, most of times, such CSs have first to be experimented: through analyses of sensibility, of robustness, etc. Hence, beside more traditional and direct mathematical researches on complex formalism, more and more modelers (working often in applied sciences) are seeking models of such complex software-based simulations [11,13,18,19]. Significantly, the effort to standardize and re-homogenize multimodels in industry had stemmed from the same previous and necessary movement of complexification, i.e. of integration and co-calculation of heterogeneous models [38].

5 Types of Simulations and Types of Emergence

As we will briefly see now, the ability to take into account some kind of emergence is a key issue for any specific standpoint on the science of complexity, particularly when we focus on CSs. But there are many definitions of emergence (see [10]).

Following Bedau [1] and [2], we can define three kinds of emergence:

1. An emergence is *nominal* when "macro-level emergent phenomena are dependent on micro-level phenomena in the straightforward sense that wholes are dependent on their constituents, and emergent phenomena are autonomous from underlying phenomena in the straightforward sense that emergent properties do not apply to the underlying entities." [2]
2. An emergence is *strong* when, contrary to what happens in nominal emergence, emergent properties have some irreducible causal power on the underlying entities. In this context, "macro causal powers have effects

on both macro and micro-levels, and macro to micro effects are termed downward causation." [2]

3. Finally, according to Bedau, *weak emergence* is a kind of *nominal emergence*. To explain it, he quotes Herbert Simon: "given the properties of the parts and the law of their interactions, it is not a trivial matter to infer the properties of the whole." [30]. Weakly emergent phenomena are those which are not easy to explain and which essentially need simulation to arise. So, all kinds of weak emergence are based on the property of "underivability without simulation". Assuming a system S, a microdynamic D on this system (*"which governs the time evolution of S's microstates"*), the locality of D (i.e. *"microstate of a given part of the system at a given time is a result of the microstates of 'nearby' parts of the system at preceding times"*), Bedau states that: *"Macrostate P of S with microdynamic D is weakly emergent iff P can be derived from D and S's external conditions but only by simulation."* [1]

According to Bedau, science is only concerned with weak emergence (WE). This question is of course still discussed. I won't enter this debate here. My aim in this section is not to state a general thesis on emergence but, more restrictively, to specify the definition of *weak emergence* by Bedau in relation with my definitions of CSs. Bedau claims that "simulation" is central for this scientific notion of emergence. But he does not give many details on what is to be understood by a simulation. So his definition can be completed.

The definition of *weak emergence* is particularly revealing when we see the importance of "simulation" and when we keep in mind what I have said of the relation between models, systems of models and their CSs. As we have seen, this relation most of the time is a delegation, removal or displacement of some degrees of complexity from a property of the initial systems of symbols to another property of another systems of symbols. As far as a formalization is concerned (and not directly the TS), we can see some kind of emergence precisely in the experience of an *obstacle* or a *difficulty* or a *non-triviality*. Simon said *"given the properties of the parts and the law of their interactions, it is not a trivial matter to infer the properties of the whole"*. We could adapt this claim to the properties of formalisms: *"given the properties of the parts [the property of the notation of symbols] and the law of their interactions [the ease of their combination], it is not a trivial matter to infer the properties of the whole [the difficulty to infer solutions]."* [30].

So, according to Bedau (following the general idea of Simon), weak emergence arises when we have $F1(S, S, C)$ and when we transform it, through simulation, in a $F2(S, S, S)$. My claim is that the variety of types and uses of CSs enable a further analysis of what is at stake in weak emergence (WE).

Bedau says that WE is founded on a CS, whatever its nature. But he seems to say too that we have WE only when we make the computer transform a $F1(S, S, C)$ into a $F2(S, S, S)$. To avoid such a risk of erroneous restriction of his definition of WE, we can say that we have WEs of at least three types. A WE of the first type (WE1) occurs when we make the computer transform a $F1(S, S, C)$ into a $F2(S, S, S)$. A WE of a second type (WE2) occurs when we make the computer transform a $F1(S, C, C)$ - e.g. multi-agents systems with complex cognitive agents - into a $F2(S, S, S)$ or when we only experiment on it. A WE of a third type (WE3) occurs when we make the computer experiment on a dynamic matrix of formalisms $[F1(S, S, S), F2(S, S, C), F3(S, C, C)...]$ (such as in pluriformalized CSs). As we have seen, in this case, it is still possible - but not necessary - to intend to simplify the computation into a calculable model of the type $Fn(S, S, S)$.

So, distinguishing between three major types of simulation has given us the opportunity to distinguish between three *types of weak emergence*. The main differences here stem from the differences between the kinds of obstacle - or difficulty or non-triviality - when facing the problem of inference or computation. In the case of Software-Based CSs, the difficulty of computation is not of the same nature than the one at stake in the case of numerical simulation. So, *the type of emergence cannot be exactly the same*. For the former, it is a problem of co-calculation of heterogeneous models. For the latter, it is a more classical problem of the effective computation of a previous analytical model.

6 Conclusion

I cannot enter here in more details about the differences between a CS and a model, in particular concerning their *routes of reference* [14] and the *entanglement* of these routes in the case of pluriformalized CSs. Let's keep in mind that the recent appearance of a variety of CSs has led to a multiplication of splits, of new frontiers between what is considered simple and what is considered complex. This frontier can be seen either only within the formalism, between the attributes of its properties, or within the different uses of a computer to treat a model or a system of models. So, the properties of models and the possibility to attribute to them the 'simple' or 'complex' predicate have been diversified and extended due to the intensive use of computers in the modeling techniques and, in particular, due to the spreading of object-oriented programming techniques.

More generally, what makes a system of models work? Most of the time, it is due to a massive discretization of parameters, variables, time and space, amount of populations, etc. In my opinion, the reason why this strategy of discretizing is a good one does not lie upon the metaphysical fact that nature really is discrete or is a computer. More pragmatically (from the standpoint

of the management of formalisms), discretization works because it causes multiple and heterogeneous occasions of choices in computations (such as in Discrete Event CS or Individual-Based modeling techniques ; see [15]). It causes decisive new degrees of freedom in a CS, especially when you see it more as an intensive computer-aided supervision of formalisms. It causes a multiplicity of instant of choices which enable a possibly constant change (1) of the status of symbols at stake in the denotational hierarchy (either sub-symbolizing, denotating or exemplifying), (2) of the ways those formalisms denote each other during the CS (for the sake of a tractable co-calculation), (3) of the ways those formalisms punctually denote levels or parts of the TS (for the sake of the external validity of the resulting CS).

Above all, the multiplication of the types of CSs has given birth to new ways of creating weak emergence with computers. Of course, the multiplication of the variety of emergence which can be simulated on computers does not solve the problem of the external validity of these new formalizations of emergence. Is the emergence in a complex CS an adequate model of a real emergence in the target system? Such a question, especially when asked at this level of generality, remains open. But, this diversification makes the models and simulations of emergence more accommodating and less constraining as diverse ways of reproducing different kinds of emergence possibly occurring in real systems become available.

References

1. Bedau, M.A.: Weak Emergence. In: Tomberlin, J. (ed.) Philosophical Perspectives: Mind, Causation and World, vol. 11, pp. 375–399. Blackwell, Malden (1997)
2. Bedau, M.A.: Downward Causation and the Autonomy of Weak Emergence. Special issue on Emergence and Downward Causation, Principia 6, 5–50 (2002)
3. Bergé, P., Pomeau, Y., Vidal, C.: Order within Chaos. Hermann & John Wiley & sons, Inc., Paris (1987)
4. Berkeley, I.: What the.. is a subsymbol? Minds and Machines 10(1), 1–14 (2000)
5. Berkeley, I.: What the.. is a symbol? Minds and Machines 18, 93–105 (2008)
6. Bhala, U.: Understanding complex signaling networks through models and metaphors. Progress in Biophysics and Molecular Biology 81(1), 45–65 (2003)
7. Braillard, P.A.: Que peut expliquer un modèle complexe et peut-on le comprendre? In: Kupiec, J.J., Lecointre, G., Silberstein, M., Varenne, F. (eds.) Modèles, Simulations, Systèmes, Syllepse, Paris, pp. 89–108 (2008)
8. Crombie, A.: Augustine to Galileo: The History of Science A.D. 400 - 1650, 1st edn. Penguin, London (1959) (revised edition 1969)
9. Deléage, J.-P.: Une histoire de l'écologie, La Découverte, Paris (1991)
10. Dessalles, J.L., Müller, J.P., Phan, D.: Emergence in Multi-agent Systems - Conceptual and Methodological Issues. In: Phan, D., Amblard, F. (eds.) Agent-Based Modeling and Simulation in the Social and Human Sciences, pp. 327–355. The Bardwell Press, Oxford (2007)

11. Duboz, R.: Intégration de modèles hétérogènes pour la modélisation et la simulation de systèmes complexes - Application à la modélisation multi-échelles en écologie marine, Thèse d'informatique, Université du Littoral (2004)
12. Duhem, P.: To Save the Phenomena: An Essay of the Idea of Physical Theory from Plato to Galileo, translated from Sauver les phénomènes - Essai sur la notion de théorie physique, University of Chicago Press, Paris, 1908 (1969)
13. Edwards, M.: Intérêt d'un modèle agrégé pour étudier le comportement et simplifier la simulation d'un modèle individu-centré de consommation couplé à un modèle de ressource en eau, Thèse d'informatique, Université Paris 6 - Pierre et Marie Curie (2004)
14. Goodman, N.: Routes of Reference. Critical Inquiry 8(1), 121–132 (1981)
15. Grimm, V.: Ten years of individual-based modelling in ecology: what we have learned and what could we learn in the future? Ecological Modelling 115, 129–148 (1999)
16. Haken, H.: Synergetics. Springer, Berlin (1977)
17. Hartmann, S.: The world as a process. In: Hegselmann, Müller, Troitzsch (eds.) Modelling and simulation in the social sciences from the philosophy of science point of view, pp. 77–100. Kluwer, Dordrecht (1996)
18. Huet, S., Edwards, M., Deffuant, G.: Taking into Account the Variations of Neighbourhood Sizes in the Mean-Field Approximation of the Threshold Model on a Random Network. Journal of Artificial Societies and Social Simulation 10(1) 10 (2007)
19. Huet, S., Deffuant, G.: Differential Equation Models Derived from an Individual-Based Model Can Help to Understand Emergent Effects. Journal of Artificial Societies and Social Simulation 11(2) 10 (2008)
20. Humphreys, P.: Extending Ourselves - Computational Science, Empiricism, and Scientific Method. Oxford University Press, Oxford (2004)
21. Jensen, H.J.: Self-Organized Criticality - Emergent Complex Behavior in Physical and Biological Systems. Cambridge Lecture Notes in Physics. Cambridge University Press, Cambridge (1998)
22. Kauffman, S.: At Home in the Universe - The Search for the Laws of Self-Organization and Complexity. Oxford University Press, Oxford (1995)
23. Laskar, J.: A numerical experiment on the chaotic behavior of the solar system. Nature 338(6212), 237–238 (1989)
24. Mainzer, K.: Thinking in Complexity - The Complex Dynamics of Matter. Mind and Mankind. Springer, Berlin (1994); 3rd revised and enlarged edition: (1997)
25. Minsky, M.: Matter, Mind and Models. In: Proceedings of the International Federation of Information Processing Congress, vol. 1, pp. 45–49 (1965)
26. Mitchell, S.D.: Biological Complexity and Integrative Pluralism. Cambridge University Press, Cambridge (2003)
27. Nicolis, G., Prigogine, I.: Exploring Complexity. W.H.Freeman & Co Ltd, New York (1989)
28. Prusinkiewicz, P., Lindenmayer, A.: The Algorithmic Beauty of Plants. Springer, New York (1990)
29. Russell, B.: On Denoting. Mind, new series 14, 479–493 (1905)
30. Simon, H.A.: The Sciences of the Artificial, 3rd edn. MIT Press, Cambridge (1996)
31. Smolensky, P.: On the proper treatment of connectionism. The Behavioural and Brain Sciences 11, 1–74 (1988)

32. Stöckler, M.: On Modeling and Simulations as Instruments for the Study of Complex Systems. In: Science at Century's End. Proc. of the Pittsburgh/Konstanz Colloquium in the Philosophy of Science, October 1997, pp. 355–373. Univ. of Pittsburgh Press (2000)
33. Toffoli, T., Margolus, N.: Cellular Automata Machines: A New Environment for Modelling. MIT Press Series in Scientific Computation (1987)
34. Ulam, S.: On some mathematical problems connected with patterns of growth of figures. In: Proceedings of Symposia in Applied Mathematics, vol. 14, pp. 215–224. American Mathematical Society (1962)
35. Varenne, F.: Du modèle à la simulation informatique, Vrin, Paris (2007)
36. Winsberg, E.: A Tale of Two Methods. Synthese (forthcoming) (2008)
37. Wolfram, S.: A New Kind of Science. Wolfram Media Inc., Champaign (2002)
38. Zeigler, B.P., Praehofer, H., Kim, T.G.: Theory of Modeling and Simulation - Integrating discrete event and continuous complex dynamic systems. Academic Press, New York (2000)

About the Predictability and Complexity of Complex Systems

Renate Sitte

Griffith University
SEET/ICT, Gold Coast Campus
Griffith University, QLD 4222, Australia
r.sitte@griffith.edu.au
r.sitte@servusresearch.net

Summary. With ever-growing complexity of systems to be modeled, there is a strong need for a proper theory of complexity, other than the algorithmic complexity known in computing. The problem is that there is no unanimous consensus as to what complexity is. Several attempts have been made, some are very promising, but a widely applicable theory and practice have not been derived. Quantification is an essential step in modeling to achieve prediction and control of a system. Quantification is also a crucial step in complexity and some complexity quantification models are emerging. In this chapter, a unifying and systematic approach to complexity is proposed. Its aim is to bring some clarity into the unknown, and a step further towards predictability. It serves as an overview and introduction, in particular to the novices on how to deal with complex system as a practical approach. Some practices summarized here are elementary and others are quite ambitious. It happens over and over that the uninitiated researchers make errors, reinvent the wheel or fall into traps. The purpose of this chapter is to offer good advice and a sense on how to avoid pitfalls.

1 Introduction

Understanding and prediction are the ultimate goals in Modeling and Simulation. In this endeavor, we are faced with ever-increasing complexity. This works twofold with ever more complex models and modeling techniques mapping to the increasingly complex phenomena that we are trying to simulate. With steadily raising computing capability, we are able to incorporate more complexity into our models and simulations. In return, this has had an impact on our thinking paradigm and capability for abstraction and modeling techniques. So, when are things complex? Our thorough analysis has revealed that complexity can mean very different things to different people, contexts and environments. In many cases, it refers to an aspect or part of the system under study, in others the magnitude and variety of the system itself.

To some complexity rather means the unknown: that what cannot be easily understood, modeled or predicted with current knowledge. In this chapter, a unifying and systematic approach to complexity is presented, to bring some clarity into the unknown, and a step further towards predictability. We will be looking at different types of complexity, ranging from simple structural and functional complexity, to those highly complex systems that change (but not necessarily evolve) through their own dynamics. With this in mind, we will be able to identify what can be predicted and what cannot. By knowing it, we will be in a better position to find the unknown, which is the most important step towards the prediction of complex systems.

Since ancient times, one of human's endeavors is taking control of what happens around us, and affects us in one or another way. To take control or even influence our surroundings, we have to anticipate what will occur and where will it have an effect. We have to be able to manipulate all or part of what we want to control. To do this we must have a mental mapping, a model of its components and their functionality, and this in turn requires some understanding causes and interactions the overall result. This works fine so long things are not getting too complex. It helps to ask *what* is complex? Is it the *process* of designing or constructing something, or is the *product* , i.e. the system to be studied, designed, simulated, or constructed the one who is complex? The human brain is limited to handle about 2.5 bits of information at a time (known as the 7 ± 2 "chunks"), for anything more, we need memorizing aids such as diagrams, itemized text, etc. For larger projects, this can grow quickly into many pages, folders, even boxes of information storing papers or other media. With the invention of modern computers, we have been able to solve ever-growing large and complex problems. So too is the complexity of the problems to solve. Since J. Forrester's World models in the early seventies, the past decades have seen cycles of ever-increasing complexity and computing power, pulling each other further. We have seen astonishing results. The word "complexity" has become a fashion statement in the research communities. Everything is "complex" now: it sounds good, adds magic and importance, and has the potential to attract funding. But, there is more to it than just fashion. There is a strong reason behind being able to understand and deal with complexity. Here is an example. Refraction through another medium is being modeled by measuring the incident angle with respect to the vertical, and then using the refraction index of the medium and exit angle in a simple equation that we learnt in high school. In ancient times, scientists knew that the medium had different refraction (indices) and tried to set up a model, but they failed because their model was highly complex. In fact, the model became very complex because they measured the incident angle with respect to the horizontal! It was many centuries later that a rotation of 90° made all the difference between a very simple and a complex model.

During humankind's different stages of intellectual evolution, different schools, insights and paradigms have emerged, such as the development of mathematics, scientific method and systems thinking that lead in one or another way to our current state of the art. The success of any of these approaches is due to a consistent clearly defined underlying definitions and theory. While we are still far away from a widely applicable definition of complexity, this chapter is to bring awareness on how to tackle complex systems as good strategies, as a practical approach rather than a prescriptive general technique.

This chapter is organized in the following way: In section 2, we have a closer look at what complexity means, and what it does not mean. It presents a fundamental classification of types of complexity and explains the need to identify how and where the complexity is located in a system under study. This is followed by section 3 that explains the need for a unifying and measurable definition of complexity. Section 4 deals with ways of measuring complexity, and a general model to define a problem and its solution that can be quantified (measured). This model helps in understanding the notion of measuring complexity. Section 5 explains successful ways of modeling complexity. Section 3.3 is on how to deal with complexity, with a recommendation for three fundamental steps as an initial approach to complexity. It then discusses complexity reduction, and this is complemented with a selection of examples. It moves then to stability and dealing with unstable systems. The section closes with a brief discussion on unsolvable systems.

Rather than following the traditional style of pointing to other references in a historical review in the introduction section, the references to other author's work are cited in a didactical and supporting context in the relevant sections of this chapter.

2 When are things complex?

We would all agree in finding that systems are complex when they undergo changes in structure, changes in functioning, when they have adaptive feedback, are evolving, or simply when they have a large number of components or relations. One of the hurdles in defining a theory of complexity, and with it developing a fundamental, helpful approach is that there is no uniformity in the meaning of complexity. Complexity means different things to different people. In a large literature review, we have found that sometimes it refers to an aspect or part of a system, other times to the magnitude and variety of the system. To others it meant the inexplicable or the emergence of unexplained behavior. To yet another group it means chaotic systems, or the difficult, the unknown: that what cannot be easily understood, modeled or predicted with current knowledge.

We have to move away from the subjective to an objective notion of complexity. We have found that perhaps the strongest characteristics of a complex system are that they have:

- Intricate interdependencies among their many parts
- Many variables operating simultaneously
- Generally nonlinear, this can include chaotic systems (butterfly effect)
- Cause and effect are not close in time and space
- Intuitive interventions do not produce the expected outcome
- Reductionist analysis fails (misleading)
- Emergence
- Self organization

In classical systems theory, emergent property means that by specific arrangements a system exhibits properties or behavior that neither of its components has. Perhaps the simplest and most striking example is water. Both, hydrogen and oxygen are gases in their natural occurrence, but by mixing them, one obtains a liquid substance: water. Another typical example is an airplane. Its isolated components cannot ignite combustion, turn turbines, or fly for a longer distance. Although emergent properties can be surprising, it must be clear that emergence does not mean the inexplicable. We use the Braitenberg [6] vehicles as an example. They have two sensors coupled to two motors and their behavior and functioning can be very well explained, yet they have an unpredictable behavior. Rather than on magic, we should think of insufficient information as we have for example with "life". We know when something is alive and the necessary conditions for life, but we have not yet been able to design and construct, or explain it because we do not have sufficient information yet. This incompleteness of information can appear anywhere ranging from minute details or in large-scale components, or their assembly.

We have to see the domain of complexity in a spatial and temporal scale. This is very helpful, but does not necessarily mean that complexity grows with time or space, at least not always. We could find as many examples as counterexamples to it. Thirty years ago, programs such as Forrester's World programs were stretching the capability of computers, but many-body interaction is still a problem with current computer power. Is this complex? No, it is limited computing capability. The models are there. There is a paradox in complex systems. For example the well-known Mandelbrot patterns are generated with very simple rules, but the patterns they generate are quite complex. Another classic example of the paradox is again the Braitenberg vehicle (with two sensors and two motors) that moves around in a seemingly complex trajectory but in reality it is a rather simple phenomenon, because when light falls on one or the other light sensor, the vehicle responds by activating the corresponding motor. It is a paradox between cause and effect, but it is not a general rule. This paradox illustrates that in a complex system not all is complex.

It is important to identify in what, where the complexity is located. This is a very important step in dealing with a complex system. We have to be aware that there are different types of complexity. These are different aspects or manifestations of complexity, but none of them is complexity per se. We can distinguish between the following fundamental types of complexity:

- Functional
  - Single or multifunctional components
- Structural
  - Dimensionality
  - Networks
  - Hierarchy
  - Levels depth breadth
- Topological
  - Connectivity
  - Relation
  - Number of relations
  - Direction of relations
- Algorithmic
- Architecture
  - Series, parallel, cascade, recursion
- Extent of ...
  - Span of orders of magnitude
  - Time frame
  - Compararative extent of ... between entities, events, operations

This list could be expanded to include more types and sub types. For example, one could argue whether we would include music into the structural types or into an additional class of say a sensorial complexity. This would lead to an argument of where to draw the boundaries between objectivity and subjectivity. We leave this matter for the moment, as it will be addressed in quantifying complexity. In any of these types, we have to consider the extent of repetition and or parallelism, and the extent of homomorphism (same structure) and isomorphisms (same functionality) that may occur. Here are some examples for extent of repetition: (a) redundant components in safety critical applications, say, an aircraft or medical software, and (b) to enhance efficiency in a grid of computers or turbines or windmills for electricity generation.

We can also see that some of these types have affinities with each other. The diagram in figure 1 shows relations between these fundamental types of complexity have.

A classification is rarely useful on its own. A complex system is likely to have more than one type of complexity. The types of complexity do not map to each other in equal proportions or weight. The way in which this taxonomy

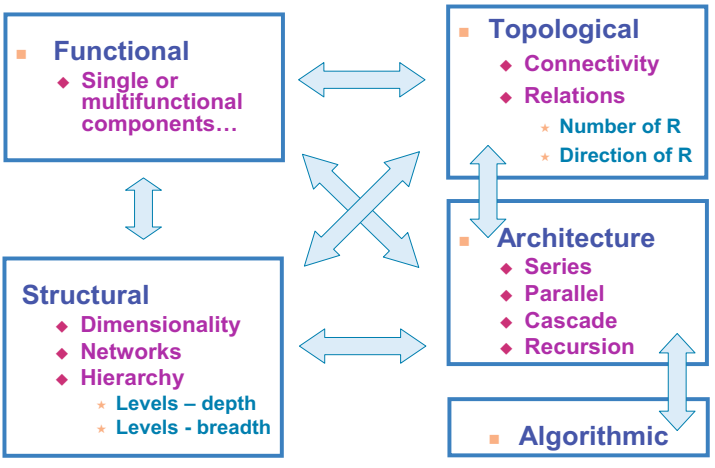


Fig. 1. Fundamental types of complexity and their affinities

approach is helpful is by identifying where the complexity is bundled. Is it because the network is complex? Is it because it is a highly multi-functional system? To identify this allows taking a better approach in handling the complex system, by using techniques that are more suitable.

If we know more about the nature of complexity, we are able to recognize patterns, derive heuristics and given time, develop a general complexity theory in its own right. A word of caution: the concept of complexity theory as used in computing (the "big O"), refers to algorithmic complexity. Its application to other complex systems is very limited.

3 The need for a unifying definition of complexity

How complex is a complex system? Is it very much, not so much, or just a little? To tackle complexity in a systematic way we need an unambiguous definition. As mentioned earlier, there is no consensus about what complexity is, because it means different things to different people. A consensus in the meaning of complexity would be helpful. We have found a whole range of conceptual descriptions of complexity in the literature, including philosophical treatises and definitions about complexity. While philosophical views have by tradition a rightful place in the scientific community, their effect is similar to what spices are in food: spices can make the food nicer, but one cannot survive by eating spices only. Even among the more pragmatic approaches, a mere concept does not bring us further in developing a formal theory. We need a clear definition from which we can derive theory and practice, paradigms for handling complexity in an efficient way that enables us to progress. For example, the Commonwealth Scientific and Industrial Research Organisation

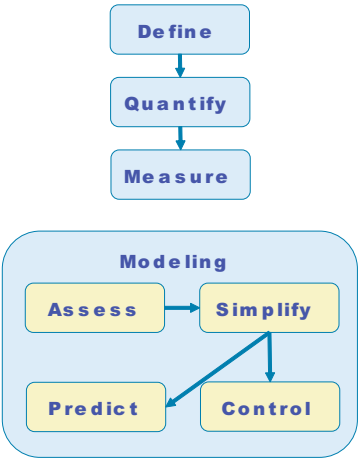


Fig. 2. Rationale for a complexity definition and quantification

(CSIRO) in Australia [8] limits the characterization of complexity to only two properties: emergence and self-organization. This seems quite elegant. The weakness with this criterion is that it does not help in distinguishing between varying degrees of complexity, because the two characteristics do not involve any quantity. Both, the rather simple substance water as a liquid made of two gases and highly complex systems such as weather systems satisfy their criterion for complexity, defying any intuitive notion about complexity itself. While this criterion helps to distinguish complicated from complex, which is part of the intention of their complexity criterion, it lacks any indication of how complex is the complexity of say - water.

A definition of complexity must be able to incorporate, derive or imply a measure; otherwise, we do not know what or how to measure complexity. We need this to be able to assess complexity, that is the degree of complexity either in a process or its representation in a model. Only when we can assess complexity quantitatively are we able to choose between alternatives, simplify and reduce complexity either in the model or in the process itself, else we do not know what or where to simplify. This rationale is schematized in figure 2

4 Measuring complexity

Before moving into measuring complexity, let us recall the vectorial model for a problem and problem solving, because it offers some intuition to what comes later in measuring complexity. Imagine that we have 300 dollars and have to pay 500 dollars. This means that we have a problem of 200 dollars. In this case, we are in one single dimension, the dollars. We could add time or interests as other dimensions. In the vectorial model of a problem, we have a current

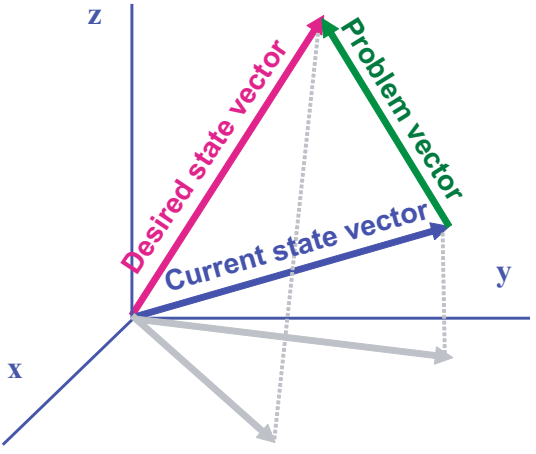


Fig. 3. Vectorial model for a problem in a 3-dim. space as the transition from a *current state* to a *desired state*

state position (represented by a vector) and a desired position (another vector). The length and direction of the vectors represent quantities that characterize these states. We consider the problem as the necessary transition to move from the current state position, to the desired state position. The necessary steps to achieve the transition are then the solution to the problem. This notion is shown in figure 3 as a 3D vectorial representation. It can be extended to a multidimensional space that is typical in dynamic systems.

If we can express complexity as a quantity, we can distinguish unambiguously between complex and simple, or to a finer scale, between highly complex, and relatively low complexity. Earlier we have seen a number of static complexities: entities, relations amongst them, hierarchy depth, etc. We are now moving into complexity of dynamics, as a continuously changing entity. We do this by bearing the problem vector in mind. It is the quantification in this model that makes it intuitive. If we can describe the complexity in terms of algorithmic operations, then we can find ways of moving from complex to simple. It is therefore that a measure of operational complexity should be sought.

4.1 Examples of complexity measures

In a major work and milestone, Hazen and his collaborators [12] have developed a definition for system complexity. The authors define *Functional Information* as the fraction of all possible configurations that sufficiently achieve a specified functionality. The notion of Functional Information is exemplified by a sequence of letters that is conceptually familiar: The four different nucleotides A, U, G, C. In a 100 nucleotides RNA strand there are 4100 different possible sequences, but not all of them are functional (i.e. some

are dummies and do not play any role). With this simple but powerful definition, and by involving the probability or potential of an entity to perform a function, the authors have derived a set of mathematical models, including advanced models for higher dimensional systems.

Another interesting model for measuring complexity is presented by Stoop & Stoop [20]. It considers computation (in a general sense) as a reduction of the difficulty of prediction, i.e. inferring the future from the past. The model derivation of the complexity measure relies on concepts from physics. The key part is the transformation relation between input and output as a mapping between the input and output string and multiplicity of operations (or iterations). It considers different observable manifestations of a complex dynamic system as different computations with their own predictability (and associated complexity). By using averages, in this case from bundles of trajectories, the model is capable of providing information about the complexity associated with the choice of elements that are involved in the computations (calculations) of predictions in noisy systems. An implicit process analogous to normalizing allows comparisons and complexity-aware choices between different of elements (sets of variables) and their predictive ability.

Basu [4] has defined a topological complexity based on their combinatorial and algorithmic properties. A bounding function of the combinatorial complexity is defined using the connectivity of their subsets. The computation (or estimate) of this bounding function poses an important problem in computational geometry i.e. the arrangements of sets of geometric objects.

A design complexity measure is proposed by Braha & Maimon [5]. The authors define a structural and a functional design complexity. The design structural complexity is associated with information content (suitable for automated design) in the form of relations, models, etc. The functional design complexity is associated with information described as the operations towards satisfaction of the (design, suitability, etc) goals. These definitions involve information content and other known concepts closely resembling the Halstaed metrics [11] as they have been defined and used in software engineering since the seventies. The level of abstraction in these definitions is suitable for intelligent computer aided design systems and IT projects, but unfortunately their generalization and application to other cases is difficult to see, and is not exemplified by the authors.

While information content is undoubtedly a valid measure, one could argue that a design structural complexity definition based on the number of components and their assembly towards a robust (i.e. non-failure) design would be better, practical and more general than focusing on its amount of information. This is particularly important when there are tradeoffs between the number of components and the easiness of an assembly process as for example

in machine design because minimizing the number of components should be part of any design goal as an inbuilt reduction of complexity. A measure of structural design complexity in terms of components and the way they are arranged (interactions) as a set or individually to perform a desired function would be more justified. In this way, a structural complexity design measure actually would complement the functional design complexity measure based on information content.

This is a good time to go back to our earlier consideration about repetition. The previous discussion questioning the justification of information content to define structural design complexity can be seen from another angle. While repetition of operators and operands can be treated as equivalent in software towards the amount of information calculation, this does not necessarily hold with hardware. Let us take our earlier example of windmills or computers arranged as a set towards more efficiency. From an information theoretical viewpoint there would be little difference in information content because 1 or 100 units of the same kind do not increase the variety of the system. It is their assembly (relations) towards an enhanced functionality, their combinations of states i.e. performance, breakdown, idle, etc. that makes the difference in complexity of the system. The assessment of this combinatorial exercise does lead us again to the definition of functional information by Hazen et al. [12] described earlier in this section. This example illustrates that independent repetition is less complex than when relations exist between entities.

5 Modelling complexity

It may be helpful to revisit the vectorial model presented at the beginning of the previous section. The importance of this model or thinking process is a marked distinction between the problem analysis and the offering of a solution that actually addresses the problem. This is the point and opportunity for lateral thinking for alternative solutions or methods, simplicity and efficient techniques etc. Proposing solutions to a nonexistent or ill-defined problem is worthless.

The ultimate purpose of modeling is prediction. If we can understand and model a process, then we are able to anticipate its behavior. This in turn enables to take rational decisions and to change the behavior of that process, if possible at all. There is no specific environment for modeling complex systems. In complex systems modeling, this often implies using a range of different models, requiring different software tools. Conventional commercial software and toolboxes can be a substantial help, in particular because of their efficient algorithms and user-friendly interfaces. They are tested, which is sometimes an underestimated advantage. However, there is no guarantee that the simulations are truthful just because the code works correctly if the

models or the results are not validated. As always in modeling and simulation, the choice of software must suit the models. Regrettably enough, too many times one can find work with wrongly applied models or inappropriate tools in the scientific literature. Some popular techniques that are used in modeling complex systems (in increasing difficulty) are dynamic systems (Matlab/Simulink^{TM}); Artificial Neural Networks, including Self Organizing Maps (SOM) for functions and clustering; Chaos, and Quantum Computing.

5.1 Abstraction

Modeling involves abstraction. This means mapping only those processes to a model to which the model is sensitive, leaving out of the modeling process the less relevant details with little or negligible influence to the overall result. This is schematized in figure 4: On the pie diagram (left side) the processes E to K appear to have little contribution contribution to the overall result and are not included in the model's subsystems schematized on the right side. The purpose of abstraction is to simplify the models, which ultimately are sets of mathematical equations in a way that that they are solvable and computable. To what extent can abstraction be done without loss or distortion by over-simplification? The answer lies "somewhere in between"; it is a compromise between the manageability of the model and its modeling accuracy. A popular approach for reducing complexity is order reduction. This comes from expressing a model as a series expansion. This can be as simple as polynomial series. One uses only the first, second, $\ldots n^{th}$ order terms, neglecting the higher order terms, as they would contribute little or negligible to the results. Caution: Not to be confused with order-predicates in logic.

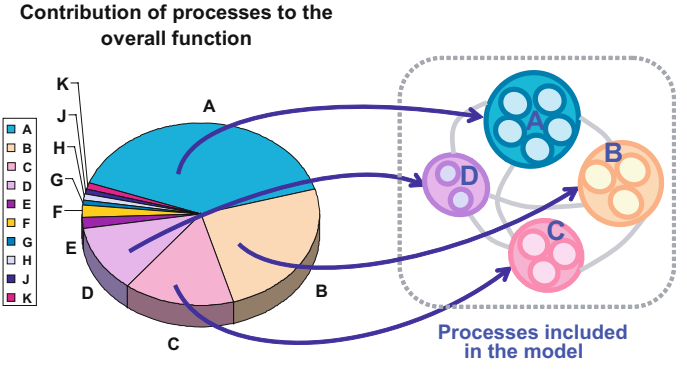


Fig. 4. The processes E to K with little contribution contribution to the overall result (pie chart) are not included in the model's subsystems (right side)

5.2 Linearity and sensitivity

An old saying tells us to "be wise, linearize". How can we do that when non-linearity is one main characteristic of complexity? But then, not all 100% in a complex system is non-linear. This means, that some of the modeling can be linearized. What is important is to find out what can be linearized and what cannot. A sensitivity analysis helps to find out, albeit that it means a substantial effort that can easily run into the combinatorial curse. A sensitivity analysis is always a good approach because it sheds light into what is of primary importance and what is less important. With a good design of experiments (DOE), it also gives a good indication of where the correlations are stronger. A good question to ask is also, does if follow a Pareto 80/20 rule? More often than not, it will. Again a word of caution: characteristics such as linearity and sensitivity can be limited to specific ranges of values. It is important to be aware of this, and find out whether and where they are. Problems can arise when, in the expectation that that a system behaves linear when in reality it does not, as this can lead to initially small errors, but the errors add up and lead to deviations of the predictions that can be substantial. In the case of chaotic systems the lessons learnt from mathematical analysis like taking limits may not necessarily reveal additional simultaneous solutions, or be possible at all, in particular when there is no certainty of the of the validity of mathematical models applied.

5.3 Uncertainty and error estimation

One tends to be satisfied with approximations, but it is paramount to estimate and know the margin of error involved. All simulations are affected with uncertainty. Schneider and Moss from Stanford University insist that "... statistical confidence levels should be attached to even the most complex scientific predictions" [10]. In dealing with complex systems, it is crucial to know where uncertainty is located. One has to find out not only how much uncertainty is involved, but also where it comes from. Schneider and Moss insist that the uncertainty of at least four sources, namely the theory applied, the observations, the model results, and the consensus should be represented graphically for a clearer information about the source of uncertainty. This can be done on a multidimensional axes (radar) graph, with four or more axes. It is easy to find many sources for errors relevant to the case, for example error in the data (missing or unreliable), root mean square errors in anything that involves fitting to a function like the training and testing error in Artificial Neural Networks. Other sources of errors are numerical method errors when solving equations. These errors can be assessed based on the way operations are implemented. Other uncertainties arise from the models, abstraction process, probabilistic estimators, and so on. All errors must be normalized to be able to compare in a plot. The error analysis and uncertainty plot allows improving the certainty of the research, or at least being able to pinpoint why the uncertainty arises, and allows comparisons between different approaches.

5.4 Network representations

Network representations are a popular choice, capable of expressing functional nodes and weighted relations that can be equally suitable modeling a range of situations from biological to industrial or social contexts, and from the rather simple to highly complex and sophisticated. Klein et al. in cooperation with Yaneer Bar-Yam et al. have produced interesting and helpful complex system research using dynamic distributed networks representations, i.e. with no centralized controller, and where global behavior is the result of localized individual actions of nodes [13]. In this work, the authors distinguish between linear and non-linear networks, i.e. single or multiple attractors, and discuss and propose suitable optimization methods. They analyze and discuss symmetry (reciprocity) in the form of dynamic attractors, where the network cycles through a set of configurations of node states instead of converging to a specific configuration of states. They also discusses a range of network techniques such as modularization and network imprinting to overcome difficulties found. Their work was done on collaborative design, but the considerations and lessons learned can be applied to other situations.

5.5 Parallelism

Parallelism is another important concept in modeling complexity, in particular when it is combined with network models. One has to distinguish between simulating parallelism i.e. simultaneous, temporarily independent events from what is known as parallel computing on a special computer with its own programming techniques. To simulate parallel events in complex systems one does not always require a parallel supercomputer. Our interest here is to focus on the modeling part of complex systems, not on the computational details. Nevertheless, modeling parallelism can be made easier by using specific software packages such as in case of discrete event simulations but it can also be represented as networks which can be networks in general or specific networks such as Petri Nets. While such systems can be large, they are not necessarily complex. It is the nature of what is being modeled that can be challenging, and there are no specific rules to help us. For example, Zhang et al. have developed a model for parallel decision making under uncertainty as a sequence of three phases in a cascade of system of systems [26]. The authors also propose a model for assessing the effectiveness of the parallel decision model.

5.6 Chaotic systems

Last not least a word on modeling chaotic systems. Chaotic systems are sometimes considered as non-deterministic. This is wrong. Chaotic systems are not random. They are deterministic systems, but they are highly sensitive to input variation, to an extent where precision becomes important. A very

small difference in the input can change the behavior of the system. This may appear as a sudden change in the behavior, shown on a plot as a set of similar yet distinguishable bundled trajectories that we can observe in the typical visual examples of attractors. This is more an effect of the limited accuracy of computers and their rounding errors than of the mathematical model or the chaotic system it describes. However, not necessarily chaotic trajectories appear as a cluster of cycles (orbits); they can be mischievous and their plots can take many shapes, like bifurcations (with two solutions) from a specific value onwards, repeating bifurcation at variable intervals. Chaotic systems have a range of properties such as self-similarity, fractal dimension, and their assessment is part of the study.

It is very important to be aware that long-term predictions of chaotic systems are unreliable due to the extreme sensitivity to initial conditions and the fast accumulating computing rounding errors. After a number of iterations, the results calculated are no longer representations of the same initial system under study, but a different one.

6 Dealing with complexity

In this section, we will discuss a selection of commendable approaches that are helpful in dealing with complexity. Systems Engineering has brought a radical change into the way we approach problems with undeniable success. However, with increasing complexity, one cannot use this approach. Complex systems do not always behave as defined and expected in classical systems theory. The danger is that the traditional breakdown into subsystems, using emergent property as the defining attribute can and does fail in complex systems. In a Systems approach one works on high, or low level models, but without mixing levels. In a complex system, the different levels may not be independent. This is because of interrelations and couplings that can act between the different levels of subsystems, thus violating a fundamental paradigm in systems theory. Based on this, Bar-Yam [2] has found that the systems engineering approach can be used in not-too-complex systems, but recommends for higher complexity systems adopting an evolutionary paradigm for complex systems engineering that involves rapid parallel exploration and a context designed to promote change through competition. In the absence of a systematic approach, we propose to start with the following three steps.

Step one: granularity

Our first important and wise step is to look at what level of resolution (granularity) our problem to solve lies. An overwhelming amount of data limits modeling detail. The following two examples illustrate scale and granularity.

Cell	→ organ	→ organism	→ ...
Group	→ organization	→ society	→ ...

Structure	→ component	→ functional unit	→ ...
Cluster	→ equipment	→ tool or wider unit	→ ...

In the first example (first two rows), our focus of interest could be at the level of cells and their interactions, for example transport of proteins inside a cell, or perhaps a like a surgeon on a conglomeration of cells as tissue or organ, or a collection of organs as an organism. On a higher level of granularity, for example, Societies do not deal with the mitochondriae or DNA in a cell. Conversely, a microbiologist developing a vaccine does not care about social or political trends.

The second example (last two rows) is a generalization and works in the same way as the first two lines. If we are interested in modeling in budget requirements to build a manufacturing plant we might be interested in what and how many processing bays we need (tool or wider unit) but we do not need to be concerned with the model or brand or the structure of the machinery, except perhaps its processing throughput capabilities or space requirements. We do not need low-level details such as the exact number of how many components there are in machines or equipment or how these components are assembled.

It is however, important to know what *range* of layers of granularity need to be considered in a model, as this is something that affects complexity of the model.

There is yet another strong reason. Doing calculations with very high numbers and very low numbers e.g. as in semiconductor manufacturing ranging in the orders of 10E-7 and 10E14, the calculations will be rounded or truncated. When we have a complex system, it is likely to be made up of several components. These components may well be at a high, low, or medium level of resolution and values in a typical range of orders of magnitude used in their computations, each of them with their own rounding and truncating error. Being numbers of similar order of magnitude makes the partial results from simulation components less affected by error, but when we put them all together, calculations can be quite distorted. One should always avoid calculations with very large and very low quantities together. One way out of this is normalizing the data, but when the models are not linear this may not be legitimate.

This does not warrant that our solution space will be in the same granularity. Granularity can involve selective detailing. This is because not all needs to be modeled and simulated in equal detail. This comes because mixed modeling

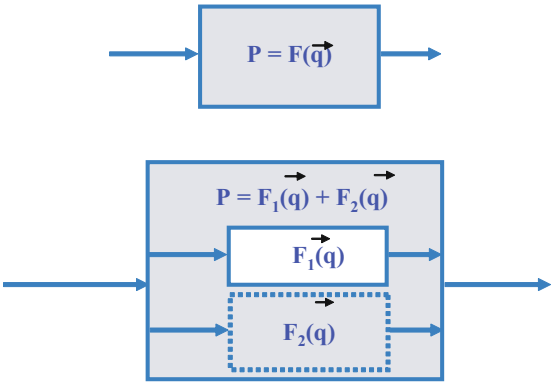


Fig. 5. "Zooming-in" applies to hierarchical systems that can be represented as a flow system

approaches can be integrated. One can model in detail those entities that need low-level detail, and all other entities at a high level under given circumstances, despite violating the rules of general systems theory. This practice of "zooming in" is explained in [19] and shown as a diagram in figure 5. It is about selectively detailing by progressively building in detail. Let us consider that P is a state function of q (they are vectorial functions). Let us consider that this function can be split as the sum of two functions, F_1 and F_2 as shown in the diagram. We can now model some of it in a large resolution (like "zooming in"), and leave all other in a lower resolution. If the original unsplit function P is embedded in another system, then the sum of the high and low-resolution functions F_1 and F_2 has to be carried out before further calculations. We have used this successfully in our dynamic system models on Economy or Energy. We used a rather high level, modular model, with several layers of hierarchy (in Simulink). At some stage, we had to introduce a highly complex function that was only available as a data collection of several decades. We could zoom in on these low-level data by training an artificial Neural Network (NN), and plugging it into our model for future predictions. Here we used the benefit of an NN, which is able to return a weighted function that is able to recognize similar data. The drawback is that the weights or coefficients cannot be interpreted and associated with a cause, as one does in physics. If this interpretation is not necessary, then a NN or other AI technique are suitable zooming in mechanisms that can be within a model of a different, higher resolution.

If a modular, e.g. a hierarchical modeling approach is used, then complexity can built in progressively wherever necessary. One has to ensure that (a) the outcome does map to data with the real world, and (b) that the overall result of each subsystem must be maintained and stable. The calls "Calibrate! " and "Validate! " must resound in our ears and actions.

Step two: time and space

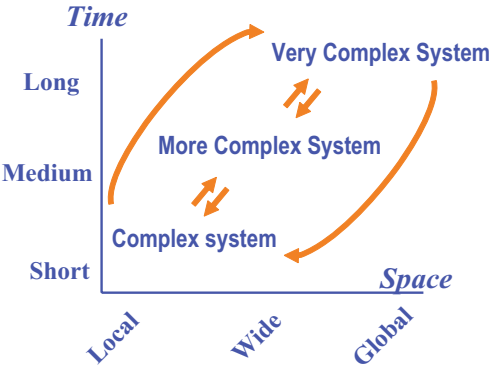


Fig. 6. Temporal and spatial complexity increase

Our second step is to try to ubicate our problem in a spatial and temporal reference. For example, we consider a daily vs. yearly weather forecast, a daily fluctuation vs. yearly stock market trend, metabolism of one cell vs. a whole organism, or a socio-economic system. Figure 6 shows how this applies. This figure is an adaptation from Wallace [24], which the author has used to show temporal and spatial scales associated with different water related issues in the GWSP-2003 project. For our case on increasing complexity, we could add a third axes with a dimensionality scale. Although this may seem a circular reference, it is not. Dimension is a quantity that can equally be represented on a scaled axis. However, complexity does not necessarily increase at higher dimensions; it actually can be reduced, as we will see later.

Step three: repetition

The third step is to identify repetition, and if applicable, one should find out whether the repetition makes the system more complex or not. The difference is that if the repetition can be bundled into one functional entity, then it is unlikely to affect the complexity. To what extent (i.e. the quantity of "how much") the performance of the system is affected can be used as an indicator, tolerance or threshold (e.g. a percentage). Longstaff [14] has made the interesting observation that by adding an element that can be duplicated to a complex system can cause a shift in the total system that is much greater than the amount added. Perhaps this sounds familiar like another version of the emergent property, but one can think of its effect in the context of some kind of information in a social system, e.g. the range of possible consequences of single or repeated broadcasts announcing a disaster evacuation warning that reaches just a few or a whole city.

6.1 Reducing Complexity

We are now going to look into ways for reducing complexity. In some cases reducing complexity can be done easily, and in other cases it cannot and should not be done because of the risk involved that non-obvious correlations are violated leading to distorted results. It can only be done with much care. There are no universal recipes for reducing complexity. In the absence of a theory on complexity, at this time it helps to look at what has worked, and what not. Decomposition and clustering can be powerful techniques in certain cases. While it is not possible to provide a simple prescription for complexity reduction, two important approaches emerge: (a) the reduction of the (often combinatorial) search space by pruning or by expanding into a higher dimensional space; and (b) focusing on the relations and their strength in the formation of hierarchies or clusters.

A wide range of papers about reducing complexity can be found in the literature. Unfortunately and perhaps too often, their application is limited to their own specific cases presented, hence a trend or general recommendation cannot be derived. However, in what follows, a selection of examples will be presented. Their didactical purpose should be seen as *what* the authors have done, rather than the application case itself. The examples should bring the reader into the mental paradigm of reading papers on complexity reduction with a "what can be done" attitude, but not "where can I find a similar work", because it might not apply as a best choice to that researcher's case. This is part of the nature of complexity.

Increasing the space

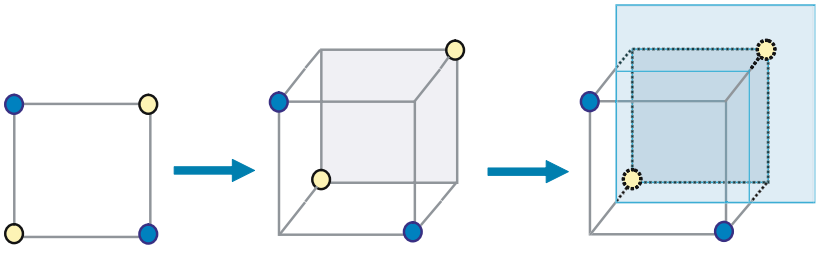


Fig. 7. Increasing the dimension will reduce complexity

A representation (transform) into higher dimensions (in general) will reduce complexity. Sharony [17] has demonstrated the general case that by switching from a 1-D network to a k-D network increases the connectivity but reduces the complexity. Sharony used this for routing in a signal-processing network.

Different ranges of frequencies were assigned to different dimensions to multiplex the signals into different channels.

While increasing the dimensionality to reduce complexity might appear astounding, the effect is that it acts as selection mechanism or filter. The notion is best illustrated graphically as shown in figure 7. Initially the light and dark balls are mixed. After increasing the dimension and applying a selection criterion, the light balls are separated on one side, and the dark balls are on the other side.

A similar notion is used by VanDenBoom et al. [23] where additional effort and expansion lead to complexity reduction. The context of their work is in process industry, dealing with multivariate systems, and typically with input and output constraints. The authors derive an input model for the optimization of a Model Predictive Controller (MPC) model that is suitable for Discrete Event Systems (DES). In this model the authors overcome a computational complexity problem posed by non-linearities and large amount of numerical integrations the object function evaluation by changing the modeling approach into a linear Max-Plus algebra optimization model. The key point in their model is a variability expansion to calculate (approximate) stochastic integrals. They achieve this by introducing a stochastic parameter that allows then applying a Taylor expansion of that parameter, which in turn allows drastically simplifying the expression. A penalty for late production is implicit in the optimization model.

Reducing the space

Yang [25] has demonstrated that construction of hierarchical systems can be drastically reduced by using only those strong relations, which span typically one order of magnitude (in a graph representation this can be a hop from a node in one level to another node in the next level). Yang's work is based on concepts and definitions of 1978 Economy Nobel Laureate H. E. Simon, in particular the hierarchic architecture of complex systems and definitions of near-decomposability of hierarchic systems. Simon's concept is based on the strength (or weakness) of interactions between subsystems from a temporal point of view. This translates into a short-term independence and/or long-term inter-dependence of subsystems. Yang uses a bottom up approach by analyzing elements for strong interactions with each other forming clusters (or subsystems), then moves one level up to analyze the interactions of the clusters (or subsystems) with each other, where those with many strong interactions in turn become a cluster or subsystem on the next level higher up, and so on.

A simpler and perhaps well-known method for reducing complexity is pruning. This can be done when the system can be modeled as a network and

contains parts that are beyond the interest or purpose of observation for whatever reason. Tan et al. [22] has developed a strategy for pruning a network (in this case an Artificial Neural Network (ANN) dynamically. What is different in this work is that this being done dynamically, based on the importance of the data. They achieve this by considering the input vs. output spaces by a mapping. At this point, the reader might recall some of the work of Stoop [20] but now in a different context.

Reduction of search space is also successfully applied by Bartlet et al. [3] to be able to interpret in a cause-effect between control mechanisms and the pulse structure in spectral phase function. The authors achieve this by stepwise parameter reduction and the effect on control pulses such that the effect can be understood and interpreted. The parameter reduction in this case is achieved by controlled physical/chemical experimentation.

Function transforms can be applied to reduce parameter space. Fast Fourier transforms is commonly applied in mathematics and signal processing, in particular sound and image processing. Modeling fishery dynamics involves estimating time-varying parameters about the growth, stock recruitment, mortality rates, fishing, and others. Each parameter varies with time, and requires one value per time-step. This means that for each time-step the parameter space increases. Subbey [21] represents time-varying parameters using functional representations together with Fast Fourier transforms. The authors have achieved reducing a total of 76 parameters to 36, which is about 47% reduction.

Clustering

Clustering appears anywhere: faults and defects in silicon wafers, software, living organisms, cars on a freeway, clouds, chaotic trajectories, etc. Modeling the dynamics of clustering is complex in any discipline. Gibbon & Titi [9] derived a mathematical model for clustering focusing on an "interface" between a cluster and its surroundings. It is based on the existence of a hierarchy of length scales (correlations, interactions), whose members are in competition, for example an ordered set of correlations or coherence lengths. Depending on parameters, the model yield sets "good" and "bad" regions. Due to intermittency (gaps) very short length scales crowd into clusters with fractal-like boundaries.

Modifying the model

Recognizing when and where a mathematical model or technique can be modified to reduce complexity, is good. For decades, chemical plants and oil refineries are known to be complex systems. They involve a considerable variety

of products, which in turn involve a huge amount of equipment. They are dangerous places where any mistake puts human lives at risk. The following example is in Operations Research and Parametric Programming. It involves Model Predictive Control (MPC). This is a specialized technique in process control that is used in chemical plants and oil refineries. MP-QP Multi Parametric-Quadratic Programming can be part of MPC. It requires vast number of calculations and data storage, and it is slow. The complexity (number of regions) for an MP-QP solution can be substantially reduced when covering a pre-specified feasible region Rossiter [16] has found that by splitting and refining the MP-QP algorithm (mathematical model) it is possible to delegate part of the online computation (that slows down the production process) to offline computation. In this way, it is possible to improve speed of online control.

Another example is on decision taking, where strategies are produced as a long-term stable state (wider picture), instead of short-term incidental decisions. In Game theory representations one has typically a dichotomy of decisions (yes/no) in a bounded rationally situation; that is with limited, incomplete or vague info. Mittone [15] has derived a model that incorporates incomplete information into a deterministic frame. To represent all possible moves requires a vast matrix of moves. The model strategy aims for a Nash equilibrium solution, or ideally Pareto equilibrium (no damage to thirds). It then applies an algorithm for maximization of payoff function. After several iterations, eventually the gain stabilizes, yielding an optimal solution. A word of caution: stability may not always be possible.

Reducing complexity can require complex algorithms to achieve success, as demonstrated in the following example. Flexible manufacturing is also considered as a complex system. It involves production with a variety of different products e.g. car models or electronic components, which require alternative combinations of routings. It involves hundreds and millions of routings and their control, whose scheduling and timing is critical. Spiliopoulos and Sofianopoulou [18] have addressed the Cell (or Cluster) Formation Problem (CFP), i.e. the problem of creating manufacturing cells with minimum interactions. The idea is to decentralize production by grouping the machines into clusters and various part types into part families, then allocate the processing of each part family to a machine cluster. This Eliminates 50% stochastic components. The authors apply first a preprocessing step that develops the tree of possibilities. It steps through the tree examining all combinations of process plans and eliminates equivalent sub-problems. It then applies a bounding technique using an Ant Colony Optimization. The second step is then a nested procedure, whose outer loop examines the cost matrice's different combinations, while an inner loop applies an algorithm that solves cell-formation problems.

Successful coalitions are important in carrying out tasks that require cooperation of its agents, i.e. which cannot be fulfilled otherwise. Aknine and Shehori [1] use a method of dependence relations to search in a combinatorial task space with the aim to form successful coalitions. The approach incorporates the prevention of conflict in the formation of coalitions, instead of disregarding conflict and dealing with its resolution a posteriori. The formation of a support tree is somewhat similar to the hierarchy construction developed by Yang [25] mentioned earlier in this section. The support tree serves for negotiations towards coalitions. The weakness of this method is its demanding computing effort.

6.2 Stability

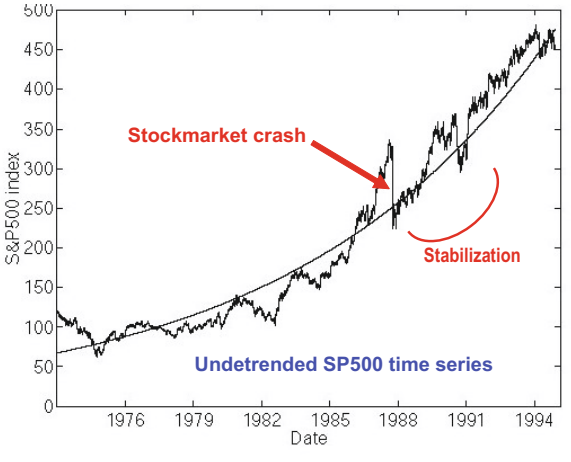


Fig. 8. Example of behavior stabilization reached after a disturbance; in this case the stockmarket crash

First of all, one has to be aware that stability means two things: (a) the behavior stabilization reached after a disturbance of what is being modeled (see figure 8), for example, an economy going through adaptation after a new taxation scheme, or a forest recovering after a bushfire; and (b) the simulation stabilization time until a model replicates truthfully enough what it is modeling (shown in figure 9). This depends strongly on the iteration cycles of calculations, on the sensitivity of the model. Feedback systems can be vulnerable to either of them.

Non-linear and continuously changing relations between many of the entities (or nodes) are typical in complex systems, and anticipating where and when and how much such changes will occur are the very substance that

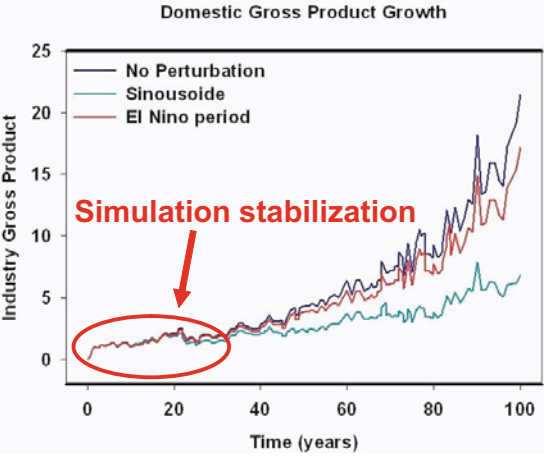


Fig. 9. Example of the unstable initial iterations of a simulation which must be discarded

makes their modeling difficult. This is turn makes it difficult or impossible to predict the system's behavior. Longstaff [14] provides some guidelines for the challenge of managing unpredictable systems. The author cites four responses of biological systems to cope with uncertainty or change; they are (a) prediction; (b) detection and response; (c) broad tolerance; and (d) prevention. These responses can occur in any way: single, as combinations, or all four, and their ratio of application or importance to each other may be case dependent or changing over time. A case of unpredictable complex systems is presented, rich in details in a realistic scenario as the power games by a typically unidentified and self rewarding minority in a society of uninformed players, with undisclosed multiple interactions and inconsistent outcomes of the players' moves, including corruption. These can be modeled with networks where different strengths of relations lead to hierarchies and coalitions. The author warns of a possible failure of attempts to make predictions based on traditional approaches e.g. statistics, boundaries, trends, etc, where the success of decisions under uncertainty cannot be guaranteed, and should be replaced by flexible plans, plans for contingencies and operating procedures in the framework of the four responses of a biological system.

Evolving systems

A special case of unstable systems are Evolving Systems. They are complex systems, which undergo change in any way. This change can be a structural change, or a control change, or any other, and combinations of them. An example are the morphing, appearance and disappearance of giant companies, political entities (e.g. countries), or large ecosystems. Growth can be internally

controlled, or externally. Sometimes these systems can be modeled as dynamic systems, but this is not always the case; in particular when the controlling agent is not clearly identifiable (e.g. a population's movement for democracy and freedom) or the agent changes. A change can be abrupt (e.g. companies) or it may be slow and not readily perceivable. It is interesting to know that non-linear dynamic systems described by three or more equations can suddenly and unpredictably turn into chaotic behavior. Additional input can bring an unstable system into a stable mode, an attractor in a chaotic system, or into deterministic mode. There is no theory about these systems, but early work by Domingo and Tonella [7] explain the requirements for such a theory.

6.3 Unsolvable Systems

In theory, unsolvable systems should not exist. In practice, thy are real problems. Here are some examples, a system can be undetermined, it can have infinite solutions, or it can be intractable, i.e. not computable, or not capable to solve with current computer power. The worst of it is that these systems tend to be mischievous.

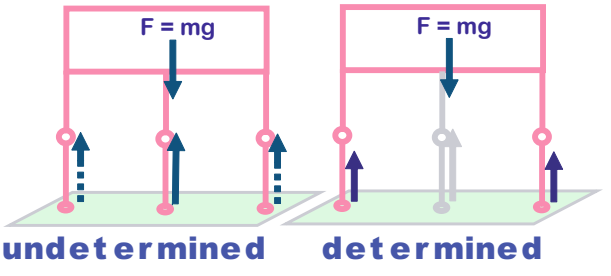


Fig. 10. Examples of an undetermined and a determined system: a robot with three legs has undetermined reaction forces

Here is an example: For decades, Engineers have been eagerly researching ways of locomotion of robots, with an impressive range of results. Keeping the balance in rugged terrain, climbing up stairs are still serious challenges for current technologies. Who would not be tempted to try to imitate nature's success? A spider? Perhaps not an eight but a three-legged robot? Wrong! It is an undetermined system (see figure 10), with too many degrees of freedom. While this is a problem that has an easy solution, others do not. One has to consider and recognize such asituation.

References

1. Aknine, S., Shehori, O.: Coalition Formation: Concessions, Task Relationships and Complexity Reduction. CoRR, volume abs/cs/0502094 (2005), `http://arxiv.org/abs/cs/0502094`

2. Bar-Yam, Y.: When Systems Engineering Fails - Toward Complex Systems Engineering. In: IEEE International Conference on Systems, Man and Cybernetics, October 5-8, vol. 2, pp. 2021–2028 (2003)
3. Bartelt, A.F., Feurer, T., Wöeste, L.: Understanding optimal control results by reducing the complexity. Chemical Physics 318, 207–216 (2005)
4. Basu, S.: On the Combinatorial and Topological Complexity of a Single Cell. In: Proc. 39th Annual Symposium on Foundations of Computer Science, November 8-11, pp. 606–616 (1998)
5. Braha, D., Maimon, O.: The Measurement of a Design Structural and Functional Complexity. IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans 28(4), 527–535 (1998)
6. Braitenberg, V.: Vehicles Experiments in Synthetic Psychology. MIT Press, Cambridge (1986)
7. Domingo, C., Tonella, G.: Towards a theory of structural change. Structural Change and Economic Dynamics 11, 209–225 (2000)
8. Finnigan, J.: The Science of complex systems. Australasian Science, 32–34 (June 2005)
9. Gibbon, J.D., Titi, E.S.: Cluster formation in complex multi-scale systems. Proc. Royal Soc. 461, 3089–3097 (2005)
10. Giles, J.: When doubt is a sure thing (Interview with S. Schneider and R. Moss, Stanford). Nature 418(6897), 476 (2002)
11. Halstead, M.H.: Elements of Software Science, Operating, and Programming Systems Series, vol. 7. Elsevier, New York (1977)
12. Hazen, R.M., Griffin, P.L., Carothers, J.M., Szostak, J.W.: Functional information and the emergenceof biocomplexity. Proceedings of The National Academy of Sciences of The United States of America 104, 8574–8581 (2007)
13. Klein, M., Faratin, P., Sayama, H., BarYam, Y.: What complex systems research can teach us about collaborative design. In: The Sixth International Conference on Computer Supported Cooperative Work in Design, July 12-14, pp. 5–12 (2001)
14. Longstaff, P.H.: Can Unpredictable Systems be Managed. In: IEEE International Conference on Systems, Man and Cybernetics, vol. 2(5), pp. 2013–2020 (October 2003)
15. Mittone, L.: The reduction of decision complexity: normative policies and the role of information. Neurocomputing 69(16-18), 2456–2460 (2006)
16. Rossiter, J.A., Kouvaritakis, B., Cannon, M.: An algorithm for reducing complexity in parametric predictive control. International Journal of Control 78(18), 1511–1520 (2005)
17. Sharony, J.: The Universality of Multi-dimensional Switching Networks. IEEE/ACM Trans-actions on Networking 2(6), 602–612 (1994)
18. Spiliopoulos, K., Sofianopoulou, S.: Manufacturing cell design with alternative routings in generalized group technology: reducing the complexity of the solution space. International Journal of Production Research 45(6), 1355–1367 (2007)
19. Sitte, R.: Zooming-In Modeling Method For Managing Complexity. International Journal of Modeling and Simulation 21(4), 104–109 (2002)
20. Stoop, R., Stoop, N.: Natural computation measured as a reduction of complexity. Chaos 14(3), 675–679 (2004)
21. Subbey, S., Howell, D., Bogstad, B., Åsnes, M.: Reducing fisheries model complexity using a finite Fourier series reparameterization. Fisheries Research 84, 390–394 (2007)

22. Tan, S.C., Rao, M.V.C., Lim, C.P.: On the reduction of complexity in the architecture of fuzzy ARTMAP with dynamic decay adjustment. Neurocomputing 69, 2456–2460 (2006)
23. van den Boom, T.J.J., Heidergott, B., De Schutter, B.: Complexity reduction in MPC for stochastic max-plus-linear discrete event systems by variability expansion. Automatica 43, 1058–1063 (2007)
24. Wallace, J.S.: HELP bridging scales in water science, management and policy. In: Proceedings of the International Congress on Modeling and Simulation, Townsville, July 14-17, pp. 428–433 (2003)
25. Jianmei, Y.: An Application of Simon's Theory on the Architecture of Complex Systems. IEEE Transactions on Systems, Man, and Cybernetics 23(1), 264–267 (1993)
26. Zhang, S.B., Huang, J.C., Zhang, W.M., Liu, Z.: Research on Parallel Decision Analyzing for Complex System of Systems. In: IEEE Proceedings of the Fifth International Conference on Machine Learning and Cybernetics, Dalian, August 13–16, pp. 1812–1817 (2006)

What Makes a System Complex? – An Approach to Self Organization and Emergence

Michel Cotsaftis

ECE
53 rue de Grenelle
75007 Paris France
mcot@ece.fr

Summary. The fast changing reality in technical and natural domains perceived by always more accurate observations has drawn the attention on a new and very broad class of systems mainly characterized by specific behavior which has been entered under the common wording "complexity". Based on elementary system graph representation with components as nodes and interactions as vertices, it is shown that systems belong to only three states : simple, complicated, and complex, the main properties of which are discussed. The first two states have been studied at length over past centuries, and the last one finds its origin in the elementary fact that when system performance is pushed up, there exists a threshold above which interaction between components overtake outside interaction. At the same time, system self-organizes and filters corresponding outer action, making it more robust to outer effect, with emergence of a new behavior which was not predictable from only components study. Examples in Physics and Biology are given, and three main classes of "complexity" behavior are distinguished corresponding to different levels of difficulty to handle the problem of their dynamics. The great interest of using complex state properties in man-made systems is stressed and important issues are discussed. They mainly concentrate on the difficult balance to be established between the relative system isolation when becoming complex and the delegation of corresponding new capability from (outside) operator. This implies giving the system some "intelligence" in an adequate frame between the new augmented system state and supervising operator, with consequences on the canonical system triplet {effector-sensor-controller} which has to be reorganized in this new setting. Moreover, it is observed that entering complexity state opens the possibility for the function to feedback onto the structure, ie to mimic at technical level the invention of Nature over Her very long history.

Keywords: Complicated Systems, Complex Systems, Self-Organization, Emergence, Invariance, Nonlinear Control of Natural and Man-made Systems

1 Introduction

After a very long period of observations, it became very slowly clear to first Human observers since prehistoric age that the phenomena they were distinguishing were due to a specific order in Nature and not to the capricious will of mysterious Gods. Since then, they have been over the centuries patiently organizing the most visible ones, and, proposing hypotheses to find this order, they finally end up on laws to represent the simplest phenomena first which have been verified in the framework of actual observations. These elementary phenomena were mainly concerning "simple" systems which can reasonably be isolated in their dynamics and their observation.In the mean time, the advance in technology and observation accuracy was driving the attention on more complicated systems with always larger number of elements, which in the 17th century have been shown by R. Descartes to be reducible in their study to the former class of simple ones[1].

Finally, the result today of this long quest was an "adapted" representation of the Universe with a broad "classical" part for "human size" phenomena, corrected by "quantum" effects at very small atomic level, and by "relativistic" ones at very large galactic level. This basic "1-dimensional" picture emerging from millennium long effort of human kind, and leaving fundamental questions only at infinitely large universe and infinitely small super elementary particle levels, is however not sufficient to represent correctly enough the present situation where systems exist over an extremely large parametric domain now easily explored with the recent development of always more performing technology over the last decades.

First observed from time to time in different domains, there is now evidence of existence of new huge class of systems, natural as well as artificial in all scientific and technical human activities, which have reached their own status by the corner of the millennium under the name of "complex" systems. They are in some sense opening a "2-dimensional" more global picture of the universe by being located along a direction "perpendicular" to previous 1-dimension line going from infinitely small to infinitely large size, see Fig.1 at the end of the text.

There is today a strong questioning about their origin and their formation[2]. Some have even suggested that previous "mechanistic" approach to the universe by researching "ultimate" fully unifying law[3] is just mistaking their very and irreducible structure and that a new "revolution" is necessary to grasp all aspects of known universe today[4].

To address these questions, a very pedestrian approach is proposed here, based on elementary source-sink model applied to the graph representing the aggregate of possible components of a system. Then it evidently appears that

system structure can be divided into three different groups, simple, complicated and complex, corresponding to specific properties. The first two groups are usual ones nicely approachable by the methods of scientific reductionism[5]. The third group, by its very global nature, cannot be just reduced to the effect of its components[6] and requires some adjustment for being correctly handled, because now the key point is the way the system behaves under (or against) the action of its environment.

It will appear that the mechanistic notion of individual "trajectory" is loosing its meaning and should be replaced by more general "manifold" entity corresponding to accessible "invariants" under environment action.

System self-organization and emergence properties are deduced and discussed. Different grades of complexity are fixed depending on specific system properties which relate to well known classes of observed phenomena. Advantages for application of complex structure to man made systems are stressed.

2 System State Analysis

Let consider first a system with a finite number **N** of identified and distinguishable components which can be represented by a graph with **N** nodes N_i . On this graph, three types of vertices can be drawn in between the **N** components i and outside sources e whenever an exchange exists between them. They correspond respectively to free flight state for vertex V_{ii}, to driven state from outer source for vertex V_{ie}, and to interactive state with other system components for vertex V_{ij}., see Fig.2.

System dynamics are usually resulting from a combination of previous three different exchanges to which, on a very general setting, there can be associated three characteristic fluxes for each system component. Their specific nature (power, information, chemical,..) is depending on the problem at hand, but is here unnecessary as long as they unambiguously characterize system components status.

The first flux corresponds to the "free" dynamics of i-th component p_{ii} along vertex V_{ii}, the second one $p_{i,e}$ to the transfer flux between outer source and system i-th component along vertex V_{ie}, and the last one $p_{i,j}$ is the characteristic and oriented flux exchange between components i and j along vertex V_{ij}, see Fig.2.

Let $p_{i,int} = \mathrm{Inf_j}\ |\mathrm{p_{i,j}}|$. Depending on the various possible ordering between the absolute value of these three fluxes, the system behaves in different ways.

When **(A)** $|p_{ii}| >> p_{i,int}, |p_{i,e}|$ the i-th component is weakly coupled to other components. Its dynamics are mainly the ones of an isolated element slightly perturbed by the other actions, and the system reduces to a set of almost independent one-component sub-systems. In this sense it can be considered as a **simple** one.

When **(B)** $|p_{i,e}| >> p_{i,int}, |p_{ii}|$ the i-th component is mainly depending on outside sources. The action of the other components creates a (weak) coupling between the components, but the system can nevertheless be decoupled, at least locally, into a set of sub-systems which can be acted upon and controlled by as many exterior sources as there are components in the system because they can still be identified. In this sense the system can be termed as **complicated** from etymologic sense (from Latin cum pliare : piled up with).

Finally when **(C)** $p_{i,int} >> |p_{ie}|, |p_{ii}|$ the i-th component is very strongly coupled to other ones and its dynamics are now mainly determined by components interaction satisfying the inequality. If apparently there does not seem to exist a dramatic difference with the previous case as concerns their dynamics, there is however a fundamental one as concerns the effect of outside action. In previous case input from the source can be tracked to the concerned component so this degree of freedom can be completely controlled from outside, whereas in present one the nature of interactions dominates and shields this tracking.

As a result, control action can only be a "global" one from other system components satisfying condition for second case (B) and system dynamics are now also driven by internal action. Consequently a totally new situation occurs for all system components passing in third state in the sense that their control cannot as in second case be fixed by only outer source action. Effectively, it is no longer possible as in second state to determine completely all system components dynamics from outside because of stronger interaction effect which now dominates the dynamics of concerned components. In a sense a self-organization has been taking place inside the system which leads to an internal control replacing the classical one from outside. So it is no longer possible to continue to manipulate inputs with as many degrees of freedom as the system initially has because of the mismatch which would result from the conflict with the internal control due to components interaction satisfying condition for the third case.

External system control dimension is thus reduced. A system in this case will be termed as **complex** in agreement with etymology (from Latin cum plexus: tied up with).

According to this definition, a very elementary test for determining if a system is passing to complex state is thus to verify that its control requires the manipulation of less degrees of freedom than the system has initially.

There exists an immense literature about complexity, its definition and its properties covering an extremely wide range of domains from Philosophy to Technology[7], especially in recent years where its role has been "discovered" in many different fields such as networks now playing a crucial role with the ascent of Information Technologies[8].

Here emphasis is more modestly put only on more restricted complex state compared to complicated state as concerns action from outside environment (ie from control point of view) onto the system.

For a system S with a finite number $\mathbf{N}$ of identified components, one can then define its index of complexity $\mathcal{C}_{\mathcal{S}} = \prime$ if it is only in complicated state, and $\mathcal{C}_{\mathcal{S}} = \backslash/\mathbf{N}$ if $n < \mathbf{N}$ components cluster in complex state by satisfying inequality (C) and become insensitive to outer action. $\mathcal{C}_{\mathcal{S}}$ measures the grasp on the system from outside action/control and the limit $\mathcal{C}_{\mathcal{S}} = \infty$ corresponds to totally autarchic system, the most complex possible structure with this number of components.

This may look paradoxical as the definition of a simple system is precisely that it can be isolated. In fact this apparently contradictory statement is resulting from the very nature of internal interactions effect which reduces the number of invariants on which system trajectory takes place.

An elementary example is given by the particles of a neutral gas for which their initial $6\mathbf{N}$ positions and velocities (the mechanical invariants of motion) are reduced to the only energy (or temperature), justifying thermodynamic representation.

In other words, when represented from outside, a system is the less complicated as it is more complex inside. Differently said, a complicated system remains complicated whether observed from inside or from outside, whereas when becoming complex it is less complicated when observed from outside, on top of the fact that being depending on a restricted number of parameters, it is more insensitive to outer action.

However, when observing from outside a system with apparent $\mathcal{C}_S = n/\mathbf{N}$, if $\mathbf{N}$ is not fixed when the system is not well known, there is no one-to-one relation in general between apparent outer behavior of the system and its initial internal structure. It is only possible to check relative variation of $\mathcal{C}_S$ when crossing condition (C) for some components.

On a general setting, it is very easy from its definition to envision complex state as a powerful way set up by Nature to generate more sophisticated

entities able to gain at the same time more independence with respect to their environment, or, in other terms, to become more robust to its variations and yet, to appear much less complicated outside, so that they can in turn link more easily to similar entities and generate a chain of nested clusters.

This seems to be an underlying trend in evolution on Earth which has been allowing the elaboration of living beings, the up-most advanced form of complex systems[9]. Summarizing, exactly like matter does appear in three states : solids, liquids and gas, depending on conditions, systems are exhibiting three states : simple, complicated and complex.

A very simple way to locate a system with respect to them is to plot for each component i the three values $[|p_{ii}|, |p_{i,e}|, p_{i,int}]$ in a three dimensional space where each direction represents a single state. Each component is then represented by a point C_i and the system by a cluster of **N** points, see Fig.3 where the three domains corresponding to satisfaction of inequalities **(A),(B)** and **(C)** are respectively delimited by prismatic domains $\mathcal{P}_A$=$[(O, |p_{ii}|), (O, \delta_1), (O, \Delta), (O, \delta_3)]$, $\mathcal{P}_B$=$[(O, |p_{ie}|), (O, \delta_2), (O, \Delta), (O, \delta_1)]$ and $\mathcal{P}_C$=$[(O, p_{i,int}), (O, \delta_3), (O, \Delta), (O, \delta_2)]$.

Depending on where they are located in this space, it is evident to figure out the status of the system with respect to inequalities **(A),(B)** and **(C)**. Moreover, it is easy to check what will be the consequence of parameter variation, especially when crossing inequality **(C)**, a point extremely important for control of man-made systems as seen later.

In this picture, components for which inequality $p_{i,int} >> |p_{ie}|, |p_{ii}|$ holds are in fact an internal sub-cluster which cannot be further split from outside observation and action. As indicated above, the first two states have been observed and studied along past centuries.

The fact is that with modern and very detailed observation diagnosis, complex state has been now very often observed in natural systems, and becomes the most common one in a broad range of phenomena.

For artificial man-made systems, even if it were sometimes crossed in the past, it is only with recent progress of advanced technology that, when pushing systems performance for higher efficiency, the threshold for overtaking of internal interactions effect corresponding to third case is also very routinely over passed now.

So complex behavior is seen in many situations, and becomes an object of study in itself, with considerable consequences in technical and industrial applications it implies in conjunction with the corresponding demand of more delegation of decisional nature to the systems for better efficiency. From

previous and relatively rough source-sink representation always applicable in a first instance, it is already possible to stress a deep difference existing between the first two states and last complex one. In first two states, the possibility exists to split the system into as many independent one-component systems in a first approximation, whereas this is impossible in third state where all interacting components have to be taken as a whole.

In mathematical terms the consequence is that usual approximation methods developed for the first two states do not straightforwardly apply and have to be revised in order to handle the global aspect of the coordinated response of components in complex state. This difficulty is at the origin of today important computerized research undertaken on the problem.

3 Emergence

When going to more specific situations, other elements are also entering the description and it is interesting from elementary global description above to recover various situations which have been observed and analytically studied for different parameters value.

First it is likely that in general the system is not always in a "pure" state and often exhibits a mixed structure where some components are in one state and others in another one.

An important example is the case of natural inhomogeneous but continuous natural systems such as fluids with non zero gradients in a domain. In these systems, fluctuations are universally observed the source of which is the free energy available in between this stationary equilibrium and complete (homogeneous) thermodynamic one.

In present case, the free energy is coming from the space gradients related to medium non homogeneity. They have in general a very large range of fluctuations (roughly because the system has a very large number of components) which can be split into two groups depending on their wavelength compared to system characteristic gradient length.

Typically, the fluctuations with large wavelength excited in the medium are in complicated state and, because they are sensitive to boundary conditions, can be observed and possibly acted upon as such from outside. Fluctuations with small wavelength on the other hand are generally in complex state due to parameter values.

So under their strong interactions, and because they are much less sensitive to boundary conditions, they are loosing their phase and globally excite a

leak out of the fluid (usually called a transport) manifested by an out-flux expressing the non equilibrium situation of the system, and which counteracts the input flux responsible of medium non homogeneity. The determination of these transports is a very important element for qualifying the behavior of the system and is an active research problem studied worldwide.

This feature is observed for all natural systems when they enter the dissipative branch. A natural system is called dissipative[10] when it is exchanging (particles and energy) fluxes with outside environment.

Evidently the channels by which internal energy sources are related to these fluxes are playing a privileged role because they regulate the energy ultimately available for the system and finally determine its self-organized state[11]. Dissipative systems more generally only exist to the expense of these fluxes, and they evolve with parameter change such as the power input along a set of neighboring states determined by branching due to bifurcations where internal structure changes in compatibility with boundary conditions and by following the principle of largest stability.

So the picture of such systems is a transport system governing flux exchanges guided by the bifurcation system which, as a pilot, fixes the structure along which these exchanges are taking place. Finding the branching pattern thus entirely defines the possible states of the system and determines the fluctuation spectrum. Branching is found as nontrivial solutions of variation equations deduced from general system dynamics equations.

Despite the physical origin of phenomena is here well identified, the analysis is still in progress in a very large number of situations, such as in thermonuclear plasma to determine correct parameter sizing for realizing "burning" conditions[12], and in neutral fluids where turbulence is not fully explained yet[13], whereas the problem has been cleared up for fluctuations in deformable solid bodies[14].

In most cases the difficulty stems from finding an adequate representation of the global effect of small wavelength fluctuations, especially when due to parameter values, they feedback on longer time scale large wavelength oscillations and modify their dynamics. In this case, there is a significant change of initial system dynamics due to passage in complex state of an internal part of the system generated by branching phenomenon.

Moreover, it is very likely that modification of system dynamics is the more important as the non homogeneity zone is the less "transparent" between the two domains above and below along power flow. This has been extended to the extreme with living cell systems which have been able to completely encapsulate within a filtering membrane (ie a steep gradient) a space domain

where very specific "memory" DNA molecules are fixing the dynamics of inside system they control, with corresponding exchange across the membrane.

It is easy to understand that conjunction of a gate and of adequate information is the most efficient way to deeply modify locally the effect of regular physics laws, as exemplified by the counter streaming motion of a piece of wood in a river with one stage locks when regulating their opening and closing.

However, dissipative behavior does not exhaust all possible situations and many other ones do not follow this pattern. Natural systems so far are exhibiting components with relatively elementary features (charge, mass, geometrical structure, chemical activity, wavelength, frequency..), but there are also cases where complex state occurs in systems with more sophisticated components, usually called "agents".

Examples are herds of animals, insect colonies, living cell behavior in organs and organisms, and population activity in an economy. In all cases, when observed from outside the systems are exhibiting relatively well defined behaviors but a very important element missing in previous analysis is the influence of the goal the systems are seemingly aiming at. Very often the components of these systems are searching through a collective action the satisfaction of properties they cannot reach alone, and to represent this situation the specific word "emergence" has been coined[15].

The point is that it is now possible to return back to previous case and in a unified picture to envision the laws of Physics themselves as emerging phenomena. For instance for an ensemble of neutral particles with hard ball interactions, and beyond the threshold of rarefied gas, (ie when the Knudsen number $K_n = \lambda/L$ is decreasing to 0 from the value 1, with λ the particle mean free path and L a characteristic length scale), the particle system is suddenly passing from a complicated to a complex state (due to the huge value of Avogadro number).

Consecutive to overtaking of collective interaction by collisions (expressed by decrease of Knudsen number), it could be said that there is emergence of a pressure and a temperature, which, from a point of view outside the gas, summarizes perfectly well the representative parameters (the invariants) describing it at this global level and amply justifies the thermodynamic representation in this case, see next Paragraph. The elementary reason is that here the fast particle dynamics described by Boltzmann equation relaxes within extremely short collision time to its equilibrium Maxwell-Boltzmann-Gibbs distribution only depending on temperature (and the potential of applied forces if any), which justifies usual stochastic methods[16].

Of course it is tempting at this point to "forget" the underlying mechanical base and to decide that perfect gas law, if any, is the emergence directly resulting from billiard ball "properties" of interacting (agents) molecules. The difficulty in this case is that application of "ergodic" hypothesis[17], even if it can mathematically be invented from scratch, rests here heavily upon thorough analysis of molecule ensemble dynamics, ie on mechanical laws.

Similar problem does exist at atom level itself where, after baryons are assembled from primitive quark particles below some threshold energy, protons and neutrons assemble in turn themselves below another lower energy threshold into ions with only mass and charge parameters, able to combine finally with electrons to create atoms.

Speaking of "emergence of atoms" from underlying components does not add any more information to classical approach. So in an opposite direction, a more conservative and probably more appropriate way to account for the new situation is to just state that when becoming complex, systems are exhibiting emergence of self-organization out of which it has been verified that their new behavior in complex state is always a consequence, without going into the details of this behavior which is depending on specific system parameters.

In fact it is quite elementary from previous source-sink model to understand that a key point is in the accuracy of modelling the components in complex state, as long as the resulting "invariants" which will grasp all system information for interaction with environment are directly depending on this modelling.

This has been at the origin of a computer "blind" search where the agents are given properties and "emerging" behavior is obtained in a bottom-up approach, sometimes in surprising compatibility with experimental observations[18], in parallel to theoretical analysis[19].

Of course following classical science reductionism, there remains to justify the choice of agent properties, and to show that they are in one-to-one relation with observed behavior. As this does not seem always possible, a different holistic approach[20] has been proposed where the research of ultimate underlying elements from which everything is constructed by as many layered shells on top of one another, is considered as unnecessary, to the price of loosing prediction power. Finally an extremely important remark is that the logical chain :

[stimuli/parameter action]
→ **[higher interactions between system components]**
→ **[passage to complex state]** → **[system self-organization]**

discussed here is nothing but the sequence leading to the final step of system evolution toward more independence, and which is the feedback of "function" onto "structure", a specific property of living organisms explaining their remarkable survival capability by structure modification.

4 Mathematical Analysis of Complex Systems

The few previous examples from common sense observation illustrate the elements which have been described above and which are providing a general base for complex system paradigm. From atomic nucleus to galaxy natural systems are seen to be constituted by aggregates of identifiable components (which, as already stressed, can be themselves, at each observation level, aggregates of smaller components) with well defined properties. These aggregates have been said to exhibit a complex behavior when interaction between the components or some of them is overtaking their interactions with exterior environment.

Similarly living beings are exhibiting the same behavior, as observed with gregarious species, and in artificial man-made systems the same phenomenon is occurring when the overtaking conditions are satisfied. This is for instance the case for high enough performance level systems because the components are then tightly packed, as for high torque compact electrical actuators. Despite an extremely large variety of possible situations there are few base interactive processes leading to complex self-organization. Three main types will be discussed: the reducible case, the instantaneous differential case, and the integro-differential case with past neighborhood memory.

4.1 A-Reducible case

It is first intuitively expectable that the approach to full problem gets easier when time-space scales are very different for components going to complex state as compared to the ones of other system components. This is mostly the situation in dissipative systems where characteristic frequency and wavelength of the spectrum of complex system part are respectively much larger and much smaller than the ones of the other (non complex) part of the system.

The main reason is as explained above that components in complex state are produced in the initial system by the effect of bifurcation determined from variation equations of the system and are corresponding to high mode numbers (and have small enough wavelength). For such modes, their saturation amplitude is fixed by their nonlinear interaction which overtakes the level they would have reached under the only effect of the source and, as well known, exhibit globally a fluctuation spectrum usually very difficult to determine analytically.

Let τ_s and τ_{cp} be respectively the characteristic response time of the non complex system part and of complex modes part with $\tau_{cp} = n\tau_s$.

The simplest situation n > 1 corresponds to the case where the interactions are strong and frequent enough to kill corresponding modes dynamics. In this case and because there are usually many modes with small amplitude, complex system part dynamics can be reduced to "thermodynamic" approximation with a random distribution of fluctuation spectrum.

Then the initial system of equations describing the full system splits into a restricted set of identified modes perturbed by a random term. This unifying and powerful model is quite successful in science and engineering[21] but it does not cover all situations.

A more difficult and yet common case is occurring when n = 1, for which small $O(\epsilon)$ modes dynamics have long enough coherence characteristic time comparable to τ_s. To account for this coherence, asymptotic expansion of small modes dynamics from complex part dynamic equations is required up to first order to be injected in non complex system part dynamics, because ϵ order term acting coherently over a time $1/\epsilon$ produces a non negligible finite order term. The system again reduces to its non complex part but now modified by coherent effects of complex system part dynamics, see Fig.4.

The unsolved situations in [12,13] are belonging to this case. However it will nevertheless be termed as reducible because full system dynamics can be projected onto non complex system subpart.

It clearly illustrates the fact that, given a system with a fixed number of degrees of freedom and associated initial and boundary conditions, this number should increase at bifurcation crossing so other conditions would be required to determine the solution.

In fact this is not the case here as new complex modes dynamics can be solved by asymptotic procedure and re-injected into non complex system part which obviously keeps the same number of degrees of freedom. The only constraint is on initial complex modes amplitude which, due to their smallness, obey central limit theorem and are randomly distributed. So the difference with previous case n > 1 is that only initial amplitudes values are random now and follow afterward their own dynamics whereas they were previously random for all time.

In both cases the number of final degrees of freedom remains the same, so despite their number has apparently increased at bifurcation, in reality the system is still of the same dimension. The extremely important consequence is that for controlled man-made systems, when their dynamics cross a

bifurcation point, the same number of controllers is still required but they should be adapted to explore a new larger function space. The main effect of the bifurcation has been to augment system response to a wider function space, and the motion takes place on a manifold of this new space so that at the end the effective number of degrees of freedom is remaining the same. As stressed earlier, this behavior is the manifestation of system self organization. The double mechanical-thermodynamic feature is extremely general and characterizes complex reducible systems where complex part is added by bifurcation crossing, see Fig.5.

Again control of such systems (whether natural or man-made) is not classical because for $n \geq 1$ the new modes created after bifurcation crossing, though sometimes observable, are not accessible from outside. Because of their definite effect on system dynamics, a new more balanced approach respecting internal system action has to be worked out, which, very generally, is aimed at replacing inability of control action by insensitivity to variation for inaccessible complex modes.

Though apparently loosing some hand on such systems, it has been surprisingly possible along this line to find explicit conditions in terms of system parameters expressing somewhat contradictory high preciseness (by asymptotic stability condition) and strong robustness (against unknown system and environment parts)[22]. In this way system dynamics are finally controlled and asymptotic stability can be demonstrated, but in general the price to pay is a not necessarily decreasing exponential asymptotic type.

4.2 B-Instantaneous Differential Case

The reducible case mainly deals with systems in partly or totally complex state, as with bifurcation crossing by some varying parameter, and with small parameter ordering between time and space scales, so it is possible to proceed to asymptotic expansion.

More generally, a system may be in complete complex state, examples of which are atomic nucleus, herd of animals, and galaxies. Despite their very different space sizes, the systems exhibit always the same base characteristic feature to finally depend on an extremely restricted number of parameters as compared to the aggregate of their initial components. Searching the way to extract directly the remaining "control" parameters of such systems from their dynamics is a fundamental issue which today motivates a huge research effort worldwide, especially in relation with information networking.

Extensive analytical and numerical study has been developed for differential systems of generic form

$$\frac{dX}{dt} = A(t)X + \lambda F(t, X, u(X,t)) + \mu S(t) \tag{1}$$

where $X = \text{col}\{X_1, X_2, ..X_n\}$ is system state space, are n-vector coupling parameters, and $A(.)$, $F(.,.,.)$, $S(.)$ are three specific $q_1 \times q_2$ matrix terms (q_1, $q_2 \leq$ n) corresponding to free flight, nonlinear internal interactions and source terms respectively (the linear and source terms in the right hand side of eqn(1) are here split apart to indicate their respective role).

The third variable $u(X,t)$ in F function accounts more generally for the possibility to feedback evolution of $X(t)$ onto its own dynamics as it often occurs in systems when splitting parameters into given and manipulated control ones. For fixed $u(.,.)$, depending on the value of, the system will be in simple, complicated or complex state described in §1.

When increasing the components of the system runs into complex state, and it has been repeatedly observed, especially on systems close to Hamiltonian ones[23] and with scalar , that system representative point in n dimensional state space follows a more and more chaotic trajectory when crossing bifurcation values and at the end fills up a complete domain[24].

Of course sensitivity is largest when the system exhibits resonances, ie is close to conservative, and adapted mathematical expansion methods have to be worked out[25].

Because systems are basically non integrable[26], this is a direct evidence of increasing effect of internal interactions which reduce system dynamics to stay on attractor manifold of degree $p < n$, so that system dynamics are now layered on this manifold. This also expresses the fact that trajectories on the remaining n−p dimension space are becoming totally indistinguishable (from outside) when taken care of by internal interactions of n−p components going to complex state.

So system trajectories reorganize here in equivalence classes which cannot be further split, a dual way to express the fact that there exists an invariant manifold on which system trajectories are lying. It is easy to understand that continuing to control these components by regular previous control[27] worked out for complicated state and specially designed for tracking a prescribed trajectory, is no longer possible and that a new approach is required which carefully respects internal system action due to complex state self-organization.

More global methods of functional analysis[28] related to function space embedding in adapted function spaces[29] by fixed point theorem[30] are now in order as shown for reducible case[31], because they are providing the correct framework to grasp the new structure of system trajectory which cannot be fully tracked as before. Basically the method is again to counteract

impreciseness in an element by robustness to its variation, a method very largely followed by living organisms.

4.3 C-Integro-Differential Case with Past Neighboring Memory

Another important dimension relates to the properties system components are given, and a very influential one is the range of inter-component interaction, because this determines completely the build up of system clustering when becoming complex. Obviously long range interactions are leading to more intricate response with more difficult analysis.

Examples are stars in a galaxy, electromagnetic interactions between ions and electrons in a plasma, animals in a herd and social behavior of human population in economic trading such as stock market with internet link.

In all cases a new element is coming from the size of the neighboring domain each system component is sensitive to, and implies a time extension to past neighboring components trajectories, see Fig.6.

So the resulting complex behavior is more generally determined by interactive component effects over a past time interval and weighted according to their importance. In this case systems are obeying for $t = t_0 > 0$ delayed equations of the form

$$\frac{dX}{dt} = F(t, X(t), I(t), u, d) \tag{2}$$

$$\frac{dI}{dt} = \int \rho(\lambda, t) G\big(t, X(t), \int_{-\infty}^{t} \phi(t') X(t - h(t', X(t'))dt', \lambda\big) d\lambda \tag{3}$$

with $X = \mathrm{col}\{X_1, X_2, ..X_n\}$, $I = \mathrm{col}\{I_1, I_2, ..I_q\}$, which clearly indicates the role of the past through intermediate "moments" $I(t)$ gathering a set of elements G with a weight (,t), each G being itself constructed on the past history of system trajectories $X(t)$ with fading weight (t), and where $F(.,.,.)$: $\mathbf{R}_1^+ \times \mathbf{R}_n \times \mathbf{R}_q \times \mathbf{U} \times \mathbf{D} \rightarrow \mathbf{R}_n$ and G(.,.,.,.) : $\mathbf{R}_1^+ \times \mathbf{R}_n \times \mathbf{C}_0(\mathbf{R}_1^-, \mathbf{R}_n) \times \Lambda \rightarrow \mathbf{R}_p$ are specific vectors.

Their expression generally include control parameters $u(.) \in \mathbf{U}$ which have to be stated explicitly by control law to fix the structure of eqns(2,3) and uncontrolled disturbing effects often represented by a noise $d(.) \in \mathbf{D}$.

Here the delay $h(., X(.)) < t$ can be state-depending as it appears in the analysis of natural (biological) systems where information from the past can be shielded by the state itself. The mixing from the breath of trajectory coverage and from past memory basically creates a "double" complexity in time and space for system components, slightly reduced by averaging effect of acting moments I(t). It indicates that component trajectory is more constrained

as it is constantly depending on neighboring ones, and a possible way is to represent this collective action as an average potential through a distribution function equation of which eqns(2,3) are the characteristics.

Apart few existence theorems[32], study of eqns(2,3) is at very preliminary stage today. Like for eqn(1), their numerical study can be undertaken by fixing design and control parameters in F and G such that some trajectory properties are met.

Here also a useful companion approach is in application of embedding theorems by fixed point methods, especially when F and G are satisfying generalized Lipschitz conditions, which often provides an embedding property in Sobolev spaces by use of substitution theorems[33]. More precise statement can be obtained when there exist upper bounding estimates in norm of the RHS of eqns(2,3).

For example, if the right hand sides of eqns(2,3) satisfy inequalities of generalized Lipschitz type[34]

$$|F|, |G| \leq a(t) + \sum_{i=1,n} \Big[b_i(t) g_i(Z(t)) + c_i(d_i(t)) k_i(Z(d_i(t))) \Big] \tag{4}$$

where $Z = \|\text{col}\{X, I\}$, $\mathrm{d_i(t) = Inf_X\{t - h_i(t, X(.))\}}$, and the various functions $\mathrm{g_i(.)}$, $\mathrm{k_i(.)}$, $\mathrm{a_i(.)}$ and $\mathrm{b_i(.)}$ are positive in their definition intervals, there exists the following bound on the solution

$$Z(t) \leq \mathbf{H}^{-1} \left\{ \mathbf{H}_0 + sgn(\mathbf{H}) \int_{t_0}^{t} \Phi(s) ds \right\} \tag{5}$$

with $H(Z) = \|\text{col}[1, g_i(Z), k_i(Z)]\|, \Phi(t) = \text{col}\big(a'(t), b_i(t), c_i(t) d_i'(t)\big)\|$, '= d/dt and $\mathbf{H}(\mathrm{x}) = \int^x ds/H(s)$, $\mathbf{H}_{-1}(.)$ its inverse function.

The solution $X(t)$ is bounded with $Z(t)$ defined over the complete time interval $t \in [t_0, +\infty] \subset \mathbf{R}_1^+$ and for all $Z_0 = Z(t_0) \in \mathbf{R}_n^+$ when $\mathbf{H}_{-1}(.)$ is bounded which generally occurs when the function is sub-linear, whereas there is a conditional bound inside the domain limited by $H_0 + \text{sgn}(H) \int^T \Phi(s) ds = 0$ in (H_0, T)-space, ie in (X_0, T)-space, when the function is super-linear.

When there exists monomial bounds $g_i(x) = \lambda_i x^p, k_i(x) = \mu_i x^p$ and $a(t) = 0$, one obtains the explicit bound

$$Z(t) \leq \left[Z_0^{1-p} - (p-1) \sum_{i=1,n} \left\{ \lambda_i \int_{t_0}^{t} b_i(s) ds + \mu_i M_i \int_{d_i(t_0)}^{d_i(t)} c_i(s) ds \right\} \right]^{\frac{1}{1-p}} \tag{6}$$

with $M_i = \text{Max}[1/d_i']$, showing a limit in $(t, Z(t_0)$-plane for p>1 when the term between bracket is 0.

Bounds in eqns(5,6) exhibit double advantage of being analytically meaningful and tractable, and to naturally create equivalence classes within which system dynamics are globally the same. This is exactly the expression of system natural robustness resulting from the passage of some components to complex state which can be here easily exploited.

An interesting observation from such an approach is again the emphasis on importance of the manifold on which system trajectories are staying, much more than one specific trajectory which has been seen above not to contain enough information for generating an outer action to control it. Another point is the way stability is now perceived in present setting: rather than researching usual specific decay by classical Lyapounov method (for asymptotic stability), it is more generally the belonging of solution manifold to a given function space determined by its properties, see Fig.7.

Obviously the bound in eqn(6) exists over the entire time interval only if the term between bracket on the right hand side of eqn(6) is > 0, which implies a relation between bounds parameters λ_i, μ_i and initial values $Z(t_0)$.

So for a given initial ball $\|Z(t_0)\|$, there exists an uncertainty ball $\|[\Delta\lambda_i, \Delta\mu_i]\|$ on system parameters within which the solution belongs to a manifold $\mathcal{M}$ defined by the RHS of eqn(6), the embedding of which in required function space S can be studied in a second step. This opens the way to get at the same time asymptotic stability and robustness, both useful properties for control of complex systems. The approach can be extended to the more difficult situation where system uncertainty ball is larger than robustness ball[35].

5 Applications to man made systems

Evidently previous properties are of up-most importance when applied to man made systems now appearing in industry. Under economic competition, more advanced systems are conceived and worked out which include an always increasing number of heterogeneous components to be operated all together for production of higher value objects.

This would imply to keep complete system mastery by efficient control, which becomes the more illusory as system dynamics cover a larger number of elements escaping from one single centralized control structure, especially by becoming complex due to higher value of coupling parameters. In parallel, to reduce system behavior fragility, it is also interesting to reduce the number

of input control parameters by transforming the system into a (partially) complex one by clustering some components into bigger parts.

A trivial illustration of this approach is given by the way shepherd dogs are acting on a herd of cattle in the meadows. If there are n animals wandering around, they represent a system with $2n$ degrees of freedom ($3n$ when counting their orientation). Clearly the dog understands that it is hopeless with his only 2 degrees of freedom to control all the animals, so his first action is to gather all of them in a restricted space so that by being close enough they have strong enough interactions transforming their initial complicated system into a complex one with dramatically less degrees of freedom, in fact only two like himself. Then he can control perfectly well the herd as easily observed. The astonishing fact is that dogs are knowing what to do (and they even refuse to do anything with animals unable to go into this complex stage, ie to develop gregarious potential) whereas today engineers are not yet able to proceed in similar way with their own constructions and to get corresponding benefits in terms of global mastery.

This is an immense challenge industrial civilization is facing today justifying if any the needs to study and to create these complex systems. On the representation of system in Fig.3, this would mean to vary adequate parameters to move the representative points along complex axis in order to decide exactly new system status.

In any case, internal non controlled dynamics are taken care of by system self organization resulting from passing to complex state, implying that precise trajectory control is now delegated to system. The challenging difficulty is that to comply with new structure, some "intelligence" has also to be delegated to the system, leading for the operator to a more supervisory position[36].

In present case, this is contribution to trajectory management by shifting usual (imposed) trajectory control to more elaborated task control [37], a way followed by all living creatures in their daily life to guarantee strong robustness while still keeping accuracy and preciseness. This illustrates if any the limited possibility of expressions from laws of Physics because they are tightly linking information flux related to the described action to power flux implied in it.

Thanks to the discovery of "memory" DNA molecules, Nature has been very early able to "escape" from the constraint of solely following Physics laws strict causality, by inventing sufficiently isolated systems manipulating information flux as well (ie in storing and releasing it according to a timing fixed by survival goal). In fact resulting entropy decrease from created order cannot be maintained for ever due to their unavoidable transport and energy loss because they cannot exist in complete isolation, so Nature has been

circumventing resulting isolated system finite life by inventing reproduction permitting species survival instead.

In some sense, Human kind today is faced to a similar problem in the research of higher efficiency for industrial systems. After first development of tools, then to machines, and later to efficient control structure, leading to Mechatronics, a new step is now under way to give man made systems more efficiency and autonomy by delegating more "intelligence" to them, see Fig.8.

This implies to search an adequate merging of information flux mastery from recent Information Technology development with power flux mastery resulting from classical long term mechanical development[38].

An elementary illustration is provided by the evolution of Robotics to comply with always more severe industrial constraints. It can be shown[14] that robotic systems are completely fixed by the three parameters $(Jr^2)_i$, ϵ_i and $(\omega_d \tau_R)_i$ for each link i, with J the inertia moment of the link (including actuator), r the gearing ratio, ϵ the compliance factor and $(\omega_d \tau_R)$ the "rigidity" factor with ω_d the characteristic deformation frequency and τ_R the characteristic mechanical frequency. First N-link robots were build up so that parameters Jr^2 , ϵ and $(\omega_d \tau_R) << 1$ for each link i.

The inequalities can be shown in full generality to decouple link interactions and to keep the system rigid, so the system was in simple state. However slowness (speed is multiplied by r) could not be compensated by higher actuator rotation speed, and passing to complicated state, use of decoupling method has been allowing direct drive actuation $r = 1$ but with preciseness problems from inadequate control.

This has been improved by developing more robust controllers in broader Mechatronics, especially to improve mastery of compliance and vibrations. Recent orientation with higher nominal parameters is toward more autonomous and more intelligent systems with larger decision delegation shifting new robots toward complex state. This new step implies the introduction of "intention" into the system and not to stay as before at simple action level of following prescribed fixed trajectory dictated by classical control. In such scheme intention is represented by the "task" communicated to the system which now determines its own trajectory. Such evolution is following the heavy trend started from beginning of human activity, as described in Fig.8.

6 Conclusion

Thorough examination of natural systems in last fifty years has confirmed the existence of systems exhibiting behaviors which do not fit with main

stream scientific laws established from patient observations of Nature over past centuries.

These laws are based on mechanistic representation of the Universe born from celestial body motion, ie from identifiable simple objects with well defined trajectories. With the ascent of modern technology, new natural systems have been studied and artificial ones have been constructed both with very intricate structure implying a large number of heterogeneous components in strong interaction.

Application of usual laws is often unable to describe their dynamics, because they stay outside the domain of complicated multi component systems only covered by use of reductionism method. The main reason is in the overtaking of component interaction strength which dominates enough over other effects to force the system to close on itself and to manifest an internal self-organization responsible of its new behavior.

Differently said in elementary terms, the new paradigm is that "increasing interactions between components lead to their isolation" as easily verifiable. Such systems are termed as "complex" from etymology, and their main feature is that components in complex state are internally ruled through this self-organization so that at the end they are less depending on environment action.

As a consequence, natural complex systems are structurally more robust than complicated ones as evidenced by observation of living organisms, the most complex existing systems. Similarly, artificial complicated systems can be made more robust by transforming them into complex ones when linking adequately and strongly enough some of their components. As always since human origin, the price to pay for this insensitivity is the needs to replace shielded outer action by larger delegation of "intelligence" to the system for its correct internal driving.

Analysis of complex systems dynamics shows, for high enough ability of system components (now called "agents"), the possibility of "emergence" of a new behavior which is not included into the set of initial components behavior. Though it could be the way of approaching the new global and larger "2-dimensional" picture of the Universe proposed here, to avoid trivial rephrasing of well known classical results, emergence in this context appears to be more justified when restricting its domain to action of agents teaming up for collective goal they cannot individually reach. The problem of representing complex system structure build up has been discussed, and three typical structures have been singled out.

A− The broad class of reducible structure to which natural dissipative systems are belonging, usually formally amenable to be analytically projected back into complicated structure thanks to very large time and space scale difference between components. Application of asymptotic expansion methods allows to directly get important system quantities such as transport coefficients.

B− The ordinary differential structure, thoroughly analyzed theoretically and by computer simulation, which shows when increasing component coupling interaction the transformation of system trajectories into finally fully chaotic and indistinguishable ones filling complete domains of state space, manifesting a layered organization of motion into equivalence classes which are the only reachable system parameters from environment.

C− The most difficult delayed structure mainly due to long range component interactions, exhibiting a double complexity in space and time together, and grasping the fundamental aspect of component self determination by properly accounting for neighboring ones within the range of interaction potential.

Origin and main properties of class **A** have been cleared up by Physics study, but transport problem is not completely solved for continuous systems such as Fluids and Plasmas.

Class **B** has been carefully analyzed from Mathematics, especially in the difficult case of resonances for Hamiltonian systems, but still resists to a full picture from analytical reconstitution of trajectory manifold.

Class **C** remains mainly open as its study did not really begin yet. Functional methods approach in adapted function spaces has been shown to be successful for reducible systems in class **A**, in that both robustness and asymptotic stability are obtained together.

Extension to classes **B** and **C** will be discussed elsewhere. As it is based on elementary source sink description from Physics of interactions between elements of a system (the agents), and between the system and its environment, complexity paradigm discussed here can be tested through a falsifiable protocol. This bottom-up analysis is directly privileging a middle (control engineering type) approach along possible action on the system from environment in between purely (speculative) mathematical and purely (contemplative) physical classical approaches, because it fits immediately with global concepts of auto-organization and adaptive robustness resulting from transformation to complex, and allows direct application of powerful and particularly adapted modern functional analysis methods.

Application to artificial man-made systems is particularly appropriate for highly performing ones, as entering complexity is opening the possibility for the "function" to feedback onto the "structure", due to emergence produced by self-organization, a very specific property of the only living systems to date guaranteeing their remarkable survival ability. Finally, possible extension of this approach to "soft" sciences could open on orientation of economic and social sciences research towards more rigorous and potentially predictive models of human conduct.

References

1. Descartes, R.: Le Discours de la Méthode. Œuvres, V. Cousin, (ed.), Paris (1824)
2. Pattee, H.H.: Hierarchy Theory, the Challenge of Complex Systems. Braziller, New York (1973)

Galbraith, J.: Designing Complex Organizations. Addison Wesley, Reading (1973)

Cottrell, A.: Emergent Properties of Complex Systems. In: Duncan, R., Weston Smith, M. (eds.) Encyclopedia of Ignorance, pp. 129–135. Pergamon Press, Oxford (1977)

Simon, H.A.: How Complex are Complex Systems? Philosophy of Science Assoc. 2, 507–522(1976)

Hayek, F.A.: The Theory of Complex Phenomena. In: Bunge, M. (ed.) The Critical Approach to Science and Philosophy. Collier McMillan, London, pp. 332–349 (1964)

Walker, C.C.: Behavior of a Class of Complex Systems: the Effect of Systems Size on Properties of Terminal Cycles. J. of Cybernetics 1, 55–67 (1971)

Ferdinand, A.E.: Complex Systems are Those that Work Even with Maximum Entropy. Intern. J. of General Systems 1, 19–33 (1974)

Nelson, R.J.: Structure of Complex Systems. Philosophy of Science Association 2, 523–542 (1976)

Casti, J.L.: Complexity, Connectivity and Resilience in Complex Ecosystems. In: IFAC Symp. on Bio and Ecosystems, Liepzig, Germany (1977)

Trappler (ed.): Power, Autonomy, Utopia: New Approaches towards Complex Systems. Plenum Press, New York (1986)

Johnson, J.: A Theory of Stars in Complex Systems. In: Casti, J.L., Karlquist, A. (eds.) Complexity, Language and Life: Mathematical Approaches, pp. 21–61. Springer, Berlin (1986)

Haken, H.: Information and Self Organization: a Macroscopic Approach to Complex Systems. Springer, Berlin (1988)

Soodak, H., Iberall, A.S.: Thermodynamics and Complex Systems. In: Yates, F.E.(ed.) Self-Organizing Systems: the Emergence of Order, pp. 459–469. Plenum Press, New-York (1987)

Dyke, C.: The Evolutionary Dynamics of Complex Systems. OUP, New York (1988)

Funtowicz, S. Ravetz, J.: Emergent Complex Systems. Futures 26, 568–582 (1994)

Wolfram, S.: Complex Systems Theory. In: Pines, D. (ed.)Emerging Syntheses in Science, pp. 183–190. Addison Wesley, Redwood City (1987)

Haken, H.: Synergetics: An Introduction, 2nd edn. Springer Series in Synergetics, vol. 1. Springer, Heidelberg (1983)

Haken, H.: Information and Self-Organization: A Macroscopic Approach to Complex Systems. Springer Series in Synergetics, vol. 40. Springer, Heidelberg(1988)

3. Weinberg, S.: Dreams of a Final Theory. Pantheon Books, New-York (1993)
4. Laughlin, R.: A different Universe: Reinventing Physics from the Bottom Down. Basic Books, New-York (2005)
5. Laplace, P.S.: OEuvres Complètes. Part 6. Gauthiers Villard, Paris (1878-1912)

   Lagrange, J.L.: Mécanique Analytique, vol. I-II. Ed. Albert Blanchard, Paris (1788)
6. Aritotle (circa 340 BC) "Physics", Part I, ch. 9

   Anderson, P.W.: More is Different. Science 177, 393–396 (1972)
7. Gell-Mann, M.: What is Complexity? Complexity 1, 16–19 (1995) Nicolis, G., Nicolis, C.: Foundations of Complex Systems. World Scientific Pub., Singapore (2007)

   Weaver,W.: Science and Complexity. American Scientist 36, 536–544 (1948)

   Nicolis, G., Prigogine, I.: Exploring Complexity : an Introduction. W.H. Freeman and Co, New-York (1989)

   Bunge, M.: The Myth of Simplicity. Prentice Hall, Englewood Cliffs (1963)

   Ashby, W.R.: Introduction to Cybernetics. Methuen, New-York (1964)

   Baldwin, M.: Portraits of Complexity. Battelle Memorial Institute, Ohio (1975)

   Traub, J.F., Wasilkowski, G.W., Wozniakowski, H.: Information, Uncertainty, Complexity. Addison Wesley, Reading (1983)

Casti, J.L.: Connectivity, Complexity and Catastrophe. J. Wiley and Sons, New-York (1979)

Cornacchio, J.V.: System Complexity - A Bibliography. Intern. J. of General Systems 3, 267–271(1977)

Warfield, J.N.: Planning, Policy, and Complexity. J. Wiley and Sons, New-York (1976)

Prigogine, I., Stengers, I.: Order Out of Chaos - Man's New Dialogue with Nature. Bantam Books, Toronto (1984)

Grunberg, E.: "Complexity" and "Open Systems" in Economic Discourse. J. of Economic Issues 12, 541–560(1978)

Brewer, G.D.: Analysis of Complex Systems: An Experiment and its Implications for Policy Making. Rand Corp. Report, P-4951, Santa Monica, CA (1973)

Van Emden, M.H.: An Analysis of Complexity. Mathematical Centre Tracts. Mathematische Centrum Amsterdam, Amsterdam (1971)

Ashby, W.R.: Some Peculiarities of Complex Systems. Cybernetic Medicine 9, 1–7 (1973)

Kemeny, J.G.: Two Measures of Complexity. The Journal of Philosophy 52, 722–733 (1955)

Winograd, S.: Redundancy and Complexity of Logical Elements. Information and Control 5, 177–194 (1963)

Hillel, B. Kasher, Y. Shamir, E.: Measures of Syntactic Complexity. In: Booth, A.D. (ed.) Machine Translation, pp. 30–50. North Holland, Amsterdam (1967)

Moshowitz, A.: Entropy and the Complexity of Graphs. Bull. Mathematical Biophysics 30, 175–204, 225–408 (1968)

Rock, R.A.: Integrative Complexity and Non-Specific Set as Determinants of Performance in a Complex Solving Task. J. Behavioral Science 1, pp. 17–25 (1969)

Zvonkin, A.K., Levin, L.A.: The Complexity of Finite Objects and the Development of the Concepts of Information and Randomness by Means of the Theory of Algorithms. Russian Math. Surveys 256, 83-124 (1970)

Beer, S.: Managing Modern Complexity. Futures 2, 245–257 (1970)

Ausiello, G.: Abstract Computational Complexity and Cycling Computations. J. Computer and System Sciences 5, 118–128 (1971)

Sober, E.: Simplicity. Clarendon Press, Oxford (1975)

Wimsatt, W.: Complexity and Organization. In: Schaffner, K., Cohen, R.(eds.) Studies in the Philosophy of Sciences, pp. 67–86. Reidel, Dordretch (1972)

Cilliers, P.: Complexity and Post-modernism. Routledge, London (1998)

Feldman, J.A.: Some Decidability Results on Grammatical Inference and Complexity. Information and Control 20, 244–262 (1972)

Seaman, J.M., Koenig, F.: A Comparison Measures of Cognitive Complexity. Sociometry 37, 375–390 (1974)

Rosen, B.K.: Syntactic Complexity. Information and Control 24, 305–335 (1974)

Morin, E.: Complexity. Intern. J. of Social Science 26, 583–597 (1974)

Lofgren, L.: Complexity of Descriptions of Systems: A foundational Study. Intern. J. of General Systems 3, 197–214 (1974)

Arbib, M.A., Manes, E.G.: Foundations of System Theory: Decomposable Systems. Automatica 10, 285–302 (1974)

Bremermann, H.: Complexity of Automata, Brains and Behavior. Lecture Notes in Bio Mathematics 4, 304–331 (1974)

Gottinger, H.W.: Complexity and Information Technology in Dynamic Systems. Kybernetics 4, 129–141 (1975)

Pippenger, N.: Complexity Theory. Scientific American 239, 90–100 (1978)

Futia, C.: The Complexity of Economic Decision Rules. J. Mathematical Economics 4, 289–299 (1975)

De Luca, A.: Complexity and Information Theory. In: Longo, G. (ed.) Coding and Complexity, pp. 207–270. Springer, Berlin (1975)

Yokoyama, T.: Complexity and Hierarchy Theory. Osaka Economic Papers 26, 26–34 (1976)

Simon, H.A.: The Architecture of Complexity. In: The Sciences of the Artificial, pp. 192–229. MIT Press, Cambridge (1981)

Sahal, D.: Elements of an Emerging Theory of Complexity Per Se. Cybernetica 19, 5–38 (1976); System Complexity: Its Conceptions and Measurement in the Design of Engineering Systems. IEEE Trans. on Systems, Man and Cybernetics 6, 440–445 (1976)

Auyang, S.Y.: Foundations of Complex-System Theories. Cambridge University Press, Cambridge (1998)

Holland, J.: Hidden Order: How Adaptation Builds Complexity. Helix Books/Addison Wesley, New-York (1995)

Waldrop, M.: Complexity: the Emerging Science at the Edge of Order and Chaos. Touchstone, New-York (1993)

McCabe, J.: A Complexity Measure. IEEE Trans. on Software Engineering 2, 308–320 (1976)

Gottinger, H.W.: Complexity and Dynamics: Applications of Dynamic Systems Theory. IEEE Trans. on Systems, Man and Cybernetics 6, 867–873 (1976)

Gaines, B.R.: System Identification, Approximation and Complexity. Intern. J. of General Systems 3, 145–174 (1977)

Rosen, R.: Complexity and System Descriptions. In: Harnett, W.E. (ed.) Systems: Approaches, Theories, Applications, pp. 169–175. Reidel, Dordrecht (1977)

Marcus, M.J.: The Theory of Connecting Networks and their Complexity: A Review. Proc. IEEE 65, 1263–1271 (1977)

Goguen, J.A.: Complexity of Hierarchically Organized Systems. Intern. J. of General Systems 3(4), 233–251 (1977)

George, B.: Tests for System Complexity. Intern. J. of General Systems 3, 253–257 (1977)

Yates, F.E.: Complexity and the Limits to Knowledge. Amer. J. of Physiology 235, R201–R204 (1978)

Solomonoff, R.J.: Complexity-Based Induction Systems: Comparisons and Convergence Theorems. IEEE Trans. on Information Theory 24, 422–432 (1978)

Klir, G.J.: Complexity: Some General Observations. Systems Research 2, 131–140 (1985)

Pearl, J.: On the Connection Between the Complexity and Credibility of Inferred Models. Intern. J. of General Systems 4, 255–264 (1978)

Hopcroft, J.E., Ullman, J.D.: Introduction to Automata Theory, Languages, and Computation. Addison Wesley, Reading (1979)

Levine, R.D., Tribus, M. (eds.): The Maximum Entropy Principle. MIT Press, Cambridge (1979)

Rouse, W.B., Rouse, S.H.: Measures of Complexity of Fault Diagnosis Tasks. IEEE Trans. on Systems, Man and Cybernetics 9, 720–727 (1979)

Gershenson, C., Heylighen, F.: How Can we Think the Complex? In: Richardson, Kurt (eds.) Managing the Complex: Philosophy, Theory and Application, vol. 1. Institute for the Study of Coherence and Emergence/Information Age Publishing (1987)

Papentin, F.: On Order and Complexity. J. Theoretical Biology 87, 1–456 (1980)

Albin, P.S.: The Complexity of Social Groups and Social Systems by Graph Structures. Mathematical Social Sciences 1, 101–129 (1980)

Dussauchoy, R.L.: Generalized Information Theory and the Decomposability of Systems. Intern. J. of General Systems 9, 13–36 (1982)

Albin, P.S., Gottinger, H.W.: Structure and Complexity in Economic and Social Systems. Math. Social Sciences 5, 253–268 (1983)

Hinegardner, R., Engelberg, J.: Biological Complexity. J. Theoretical Biology 104, 7–20 (1983)

Ziv, J.: Complexity and Coherence of Sequences. In: Durand, H., di Lullo, M., Sinclair, C. (eds.) The Impact of Data Processing Techniques on Communications, Nijhoff, Dordrecht, pp. 23–33 (1983)

Wolfram, S.: Universality and Complexity in Cellular Automata. Physica D 10, 1 (1984); Computational Theory of Cellular Automata, Comm. Mathematical Physics 96, 15–57 (1984); Cellular Automaton as Models of Complexity. Nature 311, 419–424 (1984)

Aida, et al. (eds.): The Science and Praxis of Complexity. United Nations University, Tokyo (1984)

Prather, R.E.: An Axiomatic Theory of Software Complexity Research. The Computer J. 27, 340–346 (1984)

Nerisian, A.: On a Definition of the Description Complexity of Finite Systems. Intern. Inst. for Applied Systems Analysis, Rept. WP-84-004 (1984)

Rosen, R.: Anticipatory Systems. Pergamon Press, New-York (1985)

Lovsz, L.: The Mathematical Notion of Complexity, Inst. fur Okonometrie und Operations Research Report, University of Bonn, Germany (1985)

Ho, J.K.K., Sculli, D.: System Complexity and the Design of Decision Support Systems. Systems Practice 8, 505–516 (1995)

Rosen, R.: Organisms as Causal Systems which are not Mechanisms: An Essay into the Nature of Complexity. In: Rosen, R. (ed.) Theoretical Biology and Complexity: Three Essays on the Natural Philosophy of Complex Systems, pp. 165–203. Acad. Press, London (1985)

Casti, J.L., Karquist, A. (eds.): Complexity, Language and Life: Mathematical Models. Springer, Berlin (1985)

Celluci, C.: Proof Theory and Complexity. Synthese 62, 173–189 (1985)

Bonner, J.T.: The Evolution of Complexity. Princeton Univ. Press, Princeton (1988)

Bennett, C.H.: Dissipation, Information, Computational Complexity and the Definition of Organization. In: Pines, D. (ed.) Emerging Syntheses in Science, pp. 215–233. Addison-Wesley, Redwood City (1985)

Wood, R.E.: Task Complexity: Definition of the Construct. Organizational Behavior and Human Decision Processes 36, 60–82 (1986)

Huberman, B.A., Hogg, T.: Complexity and Adaptation. Physica D 22, 376–384 (1986)

Grassberger, P.: Towards a Quantitative Theory of Self-Generated Complexity. Intern. J. of Theoretical Physics 25(9), 907–938 (1986)

Bennett, C.H.: On the Nature and Origin of Complexity in Discrete, Homogenous, Locally-Interacting Systems. Foundations of Physics 16, 585–592 (1986)

Kampis, G., Csanyi, V.: Notes on Order and Complexity. J. of Theoretical Biology 124, 111–121 (1987)

Koppel, M.: Complexity, Depth and Sophistication. Complex Systems 1, 1087–1091 (1987)

Gottinger, H.W.: Choice and Complexity. Mathematical Social Sciences 14, 1–17 (1987)

Flood, R.L., Carson, E.R.: Dealing with Complexity. Plenum Press, New-York (1988)

Mainzer, K.: Complex Systems in Nature and Society. Physik in Unserer Zeit 28, 74–81 (1997); Thinking in Complexity: The Complex Dynamics of Matter, Mind, and Mankind, 3rd edn. Springer, Heidelberg (1997)
Levesque, H.J.: Logic and the Complexity of Reasoning. J. of Philosophical Logic 17, 355–389 (1988)

Campbell, D.J.: Task Complexity: A Review and Analysis. Academy of Management Review 13, 40–52 (1988)

Turney, P.: The Architecture of Complexity: A New Blueprint. Synthesis 79, 515–542 (1989)

Stein, D.: Lectures in the Sciences of Complexity. Addison Wesley, Redwood City (1989)

Berger, E.: Computability, Complexity, Logic. North Holland, Amsterdam (1989)

Badii, R.: Unfolding Complexity in Nonlinear Dynamical Systems. In: Abraham, N.B. (ed.) Measures of Complexity and Chaos, pp. 313–325. Plenum Press, New-York (1989)

Turney, P.: Problems with Complexity in Gold's Paradigm of Induction– Part I: Dynamic Complexity. Intern. J. of General Systems 17, 329–342 (1990); Part II: Static Complexity. IBID, 343–358 (1990)

Rustem, B., Velupillai, K.: Rationality, Computability and Complexity. J. of Economic Dynamics and Control 14, 419–432 (1990)

Kindlmann, P.: Stability vs. Complexity in Model Computational Communities. Lecture Notes in Bio-Mathematics 54, 191–207 (1984)

d'Alessandro, G., Politi, A.: Hierarchical Approach to Complexity with Applications to Dynamic Systems. Phys. Rev. Letters 64, 1609–1612 (1990)

Banks, J.F., Sundaram, R.K.: Repeated Games, Finite Automata and Complexity. Games and Economic Behavior 2, 97–117 (1990)

Machlin, R.: The Complex Behavior of Simple Machines. Physica D 42, 85–98 (1990)

Gorodkin, J., Sorensen, A., Winther, O.: Neural Networks and Cellular Automata Complexity. Complex Systems 7, 1–23 (1993)

Lloyd, S.: Physical Measures of Complexity. In: Jen, E. (ed.) Lectures in Complex Systems, pp. 67–73. Addison Wesley, Reading (1990)

George, D.A.R.: Chaos and Complexity in Economics: Review Article. J. of Economic Surveys 4, 397–404 (1990)

Bennett, C.H.: How to Define Complexity in Physics, and Why. In: Zurek, W.H. (ed.) Complexity, Entropy and the Physics of Information, pp. 137–148. Addison Wesley, Redwood City (1990)

Balcazar, J.L.: Structural Complexity. Springer, Berlin (1990)

Baker, G.L.: Chaotic Dynamics. Cambridge Univ. Press, Cambridge (1990)

Traub, J.F., Wozniakowski, H.: Information-Based Complexity: New Questions for Mathematicians. Mathematical Intelligencer 13, 34–43 (1991)

Zuse, H.: Software Complexity. Walter de Gruyter, Berlin (1991)

Lakshmanan, K.B., Jayaprakash, S., Sinha, P.K.: Properties of Control-Flow Complexity Measures. IEEE Trans. on Software Engineering 17, 1289–1295 (1991)

McShea, D.: Complexity and Evolution: What Everybody Knows. Biology and Philosophy 6(3), 3003–3024 (1991)

Kampis, G.: Self-Modifying Systems in Biology and Cognitive Science: A New Framework for Dynamics. Information and Complexity. Pergamon Press, Oxford (1991)

Grassberger, P.: Information and Complexity Measures in Dynamical Systems. In: Atmanspacher, H., Scheingraber, H. (eds.) Information Dynamics, pp. 15–33. Plenum Press, New-York (1991)

Woodward, D.: Understanding Complexity: A Critique and Synthesis. Henley Management College, Henley on Thames, England (1991)

Atmanspacher, H., Kurths, J., Scheingraber, H., Wackerbauer, R., Witt, A.: Complexity and Meaning in Nonlinear Dynamical Systems. Open systems and Information Dynamics 1, 269–289 (1991)

Heylighen, F.: Self-Organization, Emergence and the Architecture of Complexity. In: AFCET (eds.) Proc. European Congress on System Science, Paris, pp. 23–32 (1992)

Crutchfield, J.P.: Knowledge and Meaning... Chaos and Complexity. In: Lam, L., Naroditsky, V. (eds.) Modeling Complex Phenomena, pp. 66–101. Springer, Berlin (1992)

Casti, J.L.: Reality Rules: Part I. Picturing the World in Mathematics: The Fundamentals, Part II. Picturing the World in Mathematics: The Frontier. J. Wiley and Sons, New-York (1992); The Simply Complex: Trendy Buzzword or Emerging New Science. Bull. of the Santa Fe Institute 7, 10–13 (1992)

Badii, R.: Complexity and Unpredictable Scaling of Hierarchical Structures. In: Bountis, T. (ed.) Chaotic Dynamics, Theory and Practice, pp. 1–19. Plenum Press, New-York (1992)

Kauffman, S.: At Home in the Universe: The Search for Laws of Self-Organization and Complexity. Oxford Univ. Press, New-York (1995)

Auger, P.M.: Hierarchically Organized Systems The Response to Complexity. J. of Scientific and Industrial Research 51, 725–735 (1992)

Ruthen, R.: Adapting to Complexity. Scientific American 268, 130–135 (1993)

Kauffman, S.A.: The Origins of Order. Oxford Univ. Press, New-York (1993)

Bates, J.E., Shepard, H.K.: Measuring Complexity Using Information Fluctuation. Phys. Letters A 172, 416–425 (1993)

Holm, H.J.: Complexity in Economic Theory, University of Lund: Lund Economic Studies, Lund, Sweden (1993)

Arthur, W.B.: Why Do Things Become More Complex? Scientific American 268, 92 (1993)

Yates, F.E.: Order and Complexity in Dynamical Systems Homeodynamics as a Generalized Mechanics for Biology. Mathematical and Computer Modelling 19, 49–74 (1994)

Ray, T.S.: Evolution, Complexity, Entropy and Artificial Reality. Physica D 75, 239–263 (1994); Evolution and Complexity. In: Cowan, G., Pines, D., Meltzer, D. (eds.) Complexity: Metaphors, Models and Reality, pp. 161–173. Addison Wesley, Reading (1994)

Reggiani, A., Nijkamp, P.: Competition and Complexity in Spatially Connected Networks. System Dynamics Review 11, 51–66 (1995)

Poon, L., Grebogi, C.: Controlling Complexity. Phys. Rev. Letters 75, 4023–4026 (1995)

Horgan, J.: From Complexity to Perplexity. Scientific American 272, 74–79 (1995)

Badii, R., Politi, A.: Complexity: Hierarchical Structures and Scaling in Physics. Cambridge Univ. Press, Cambridge (1997)

Khalil, E.L.: Organizations, Naturalism and Complexity. Review of Social Economy 53, 393–419 (1995)

Zeigarnik, A.V., Temkin, O.N.: A Graph Theoretic Model of Complex Reaction Mechanisms: a New Complexity Index for Reaction Mechanisms. Kinetics and Catalysis 37, 372–385 (1996)

Helbing, D., Schweitzer, F. (Guest eds.): Complex Dynamics in Economics. Advances in Complex Systems 4(1), 1–176 (2001)

Casti, J.L.: Connectivity, Complexity and Catastrophe in Large-Scale Systems. Intern. Ser. on Applied Systems Analysis, vol. 7. J. Wiley and Sons, Chichester (1979)

Feldman, D.P., Crutchfield, J.P.: Measures of Statistical Complexity: Why? (November 11, 1997), `www.santafe.edu/~cmg/papers/mscw.pdf`

Bennett, C.H.: How to Define Complexity in Physics and Why. In: Gregersen, N.H. (ed.) From Complexity To Life: On the Emergence of Life and Meaning. Oxford Univ. Press, New-York (2003)

Lòpez-Ruiz, R., Mancini, H.L., Calbet, X.: A Statistical Measure of Complexity. Phys. Lett. A 209, 321–326 (1995)

Feistel, R., Ebeling, W.: Evolution of Complex Systems: Self-Organization, Entropy and Development, Mathematics and its Applications. East European Series, vol. 30. Kluwer Acad. Publ., Dordrecht (1989)

Frenken, K.: A Complexity Approach to Innovation Networks: The Case of the Aircraft Industry (1909-1997). Research Policy - A Journal Devoted to Research Policy, Research Management and Planning 29, 257–272 (2000)

Lòpez Ruiz, R.: Shannon Information, LMC Complexity and Rényi Entropies: a Straightforward Approach (2003), http://arxiv.org/PS$_$cache/nlin/pdf/0312/0312056.pdf

Calbet, X., López-Ruiz, R.: Tendency towards Maximum Complexity in an Isolated Non-equilibrium System. Phys. Rev. E 60(9), 6–6116 (2001)

Shiner, J.S., Davison, M., Landsberg, P.T.: Simple Measure for Complexity. Phys. Rev. E 59, 1459–1464 (1999)

Crutchfield, J.P., Feldman, D.P., Shalizi, C.R.: Comment on Simple Measure for Complexity. Phys. Rev. E 62, 2559–2564 (2000)

Shallit, J., Elsberry, W.: Playing Games with Probability: Dembski's Complex Specified Information. In: Young, M., Edis, T. (eds.) Why Intelligent Design Fails: A Scientific Critique of the New Creationism, Piscataway, N.J (2004)

Nicolis, G., Prigogine, I.: Exploring Complexity: an Introduction. W.H. Freeman and Co, New York (1989)

8. Albert, R., Barabasi, A.L.: Topology of Evolving Networks: Local Events and Universality. Phys. Rev. Letters 85, 5234–5237 (2000)

Albert, R., Barabasi, A.L.: Statistical Mechanics of Complex Networks. Reviews of Modern Physics 74, 47–97 (2002)

Albert, R., Jeong, H., Barabasi, A.L.: Diameter of the World Wide Web. Nature 401, 130–131 (1999)

Albert, R., Jeong, H., Barabasi, A.L.: Error and Attack Tolerance of Complex Networks. Nature 406, 378–382, 409–542 (correction) (2000)

Barabasi, A.L.: Linked: The New Science of Networks. Perseus Publishing, Cambridge (2002)

Barabasi, A.L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., Vicsek, T.: Evolution of the Social Network of Scientific Collaborations. Physica A-311, 590–614 (2002)

Bartholomay, A.F., Bianconi, G., Barabasi, A.L.: Competition and Multi-Scaling in Evolving Networks. Europhysics Letters 54, 436–442 (2001)

Bartholomay, A.F., Bianconi, G., Barabasi, A.L.: Bose-Einstein Condensation in Complex Networks. Phys. Rev. Letters 86, 5632–5635 (2001)

Bornholdt, S., Schuster, H.G. (eds.): Handbook of Graphs and Networks: From the Genome to the Internet. Wiley VCH Verlag, Weinheim (2003)

Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., Wiener, J.: Graph Structure in the Web. Computer Networks - the Intern. J. of Computer and Telecom. Networking 33, 309–320, See also special issue; Proc. 9th. Intern. World Wide Web Conference, the Web: the Next Generation, Amsterdam, May 15-19 (2000); Computer Networks - the Intern. J. of Computer and Telecom. Networking 33(1-6), v-viii, 1-843 (2001)

Buchanan, M.: Nexus: Small Worlds and the Groundbreaking Science of Networks. W. W. Norton and Company, New-York (2002)

Dorogovtsev, S.N., Mendes, J.F.F., Samukhin, A.N.: Principles of Statistical Mechanics of Uncorrelated Random Networks. Nucl. Phys. B 666, 396–416 (2003)

Brossel, B., McKane, A.J.: Modelling Food Webs. In: Bornholdt, S., Schuster, H.G. (eds.) Handbook of Graphs and Networks: From the Genome to the Internet, pp. 218–247. Wiley VCH Verlag, Weinheim (2003)

Ebeling, W., Sonntag, I., Schimansky-Geier, L.: On the Evolution of Biological Macromolecules II: Catalytic Networks. Studia Biophysica 84, 87–88 (1981)

Erdös, P., Rényi, A.: On Random Graphs. I. Publicationes Mathematicae (Debrecen) 6, 290–297 (1959-1960); On the Evolution of Random Graphs. Publications of the Math. Institute of the Hungarian Academy of Sciences 5, 17–61 (1959-1960)

Faloutsos, M., Faloutsos, P., Faloutsos, C.: On Power-Law Relationships of the Internet Topology. Computer Communications Review 29, 251–262 (1999)

Sonntag, H.I.: Analysis of the Selection Equations for a Multivariable Population Model: Deterministic and Stochastic Solutions and Discussion of the Approach for Populations of Self-Reproducing Biochemical Networks. J. Theoretical Biology 93, 325–361 (1981)

Huberman, B.A.: The Laws of the Web: Patterns in the Ecology of Information. The MIT Press, Cambridge (2001)

Jain, S., Krishna, S.: A Model for the Emergence of Cooperation, Interdependence and Structure in Evolving Networks. Proc. National Acad. Sciences of the U.S.A. 98, 543–547 (2001)

Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., Barabasi, A.L.: The Large-Scale Organization of Metabolic Networks. Nature 407, 651–654 (2000)

Kirman, A.: Economic Networks. In: Bornholdt, S., Schuster, H.G. (eds.) Handbook of Graphs and Networks: From the Genome to the Internet, pp. 273–294. Wiley-VCH Verlag, Weinheim (2003)

Klimontovich, Y.L.: Statistical Theory of Open Systems. In: Part 1: A Unified Approach to Kinetic Description of Processes in Active Systems. Fundamental Theories of Physics, vol. 67. Kluwer Acad. Publishers, Dordrecht (1995)

Leydesdorff, L.: A Sociological Theory of Communication: The Self Organization of the Knowledge Based Society, 2nd edn. Universal Publishers, Parkland (2003)

Maurer, S.M., Huberman, B.A.: Competitive Dynamics of Web Sites. J. Economic Dynamics and Control 27, 2195–2206 (2000)

Juillard, M. Marcet, A. (guest eds.): Special issue, Computing in Economics and Finance. In: Proc. 6th Intern. Conf. of the Society for Computational Economics, Barcelona, Spain, July 6-8; J. Economic Dynamics and Control 27(11-12), 1939–2265 (2000)

Milgram, S.: The Small World Problem. Psychology Today 1, 60–67 (1967)

Montoya, J.M., Solé, R.V.: Small World Patterns in Food Webs. J. Theoretical Biology 214, 405–412 (2002)

Nachtigall, C.: Selbstorganisation und Gewalt. Internationale Hochschulschriften, vol. 259. Waxmann Verlag, New York (1998)

Newman, M.E.J.: Scientific Collaboration Networks: Part I. Network Construction and Fundamental Results. Phys. Rev. E 64, 016132, 8 (2001); Part II. Shortest Paths, Weighted Networks, and Centrality. Phys. Rev. E 64, 016132, 7 (2001)
Newman, M.E.J.: Models of the Small World. J. Statistical Physics 101, 819–841 (2000)

Newman, M.E.J., Watts, D.J.: Scaling and Percolation in the Small-World Network Model. Phys. Rev. E 60, 7332–7342 (1999)

Pyka, A., Küppers, G. (eds.): Innovation Networks: Theory and Practice. New Horizons in the Economics of Innovation Series. Edward Elgar Publ., Cheltenham (2002)

Rechenberg, I.: Evolutionsstrategie 1994, Werkstatt Bionik und Evolutionstechnik, vol. 1. Frommann-Holzboog-Verlag, Bad Cannstatt (1994)

Scharnhorst, A.: Complex Networks and the Web: Insights from Nonlinear Physics. J. Computer Mediated Communication 8(4) (2003), http://www.ascusc.org/jcmc/vol8/issue4/scharnhorst.html

Schwefel, H.P., Ebeling, W., Rosé, H., Weiss, O.: Optimization of Road Networks Using Evolutionary Strategies. Evolutionary Computation 5, 419–438 (1997)

Scott, J.: Social Network Analysis: A Handbook, 2nd edn. SAGE Publ., London (2000)

Solé, R.V., Montoya, J.M.: Complexity and Fragility in Ecological Networks. Proc. Royal Society London, Series B - Biological Sciences 268, 2039–2045 (2001)

Sonntag, I.: Random Networks of Catalytic Biochemical Reactions. Biometrical J. 26, 799–807 (1984); Application of the Percolation Theory to Random Networks of Biochemical Reactions. IBID, 809–813 (1984)

Sonntag, I., Feistel, R., Ebeling, W.: Random Networks of Catalytic Biochemical Reactions. Biometrical J. 23, 501–515 (1981)

Strogatz, S.H.: Exploring Complex Networks. Nature 410, 268–276 (2001)

Temkin, O.N., Zeigarnik, A.V., Bonchev, D.: Chemical Reaction Networks: A Graph Theoretical Approach. CRC Press, Boca Raton (1996)

Wagner, A., Fell, D.: The Small World Inside Large Metabolic Networks. Proc Royal Society London, Series B - Biological Sciences 268, 1803–1810 (2000)

Watts, D.J.: Small Worlds: The Dynamics of Networks between Order and Randomness. Princeton Studies in Complexity. Princeton Univ. Press, Princeton (1999)

Watts, D.J., Strogatz, S.H.: Collective Dynamics of 'Small World' Networks. Nature 393, 440–442 (1998)

Armbruster, D., Kaneko, K., Mikhailov, A.S.: Networks of Interacting Machines, Singapore. World Scientific Lecture Notes in Complex Systems, vol. 3 (2005)

9. McShea, D.: Metazoan Complexity and Evolution: is There a Trend? Evolution 50, 477–492 (1996)

Saunders, P.T., Ho, M.W.: On the Increase of Complexity in Evolution. J. Theoretical Biology 90, 515–530 (1981)

Goodwin, B.: How the Leopard Changed Its Spots. In: The Evolution of Complexity. Weidenfield and Nicholson, London (1994)

Edmonds, B.: What is Complexity? the Philosophy of Complexity perse with Application to Some Examples in Evolution. In: Heylighen, F., Aerts (eds.) The Evolution of Complexity. Kluwer, Dordrecht (1994)

Steward, R.: Self Organization in Evolution: a Mathematical Perspective. Phil. Trans. Royal Soc., London A361, 1101–1123 (2003)

Gregersen, N.H. (ed.): From Complexity to Life: On the Emergence of Life and Meaning. Oxford Univ. Press, New-York (2003)

Heylighen, F.: The Growth of Structural and Functional Complexity during Evolution. In: The Evolution of Complexity. Kluwer Acad. Publ., Dordrecht (1996)

Ebeling, W., Jimenez Montano, M.A.: On Grammars, Complexity and Information Measures of Biological Macromolecules. J. Math. Bioscience 52, 3–71 (1980)

Cohen, B., Harwood, W.T., Jackson, M.I.: The Specification of Complex Systems. Addison Wesley, New-York (1986)

Flood, R.L., Carson, E.R.: Dealing with Complexity, 2nd edn. Plenum Press, New York (1993)

Puri, S., Datta, S.: Nonlinearities in Complex Systems. Norosa Publ. House, Indian Institute of Advanced Studies (1997)

10. Prigogine, I.: Introduction to Thermodynamics of Irreversible Processes. American Lecture Series, vol. 185. Ch.C. Thomas Publisher, Springfield, 3rd edn., Interscience Publ. New-York (1967)

Prigogine, I., Nicolis, G.: Self Organization in Non Equilibrium Systems: From Dissipative Structures to Order through Fluctuations. J. Wiley and Sons, New-York (1977)

Klimontovitch, Y.L.: Theory of Open Systems. Kluwer, Dordrecht (1995)

Kleidon, A., Lorenz, R.D. (eds.): Non-equilibrium Thermodynamics and the Production of Entropy: Life, Earth, and beyond. Springer, Heidelberg (2005)

11. Kadomtsev, B.B.: Self Organization and Transport in Tokamak Plasma. Plasma Phys. and Nucl. Fus. 34(13), 1931–1938 (1992)

Bautay, P., Janosi, I.M.: Self Organization and Anomalous Diffusion. Physica A 185(1-4), 11–18 (1992)

Biebricher, C.K., Nicolis, G., Schuster, P.: Self-Organization in the Physico-Chemical and Life Sciences, Vol. 16546, EU Report, Brussels (1995)

Schweitzer, F.: Multi-Agent Approach to the Self- Organization of Networks. In: Reed-Tsochas, F., Johnson, N.F., Efstathiou, J. (eds.) Understanding and Managing Complex Agent-Based Dynamical Networks. World Scientific, Singapore (2007)

Ebeling, W., Schweitzer, F.: Self-Organization, Active Brownian Dynamics, and Biological Applications. Nova Acta Leopoldina NF 88(332), 169–188 (2003)

12. Strand, P.I.: Predictive Simulations of Transport in Tokamaks, PhD Thesis, Dept. of Electromagnetics, Chalmers Univ. of Technology (1999)

Perkins, F., et al.: Nondimensional Transport Scaling in the Tokamak Fusion Test Reactor: is Tokamak Transport Bohm or Gyro Bohm? Phys. Fluids B 5, 422–498 (1998)

Redd, A.J., Kritz, A.H., Bateman, G., Kinsey, J.E.: Sensitivity of Predictive Tokamak Plasma Transport Simulations. Phys. of Plasmas 4, 2207–2214 (1997)
13. Zakharov, V.E., L'vov, V.S., Falkovich, G.: Kolmogorov Spectra of Turbulence. Springer, New York (1992)

Farge, M., Schneider, K., Pellegrino, G., Wray, A.A., Rogallo, R.S.: Coherent Vortex Extraction in 3D Homogeneous Turbulence: Comparison between CVS-wavelet and POD-Fourier Decompositions. Phys. Fluids 15(10), 2886–2896 (2003)

Temam, R.: The Navier-Stokes Equations. North Holland, Amsterdam (1978) Klimontovich, Y.L.: Turbulent Motion and Structure of Chaos. Kluwer, Dordrecht (1991)

Frisch, U.: Turbulence. Cambridge Univ. Press, Mass (1995)

Nikolaenko, B., Foias, C., Temam, R.: The Connection between Infinite Dimensional and Finite Dimensional Dynamical Systems, AMS, Providence, RI (1992)
14. Cotsaftis, M.: Lectures on Advanced Dynamics, Taiwan Univ., Taipeh, R.O.C (1993)
15. Standish, R.K.: On Complexity and Emergence. Complexity International 9, 1–6 (2004)

Fromm, J.: The Emergence of Complexity. Kassel Univ. Press (2004) Schweitzer, F.: Self Organization of Complex Structures: From Individual to Collective Dynamics. Gordon and Breach Science Publ., Amsterdam (1997)

Johnson, S.: Emergence. Penguin Books, New-York (2001)

Bedau, M.A.: Weak Emergence. In: Tomberlin, J. (ed.) Philosophical Perspectives: Mind, Causation, and World, vol. 11, pp. 375–399. Blackwell, Malden (1997)

Strogatz, S.: Sync: The Emerging Science of Spontaneous Order. In: Hales, D., et al. (eds.) MABS 2003. LNCS (LNAI), vol. 2927. Springer, Heidelberg (2003)

Ferber, J., Gutknecht, O., Michel, F.: From Agents To Organizations: An Organizational View of Multi Agent Systems. In: Giorgini, P., Müller, J.P., Odell, J.J., et al. (eds.) AOSE 2003. LNCS, vol. 2935, pp. 214–230. Springer, Heidelberg (2004)

Bernon, C., Gleizes, M.-P., Peyruqueou, S., Picard, G.: ADELFE, a Methodology for Adaptive Multi-Agent Systems Engineering. In: Petta, P., Tolksdorf, R., Zambonelli, F., et al. (eds.) ESAW 2002. LNCS, vol. 2577, pp. 156–169. Springer, Heidelberg (2003)

Fromm, J.: Types and Forms of Emergence, http://arxiv.org/abs/nlin.AO/0506028

Bresciani, P., Perini, A., Giorgini, P., Giunchiglia, F., Mylopoulos, J.: A Knowledge Level Software Engineering Methodology for Agent Oriented Programming. In: Proc. 5th Intern. Conf. on Autonomous Agents, pp. 648–655. ACM Press, New York (2001)

Lind, J.: Iterative Software Engineering for Multiagent Systems. LNCS, vol. 1994, pp. 648–655. Springer, Heidelberg (2001)

Jennings, N.: On Agent Based Software Engineering. Artificial Intelligence 117(2), 277–296 (2000)

Kephart, J.O.: Research Challenges of Autonomic Computing. In: Proc. ICSE 2005, pp. 15–22 (2005)

Kolp, M., Giorgini, P., Mylopoulos, J.: A Goal Based Organizational Perspective on Multi-Agent Architectures. In: Meyer, J.-J.C., Tambe, M. (eds.) ATAL 2001. LNCS (LNAI), vol. 2333, pp. 128–140. Springer, Heidelberg (2002)

Maes, P.: Modelling Adaptive Autonomous Agents, http://citeseer.ist.psu.edu/42923.html

Sycara, K.P.: Multi Agent Systems. AI Magazine 19(2), 79–92 (1998)

Abbott, R.: Emergence Explained (to be published), http://cs.calstatela.edu/~wiki/~images/9/90/Emergence.Explained.pdf

Yamins, D.: Towards a Theory of "Local to Global" in Distributed Multi-Agent Systems. In: Proc. Intern. Joint Conf. on Autonomous Agents and Multi Agent Systems (AAMAS 2005) (2005)

Vogels, W., van Renesse, R., Birman, K.: The Power of Epidemics: Robust Communication for Large Scale Distributed Systems. ACM SIGCOMM Computer Communication Review 33(1), 131–135 (2003)

Zambonelli, F., Jennings, N.R., Wooldridge, M.: Developing Multi Agent Systems: The Gaia Methodology. ACM Trans. on Software Engineering and Methodology 12(3), 317–370 (2003)

16. Schimansky Geier, L., Pöschel, T. (eds.): Stochastic Dynamics. Springer, Berlin (1997); Doob, J.: Stochastic Processes. J. Wiley and Sons, New York (1953)

Gross, D.H.E.: Micro Canonical Thermodynamics. World Scientific, Singapore (2001)

Zubarev, D.N., Röpke, G., Morozov, A.A.: Statistical Mechanics of Non Equilibrium Processes, vol. I & II, IBID. V.C.H. Viley, Weinheim (1996/1997)

Yulmetyev, R., Hänggi, P., Gafarov, F.: Statistical Theory of Non Stationary Time Correlation in Complex Systems with Discrete Time. Nonlinear Phenomena in Complex Systems 5(2), 129–136 (2002)

17. Arnold, V.I., Avez, A.: Ergodic Problems of Classical Mechanics. Advanced Book Classics. Addison Wesley, Reading (1989)

Jones, R.L., Ostrovskii, I., Rosenblatt, J.: Square Functions in Ergodic Theory. Ergodic Theory Dynamical Systems 16, 267–305 (1996)

Kachurovskii, A.G.: The Rate of Convergence in Ergodic Theorems. Russian Math. Surveys 51(4), 653–703 (1996)

Sinaï, Y.G.: Introduction to Ergodic Theory. Princeton Univ. Press, Princeton (1976)

Cornfeld, I.P., Fomin, S.V., Sinai, Y.G.: Ergodic Theory. Springer, Berlin (1982)

Petersen, K.: Ergodic Theory. Cambridge Studies in Advanced Math. Cambridge Univ. Press, Cambridge (1983)

Pesin, Y., Barreira, L.: Lyapounov Exponents and Smooth Ergodic Theory. University Lecture Series no 23. American Math. Soc., Providence (2001)

Bedford, T., Keane, M. (eds.): Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces. Oxford Univ. Press, Oxford (1991)

Ollagnier, J.M.: Ergodic Theory and Statistical Mechanics. Lecture Notes in Math. No 1115. Springer, Berlin (1985)

Lebowitz, J., Penrose, O.: Modern Ergodic Theory. Phys. Today 26, 155–175 (1973)

Ruelle, D.: Ergodic Theory of Differentiable Dynamical Systems. Public. Mathématiques IHES 50, 27–58 (1979)

Pugh, C., Shub, M.: Stable Ergodicity. Bull. Amer. Math. Soc. 41, 1–41 (2004)

Mackey, G.W.: Ergodic Theory and its Significance for Statistical Mechanics and Probability Theory. Advances in Mathematics 12(2), 178–268 (1974)

Birkhoff, G.D.: Proof of the Ergodic Theorem. Proc. National Acad. Sc. USA 17, 656–660 (1931)

Kac, M.: Probability and Related Topics in Physical Science. Lectures in Applied Math. Series, vol. 1a. Amer. Math. Soc. (1957)

Penrose, O.: Foundations of Statistical Mechanics. Rept. Progress Physics 42, 1937–2006 (1979)
18. Camazine, S., Deneubourg, J.L., Franks, N.R., Sneyd, J., Theraulaz, G., Bonabeau, E.: Self Organization in Biological Systems. Princeton Univ. Press, Princeton (2002)

Shen, W.-M., et al.: Hormone Inspired Self Organization and Distributed Control of Robotic Swarms. Autonomous Robots 17, 93–105 (2004)

Bonabeau, E., Dorigo, M., Theraulaz, G.: Swarm Intelligence: from Natural to Artificial Systems. Oxford Univ. Press, New-York (1999)
19. Eigen, M.: The Self-Organization of Matter and the Evolution of Biological Macromolecules. Die Natur-wissenschaften 58, 465–523 (1971)

Eigen, M., Schuster, P.: The Hypercycle: A Principle of Natural Self-Organization. Part A: Emergence of the Hypercycle. Die Naturwissenschaften 64, 541–565 (1977) and Part C. The Realistic Hypercycle. IBID, 65, 341–369 (1978)

Yates, F.E. (ed.): Self Organizing Systems: The Emergence of Order. Plenum Press (1987)

Lewin, R.: Complexity Life at the Edge of Chaos. MacMillan, Basingstoke (1993)

Lerner, A.Y. (ed.): Principles of Self Organization. Mir, Moscow (1966)

Nolfi, S., Floreano, D.: Evolutionary Robotics: the Biology, Intelligence and Technology of Self-Organizing Machines. The MIT Press, Cambridge (2000)

Serra, R., Andretta, M., Compiani, M., Zanarini, G.: Introduction to the Physics of Complex Systems (the Mesoscopic Approach to Fluctuations, Nonlinearity and Self-Organization). Pergamon Press, Oxford (1986)
20. Weidlich, W.: Socio-Dynamics: A Systematic Approach to Mathematical Modelling in the Social Sciences. Harwood Academic Publishers, Amsterdam (2000)

Weidlich, W., Haag, G.: Concepts and Models of Quantitative Sociology: The Dynamics of Interacting Populations. Springer Series in Synergetics, vol. 14. Springer, Heidelberg (1983)
Wolfram, S.: A New Kind of Science. Wolfram Media (2002)

Bohm, D., Hiley, B.J.: The Undivided Universe. Routledge, New-York (1993)

Bellouquid, A., Delitala, M.: Mathematical Modelling of Complex Biological Systems - a Kinetic Theory Approach. Birkhauser, Boston (2006)
21. Chandrasekhar, S.: Stochastic Problems in Physics and Astronomy. Rev. Mod. Phys. 15(1), 1–89 (1943)

Risken, H.: The Fokker-Planck Equation, 2nd edn. Springer, Berlin (1996)

22. Cotsaftis, M.: A Passage to Complex Systems. In: Proc. CoSSoM 2006, Toulouse, France, September 24-27, 2006. Springer Series Understanding Complex Systems, Emergent Properties in Natural and Artificial Dynamical Systems (2006)
23. Salamon, D.: The Kolmogorov-Arnold-Moser Theorem, Preprint, ETH-Zürich (1986)

Salamon, D., Zehnder, E.: KAM Theory in Configuration Space. Comment. Math. Helv. 64, 84–132 (1989)

Sevryuk, M.: KAM-Stable Hamiltonians. J. Dyn. Control Syst. 1, 351–366 (1995)

Pöschel, J.: A Lecture on the Classical KAM Theorem. In: Proc. Symp. Pure Math., vol. 69, pp. 707–732 (2000)

Hamilton, R.S.: The Inverse Function Theorem of Nash-Moser. Bull. A.M.S. 7, 65–222 (1982)

Alinhac, S., Gerard, P.: Opérateurs Pseudo-Différentiels et Thèoreme de Nash-Moser, Inter-Edition et Edition du CNRS (1991)

Arnold, V.I.: Mathematical Methods of Classical Mechanics, 2nd edn. Springer, Heidelberg (1989)

Arnold, V.I.: Dynamical Systems. Encyclopaedia of Math. Sci, vol. 3. Springer, Heidelberg (1988)

Moser, J.: On Invariant Curves of Area Preserving Mappings of an Annulus. Nachr. Akad. Wiss. Gött, Math. Phys. Kl, pp. 1–20 (1962)

Markus, L., Meyer, K.: Generic Hamiltonian Systems are neither Integrable nor Ergodic. Mem. Am. Math. Soc. 144 (1974)

Moser, J.: Convergent Series Expansions for Quasi-periodic Motions. Math. Ann. 169, 136–176 (1967)

Moser, J.: Stable and Random Motions in Dynamical Systems. Princeton Univ. Press, Princeton (1973); Zehnder, E.: Lectures on Hamiltonian Systems, Lecture Notes, ITCP (1973)
24. Tabor, M.: Chaos and Integrability in Nonlinear Dynamics, an Introduction. Wiley and Sons, New-York (1989)

Stewart, I.: Does God Play Dice? The New Mathematics of Chaos. Penguin, Harmondsworth (1997)

Abraham, N.B., Albano, A.M., Tufillaro, N.B.: Complexity and Chaos. In: Abraham, N.B. (ed.) Measures of Complexity and Chaos, pp. 1–25. Plenum Press, New Jersey (1989)

Alliwood, K.T., Sauer, T.D., Yorke, J.A.: Chaos. An Introduction to Dynamical Systems. Springer, New York (1997)

Moon, F.C.: Chaotic Vibrations. Wiley Interscience, New York (1987)

Ott, E.: Chaos in Dynamical Systems. Cambridge Univ. Press, Cambridge (1993)

Hilborn, R.C.: Chaos and Nonlinear Dynamics. Oxford Univ. Press, Oxford (1994)

Prigogine, I.: Les Lois du Chaos, Nouvelle Bibliothéque Scientifique, Flammarion, Paris (1993)
Chen, G.: Controlling Chaos and Bifurcations in Engineering Systems. CRC Press, Boca Raton (2000)

Wiggins, S.: Chaotic Transport in Dynamical Systems. Springer, New York (1991)

Strogatz, S.H.: Nonlinear Dynamics and Chaos with Applications to Physics, Biology, Chemistry, and Engineering. Addison-Wesley, Reading (1994)

Strogatz, S.H.: Sync: The Emerging Science of Spontaneous Order, Theia (2003)

Cvitanovic, P.: Universality in Chaos. Adam Hilger, Bristol (1984); Hilborn, R.C.: Chaos and Nonlinear Dynamics. Oxford Univ. Press, Oxford (1994)

Magnitskii, N.A., Vasilevich, S.V.: New Methods for Chaotic Dynamics, Singapore. World Scientific Series on Nonlinear Science, Ser. A (2006)

Kapitaniak, T.: Chaos for Engineers: Theory, Applications and Control. Springer, Berlin (1998)

Sagdeev, R.Z., Usikov, D.A., Zaslavsky, G.M.: Nonlinear Physics; from Pendulum to Turbulence and Chaos. Harwood Academic, New York (1988)
25. Poincaré, H.: Les Méthodes Nouvelles de la Mécanique Céleste, Gauthier Villars, Paris, vol. 3 (1892-1999)

Arnold, V.I.: Small Denominators and Problems of Stability of motion in Classical and Celestial Mechanics. Russian Math. Surveys 18(6), 85–191 (1963)

Arnold, V.I.: Mathematical Methods of Classical Mechanics. Springer, New York (1978)

Arnold, V.I., Koslov, V.V., Neishtadt, A.I.: Dynamic Systems III: Mathematical Aspects of Classical and Celestial Mechanics. Springer, Berlin (1993)

Bambusi, D.: Nekhoroshev Theorem for Small Amplitude Solutions in Nonlinear Schrdinger Equations. Math. Zb. 230(2), 345–387 (1999)

Benettin, G., Gallavotti, G.: Stability of Motion near Resonances in Quasi Integrable Hamiltonian Systems. J. of Statistical Physics 44(3-4), 293–338 (1986)

Bierstone, E., Milman, P.D.: Semi Analytic and Sub Analytic Sets. Publ. Math. IHES 67, 5–42 (1988)

Broer, H.W., Huitema, G.B., Sevryuk, M.B.: Quasi Periodicity in Families of Dynamical Systems: Order amidst Chaos. LNM, vol. 1645, pp. 5–42. Springer, New York (1996)

Katok, A., Hasselblatt, B.: Introduction to the Modern Theory of Dynamical Systems. Cambridge Univ. Press, Mass (1996)

Bogoliubov, N.N., Mitropolskii, Y.A., Samoilenko, A.M.: Methods of Accelerated Convergence in Nonlinear Mechanics. Springer, Berlin (1976)

26. Hirsch, M., Pugh, C., Shub, M.: Invariant Manifolds. Lecture Notes in Math, vol. 583. Springer, Berlin (1977)

Tabor, M.: Chaos and Integrability in Nonlinear Dynamics, an Introduction. Wiley and Sons, New-York (1989)

Goriely, A.: Integrability and non Integrability of Dynamical Systems. World Scientific Publ., Singapore (2001)

27. Lefschetz, S.: Stability of Nonlinear Control Systems. Academic Press, New-York (1965)

Aizerman, M.A., Gantmacher, F.R.: Absolute Stability of Regulator Systems. Holden Day, San Franscisco (1964)

Dullerud, G.E., Paganini, F.: A Course in Robust Control Theory: a Convex Approach. Springer, New York (2000)

Hsu, J.C.S., Meyer, A.U.: Modern Control Principles and Applications. McGraw Hill, New-York (1968)

Arimoto, S.: Control Theory of Nonlinear Mechanical Systems: a Passivity Based and Circuit Theoretic Approach. Oxford Univ. Press, Oxford (1996)

Russell, S., Norvig, P.: Artificial Intelligence: a Modern Approach. Prentice Hall, Englewood Cliffs (1995)

Burl, J.H.: Linear Optimal Control, H_2 and H_∞ Methods. Addison-Wesley Longman, Menlo Park (1999)

Leonov, G.A., Ponomarenko, I.V., Smirnova, V.B.: Frequency Domain Methods for Nonlinear Analysis: Theory and Applications. World Scientific Publ., Singapore (1996)

Astrom, K.J.: Control of Complex Systems. Springer, Berlin (2000)

Antsaklis, P.J., Passino, K.M.: An Introduction to Intelligent and Autonomous Control. Kluwer Acad. Publ., Norwell (1993)

Wang, L.X.: Adaptive Fuzzy Systems and Control: Design and Stability Analysis. Prentice Hall, Englewood Cliffs (1994)

Yudin, D., Goryashko, A.: Control Problems and Complexity Theory, Part I. Engineering Cybernetics (USSR) 12, 10–24 (1974); Part II, IBID, 13, 1–13 (1975)

Morris, K.A.: Design of Finite-dimensional Controllers for Infinite-dimensional Systems by Approximation. J. Mathematical Systems, Estimation and Control 4(2), 1–30 (1994)

Brown, M., Harris, C.: Neuro Fuzzy Adaptive Modelling and Control. Prentice Hall, Englewood Cliffs (1994)
Kosko, B.: Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence. Prentice Hall, Englewood Cliffs (1991)

Grigorenko, I.: Optimal Control and Forecasting of Complex Dynamical Systems. World Scientific, Singapore (2006)

Byrnes, C.I., Martin, C.F., Saeks, R.E. (eds.): Analysis and Control of Nonlinear Systems. North-Holland, Amsterdam (1988)

Conte, G., Moog, C.H., Perdon, A.M.: Nonlinear Control Systems - An Algebraic Setting. Springer, London (1998)

Ng, G.W.: Application of Neural Networks to Adaptive Control of Nonlinear Systems. Research Studies Press, London (1997)

Francis, B.A.: A Course in H_∞ Control Theory. Springer, Berlin (1987); Dorf, R.C., Bishop, R.H.: Modern Control Systems. Addison Wesley, Reading (1995)

Helton, J.W., James, M.R.: Extending H_∞ Control to Nonlinear Systems. Frontiers in Appl. Sci. SIAM, Philadelphia (1999)

Ionescu, V., Stoica, A.: Robust Stabilization and H_∞ Problems. Kluwer Acad. Publ., Dordrecht (1999)

Emelyanov, S.V., Burovoi, A., Levada, F.Y.: Control of Indefinite Nonlinear Dynamic Systems: Induced Internal Feedback. Springer, London (1998)

Dullerud, G.E., Paganini, F.: A Course in Robust Control Theory: A Convex Approach. Springer, New York (2000)

Fradkov, A.L., Miroshnik, I.V., Nikiforov, V.O.: Nonlinear and Adaptive Control of Complex Systems. Kluwer Acad. Publ., Dordrecht (1999)

Krstic, M., Kanellakopoulos, I., Kokotovic, P.V.: Nonlinear and Adaptive Control Design. J. Wiley and Sons, New York (1995)

Francis, B.A., Tanenbaum, A.R. (eds.): Feedback Control, Nonlinear Systems and Complexity. Springer, New York (1995)

Schroeder, D. (ed.): Intelligent Observer and Control Design for Nonlinear Systems. Springer, New York (2000)

Solodovnikov, V.V., Tumarkin, V.I.: Complexity Theory and the Design of Control Systems, Theory and Methods of Systems Analysis, vol. 28, Nauka, Moscow (1990)

Wang, H.O., Tanaka, K., Griffin, F.: An Approach to Fuzzy Control of Nonlinear Systems: Stability and Design Issues. IEEE Trans. On Fuzzy Systems 4(1), 14–23 (1996)

Gopal, M.: Modern Control System Theory, 2nd edn. J. Wiley and Sons, Toronto (1993)

Wang, L.X.: Adaptive Fuzzy Systems and Control. Prentice Hall, Englewood Cliffs (1994)

28. Deimling, K.: Nonlinear Functional Analysis. Springer, Berlin (1985)

Krasnoselskii, M.A.: Asymptotics of Nonlinearities and Operator Equations. Birkhäuser, Boston (1995)

Schwartz, J.: Nonlinear Functional Analysis. Gordon and Breach, New York (1969)

Vainberg, M.M.: Variational Methods for the Study of Nonlinear Operators. Holden Day, San Francisco (1964)

Zeidler, E.: Nonlinear Functional Analysis and its Applications, vol. I-III-IV. Springer, New York (1985-1986-1988)

Bendat, J.S.: Nonlinear Systems Techniques and Applications. J. Wiley and Sons, New York (1998)

Bobylev, N.A., Burman, Y.M., Korovin, S.K.: Nonlinear Analysis and Applications. Walter de Gruyter, Berlin (1994)

Bogaesvski, V.N., Povzner, A.: Algebraic Methods in Nonlinear Perturbation Theory. Springer, New York (1991)

Lusternik, L.A., Sobolev, V.I.: Elements of Functional Analysis. Ungar, New York (1961)

Hale, J.K., Verduyn Lunel, S.M.: Theory of Functional Differential Equations. Springer, New York (1993)

29. Corduneanu, C.: Almost Periodic Functions. Interscience Tracts in Pure and Applied Maths, vol. 22. Interscience, New-York (1961)

Dubinskij, J.A.: Sobolev Spaces of Infinite Order and Differential Equations. Teubner, Leipzig (1988)

Amerio, L., Prouse, G.: Almost Periodic Functions and Functional Equations. Van Nostrand-Reinhold, New-York (1971)

Mazja, V.G.: Sobolev Spaces. Springer, New-York (1985)

Rao, M.M., Ren, Z.D.: Theory of Orlicz Spaces. Marcel Dekker, New-York (1991)

30. Dugundji, J., Granas, A.: Fixed Point Theory, vol. I. Polish Scientific Publ., Warsaw (1982)

Jiang, T.H.: Fixed Point Theory. Springer, New-York (1985)

Zeidler, E.: Nonlinear Functional Analysis and its Applications. Springer, New-York (1986)

Smart, D.R.: Fixed Point Theorems. Cambridge Univ. Press, Mass. (1980)

Burton, T.A.: Stability by Fixed Point Theory for Functional Differential Equations. Dover Publ., New-York (2006)

31. Cotsaftis, M.: Recent Advances in Control of Complex Systems. In: Survey Lecture, Proc. ESDA 1996, Montpellier, France, ASME, vol. I, p. 1 (1996)

Cotsaftis, M.: Comportement et Controle des Systemes Complexes. Diderot, Paris (1997)

Cotsaftis, M.: Popov Criterion Revisited for Other Nonlinear Systems. In: Proc. ISIC 2003 (International Symposium on Intelligent Control), Houston, October 5-8 (2003)

32. Renouzenko, A.V.: Partial Differential Equations with Discrete and Distributed State Dependent Delays. J. Math. Anal. Applic. 326(2), 1031–1045 (2007)

33. Appell, J., Zabrijko, P.P.: Nonlinear Superposition Operators. Cambridge University Press, Mass. (1990)

34. Pachpatte, B.G.: Mathematical Inequalities. North Holland/Elsevier, Amsterdam (2005)

Bainov, D., Simeonov, P.: Integral Inequalities and Applications. Kluwer Academic Press, Dordrecht (1992)

Kim, Y.H.: Explicit Bounds on Some Nonlinear Integral Inequalities. J. Inequalities in Pure and Applied Mathematics 7(2) (2006)

Mitrinovic, D.S., Pecaric, J.E., Fink, A.M.: Inequalities Involving Functions and Their Integrals and Derivatives. Kluwer Acad. Publ., Dordrecht (1991)

35. Cotsaftis, M.: Robust Asymptotically Stable Control for Intelligent Unknown Mechatronic Systems. In: Proc. ICARCV, Singapore, December 5-8 (2006)
36. Cotsaftis, M.: Beyond Mechatronics, Toward Global Machine Intelligence. In: Proc. ICMT 2005, Kuala-Lumpur, December 6-8 (2005)
37. Cotsaftis, M.: From Trajectory Control to Task Control - Emergence of Self Organization in Complex Systems. In: Aziz-Alaoui, M.A., Bertelle, C. (eds.) Emergent Properties in Natural and Artificial Dynamical Systems, pp. 3–22. Springer, Heidelberg (2006)
Cotsaftis, M.: On the Definition of Task Oriented Intelligent Control. In: Proc. ISIC 2002 Conf., Vancouver, October 27-30 (2002)
38. Cotsaftis, M.: Merging Information Technologies with Mechatronics - The Autonomous Intelligence Challenge. In: Proc. IEECON 2006, Paris, November 6-10 (2006)

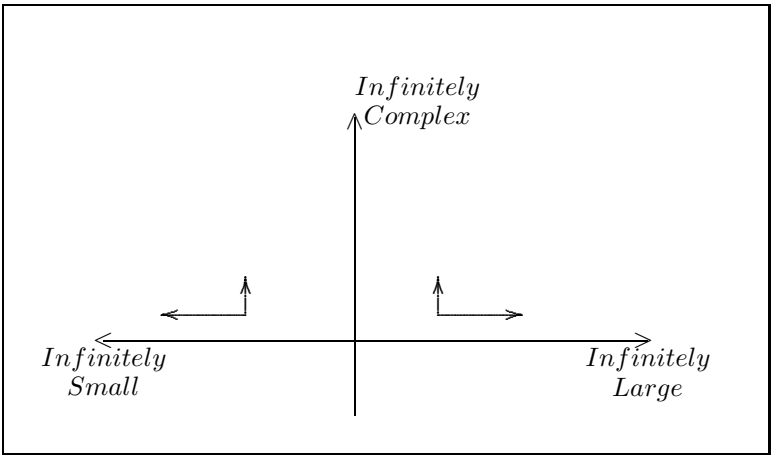


Fig. 1. New 2-D Representation of Universe with Arrows along Recent Trends

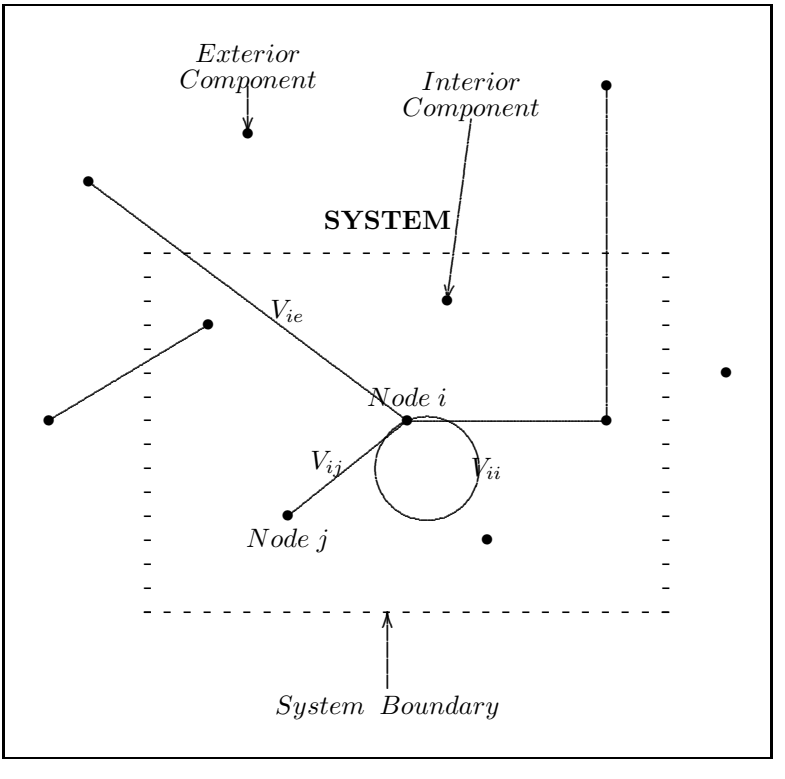


Fig. 2. Graph Representation of System with its Three Exclusive Types of Vertices V_{ii}, V_{ie} and V_{ij}

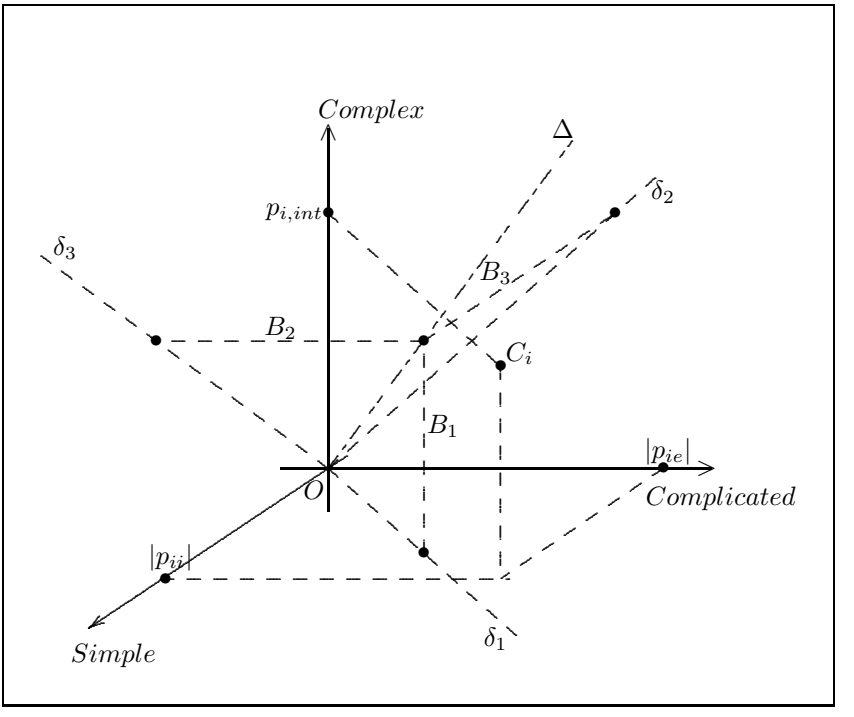


Fig. 3. System Representation in [Simple-Complicated-Complex] State-Space with Delimiting Domains : Component C_i is in Complicated State

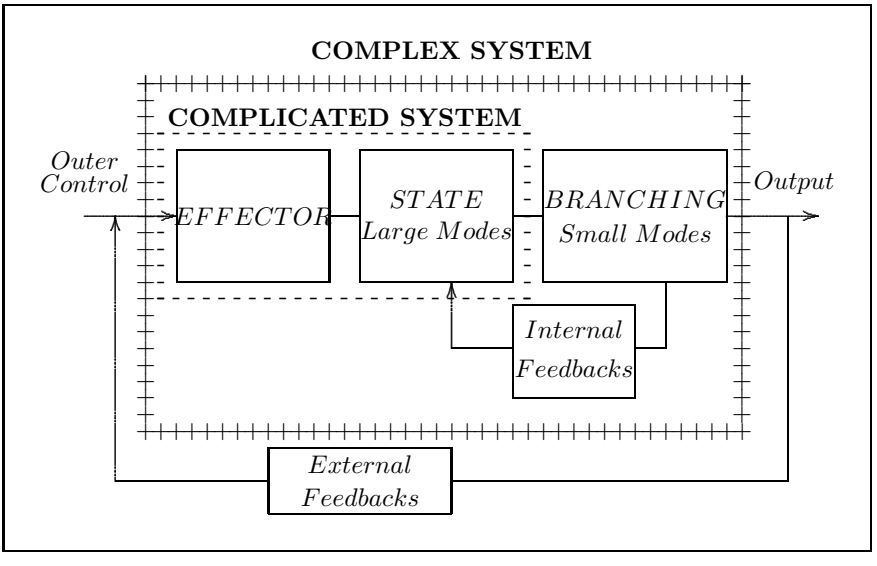


Fig. 4. Schematic Block Representation of (Controlled) Complex System with Inner Complicated Part

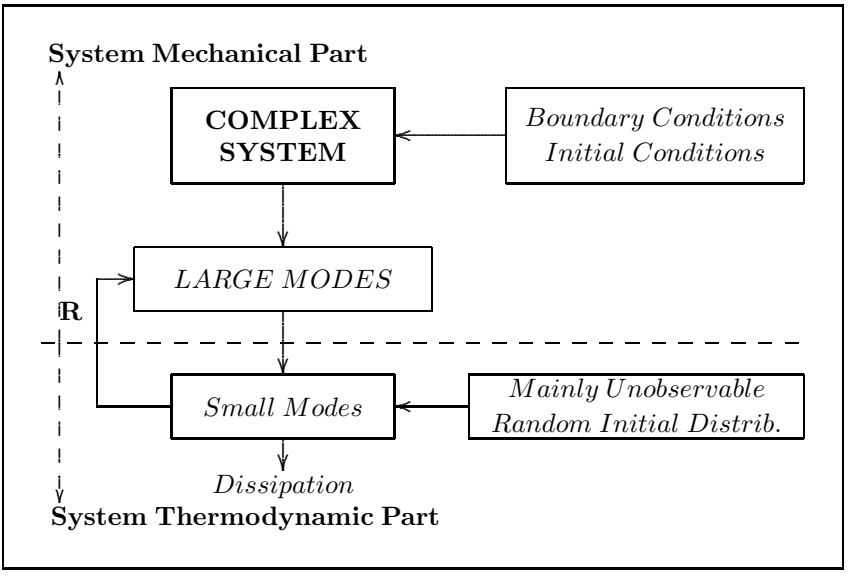


Fig. 5. Sketch of General Complex System Block Scheme

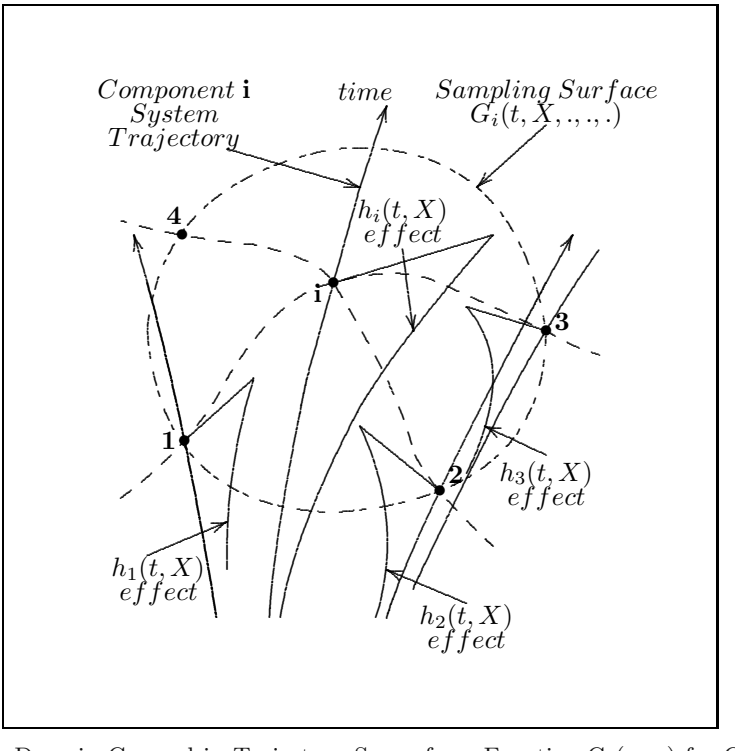


Fig. 6. Domain Covered in Trajectory Space from Function G_i(— -) for Components 1,2,3,4 and Past Trajectory Memory Effect Due to Delays h_j

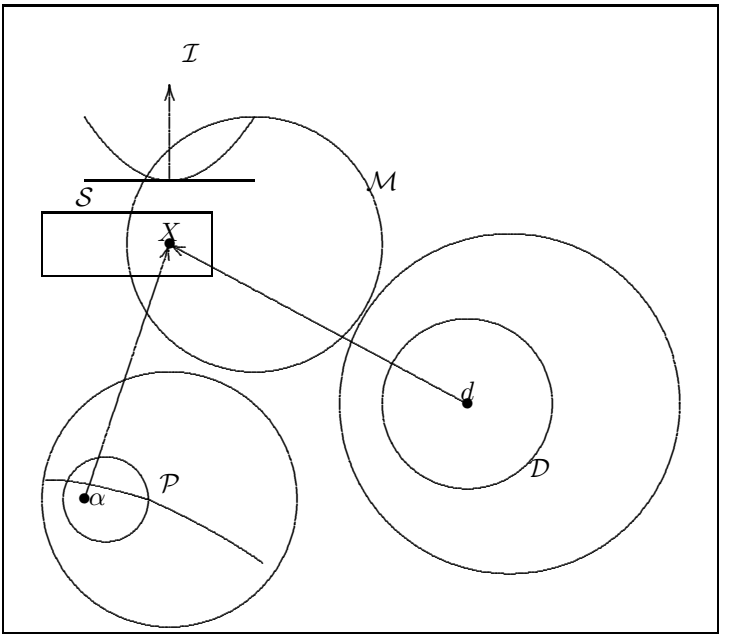


Fig. 7. Embedding of System Solution in Function Space $\mathcal{S}$ while Optimizing Trajectory Functional $\mathcal{I}$

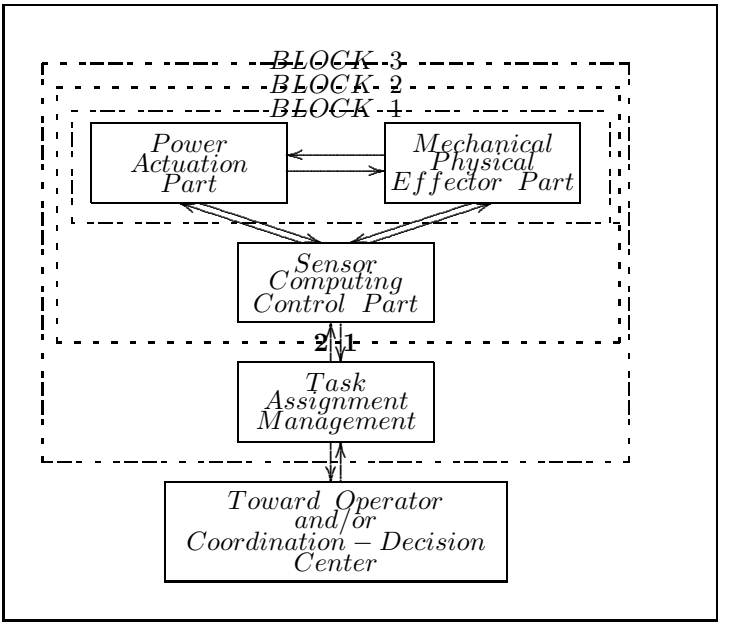


Fig. 8. System Structure Evolution with Main Component Parts - Flux 1 for Useful Information Transfer, Flux 2 for Intelligence Delegation -

A formalism for multi-level emergent behaviours in designed component-based systems and agent-based simulations

Chih-Chun Chen[1], Sylvia B. Nagl[2], and Christopher D. Clack[3]

[1] Department of Computer Science, University College London
[2] Department of Oncology and Biochemistry, University College London
[3] Department of Computer Science, University College London

Summary. There currently exists no means of specifying or analysing specific emergent behaviours in designed multi-component systems. For this reason, important questions about the lower level mechanisms giving rise to emergent behaviours cannot be resolved.
We provide a compositional definition of behaviours in terms of complex events, which can be defined at multiple levels of abstraction and related hierarchically. Based on existing theories of emergence, we also distinguish complex events that constitute emergent behaviours and those that do not. We describe how such emergent behaviours can be analysed by decomposition in terms of their underlying mechanisms.

1 Introduction

Emergent behaviours and functions in decentralised multi-component systems are of fundamental interest to both Complex Systems engineers and scientists. In both complex system engineering and agent-based modelling, there is a need to understand how certain interactions between designed or modelled components can group together to perform higher level functions [9], [20]; there is also a corresponding need to determine the lower level mechanisms underlying undesired behaviours.
There currently exists no theory of emergent behaviours in *designed or modelled* multi-component systems that allows *specific* behaviours to be described. Although general definitions of emergence in such systems exist, they only serve to distinguish emergent behaviours from non-emergent behaviours and do not address the practical issue of specifying these behaviours for empirical investigation.
To understand how behaviours at different levels relate to one another empirically e.g. which lower level interactions or mechanisms give rise to higher level behaviours, we need to be able to specify and classify them. This can be

done using a type system. Hierarchical relationships between behaviours at different abstraction levels must also be clarified so that subtype-supertype and part-whole relationships are taken into account when interpreting empirical results.
To represent behaviours in a multi-component system, we use *complex events*, which are combinations of interrelated events. Complex event *types*, can be specified to stand for particular categories of behaviours and these can be defined at any level of abstraction. Once specified, these behaviours can be identified in a dynamically executing system. This gives us a means of determining the mechanisms and interactions underlying a higher level behaviour. We show how the compositional nature of complex events allows strict hierarchical relations to be analytically established between them.
Based on previous theories of emergence we define a set of criteria with which to distinguish between complex events that represent *emergent* behaviours and those that represent non-emergent behaviours. Our work is grounded in two classes of emergence theories:

1. Definitions of emergence based on level of observation, for example information theoretic definitions, which formalise the fact that emergent properties are those that result from viewing the system at different levels of abstraction. For example, a sequence of events can be statistically significantly related to another sequence of events. However, it is often the case that only a subset of the events in the first sequence are related to a subset of the events in the sequence; these can be seen to form a minimal causal structure - other events are then (statistically) 'irrelevant'.
2. Emergence theories addressing designed or modelled systems also take into account the fact that some behaviours arise as a direct consequence of the design of each of the components while others do not. These latter behaviours arise through interactions between the components and are considered to be emergent.

1.1 Outline of the paper

The remainder of the paper is organised as follows:

- Section 2 outlines our assumptions on multi-component systems and hierarchical relationships, and also briefly reviews related work on emergence theories in multi-agent systems.
- Section 3 introduces the formalism and describes how complex event types can be used to represent behaviours at any level of abstraction that can be realised by the system or simulation.
- Section 4 expresses certain terms for emergence-related phenomena in terms of complex event types and the relationships between them. These include emergent 'laws', multi-functionality and top-down 'causation'. By expressing these in terms of relationships between complex event types, we

are able to establish empirically whether such phenomena have occurred in an executing system.
- Section 5 summarises and concludes the paper.

2 Related work: Macro-properties and emergence in multi-component systems

This section aims to clarify what we mean by different levels of abstraction and hierarchy in multi-component systems based on existing definitions of macro-properties. This is used as the basis for defining hierarchical relationships between different complex event types and behaviours at different levels. We also consider related work on defining emergence in these systems.

2.1 Multi-component systems

Before introducing the theory of hierarchy on which our work is based, it is important to first clarify our assumptions about the multi-component systems we are studying. By a multi-component system, we mean any system in which there are distinct components each of which behaves according to a set of state transition rules. A state transition rule is executed when a particular condition is satisfied. This condition might be dependent on the component's own state and/or the state(s) of other components in the system. When a rule is executed by a component, the component causes a change in state in the system (usually locally); this is known as a state transition. In summary, a state transition rule consists of (i) a condition and (ii) a state transition function, $q_1 \rightarrow q_2$, which maps an initial state q_1 to a target state q_2. In this paper, we refer to state transitions in a dynamically executing system as events, and state transitions resulting from a single state transition rule as simple events (see Section 3.2 below for a more detailed account of simple events).
A multi-agent simulation of an agent-based model is an example of a multi-component system. In an agent-based model, as in object-oriented programming, different types of components (agents or objects) are specified (designed), with each type having its own set of state transition rules. In the dynamically executing system or simulation, more than one instance of the same type of component can be instantiated. Instances of the same type would then behave in identical ways given the satisfaction of the same condition.
Figure 1 illustrates the relationship between component type and components, and between state transition functions and simple events in the executing system.

2.2 Macro-properties: Alpha and Beta hierarchy

In [10] and [11], two categories of hierarchy are described (see Figure 2):

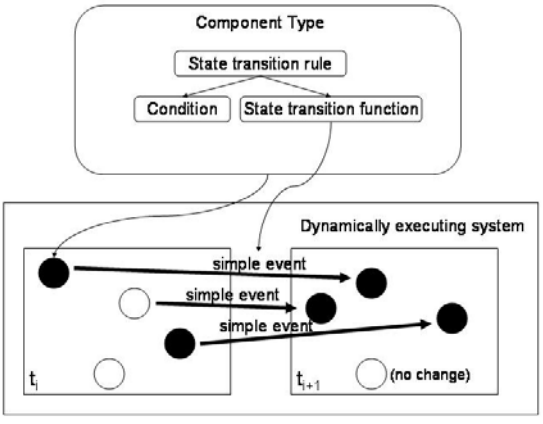


Fig. 1. Components instantiate component types in the real system or simulation while simple events can be said to instantiate state transition functions.

1. Compositional hierarchy, where lower level properties are constituents of higher level properties. This can be seen to correspond to α-aggregation, the AND relationship, or part-whole;
2. Specificity or type hierarchy where higher level properties are defined at a lower resolution than lower level properties. This can be seen to correspond to β-aggregation [10, 11], or the OR relationship.

We can relate these two categories of hierarchy to the account of micro-macro-property relationships given in [15], which defines a property P_1 to be a macro-property of another property P_2 if:

- P_2 has a greater scope than P_1;
- P_2 has a lower resolution than P_1; or
- both.

The scope of a property is the set of constituents required for the property to exist; for example, the property of being a flock requires a minimum number of birds. On the other hand, the resolution is the set of distinctions that have to be made to distinguish the property; for example, to identify a colour, one needs to be able to distinguish between a ranges of wavelengths. Information-theoretic interpretations of emergence in dynamic systems are based on the idea that often, when we are considering a greater scope, we are willing to accept some loss of accuracy or a lower resolution when predicting future behaviour (see, for example [18] and [2]).

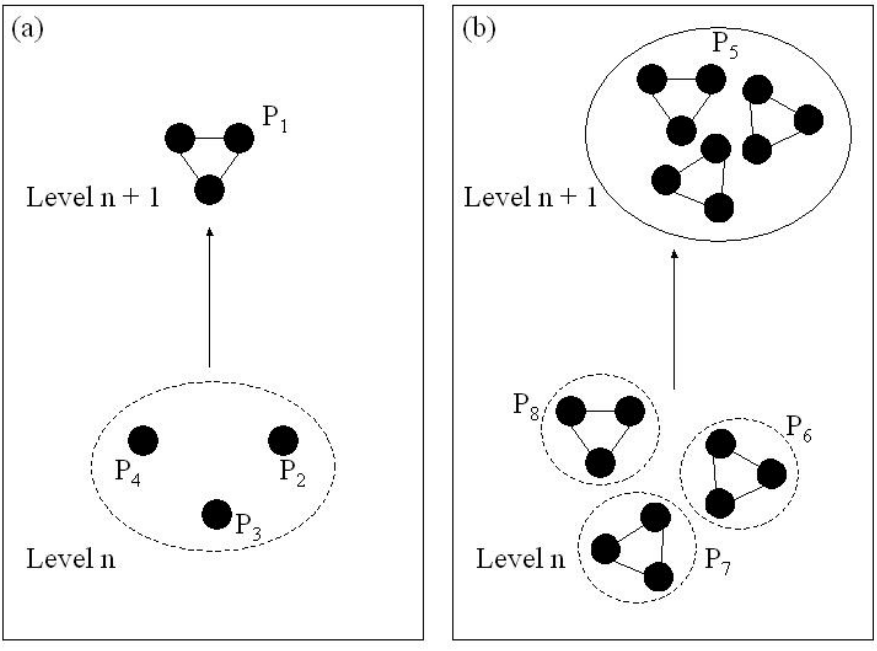


Fig. 2. Two categories of hierarchy. (a) Compositional hierarchy/α-aggregation: P_2, P_3 and P_4 are constituents of P_1. (b) Type hierarchy/β-aggregation: P_6, P_7 and P_8 fall in the set defined by P_5.

2.3 Theories of emergence in designed multi-component systems

Several theories of emergence specifically addressing designed multi-component systems have been proposed (see [5] for a review). These tend to be grammar-based definitions (see, for example, [4], [6], [13]), which attempt to formally relate the specified (designed or modelled) aspects of the system to those that are said to be emergent. A grammar L_{WHOLE} is associated with certain properties that arise only when the system operates as a whole, and this grammar is 'more than' the grammar L_{PARTS} associated with the specifed aspects of the system. In the context of multi-component systems, L_{PARTS} can be seen to represent the component specifications with their state transition rules while L_{WHOLE} can be seen to include properties and behaviours that are not contained in the component specifications. Where definitions differ is in how they express L_{PARTS} and L_{WHOLE}.

In [13], for example, L_{PARTS} is obtained by allowing components to evolve independently and then put the results together afterwards while L_{WHOLE} is obtained by allowing the components to evolve together. On the other hand, Demazeau's pseudo-equation is more specifically aimed at multi-agent systems:

$$MAS = A + E + I + O + emergence,$$

where A stands for Agents; E for Environment; I for Interactions; and O for Organisations. L_{WHOLE} can then be equated with $L(MAS)$, which is the language that can be generated by the multi-agent system when considered as a whole, while L_{PARTS} can be equated with $L(A) + L(E) + L(I) + L(O)$. Emergent aspects of the system can be seen to be those that can be generated by L_{WHOLE} but not by L_{PARTS}.

While several other definitions of emergence also exist, these tend to either be subsumed under the micro-macro relationship that we have already considered or are associated with more specific emergence-related terms, such as downward causation (e.g. [17]), which we address later on in Section 4.

2.4 Identifying emergent behaviours in multi-component systems

Two popular examples of multi-component systems are swarm systems and multi-agent simulations.

The aim of swarm systems engineering is to design agent components so that they collectively give rise to emergent behaviours (e.g. [7]) while in multi-agent simulations, agent components are modelled so that they collectively give rise to behaviours that can be seen to represent emergent phenomena in the real world [19]. Both require the ability to determine whether and how often a particular emergent behaviour has occurred.

In agent-based modelling, a variable is usually introduced that is believed to reflect the occurrence of an emergent behaviour (e.g. [1], [21], [14]), and this is tracked through the course of the simulation. Changes in the variable's value then reflect the different frequencies of the behaviour's occurrence. However, because the structure of behaviour is lost, this method fails to address situations where more than one emergent behaviour can give rise to the same changes in the variable's value.

In swarm systems engineering, the methods typically focus on particular patterns of interaction (e.g. [20]) but tend to focus on the observed behaviour while ignoring the designed aspects of the system since no distinction is drawn between two identical behaviours arising from different rule executions.

In general, there currently exists no means of specifying or analysing specific emergent behaviours in designed multi-component systems. Specifically, existing methods do not take into account both the macro-micro relationship between behaviours and the designed-observed distinction. For this reason, important questions about the lower level mechanisms giving rise to emergent behaviours cannot be resolved.

3 Multi-level and emergent behaviours as complex events

In this section, we introduce our complex event formalism and briefly describe how complex event types can be used to represent multi-level behaviours in agent-based simulations and systems. We then relate compositionality in complex event types to scope, and specificity to resolution. The formalism is also used to express important emergence constructs and definitions more precisely. This allows complex event types to be classified into those that correspond to emergent behaviours and those that do not.

3.1 Events

We take as a starting assumption the fact that behaviours in multi-component systems are events, which are defined as changes in state (see Definition 1).

Definition 1. *Event. An event is a state transition defined at a particular level of abstraction:*

$$e \equiv q_{1L} \Rightarrow q_{2L},$$

where

- q_{1L} *is the initial state described at level* L*;*
- q_{2L} *is the target state described at level* L*; and*
- $\Rightarrow$ *is a state transition function that results from the execution of one or more state transition rules.*

In many cases, state transition functions can also be decomposed into lower level state transition functions. For example, a state transition function that maps a source subsystem state q_{Sub} to another subsystem state q'_{Sub} (the target state) might be decomposable into a set of lower level subsystem state mappings $\{(q_{Sub}* \rightarrow q_{Sub}*')\}$, which can be further decomposed into a set of component state mappings $\{(q_C \rightarrow q'_C)\}$. If variables are the lowest level of state representation, every state transition function would ultimately be reduced to a set of state transitions mapping a variable value to a new variable value $\{(var \rightarrow var')\}$.

3.2 Simple events and complex events

We define a *simple* event as a state transition that results from the execution of a single state transition rule, and this state transition function can be defined at any level of abstraction.

Definition 2. *Simple event. A simple event se is a state transition defined at some level of abstraction that results from the execution of a single state transition rule:*

$$se \equiv q_{1L} \rightarrow q_{2L},$$

where

- q_{1L} *is the initial state described at level* L*;*
- q_{2L} *is the target state described at level* L*; and*
- $\rightarrow$ *is a state transition function that results from the execution of a single state transition rule.*

A *complex* event ce is defined as either a simple event se or two complex events linked by the non-commutative relationship $\bowtie$ (see Definition 3).

Definition 3. *Complex event (recursively defined). A complex event is either a simple event* se *or two complex events* ce_1 *and* ce_2 *satisfying a set of constraints with respect to each other:*

$$ce \;::\; se \mid ce_1 \bowtie ce_2$$

In a dynamically executing multi-component system, the relationship $\bowtie$ might, for example, be a temporal operator $\otimes$ optionally followed by descriptions of (i) space constraints and (ii) constraints pertaining to the variables or components of the two related complex events. In this example, the syntactic pattern for a complex event relationship $\bowtie$ would be given by:

$$e_1 \;\bowtie\; e_2 \;::\; e_1 \;\otimes\; [space]\;[var]\;e_2$$

where

- **The temporal constraint '$\otimes$':** defines the temporal relationship between $e1$ and $e2$;
- **The spatial constraint '*space*':** defines the space within which e_2 should occur relative to e_1; and
- **The component or variable constraint '*var*':** defines the relationships between variables or components of the two events e_1 and e_2.

More detailed examples can be found in [3]. For example, temporal constraints can be specified using various temporal operators such as '$e1 \| e2$' to indicate that $e2$ is initiated at the same time as $e1$; '$e1 \prec e2$' to indicate that $e2$ occurs after $e1$; or 'e1 *; e2*' to indicate that $e2$ immediately follows $e1$.

3.3 Simple and complex event types

We can define a simple event type SET by the two-tuple:

$$SET = (\rightarrow, Level),$$

where

- $\rightarrow$ is a state transition function that results from the execution of a particular state transition rule; and
- *Level* is the level of abstraction at which the state transition is described.

A simple event se has the type SET_X *iff*:

- $se = q_{1L} \rightarrow q_{2L}$;
- $\rightarrow = A$; and
- $L = B$,

where $SET_X = (A, B)$.

$\rightarrow$ allows us to distinguish between events resulting from different state transition rules, while *Level* allows us to distinguish between different levels of observation. These distinctions become important when defining emergent behaviours (see Section 3.5) and micro-macro relationships between behaviours (see Section 3.4).
A complex event type CET can be defined by the four-tuple:

$$CET = (\{SET_i\}, \{CET_j\}, \{\bowtie_k\}, \{CET_l\}),$$

where

- $\{SET_i\}$ is a set of simple event types;
- $\{CET_j\}$ is a set of complex event types;
- $\{\bowtie_k\}$ is a set of non-commutative location constraint relationships; and
- $\{CET_l\}$ is a set of complex event types.

A complex event ce is then said to have the type CET_X *iff*:

1. $ce = se$ and the type of $se \in A$; or
2. - $ce = ce_1 \bowtie ce_2$;
   - the type of $ce_1 \in B$; and
   - $\bowtie \in C$; and
   - the type of $ce_2 \in D$,

where $CET_X = (A, B, C, D)$.
The type of a complex event is therefore determined both by the types of its constituent events and the relations that hold between them. This definition of a complex event type reflects the hierarchical structure of a complex event since two complex events are only of the same type if they have the same structure.

3.4 Hierarchy, Scope and Resolution of complex event types

As described in Section 2.2, two types of macro-micro relationships can be distinguished between properties. The first is the compositional relationship or α-aggregation, where a micro-property is a constituent of a macro-property. The second is the subtype-supertype relationship or β-aggregation, where a micro-property is a subtype of a supertype.

In the context of complex event types, we can say that a complex event type CET_A is a *constituent* of a complex event type CET_X if:

$$CET_X = CET_A \bowtie CET_B$$

or

$$CET_X = CET_B \bowtie CET_A.$$

CET_X is also said to have a greater scope than CET_A.
To say that a complex event type CET_A is a *subtype* of a complex event type CET_X, the set of events that can be classified as CET_A must be a subset of CET_X:

$$E_{CETA} \subseteq E_{CETX}.$$

CET_X is then said to have a lower resolution than CET_A.

3.5 When do complex event types represent emergent behaviours?

In Section 2.3, we reviewed emergence theories for designed multi-component systems and distinguished between the designed aspects of the system (L_{PARTS}) and those aspects of the system that are not designed i.e. can not be generated solely by L_{PARTS}.
For behaviours, L_{PARTS} can be seen to correspond to the state transition rules. For a complex event type to represent an emergent behaviour, we therefore require that it contains some constraint that is not included in the state transition rules. This is the case when the complex event type can not be expressed in terms of a single simple event type i.e. it can not be generated from the execution of a single state transition rule.[4]

4 Empirical investigation of relationships between multi-level behaviours using complex event types

In the previous section, we defined two categories of micro-macro relationship between complex event types, which (given that complex events represent behaviours) correspond to two categories of relationship between behaviours at different levels in a multi-component system. In this section, we describe more specific relationships between complex event types that can be used

[4] An important point to note here is that the distinction between simple and non-simple complex events lies not in the scopes and resolutions of their state transitions, but in their source or origin. Whereas simple events are those that arise from the application of a single state transition rule, the non-simple complex events are those that are either defined at a lower resolution (i.e. include more than one type of simple event) or result from the execution of more than one state transition rule.

to formalise emergence-related phenomena, such as multi-functionality and downward causation. In each case, by giving a set of criteria that must be satsified by the relationship between a pair of complex event types, we can establish empirically whether that phenomenon has occurred.

4.1 Emergent 'laws' and top-down 'causation'

A popular application of the multi-component paradigm is in multi-agent simulations of complex systems to understand their properties. Scientists working in this area are particularly interested in establishing rules at higher levels, which they often call emergent laws. Top-down constraints (also known as top-down 'causation'[5]) and feedback are also effects that are sought in simulations. While simple statistical techniques can be used to detect these effects at the global level, a hierarchical approach has yet to be established. Definitions 4 and 5 support such an approach by stating in terms of complex event types the criteria for behaviours at any two levels to be related in these two ways.

Definition 4. *An emergent law exists between two complex event types CET_x and CET_y when the occurrence of an event of type CET_x implies the occurrence of some complex event of the type CET_y i.e. $CET_x \rightarrow CET_y$.*

A weaker version of Definition 4 allows that CET_x only increases the probability of CET_y or that there is a correlation between CET_x and CET_y.

Definition 5. *A top-down constraint effect exists between two complex event types CET_M and CET_m when an emergent law $CET_M \rightarrow CET_m$ holds and CET_m is a lower level complex event type with respect to CET_M.*

4.2 Multi-functionality

Finally, we can use the complex event type formalism to express multi-functionality. A complex event type can be decomposed into different combinations of constituent events (see also the decomposition of property C in Figure 3), e.g.:

$$CET_A = CET_{x1} \bowtie_1 CET_2 = CET_{x1} \bowtie_2 CET_3 \tag{1}$$

where $\bowtie_1$ and $\bowtie_2$ are different sets of configuration constraints.
The overlapping of constituent properties gives us a criterion by which to establish multi-functionality.

[5] Here, we make no assumptions about the metaphysical status of top-down constraints i.e. we take an agnostic stance on whether these constraints have real causal power, supervene on lower level laws, or are epiphenomena. This has been a long-standing debate in the Philosophy of Science, see, for example [8], [16], [12]

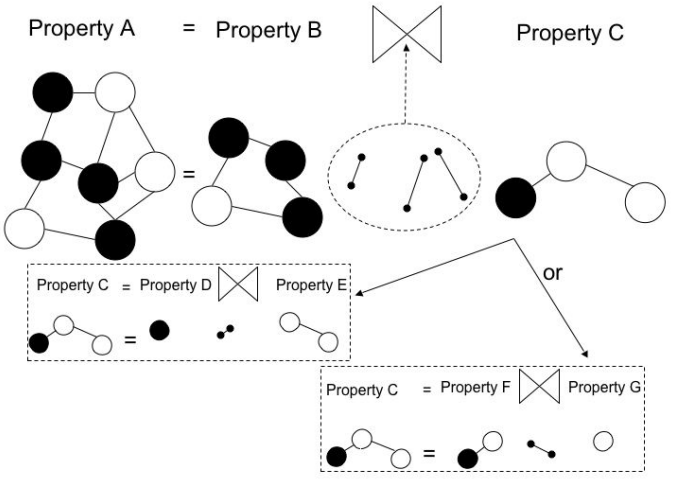


Fig. 3. Property A is composed of Property B and Property C in some configuration. B and C can be further decomposed, as can their constituents. $\bowtie$ denotes the set of constraints that determine the configuration; x, y and z are used to distinguish between two different sets of constraints. The alternative decompositions of property C illustrate the fact that a property can be decomposed in more than one way, and that a constituent can have more than one role or function. Property D, for example, might have an independent role as in the leftmost decomposition, while in the righthand decomposition, it is a constituent of property F.

Definition 6. *A complex event type CET_x is multi-functional if it is a constituent event in more than one complex event type.*

Example 1. If

$$CET_A = CET_x \bowtie CET_y \tag{2}$$

and

$$CET_B = CET_x \bowtie CET_z \tag{3}$$

we say that CET_x is multi-functional and that it has a role in CET_A and CET_B

Multifunctionality can also apply to system components and agents. If a component participates in more than one typed complex event at the same time, we say that it plays more than one role and is hence multifunctional. The complex event types in which it plays specific roles determine its function.

5 Summary and Conclusions

This paper has introduced a formalism for describing emergent behaviours in multi-component systems at any distinguishable level of abstraction.

In the first part of the paper, we introduced a compositional approach to analysing multi-level properties in terms of the hierarchical relations that hold between them. In α-hierarchy, a higher level property consists of lower level constituent properties that are organisationally related. In β-hierarchy, a higher level property consists of a set of lower level properties that can all be treated as exemplars of this property.
This general compositional approach was then applied more specifically to behaviours using complex event types to represent different categories of multi-level behaviours. By categorising different sub-trajectories of system and sub-system evolution, we are able to understand the mechanisms at work at multiple system levels. Furthermore, we are able to establish both empirical and analytical relationships between behaviours at different levels. Analytical relationships include subtype-supertype and part-whole relationships while empirical relationships include correlative and associative relationships. Our method allows for the decomposition of higher level emergent behaviours to smaller and smaller 'motifs' of behaviour and ultimately to state transition rule executions (simple events). The discovery of empirical associations between behaviours at different levels would give us a sound basis for making engineering decisions and for modifying multi-component systems in order that they should exhibit (or not exhibit) certain emergent behaviours.

Acknowledgements

The work presented in this paper is the result of a multi-disciplinary collaboration between the departments of Computer Science and Oncology at UCL. Specifically, Chih-Chun Chen contributed expertise in complexity science and multi-agent systems; Sylvia Nagl contributed expertise in complexity science, systems biology, and philosophy of science; and Christopher Clack contributed expertise in type systems, logics, rule-based and adaptive systems.

References

1. Bankes, S.C.: Agent-based modelling - a revolution? PNAS 99, 7199–7200 (2002)
2. Bonabeau, E., Dessalles, J.L.: Detection and emergence. Intellectica 2(25), 85–94 (1997)
3. Chen, C.-C., Nagl, S.B., Clack, C.D.: Specifying, detecting and analysing emergent behaviours in multi-level agent-based simulations. In: Proceedings of the Summer Simulation Conference, Agent-directed simulation. SCS (2007)
4. Dassow, J., Freund, R., Paun, G.: Cooperating array grammar systems. International Journal of Pattern Recognition Artificial Intelligence 9, 1029–1053 (1995)
5. Deguet, J., Demazeau, Y., Magnin, L.: Elements about the emergence issue - a survey of emergence definitions. ComPlexUs 3, 24–31 (2006)

6. Demazeau, Y.: Steps towards multi-agent oriented programming. In: First International Workshop on Multi Agent Systems, Boston, Mass (1997)
7. Dorigo, M., Tuci, E., Trianni, V., Gross, R., Nouyan, S., Ampatzis, C., Labella, T.H., O'Grady, R., Bonani, M., Mondada, F.: SWARM-BOT: Design and Implementation of Colonies of Self-assembling Robots, pp. 103–135 (2006)
8. Friedman, M.: Explanation and scientific undersatnding. Journal of Philosophy 71, 5–19 (1974)
9. Gleizes, M.-P., Camps, V., George, J.-P., Capera, D.: Engineering systems which generate emergent functionalities. In: Weyns, D., Brueckner, S.A., Demazeau, Y. (eds.) EEMMAS 2007. LNCS, vol. 5049, pp. 58–75. Springer, Heidelberg (2008)
10. Johnson, J.: Hypernetworks for reconstructing the dynamics of multilevel systems. In: Proceedings of European Conference on Complex Systems (November 2006)
11. Johnson, J.: Multidimensional Events in Multilevel Systems, pp. 311–334. Physica-Verlag HD (2007)
12. Kim, J.: Downward causation. In: Emergence or reduction?, vol. 1, pp. 119–138. Walter de Gruyter & Co. (1992)
13. Kubik, A.: Toward a formalization of emergence. Artificial Life 9, 41–66 (2003)
14. Parisey, N., Beurton-Aimar, M., Lales, C., Strandh, R., Mazat, J.P.: Towards modelling the q cycle by multi agent systems. In: ECAL 2005 (2005)
15. Ryan, A.: Emergence is coupled to scope, not level. Nonlinear Sciences (2007)
16. Salmon, W.: Scientific explanation and the causal structure of the world. Princeton University Press, Princeton (1984)
17. Sawyer, R.K.: Simulating emergence and downward causation in small groups. In: Moss, S., Davidsson, P. (eds.) Multi Agent Based Simulation: Proceedings of the Second International Workshop on Multi-Agent Based Simulation (MABS 2001), pp. 49–67. Springer, Heidelberg (2001)
18. Shalizi, C.: Causal Architecture, Complexity and Self-Organization in Time Series and Cellular Automata. PhD thesis, University of Michigan (2001)
19. Shalizi, C.R.: Methods and Techniques of Complex Systems Science: An Overview, pp. 33–114. Springer, New York (2006)
20. Sudeikat, J., Renz, W.: Toward systemic mas development: Enforcing decentralised self-organisation by composition and refinement of archetype dynamics. In: Engineering Environment-Mediated Multiagent Systems (EEMAS 2007). LNCS. Springer, Heidelberg (2007)
21. Walker, D.C., Southgate, J., Hill, G., Halcombe, M., Hose, D.R., Wood, S.M., Neil, M., Smallwood, R.H.: The epitheliome - agent-based modelling of the social behaviour of cells. BioSystems 76, 89–100 (2004)

Emergence of Chaos and Complexity During System Growth

Andrzej Gecow*

Institute of Paleobiology Polish Academy of Science, Warsaw Poland
gecow@twarda.pan.pl

Summary. The main topic of this article is the emergence of chaos in networks describing adaptive systems. We investigate this process mainly during the system growth in dependency on network size when other parameters of the network do not change. However, we also compare the degree of chaos for different parameters and network types including random Erdős-Rényi and BA scale-free networks. We use Kauffman networks, and we follow Kauffmann using their parameters and the notion 'chaos' for them. However, we use more than two signal variants which we assume to be equally probable, therefore the Kauffman networks considered can become different from the Boolean networks. The terms 'complex system' and 'complex network' are commonly used but they have no common established definitions. We find that chaotic properties of networks well meet our intuition of complexity and that the appearance of chaotic features during system growth can be treated as complexity threshold. Crossing this threshold defines certain properties of system and mechanisms which create 'structural tendencies'. These interesting phenomena, however, are described in another article in this book.

Keywords: Kauffman network; Boolean network; damage spreading; chaos; adaptive system.

1 Introduction

The main topic of this article is the emergence of chaos in networks describing adaptive systems, however, in this article adaptive condition is not used yet. It will be the basic assumption in my next article in this book which is a continuation of this one.

We use Kauffman networks [19, 21] and similar functioning networks named aggregate of automata [9, 11, 10, 12]. We also follow Kauffman in naming their parameters and using the notion 'chaos' which can be defined as a large

* Research financed by government grant.

probability of damage avalanche. Damage is the difference in function between two identical systems which appears as an effect of some typically small disturbance in one of these systems [29].

The existence of an equilibrium level as the limit of damage growth is the main difference between this 'chaos' and the more commonly used definition [30]. High stability of ordered systems does not allow for damage to evolve into an avalanche.

We use more than two signal variants. We suggest it as much more adequate (ch.2) than only two variants. One of the typical ways leading to two alternatives is our concentration on one particular event and collecting all the remaining events as the second alternative. With more than two signal variants, the Kauffman networks considered here become different from the Boolean networks which Kauffman used (RBN-familly containing CRBN [44], SFRBN,EFRBN [18]). For all these signal variants we assume identical probability and then we denote their number as s. Such networks are also different than RNS [35, 47] and RWN [36, 8]. Using equal probability is similar to the Derrida and Kauffman method which we expand for $s > 2$. With such an assumption we should expect the system to be chaotic. We show that the case $s = 2$ is especially extreme. Only for this case can an ordered system be obtained. We assume for the whole of our investigation that the 'internal homogeneity' of function [21] is minimal.

We estimate that adaptive systems - the typical living objects or systems designed by human - are chaotic,we don't believe that they evolve on the edge of chaos. In section 2 we collect some heuristic observations and remarks which provide an intuitive base for such a view. For adaptive systems we observe a large set of really random changes which with high probability can cause a large avalanche of damage. There exists a certain strange exception: the gene regulatory network described by the Kauffman model [20, 34, 31, 32] which successfully uses Boolean networks; however lots of other, less unusual examples can be found, e.g. [37, 11].

In ch.3 we describe some theoretical expectations. Expanding the calculation [21] (prepared for the case $s = 2$, commonly [18] named 'Derrida plot' and based on annealed approximation model) for cases $s > 2$ we obtain theoretical levels of damage equilibrium (ch.3.1). These levels for $s > 2$ are much higher than for $s = 2$ which underlines the importance of parameter s. In the next ch.3.2 we define a simple intuitive coefficient of damage propagation which gauges the ability of damage to explode.

The theoretical expectations shown in section 3 are independent of network types. In section 4 we define five network types which we will later investigate using simulation. In this set there are the old random Erdős-Rényi network

on which Kauffman works and the new Barabási-Albert scale-free network including our aggregate of automata. Following Kauffman we use a constant number K of node inputs and a flexible number k of node outputs. In the next chapter (section 5) our special algorithm for simulation is described. It is a simplified algorithm dedicated for statistical investigation of damage spreading in chaotic systems. We will use it in the next article (in this book) as a key allowing us to simulate 'structural tendencies'.

When the theoretical expectation, set of network types and algorithm are clear we perform simulation experiments for checking this expectation for the defined set of network types and for observing some differences in degree of chaos between network types. The results of these experiments are described in section 6. The main effect of this investigation is the definition of sequence of network types according to growing chaos degree and the mechanism of this differentiation. This mechanism is connected with the dispersion of node degree k where hubs which collect a big part of links decrease the local coefficient of damage propagation in remaining areas of the system. Using the same experimental data we compare the influence on chaos degree between the old parameter K and new s, also we find a specific role of parameter s.

The next step is an investigation of emergence of chaos by watching the maximal size of damage during network growth i.e. in dependency on network size which was measured using the number N of nodes in the network. Above we have limited ourselves to watching damage only inside of a network (on the nodes' outputs), now we also introduce external outputs of the system which are more interesting. In the next article they will be used for the definition of fitness which will be the main focus of the article.

To understand why we are going to collect particular data in simulation (section 8.3) we first explain the mechanisms of damage propagation and prepare intuition for the main notions and the parameters which approximate them (section 7).

We start from the simplest and most intuitive terms like 'functional sequence' and connected to it 'cone of influence' which can be expected to be non-applicable to randomly growing complex networks containing lots of feedback loops. Analysis of such simple terms gives us a certain phenomenon and its mechanism, which well describes the intuitive meaning of the term 'complexity of system' and defines an useful to detect threshold upon which a system should be treated as complex. Generally, it is the ability of typical damage observed on system outputs to reach the full level of damage equilibrium, i.e. it is the acquirement of chaotic characteristics by a growing system. It needs a particular size and structure of the system. In the first part of this article (section 6) we observed a dependency of chaos degree on the network type which is connected to structure. Near the complexity threshold in the

distribution of damage size there appear two clear, already known peaks. One of them occurs for very small changes of real fade out and the second one occurs for very large changes (pseudo-fadeout), but 'medium' changes disappear. Such a picture occurs also in percolation theory: upon percolation point large percolating cluster (right peak) grows and small clusters decrease (left peak) but there are no medium clusters.

This phenomenon is closely connected to the mechanisms of structural tendencies in adaptive evolution of complex systems. These mechanisms are investigated in the next step (in the next article in this book). These tendencies are known in the classic comparative embryology and are commonly observed in human activity but up till now they have no explanation.

As was mentioned above, now we add inputs and outputs to the system. In the earlier investigations we have considered damage in autonomous systems. Such a step is natural if we are going to consider a clear cone of influence, which needs excluding feedback loops. Accomplishing of network with external inputs and outputs allows us to keep the natural assumption that all nodes may have inputs and outputs. When feedbacks are absent we can precisely define the intuitive and useful term 'functional sequence' (functional order) which in turn allows to define 'cone of influence'. When feedbacks become allowed, cone of influence and functional order become blurred, but not fully. The presence of input and output of the system allows us also to distinguish areas inside the system using some distance to inputs or outputs. The mechanism of our complexity threshold is connected to such a distance of damage source to system outputs. This distance, called 'depth', is an approximation and substitution of functional sequence. In the next article the state of system outputs is a basis for a measure assessing the system function - fitness parameter. Adaptive changes (which must not decrease fitness) are very small and appear in the range of the first peak in the distribution.

Complexity is an intuitive notion without one general, commonly used, scientific definition. It is much poorer defined than chaos. There are as many different definitions [41, 5, 52] as different aspects and applications of this notion. On the Internet we can find 'Complexity_Zoo' with 468 contemporary classes. Most of the authors use a certain measure of complexity when discussing complex systems (e.g. algorithmic complexity [23], or Solomonov-Kolmogorov-Chaitin complexity [38]). This is different from our own approach and has different goals. We consider a certain 'phase transition' in the system maturation process during growth, but it does not depend on one parameter only. In research about complex systems there often is a reference to a critical value for a given parameter. For example in scale-free networks [1, 14], above a critical network size, the average length of the path between two nodes will not change with the growth of the network. Similarly, in self-organising criticality [7,6] after some parameters reach a critical value,

the system will spontaneously exhibit behaviour characterized by power laws. However, it is important to clearly know, what the term 'complex' means if it appears in a title of publication [12] and this parameter (level of complexity) can have a significant influence on the considered phenomenon.

The complexity threshold proposed in this article is a real phenomenon, however, it is not a critical point but a smooth change (this is the phenomenon property) which can creates some criticism. Criterion of zero appearance creates a practical method, not a particular critical value because it a little depends on number of events. It is also not general definition and threshold of complexity, e.g. it does not concern ordered systems (opposite to chaotic ones), which in some cases we like to describe as complex, it also cannot be applied to static networks (not functioning). The definition introduced here can be simplified for autonomous systems i.e. limited to damage size distribution defined inside the system, like in the first part of this article, but typically systems are observed from the outside, using a limited number of outputs. Chaotic systems with lots of outputs form the main range of applications of our complexity threshold.

2 Interpretational Estimation of Chaos and Number of Equally Probable Signal Variants for Real Adaptive Systems

The Kauffman network [19,21] is the most adequate network type to describe a living object or a system designed by human, whose main assessed properties are effects of its function. However, in the Kauffman model there are two assumptions, which we do not share in our estimation of model parameters especially of the typical living and human-designed systems. The first of these assumptions is that spontaneous order typically has a large significance in such systems and the second one - that the number two of signal variants, especially when they are equally probable, is adequate to describe function and behaviour of such systems. These discrepancies are not independent in the case of useful equality of probability of the variants - 'order' appears only for two variants (when internal homogeneity P of function is minimal which we assumed for the whole of this article). However, there is one strange exception from our view - the main application of Kauffman model. Gene regulatory networks [20] are described successfully [34, 31, 32] using Boolean networks, but this is an exception which proves the statistical rule - it, however, neglect the basic stability of living objects based on negative feedbacks.

The terms: 'Kauffman networks' and 'Boolean networks' were synonymous. For more than two variants of signal which we introduce such networks cannot be 'Boolean' anymore but they can and should remain 'Kauffman networks'. Systems created or maintained by humans and living systems grow or are

designed under adaptive conditions. We name them 'adaptive systems'. We distinguish this set of systems because of the aspect of purposefulness which is typically present when such systems are described. Summarizing our observations of these systems we have collected a few remarks or intuitive estimations. We will not prove them but we will agitate. These remarks will build an intuitive base of assumptions we will introduce, which should have some base.

2.1 Typical Living Object or a System Designed by Human are Chaotic

If a really random change happens, as opposed to a predicted event to which the system is adapted, then with high probability we can expect a large avalanche of damage which typically leads to a critical malfunction of the system. This is chaotic behaviour in the meaning of Kauffman model [21]. Note, we exclude here the large set of random changes which are 'predicted' by the adaptive process of designing of these systems because the system's reactions to such changes are not random but special. We know which changes are not random in the human-designed system but in a living system we only know that it is an effect of the Darwinian mechanism and we expect that lots of reactions to random-like changes are not really random. The existence of a large set of changes which create damage avalanches is enough to treat such systems as chaotic. This important estimation will be a part of the basis of adequateness of our assumption of more than two equally probable signal variants for the considered set of systems because this assumption leads to chaos. Using this estimation we will limit our self to potentially chaotic systems investigating complexity notion and threshold of complexity which we are going to use in investigation of structural tendencies in adaptive evolution, however this last exciting theme exceed range of this article.

Typically we find mechanism of such a not random reaction which is based on negative feedbacks. It is homeostatic in the typical meaning of this term and it is 'ultrastable' using Kauffman's term [21]. Kauffman uses the 'homeostatic stability' term for spontaneous resistance of system to disturbance which may also result from adaptation but it is an effect of system type founded by adaptation which is 'ordered'. Damage avalanches on great scale are impossible in ordered systems, they should fade out quickly or stay on low level.

Examples of such really random changes occur when control parameters rise above their critical levels [37] which are the boundary of tested (i.e. predicted) range of these parameters or in other cases which are so strange [11] (i.e. have especially small probability or are especially complicated which also leads to small probability) that they did not occur often enough or due to complication - not solved yet or possible mechanisms are too expensive. For living objects one mechanism is always standing by - it is reproduction, it may be cheaper.

Do you believe us that you are chaotic? If not, then imagine, please, that you are a patient. Can a surgeon expect during a medical operation on your body that his mistakes will be neutralized by your 'homeostatic stability' (in Kauffman sense)?

2.2 Case 'Two Variants of Signal' is Extreme, May Lead to Another Phase

A methodological and philosophical reason for more than two variants of signal is that the case of 'two variants' is an extreme one - there is no smaller sensible value for number of variants, only higher values. When we model a real system and we do not know the value of a parameter we should take a middle value, i.e. probably near the average value. This is a known, safe method leading to more adequate models. It assumes that the more average value is typically the more probable one which is suggested by the typical Gaussian distribution. Discontinuities can appear anywhere and for more probable values of a parameter we have no way to avoid it but for extreme values, which therefore are special and more naturally lead to special effects, we can and should avoid them due to the expectation of their low probability.

This is philosophy, however, the extreme case $s = 2$ (two equally probable signal variants) together with the similar extreme case $K = 2$ (two inputs per node) really leads [13] to an especially extreme case - crossing of phase transition from chaos (which occurs for all others s and K) to order (which occurs only for such a combination). Then this philosophy works. For $s = 2$ and $K > 2$ there are no such special cases and $s = 2$ seems safe in this area but 'working philosophy' suggests not to use such an extreme value, but a more probable higher value instead. We neglect the strange case $K = 1$ however we can find it in the literature [21,34]. Note that using such a suggestion we are coherent with the above estimation that modelled adaptive systems should be chaotic.

2.3 Alternatives Are Typically not Equally Probable

Typically in the description of adaptive systems we encounter two alternatives. However, the assumption of typical occurrence of equal probability of such alternatives seems a great simplification. We do not like to use description notions such as: 'correct', 'proper' or 'special' for an alternative, (such terms are defined using fitness in the set of systems which adapt) but everybody agrees that the 'correct' alternative is typically much less probable. Only novice gamblers do not agree but they will agree. This expected inequality of two variants of signal was described using probability p for one of the alternatives [13, 2]. We will propose another solution.

2.4 Introduction of More than Two Equally Probable Signal Variants ($s > 2$)

Why do we use two alternatives which is an extreme case? What about more than two alternatives which is more safe? One of the typical ways leading to two alternatives is our concentration on one particular specific event and collecting all the remaining events as the second alternative. This occurs especially in adaptive systems because of the aspect of purposefulness of considered alternatives. One of them is 'proper', 'correct' or 'special' as we mention above. There are lots of alternatives in such a case in the reality but we are only interested in one of them. We use this 'interesting one' and NOT this 'interesting one', and we obtain two alternatives. Using more than two alternatives for the description of mechanism of such a case is the only correct alternative.

For the typical adaptive system such a case is much more adequate. E.g. when we are going to describe the long process leading from genes to some properties directly assessed using fitness we should remark that there are 4 nucleotides, 20 amino acids and other unclear spectra of alternatives. In this set of spectra really two alternatives seem to be an exception. Sousa in [48] considers the scale-free network and more than two different opinions and he obtains a vote distribution in better agreement with reality. Similarly Stauffer et al. [49, 28, 50] consider Q opinion states. Luque and Ballesteros [36, 8] also have a doubt about the adequateness of two signal variants when they also similarly introduce (RWN) more than two variants. Luque with Sole proposed more variants in RNS [35, 47], however, in [8, 9] it also can be found but in form which we develope in this paper, other than RNS and RWN.

We should not expect that in a real large network the alternatives coding different meanings for each node always have exactly equal probabilities and that the numbers of them are the same. Using equal probability of these alternatives is the typical simplification, however, it is useful and maybe it is the only way to define the probability needed for prediction and calculation. We know that all nucleotides and amino acids have probabilities not exactly equal but similar and such a simplification can be assumed for more general qualitative models.

We denote the number of equally probable signal variants by s. Note, such a description contains the assumption of equal probability.

2.5 Other Similar Aspects and Their Parameters

Case $s > 2$ differs from the one described by p in the statistical mechanism and its result. For extreme p and small $K > 2$ order is expected [13, 2] but for $s > 2$ chaos is always expected - damage should grow up to an equilibrium

which our simple coefficient of damage propagation introduced in the next chapter shows easily.

Another parameter P, named 'internal homogeneity of Boolean function' [21] is also used for certain problems connected to the inequality of probability of signal variants. It also describes a different aspect of this idealisation. Parameters s and P work in opposite direction when they differ from their typical value - the smallest one. Higher s causes chaos but higher P allows to avoid it.

Summarizing, we use $s > 2$ (more than two equally probable signal variants) and we hope that this is much more adequate for description of the typical adaptive system. Such an assumption leads to chaotic systems which we expect from our interpretation and observation, and it explains the observed inequality of probability of the two alternatives. It is a different assumption and mechanism than the one leading to known p or P parameters.

Opposite $s = 2$ is typically used in Kauffman model [21]. Its basic application is the genetic regulatory network [20, 33] where 1 is interpreted as an active gene and 0 as an inactive one. This strange case seems adequate and gives results close to the experimental data [34, 31, 32] although it describes an adaptive system (but without typical for such systems high concentration of negative feedbacks).

The assumption of $s = 2$ is also used in a wide range of similar models like e.g. cellular automata, Ising model or spin glasses [29]. It is typically applied as safe, useful simplification which should be used for preliminary recognition. However similarly to Boolean networks this assumption may not be so safe and should be checked carefully. In the original application of Ising model and spin glasses to physical spin it is obviously correct, but these models are nowadays applied to a wide range of problems, from social (e.g. opinion formation [25]) to biological ones, where such assumption is typically big simplification.

The number s of equally probable variants of signals is the next main parameter of system, like Kauffman's K - number of element's inputs and P - the internal homogeneity in Boolean functions and p - probability of one of two alternatives, which define a system as chaotic or ordered.

3 Theoretical Expectation of Damage

3.1 Kauffman's Expectations for $s = 2$ Expanded to Case $s > 2$

Theoretical Expectations of Damage are shown in fig.1.

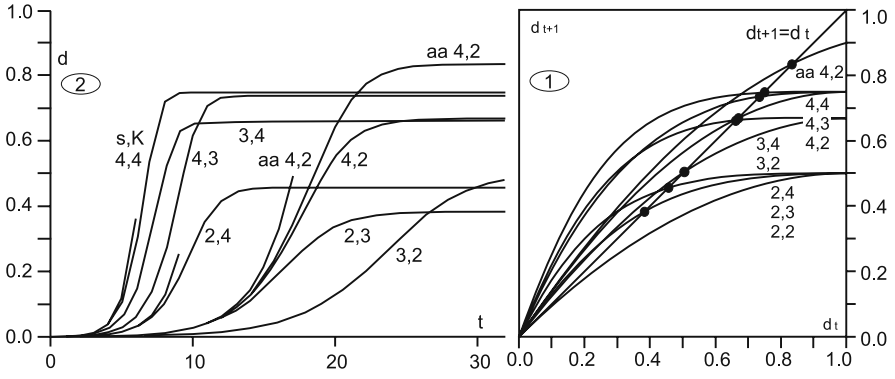


Fig. 1. Theoretical damage spreading calculated using method described in Kauffman's book [21]. (1) 'Derrida plot' - damage change in one time step in synchronous calculation. It is fig.5.8 in [21] extended for $s > 2$ and for aa network type. The crossing of curves $d_{t+1}(d_t)$ with line $d_{t+1} = d_t$ shows equilibrium levels dmx up to which damage can grow. These levels are reached in (2) on the left which shows damage size in time dependency. Approximation $d(t) = d_0 w^t$ for first critical period of small d is shown (three short curves)

This figure commonly named 'Derrida plot' is calculated in a theoretical way based on annealed model [13] described in the Kauffman book [21] on the page 199, expanded to the case $s > 2$: If a denotes a part of system B at a time t with the same state of nodes as an undisturbed system A, then a^K is the probability that the node has all its K inputs with the same signals in both systems. Such nodes will have the same state after the time step. The remaining $1 - a^K$ part of nodes will have a random state, which will be the same as in the second system A with probability $1/s$. The part of system's nodes which does not differ in time $t+1$ is therefore $a^K + (1-a^K)/s$.

The main conclusion of these calculations and fig.1 is that equilibrium levels of damage for $s > 2$ are significantly higher than for $s = 2$. This remark shows the importance of correct choice of the s parameter. The K parameter has a much weaker influence on these levels.

3.2 Coefficient w of Damage Propagation

Let us define a simple and intuitive indicator of system characteristic (chaotic or ordered), a coefficient w of damage propagation. For this let's assume that functions of nodes are correctly randomly drawn. Above we have considered number K of node inputs which we fixed as constant, as Kauffman does [21]. For the number of node outputs we will use the symbol k. The average $< k >$ is equal to K for networks which we will consider. Coefficient $w = k*(s-1)/s$ shows how many output signals of node are changed on the average if one (or

more) of its input signals is changed [8,9,35]. For the case of a single changed input signal it describes the damage multiplication on one node. If $w > 1$ and there is room for damage growth then damage should statistically grow and should create an avalanche which spreads onto a large part of the system. It resembles the coefficient of neutron multiplication in a nuclear chain reaction - if it is less than one then we have a nuclear power station, if it is greater than one then an atomic bomb explodes.

Using this coefficient we can easily see that $w \leq 1$ may occur only for the case $k = 2$ and $s = 2$. We have considered this extreme case above. The value $k < 2$ is sensible for a particular node but not as the average value in the whole, typical, randomly build network.

The damage $d = 1 - a$. For $K = 2$ we obtain $d_2 = d_1 * w - d_1^2/2 * w$ where for small d_1 we can neglect the second element obtaining $d(t) = d_0 w^t$ which is also shown in fig.1.2. If damage is still small then the probability of more than one changed input signal is also small and therefore damage is well described by w^t. In this critical period of time t the probability of damage fade out is not to be neglected. Later it practically cannot fade out but more and more often it occurs that more than one input signal is changed and the real multiplication of damage becomes smaller and smaller up to the moment of achieving a stable level of damage (fig.1.1 and fig.1.2).

The coefficient $w = 1.5$ is common for cases 2,3 and 4,2 (s, K) and for these cases it describes the first crucial period of damage spreading but the levels of damage equilibrium are significantly different which is depicted in fig.1.2.

4 Basic Network Types

The above theoretical expectations are general and independent of the network type but we can expect that they may differ accordingly for different specific network structures and dynamics and their parameters. This is an interesting problem, especially in the range of dependency of our new parameter s. We will explore it using simulation. For this task we will determine the set of interesting network types in the first step. Next we will define a useful special algorithm which we will use to simulation in last step of this project. As you can see in ch.6, the results really exhibit significant differences which depend on network types and the parameter s. Some observations and remarks made during these simulation lead to the definition of 'complexity threshold' in the next part of this article. This threshold occurs during network growth in different sizes of networks in different network types and is connected to the damage mechanism.

4.1 Kauffman Networks

As we remark above in the ch.2 Kauffman networks are the most adequate to describe adaptive systems. Kauffman and many researchers who followed him in the last two decades of 20th century use Erdős-Rényi random networks [19, 21, 44, 18] (CRBN - Classical RBN). We will denote this network type as er. The function formula of Kauffman network gives us a useful ability to differentiate k within the network. This is necessary in er network but it also allows to investigate different types of networks which differ in the distribution of node degree $P(k)$ like e.g. nowadays famous Barabási-Albert scale-free networks which we will denote as sf. This is because the definition of function does not change if k changes and K is fixed. Therefore as the degree of node only the k parameter was typically used. The sf networks seem to be more adequate to describe reality [3,10,24,25,12,16] than the old er networks and now they should be investigated as first (e.g. FSRBN [18]). This network type has a characteristic 'preferential attachment' pattern for growth which leads to power law distribution of node degree. For comparison with the old researches (for CRBN) the er network should also be used. Its node degree distribution $P(k)$ is a bell-like curve which practically ends quickly for relatively small k. Between these two types lies the next candidate for simulation - 'single-scale' [1] network type ss which grows without preferences. Its node degree distribution $P(k)$ decreases faster (exponentially, see also e.g. EFRBN [18]) and is less extreme than for sf networks.

4.2 Aggregate of Automata - the Simplest Case of Network for 'w'

In the Kauffman networks all k outputs of a node transmit the same signal - it is the state of the node, the value of its function (fig.2.1). To understand the coefficient w of damage multiplication we must average the k and the change of signals on them for lots of nodes. It is much simpler and more intuitive (which is important for introducing such a method into biology) if each output of a node has its own signal to transmit, which need not be the same as on other outputs in the same node. The averaging of the change is possible in one such node. If we fix k then we immediately have $< k >= k = K$ and the node function is easy to describe: the function argument and function value are K-dimensional vectors (fig.2.2.). I have introduced such a network in [9, 10, 11, 12] where I have named it 'aggregate of automata', therefore we name it here aa.

For this network if $K = 2$ then $d_2 = d_1 * w - d_1^2 * (s-1)^2/(s+1)/s$ which is obtained similarly as above. Note, that for small d_1 we can neglect the element with d_1^2. Theoretical curves for aggregate of automata for case $s = 4$ and $K = k = 2$ are also included on fig.1. These figures show that the level of damage equilibrium for aggregate of automata is much higher than for Kauffman networks. It is an effect of the larger range of node state value

which has s^k variants. For comparison to the Kauffman networks described above, we define ak network type which is devoid of this discrepancy and follows the Kauffman network rule, but when other features are considered, it behaves like an aa network.

4.3 Construction of Networks

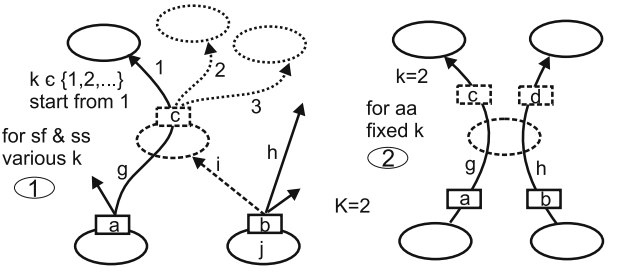


Fig. 2. The basic elements and growth patterns of Kauffman network (1) and aggregate of automata - aa network type (2). Nodes - ovals, signals - rectangles, links - arrows. Each node transforms incoming (input) signals into output signals using a function, these signals are transmitted through links to the next nodes as their input signals. K - number of input signals (or links) of a particular node. k - number of output links of a particular node (node degree). For a particular node of Kauffman network (case (1) on the left) there is one output signal (state of node) which is sent by k output links. s - number of equally probable variants of signal values (in Boolean network $s = 2$, i.e. true and false). In the case (2) of the aa network k is fixed and each output link has its own signal, possibly different from others. Patterns of node addition: Links g and h (and function) of node are drawn. Node j is drawn directly instead of link h for ss. For $K > 2$ additional inputs are constructed like the right ones (h or j). The ak network is maintained as aa but there is only one output signal c (d=c).

We have chosen five types of autonomous networks for simulation of damage spreading: 'er' - random (Erdős-Rényi), 'sf' - scale-free (Barabási-Albert), 'ss' - single-scale, 'aa' - aggregate of automata and 'ak' - a network similar to aa with fixed $K = k$, but using Kauffman formula where one state of a node is transmitted by all its outputs.

The simulation of the networks has two stages: construction of the network and damage investigation in a constant network. Construction of the network depends on the networks type. Except for the type 'er' - random networks, all networks have a growth pattern. Aggregates of automata 'aa' and 'ak' need to draw K links in order to add a new node. These links are broken: their

beginning parts become inputs to the new node and their ending parts become its outputs (fig.2.2).

For 'ss' - single-scale network the new node is connected to one of the nodes already present in the network with equal probability for each of these nodes. For 'sf' - scale-free network the new node is connected with another already present node with probability proportional to its node degree, i.e. to the k of the present node. For sf type at least one output is necessary to further participate in the network growth but we draw one link first and we break it like for aa to define one output and its destination node for both types sf and ss. This way we also define the new node's first input. Later we draw the remaining inputs according to the rules described above - for ss by drawing the node directly, for sf by drawing a link and using its source node (fig.3.2.). If $K = 2$ then only one input follows the rules, but it is enough to obtain the correct $P(k)$ distribution characteristic for these network types.

We have defined network types for our damage investigations using simulation. We determine sf network type using Kauffman formula of function as the most interesting one. Damage can spread in various environments, e.g. in computational networks [39] but typically damage in scale free networks describes: epidemic spreading [24], opinion formation [16, 25, 48] or attack and error effects [12, 17]. However, these networks typically are not directed networks and their important aspect is the spatial description which uses a particular lattice shape. They can also use methods other than preferential attachment to create a network [24, 25]. The Kauffman networks, which are interesting for us here, have an important aspect of function. A network can be 'scale-free' and directed in certain aspects and not directed in the aspect of function. Such a network was used in [49, 28, 50]. These networks describe opinion agreement process. The direction of links is used for construction of network and contact initiative but during opinion exchange information flows in both directions and in this aspect this network is undirected. This second aspect is more similar to signals flow in Kauffman network. The scale-free Boolean networks were studied by Aldana [2] and Kauffman [22], now Iguchi et al. [18]. These authors, similarly to us, look for the difference between the dynamics of er (here called: RBN) and the scale-free random Boolean network (SFRBN). However, here s=2, flexible k and K are used, therefore those networks differ from our sf.

5 Simplified Algorithm of Damage Spreading in Chaotic Systems

5.1 Assumptions and Limitations

The damage simulation in the classic case uses two full processes of systems states evolution which are compared: A for an unchanged system and B for

a system with damage initiation [29]. We perform one partial process - only damage spreading, but this process is only statistically correct in a limited range of situations.

The main assumption is: We consider chaotic systems where damage can fade out only when it is still small, but when it becomes large, then it grows up to an equilibrium level where it fluctuates around this level infinitely. In our algorithm it also stops when it reaches the equilibrium level - it is the pseudo fade out. These two cases of fade out are mixed, however they have different interpretations. Using percolation theory terms a large cluster (which percolate) corresponds with damage after avalanche on it equilibrium level and small clusters correspond with damage which really fadeout.

Cases which stop between real fadeout and pseudo-fade out with a middle damage d have no interpretation and can be permissible only in negligible frequency. Such cases occur only for $s = 2$ (sf 2,4 3,2 and ss 2,3 in small but visible level and especially for sf 2,3 network (fig.4.1) in high level) which confirms that $s = 2$ is an extreme case. For every network type the case $s, K = 2, 2$ is out of range of permissible frequencies of middle damage. We cannot use our algorithm for investigation of this case. It has very low damage equilibrium levels and consists mainly of real fade out cases predicted in fig.1.1. Its long tail for higher d is too short, therefore strongly incorrect, in the results of our algorithm.

In such a view the real fade out is connected to ordered behaviour and pseudo fade out - to chaotic behaviour. The shares of these both types of fade out can be used as a measure of the degree of chaos or order for particular network type and parameters of this network.

5.2 The Main Rule

Signals are memorized while a node is calculated. They are used to define the changed input signals. We calculate only the nodes with changed input signals [10]. If a node is reached by damage, which means that at least one input signal is changed, then the node function is calculated using 'old' remaining input signals, but only once. We do not care what remaining input signals are. They can be changed before or after the calculation of this particular node, e.g. as effect of feedbacks loop. In this article we also do not use concrete functions for nodes. If the input state is changed, then the output state is random.

This calculation gives an answer, whether output signals of this node have undergone any changes. If its input signals change later then it will not be recalculated next time - for statistically correct damaged area it is not necessary. Any initiation of a particular node in a particular network should statistically lead to the same damaged area but in each particular case it

may be different. We, however, are not interested in a particular case but in a statistical result. Such an algorithm works fast and gives correct statistical effects. For certain problems of adaptive evolution described in next article it is the key to possibility of simulation - it allows to omit the problem of periodic attractors and their comparison for fitness definition.

5.3 Intuitions and Feedbacks

To agree with the proposed assumption certain intuitive help is useful. Such intuition behind this algorithm can be found when you consider a special network without feedbacks, where each signal on the node output is equal to the value of the function of current signals on the node inputs. It is not a typical system state - in the next time step in the synchronous mode nothing will be changed. Let's change one node function as damage initiation. In such a case for calculation of node with a changed (as damage) input signal we can use the old signals on the remaining inputs if for a given node they do not depend on the remaining nodes waiting for calculation. Such a node will always exist because a node does not depend on itself. Input signals of the calculated node will not change later. After a finite number of time steps the process will stop. The damaged part will become a clear tree. As was the case at the beginning all the node states will be equal to the function value of current node inputs.

In the case with feedbacks sometimes an already calculated node gets a damaged input signal for a second time. For measuring the statistical effect only it is not necessary to examine its initiation for the second time. This abandonment leads to pseudo fade out - to stopping the process. If such second initiation will be processed, then the process may never stop - this is fluctuation on the equilibrium level of damage. However, our process stops using the same mechanism which keeps the damage at the equilibrium level. It gives one particular damage size which also fluctuates about this level but it needs lot of particular cases of processes and their stops.

5.4 Damage History

Growth of the networks is an external aspect of the algorithm described in this chapter. This aspect was described earlier, however during growth each addition of a node is an initiation of damage and is calculated using described rules of this algorithm. Such damage is investigated in another article. In this part of article we investigate the damage in a system of a particular, fixed size.

When a network achieves the assumed number N of nodes we stop the growth and we start to initiate damage: we change the output state of each node into all remaining variants. This gives $N*(s-1)$ different changes. One such change is a damage initiation. It is the smallest initiation and in the first few steps the damage can fade out. It is a real fade out of damage. In this

short way damage can encounter an already damaged node. Such node will be not calculated for the second time. This abandonment helps damage to fade out, however, such an event has a very small probability. We assume that if damage fades out when it is small, then it is not due to encountering an already calculated node. This is a simplification of our algorithm. In this case the number of damaged nodes is interpreted as the number of damaged nodes during the whole process from initiation to real fade out. The real fade out means that damage d at the time of fade out is zero.

If coefficient $w > 1$ then on average the damage grows. If damage is great, i.e. the number of nodes with changed output state is large, then it practically cannot fade out. I.e. probability of such events is very low, we neglect them. The number of nodes which were calculated again because their input state was changed is also large, which influences the pseudo fade out frequency. During this damage growth there are less and less nodes which are not reached by damage yet. Therefore the avalanche of damage must slow down and stop (the growth). It looks like a fadeout, but it is equivalent to the achievement of the stable level by the damage which appears at the end of curves in fig.1.2 or on cross of curves with line '$d_{t+1} = d_t$' in fig.1.1 This level is an equilibrium state, as fig.1.1 shows. In our simplification the process stops at this level due to the 'pseudo-fade out' on already damaged nodes. Now the number of damaged nodes is interpreted as the damage equilibrium level. It describes the statistical state of the system at one specific time step in opposition to previous case.

The damage size d is the number of nodes whose output state changed anywhere in the process (i.e. number of damaged nodes) divided by N. It is despite the fact that they are damaged during the whole process (using our algorithm), not only in the last time step. It takes into consideration the fact that we do not calculate a damaged node next time when it obtains changed input for a second time.

We calculate the damage using a fi-fo (first in-first out) queue for nodes with changed input signals waiting for calculation. The time step number t is defined observing this queue but it is not used for control of the process.

6 Degree of Chaos for Different Network Types in Simulation

6.1 Goals and Parameters of Simulations

The main goal of our simulation experiments is to examine the degrees of chaos and order for different network types in dependency of s and K parameters. Degree of chaos or order is defined in ch.5.1 as the share of pseudo fade out of damage on equilibrium level or real fade out respectively. For such

a task the area near phase transition from chaos to order is the most interesting one. Therefore we use the first three smallest values for s and K from 2 to 4 in all the combinations. Our second goal is to check the importance of the new parameter s and compare it to the K parameter in real mechanisms observed in different network types.

The above degree of chaos is determined in the several first time steps when damage is still small because this is the only period when damage can fade out for $w > 1$. In this period the coefficient w, which is practically defined using s and K, theoretically simply approximates damage as $d = w^t$. As we will see, the parameters s and K influence the experimental results for different network types not only through the coefficient w. The parameter t of damage history is practical, natural and obvious. We will start using it, however it happens not to be the best for a clear description of the process and its mechanisms.

Each simulation of a particular case s, K and one of five network types described above: sf, ss, er, ak, aa consists of 600 000 damage initiations. Generally we use a network of $N = 2000$ nodes but results depicted in fig.5 are checked also for $N = 3000$ nodes in the network and they are exactly the same. For the optimisation of the simulation time all possible initiations are imposed in each built network - each node state is initialised changing into all remaining variants. Therefore e.g. for sf 2,3 and $N = 2000$ we need 300 different networks but for aa and $s = 4$ we use 20 networks with the output state changed 15 times.

6.2 Time t Dependency and Process Speed

Fig.3.4-7 (right column) shows the distribution of time of damage fadeout in both 'real' and 'pseudo' cases. There are two peaks on this distribution: one for real fadeout in the first steps (early fadeout) and the second for 'pseudo-fadeout' when damage reaches equilibrium level at the last time steps. For the network cases with wide range of node degrees like sf and ss with great fraction of $k = 1$ the probability of early fadeout is much greater especially for small $s = 2$. If $K = 3$ then 60% nodes for sf and 33% for ss have $k = 1$ but there are 11% and 20% nodes of $k > 4$ which have 55% and 46% outgoing links. If $K = 2$ then 67% nodes for sf and 50% for ss have $k = 1$, there are 7% and 6% nodes of $k > 4$ which have 34% and 19% outgoing links. For $s = 2$ nodes with $k = 1$ have $w = 1/2$ and early fadeout is easier than for $s = 4$ where $w = 3/4$. Here hubs are present. The biggest hub ($k = 955$) appears in sf when $K = 4$, for $K = 3$ it reach $k = 520$. This single hub takes 12% (the second 9%) of all the outgoing links. Hubs decrease the average k and in effect: average w for remaining nodes, this helps damage to fade out before the first hub is achieved. For er network even $k = 0$ happened but nodes with $k < 2$ constitute less than $1/4$ of all the nodes. If s is small, e.g.

$s = 2$, then the coefficient w is locally especially low. Note that we have used local coefficient w for explanation. On the opposite end (only of Kauffman mode) there lies the network type ak where $k < 2$ and hubs are absent and coefficient w of damage propagation is high and equal for all nodes. In such a case early fadeout is very small and most of the damage grows until the equilibrium level is reached.

In the assumptions for our algorithm we need that real and pseudo fadeout be clearly separated. In the considered figure (fig.3 right column) two peaks are typically not separated, only for the last shown case of ak 4,3 there is a gap of zero frequency between peaks - the first case in the set of Kauffman networks. We move from the phase transition from order to chaos (occurring for $s, K = 2, 2$) to higher values of s and K. Zero frequency between peaks occurs first for aa (3,2 and 2,3) next for ak (4,2) later for er (2,4). Networks ss and sf do not reach it for 4,4. However, both peaks are wide when they are in dependency on the parameter t. Different speed of damage spreading is the cause of this width of peaks, especially different time of beginning of an avalanche. This phenomenon is clearly visible in fig.4 where damage d in dependency of t as in fig.1.2 is depicted for sf network type and different s, K each for one network and 2000 initiations. Beginning of avalanche can be delayed because in an area of small local coefficient w the damage (still small) can take some time before it reaches a hub (which causes many new nodes to be affected by the damage). This phenomenon is the strongest for sf network and small $K = 2$. If s or especially K grow, or distribution of node degree k is more uniform, then the dispersion of time of avalanche beginning and the width of such picture decreases and the avalanches look more similar.

6.3 Fadeout Frequency in Damage Size Dependency

Different speed of damage spreading and its effect - the width of peaks and lack of sharp boundary between them in fig.3.4-6 suggests that variable t - time of damage fadeout is not the best choice. However, variable t is interesting in practice and therefore often used [29]. Similar distribution of damage fadeout in the variable: damage size d, shown in fig.3.1-3 (left column) appears much better suited for the description and understanding of underlying mechanisms. Using such a variable we also obtain the same two peaks: the left one for real fade out and the right one for pseudo fade out, but this time they are very narrow and a big segment of exact zero frequency lies between them typically. Only for the extreme case of sf 2,3 this rule does not work (fig.3.1) but we have discussed the causes of this exception above. The cases of sf 3,2 (fig.3.2.), sf 2,4 (fig.4) and ss 2,3 follow the rule but the second peaks are not very narrow and between the peaks we can find many single counts. (These counts haven't interpretation in our algorithm.) All remaining

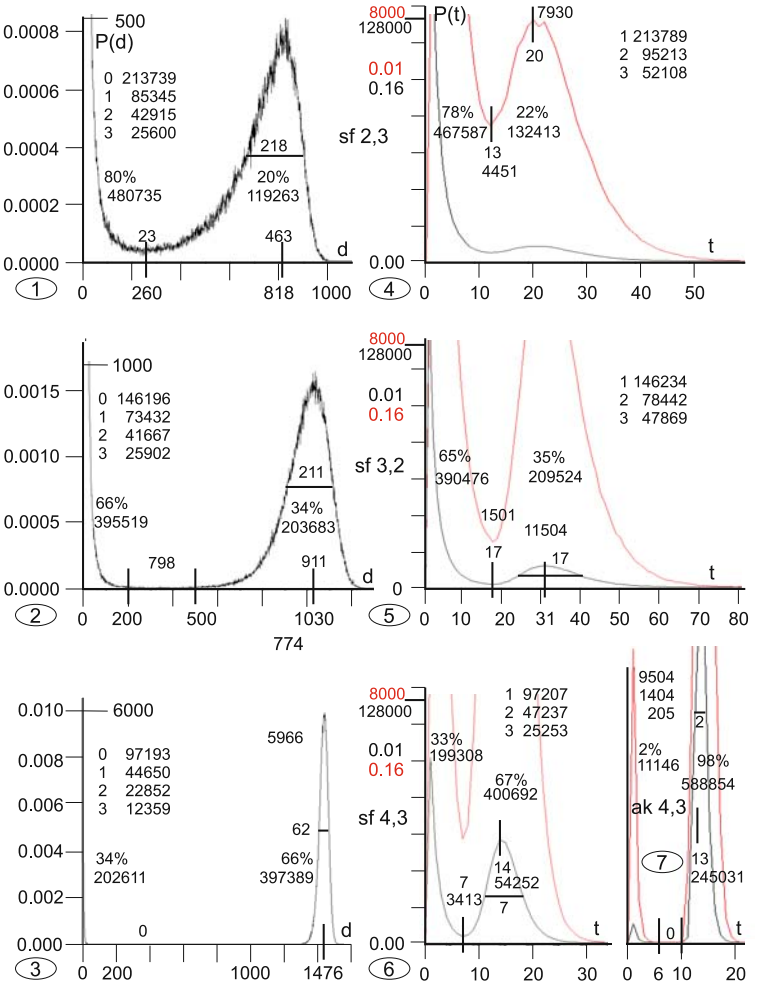


Fig. 3. Distribution of damage size (left column) and time (right column) when damage really or pseudo fade out. In the lowest row typical chaotic form of these distribution is shown, but higher rows, for sf 2.3 (minimum of chaos) and sf 3,2 networks are not so chaotic. All networks contain 2000 nodes. Each distributions are obtained from 600000 events of damage initiations. Positions and values of minimum between peaks and right maximum are shown. The width of the right peaks at half of their height is also shown.. For left peaks a few of the first values are noted. The number of events in both peaks and the percent of all the 600000 events in each peak are shown - this important information is hard to estimate only from shown figures. E.g. in (3) left peak exists, it contains 34% of the events but it is hard to see.

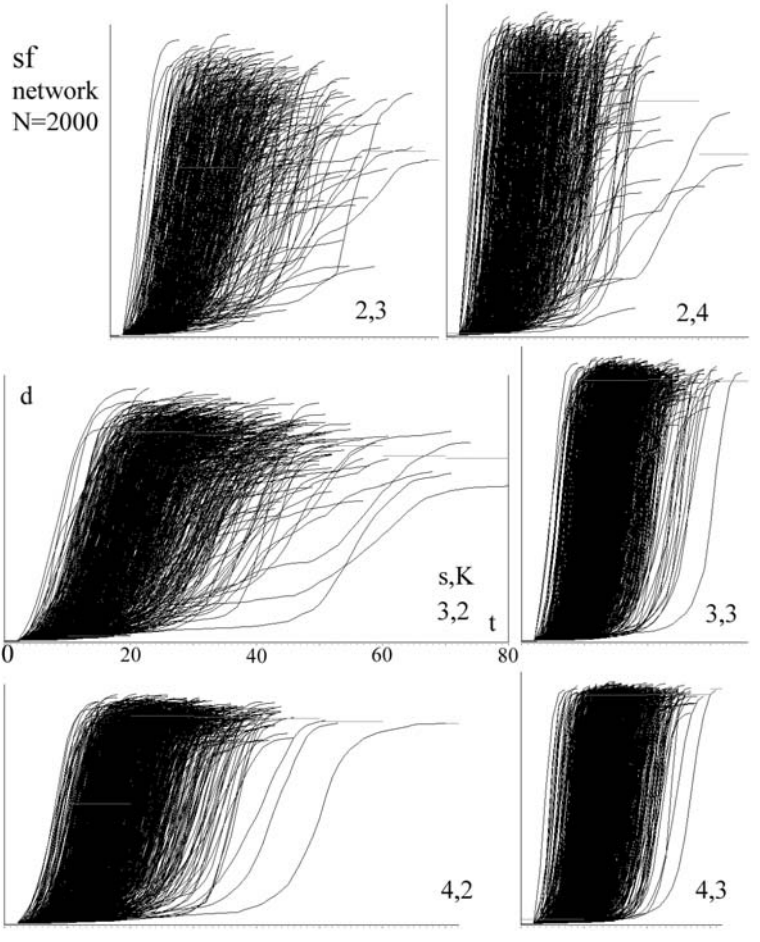


Fig. 4. Damage size d in dependency of time step t for sf network and several combinations of s, K. Each case for particular s, K has the same scale described for t only for $s, K = 3, 2$ and consists of 2000 initiations of damage in one particular network of $N = 2000$. Similar to theoretical fig.1.2. Note, that the periods, when damage avalanche begins, can differ significantly, especially for smaller K.

cases are similar to the last sf 4,3 shown in fig.3.3, small differences concern proportion of both peaks and the peaks' width.

Position of maximum of the second peak is exactly equal to the theoretical point of equilibrium of damage size (dmx).

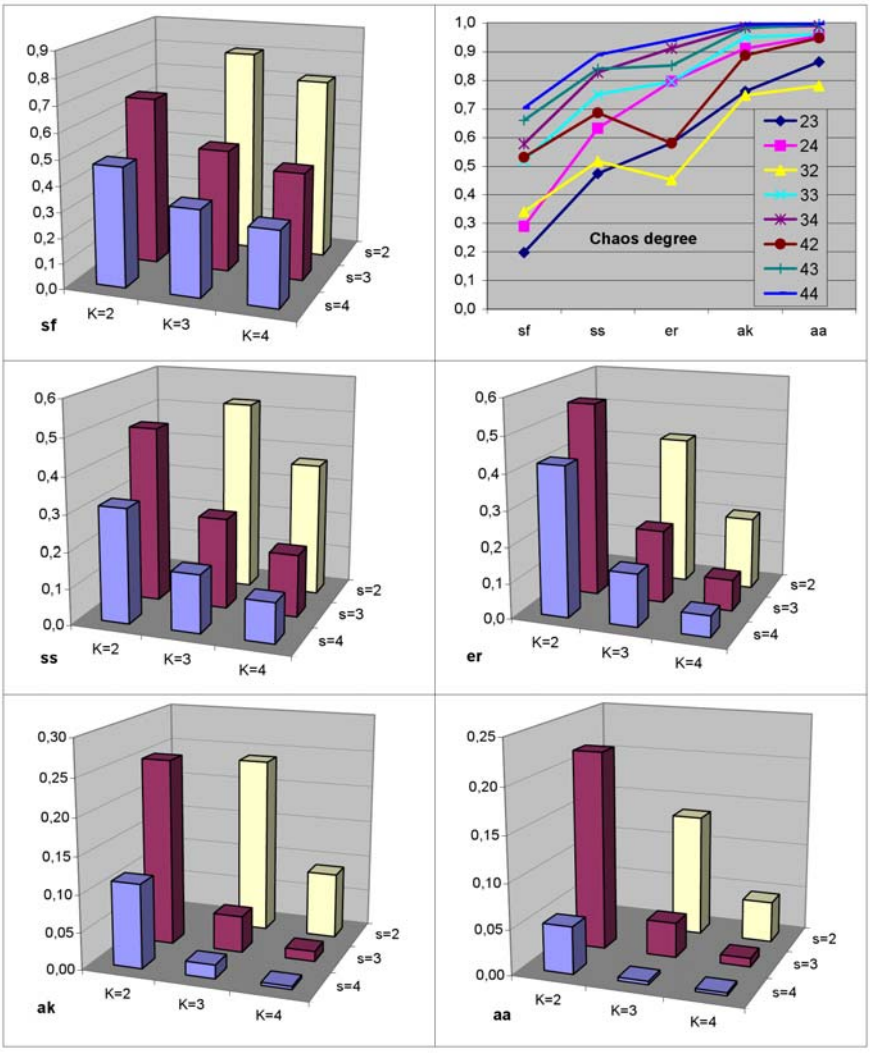


Fig. 5. Real fadeout, i.e. degree of order, as part of all initiated processes for five network types and smallest three values of parameters s and K. Degree of chaos shown in right upper corner equals one minus the degree of order. These figures show dependencies of these degrees on s and K and allow us to compare influence of both these parameters which differ especially for sf,er and aa networks. The data have 3 decimal digits of precision.

6.4 Degrees of Chaos and Order for Different Network Types

In the above description we have several times defined a sequence of network types connected to the dispersion of node degree k. It is the following sequence: sf, ss, er, ak which are Kauffman networks and aa which has s^K variants of node state. It was connected with the gap between peaks in both columns of fig.3, with the delay of avalanche beginning and with areas of smaller local w. Generally it concerns the degree of chaos which grows according to the network type's order in this sequence. (There occurs an exception for er network. The mechanism of this exception is an effect of $k = 0$.) It defines an order in the set of network types which allow us to use it like a directed axis. The degree of chaos is mainly depicted in right upper corner of fig.5 and earlier denoted in percents in the middle of right peaks in fig.3. It is defined in ch.5.1 as the share of pseudo fade out of damage in all initiated processes. The second part of these processes really fade out which is ordered behaviour and their share in all the initiated processes is the degree of order for this particular network. The greatest share (80%) of real fade-out of damage was observed in the network sf 2,3 (fig.3.4 and fig.5). For ak 4,3, however, only 1.8% of the processes ended in real fade out and for 4,4 this was only 0.4%.

We consider other two axes - of s and K parameters, now with normal numbers as values. Their growth also indicates a growth of chaos degree and a decrease of order degree as shown in fig.5. In the next chapters we will investigate the growth of chaos degree in dependency on N. The stages of growth of chaos degree in axis dependency of any parameter look similarly.

These investigations using simulations of different network types are also designed to show that parameter s is important and we cannot limit our self to parameter K only. Dependency on s is strong similarly to dependency on K but it also differs from dependency on K for different network types. In the aggregate of automata the state of a node has s^K variants and this network type has obviously stronger and different dependency on these parameters than Kauffman networks. The ss and ak networks exhibit symmetrical dependency in s and K but for the most interesting sf and er network types there is no symmetry which is depicted in the fig.5. For sf dependency on s is stronger but for er - weaker than dependency on K. These differences are not big but may be important.

The significantly lower chaos degree for sf network obtained in our simulations is known [17, 12] as higher tolerance of a scale-free network to attack.

7 Damage Size *L* on System Outputs and Its Dependency on Depth *D*

7.1 Functional Order and Cone of Influence

Considering the flow of signals through a directed network we operate in terms of cause and effect. The signal appearing on one of node inputs is one of causes of transformation effects. A node and function connected to this node represent this transformation. Function value is an effect, which is transmitted farther as the cause of other transformation. The sequence of node on signal way is a functional order. It has direction of signal flow depicted as arrow for directed network. This direction define 'earlier' and 'later node' in network structure than shown one. If some node is neither early nor later, then it is independent. Such definitions create structure named 'cone of influence' (fig.6) because of similarity to cone of time in Einstein theory of relativity.

In general, sets of later and earlier nodes are not separated. Nodes in the loops of feedbacks are simultaneously earlier and later which destroys clear and simple mathematical description as directed set. Networks, however, still stay directed because signals flow through links in one direction and function still transforms input signals to output signals. Feedbacks do not destroy casualness. Locally, in the area not containing close loops, functional order became still clear-cut. Totally, for networks with feedbacks the functional order and cone of influence dramatically lose their focuses but not fully.

Clear, simple and mathematically correct directed set of node, functional order and cone of influence exist in network without feedbacks. Let us consider such a network in the beginning to get accustomed to notions, which later despite the loss of their focuses when feedbacks are present, give connection to intuition and help to understand the phenomena and their causes.

If we demand constant number K of node input in the network without feedbacks, then this network cannot be autonomous, it must contain inputs from environment of such a system. Similarly in the case of output. Let us start our investigation using the simplest network - aggregate of automata. For aggregate of automata the coefficient of damage multiplication w introduced in ch.3.2 (average number of changed output signals if one input signal is changed) can be investigated on one, typical node. The name 'aa' is reserved for a special case of general aggregate of automata network - randomly growing and containing feedbacks. Now we will introduce a few simpler cases.
A drawing of cone of influence is practically possible only in the simplest and the most ordered case of such a network (fig.6). We denote the network shown in fig.1 as 'lw'. To get rid of not natural and needles problem of edge

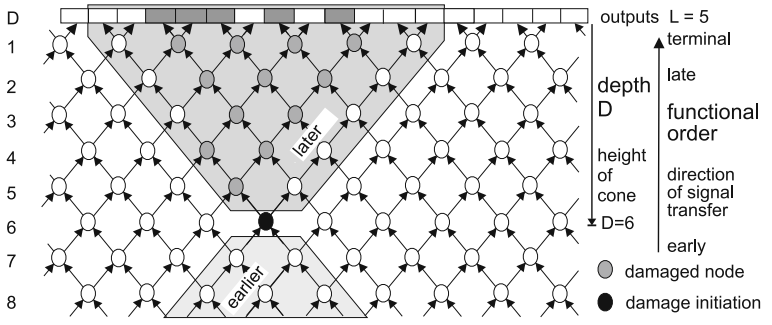


Fig. 6. Cone of influence in an extremely simple network without feedbacks (lw). The shown node (black one) separates the node set into three parts: earlier, later (which forms the cone) and the rest - independent. If this black node changes its state, then it becomes a source of change which as damage flows up to outputs, if $w > 1$, usually also as cone of damage (grey nodes). Depth of damage initiation is a height of the cone and suggests the average change size L. It is connected to the functional order as shown.

let us roll up this network to a cylinder. The circumference of this cylinder is $m = 64$ outputs or 32 of node, each of $K = 2$ inputs and $k = 2$ outputs.

The set of later node than the selected one is the most interesting part of cone of influence. Similar construction is investigate in [26]. In fig.6 it really looks like cone: its base there is on the top and shows which outputs can be reached by damage if node on the cone point is damaged. Size of the cone base measured in number of outputs depends on cone height. Because base is on the top, cone height is a depth, we denote it as D. In storeyed network in fig.1 we can measure depth D using number of levels. Depth is an approximation of functional sequence, which is a simplification, because set of network nodes cannot be numbered in such a sequence. It considers particular signal ways between the selected node and outputs. However such an approximation is useful for intuition and clear in fig.6. In less ordered networks depth can be defined using different method, the most general is the shortest way to output. Depth defines terms early, late and terminal in a clearer and general way than original functional order.

7.2 Distributions $P(L|D)$ and $P(L)$ for Simple Cases

Following [21] we have measured the size of damage $d \in < 0, 1 >$ as a share of nodes with damaged output state in all nodes of system. But this parameter is typically hard to observe for real systems and we are going to observe damage outside the system, on its outputs. The number L of damaged output signals is the simplest definition of damage size on system outputs. This is a Hamming distance of system output signal vectors for a control system and a

damaged one. The system has number m of output signals that $L \in < 0, m >$. We fix $m = 64$ for model and simulations. We can expect, that distributions $P(d)$ and $P(L)$ should be similar. In fact, asymptotic value of d which we named dmx and asymptotic value of L named Lmx are simply connected: $dmx = Lmx/m$ but such a connection is not true during the system growth and L is less than we should expect.

The lw network is extremely ordered in the aspect of structure. This order is an opposition to 'complexity' and is not simply connected to notion 'order' as an opposition to 'chaos'. It mainly considers the aspect of structure, not the aspect of damage spreading which is a function. In the current two chapters we investigate the connectivity of these two notions of 'order' and their oppositions: 'complexity' and 'chaos'. In a chaotic network two very similar initial network states (output states of all nodes and input signals of the system) typically lead to very different states later, after not a long period of time of the system's function. It can be simply translated to typical explosion of damage initiated as a small disturbance, e.g. a change of state of one element of a system. Ordered systems (as opposed to chaotic systems) exhibit 'homeostatic stability' (in the Kauffman sense [21], see section 2.1). This homeostatic stability does not allow damage to transform into an avalanche. Now we consider order of structure and we limit ourselves to chaotic systems using assumption of more than two equally probable signal variants (section.2).

Let us subtract some order from lw network allowing random connectivity of nodes from consecutive levels, but still only between consecutive levels. Let us denote such a network as 'lx'. Both of these network are storeyed but 's' letter is reserved, therefore I use the letter 'l' from 'levels' as the first. Second letter should be different for any network type. For lw network, as it is visible in fig.6 and fig.7, connection between levels looks like 'w' letter, for lx network lots of links are crossed, therefore there is an 'x'. I do not find better mnemonic rule. Networks lw and lx are the cases of aggregate of automata due to their constant $K = k = 2$ and 2-dimensional value of function.

Let us investigate and compare distribution $P(L)$ for both storeyed networks lw and lx in dependency of level, i.e. of depth D, on which damage is initiated, that is $P(L|D)$ distribution. It is obvious, that for small depth D the distribution can be different from zero only for small L due to the cone of influence containing in such a case only few outputs. When we increase depth D of damage initiation, cone of influence include more outputs but in lw and lx networks this increase is different. If we increase D by one, then for lw network the number of outputs in the cone increases only by two, and this is so $m/2$ times but for lx network in the beginning the number of outputs increases twice on average. It is much faster. The second source of difference in $P(L|D)$ is the number of ways in which damage can achieve a particular

output - for lx it is similar for different outputs, for lw it differs very much between outputs.

Fig.7 shows the simulation results for $s = 4$ and $m = 64$ outputs . Distributions $P(L|D)$ excluding very small D have two peaks. The first peak has maximum in $L = 0$ (or near), and quickly falls to zero. Second peak has a bell shape, its maximum moves to higher L when D increases. Speed of this move to the right radically depends on the network type - for lw it is slow, for lx - quick. In fig.7 distributions for lw and lx are superimposed and depth is selected so that maximum of the second peak is on the same positions. As it is visible, network lx with less order of structure can be much smaller (needs smaller D) than lw to achieve similar state of distribution $P(L|D)$. Reaching the size (number of levels) by growing network lx, where the position of right maximum does not change, occurs for lw for $D = 78$ levels but for lx $D = 26$ is enough. When we increase s from 4 to 8, we also decrease order in direction of higher complexity. For $s = 8$, i.e. twice higher value, the right maximum reaches a stable position for twice smaller depth - $D = 46$ for lw and $D = 13$ for lx.

To obtain global distribution $P(L)$ we must summarize $P(L|D)$ by levels D, which are equally probable for random initiation of damage here. Let us compare $P(L)$ for both network of equal size, greater than threshold of the practical immobilizing right maximum for both networks. (fig.8 d - lw, e - lx for $D = 120$). For lw network a segment between both peaks is filled and the right peak presents itself poorly. But for lx network influence of moving to the right peaks from section of D from 1 to 26 is small in comparison to the superimposed stable peak from the three times longer segment of higher D from 26 to 120. In result, for lx segment between the peaks has much smaller values and the right peak - much higher. Further growth of both networks causes decrease of these differences.

7.3 How Deep is Network with Feedbacks?

The next investigated network is 'an' - randomly growing aggregate of automata without ordered levels like in case of lw and lx, but still without feedback loops. For $s = 4$ and $m = 64$ as in fig.7 and fig.8, maximum of right peak occurs in $L = 34$ for network containing $N = 512$ nodes. We can assume, as we later argue, that this is the beginning of immobilizing of right peak. For aa networks containing feedbacks this stage occurs in period of N 256-384 nodes. For comparison, this stage in lx network needs $N = 320$ but in lw about $N = 1700$. Shape of distribution for an network simulated only to $N = 512$ is closely similar to distribution for lw in fig.3, but the first peak is wider and similar to the case of c. High level of probability between peaks is an effect of short period after right peak immobilizing.

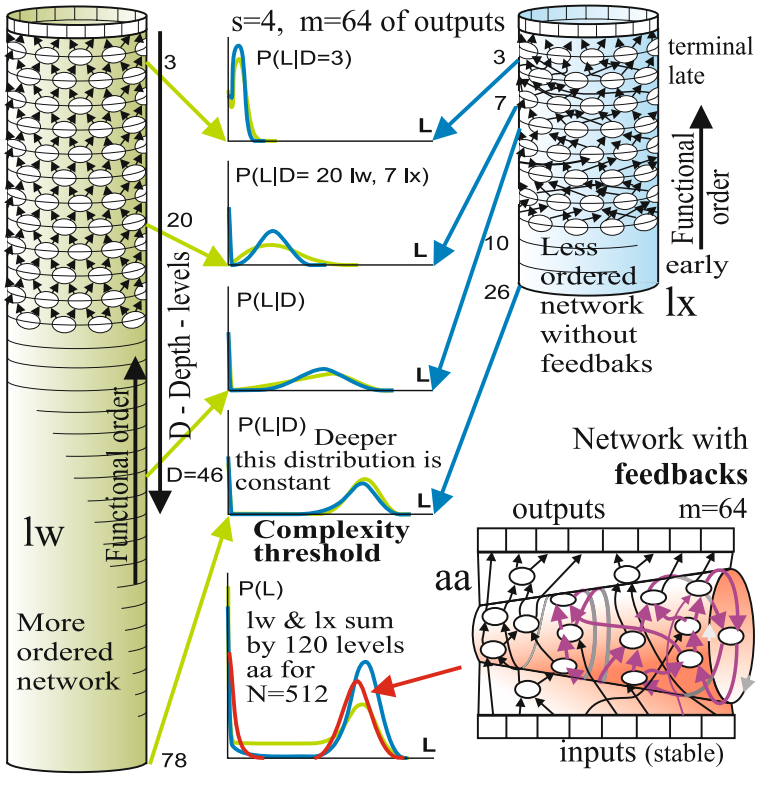


Fig. 7. Dependency of damage size distribution on depth $P(L|D)$ in different network types from the most ordered lw and less ordered lx storeyed networks without feedbacks to fully random growing aa network with feedbacks. In lw and lx networks the functional order and depth D as levels measuring way to outputs are well defined but in aa network the feedbacks (coloured arrows) can loop undefined - may be - infinite number of times making their way to outputs very long. $P(L)$ is a sum of $P(L|D) * P(D)$ by all available D. For small D the right peak of $P(L|D)$ distribution is not on its last, stable position, which it reaches for larger D. Reaching of this stable position is a complexity threshold. For more ordered networks higher D is needed to reach this position. lw needs $D = 78$, lx needs $D = 26$ but aa needs about 256 nodes which is equivalent of $D = 8$ levels of storeyed networks.

In the considered an and aa networks there are no levels and defining depth D is not as clear as for lw and lx. Exactly for this reason the size of networks with constant $m = 64$ is shown in number N of nodes. For an cone of influence is clear due to the lack of feedbacks but there is problem to draw it, similarly for lx. Whereas for lx there is no problem to define depth, for an it is not trivial, but there still exists a limitation of the longest and the shortest signal way to outputs. Typically limitation of the longest way does not exist for nodes in aa network - signals can loop in feedback loops ad infinitum

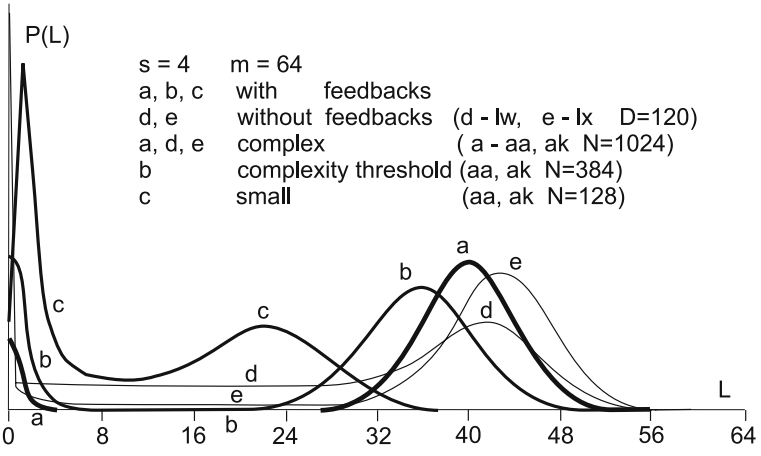


Fig. 8. Complexity threshold. $P(L)$ distribution as the effect of random changes in networks with $m = 64$ output signals, for $s = 4$. The base curve a - for complex network with feedbacks (aa and ak). It has two peaks and exact zero in-between in measured frequency. The same networks on complexity threshold (curve b), and below complexity threshold (curve c), when it was still small. For networks without feedbacks - curves d for lw and e for lx at $D = 120$ ($N = D * 32$), space between peaks has small but not a zero value. Network d is more ordered than e.

and send damage to outputs in any lap. In such a case no single output signal vector exists, but a cyclic attractor. However, damage grows up to some equilibrium level [21] (fig.1) which using our simplified algorithm (ch.5) looks like a fadeout (pseudo-fadeout) and we obtain some particular, statistically correct L or d despite the fact that we cannot show, which particular output signals or nodes are really damaged.

Let us come back to the definition of depth D. Generally we can use the shortest distance to outputs but for lots of reasons such a measure is not fully adequate. Existence of many different ways is the main passed over aspect, especially when they are similarly long. For aa simulations (when $K = 2$) we have used some more adequate definition of depth [10,11,12] (see my next article in this book) taking into consideration this aspect but this definition can be used only for aggregate of automata or ak and $K = 2$. For network with flexible node degree k it is hard to extend. For such networks (in this article we consider ss - single-scale, sf - scale-free and er - classic random networks) number of different ways additionally differ significantly for extreme node degree k, which can be less than two or large for hubs. In such a reason the shortest way is the simplest and the most general definition for depth, it is enough for this first exploration of this new area of phenomena, but it is unsatisfying.

Comparing $P(L)$ for different network types in terms of dependency on their size we should focus our attention on shape of distribution when feedbacks are present. When such a network is small a distribution is similar to small storeyed networks lw and lx, but for $s = 4$ when the second peak immobilizes the probability between peaks practically reaches zero on large segment of damage size d or L. For networks without feedbacks this segment always has non-zero values, however they decrease when network grows. This significant difference is an effect of lack of limitation for the longest way to output for networks with feedbacks. Damage initiated even 'near' the outputs (in the meaning of the shortest way) does not have a small probability of reaching a loop and using it - reaching an area far from outputs and other loops. If height of the cone of influence and the cone of damage inside it, is so big that a middle damage size L is possible, it also has 'volume' that is big enough to contain nodes participating in loops. For this reason is hard to talk about middle depth in networks with feedbacks. Such networks, despite the limited size measured in N, observed in their $P(L)$ distribution look especially big due to their average way to the output, which like the longest way may be infinite. As it is visible, the shortest way does not describe this important aspect correctly.

8 Complexity Thresholds

8.1 Evolution Stages of Damage Size Distribution

Stabilization of position of the second peak of distribution $P(L)$ is a clear qualitative boundary, however, it is an asymptotical scale-free phenomenon and needs arbitrarily chosen value of practical level of exactness. As we have shown, it is connected not only to network growth, but also to structural order degree and dynamics parameter s. Summarizing these remarks the qualitative threshold shown is connected to intuitive meaning of system complexity. If similar states of system create very different effects, we must know much more to predict these effects. We connect such a feature with the complexity of system but this is exactly chaos. Let us call this qualitative threshold the 'complexity threshold' of system in such a sense that a system should be treated as a complex upon this threshold, but below it - not yet. This name seems adequate in the described aspects.

Attractiveness of such a criterion of complexity mainly results from: 1- its direct connectivity to mechanisms of different phenomena which we are going to explain and from: 2 - basing it on observable phenomenon which gives hope for experimental use for comparison theoretical and simulation investigation to reality. So defined complexity threshold considers networks with - as well as without feedbacks.

As we have already discussed, there also is the other phenomenon in networks with feedbacks -large segment of exact zero frequency occurs between the peaks. For $s = 4$ it occurs nearly simultaneously with immobilization of the second peak in stable position. This suggests that it be a property of the above defined complexity threshold. Unfortunately, for $s > 4$ or $K > 2$ zero between peaks occurs earlier than immobilization of the second peak, which occurs also earlier - for lower N, but this movement is slower. Occurrence of exact zero between peaks is also a particular phenomenon connected to complexity threshold but, unfortunately, it is not exactly the same threshold generally. It is earlier threshold defined only for networks with feedbacks. It also need certain arbitrarily defined value - it can be number of collected events (if such a number of events radically grows, then point of occurrence of zero frequency slowly and not significantly grows) or e.g. level of probability, which is more 'correct' solution. We later use 'zero occurrence' as symbol of whole set of such definitions.

We can also indicate the third phenomena of similar character: it is the appearance of the second peak. It appears from the tail of the first peak at a very early stage of network growth, when complexity is still very low. However, it is not a stage of interest for our investigation.

Consecutive stages of the shape transformation of distribution $P(d)$ and $P(L)$ during system growth are shown in fig.9. Distribution has one peak for very small systems. Its maximum is in zero or in one. Right slope is smooth and steep. Range quickly grows to the right as a long tail. Ordered systems (opposite to chaotic), where $s = 2$ and $K = 2$, also have such a distribution (fig.9.1). On the end of this stage the power-low distribution occurs and this is the percolation critical point. A large cluster appears first time and later it appears more and more frequently. In the effect the second peak appears (fig.9.2) which describe size of percolating cluster. Minimum between the peaks becomes deeper and deeper and the second peak quickly goes to the right. The minimum does not reach zero in a long period of time, it is narrow when it is shown in large exactness but when we examine the whole distribution in a linear scale the minimum is wide, flat and close to zero. Close similarity of the effects of degree of chaos can be found in fig.4 for different network types and their parameters instead of the number N (network size) or time of network growth. Tempo of movement to the right of the second peak slows down and the peak immobilizes. In a short period of growth the minimum between the peaks reaches zero on a long segment. It is observed in frequency distribution, despite large statistics. Note, it does not mean that probability reaches exact zero but it dramatically falls down. For much higher number of events the point of this 'zero occurrence' a little moves into right. In the end, the right peak narrows and its maximum does not move any more.

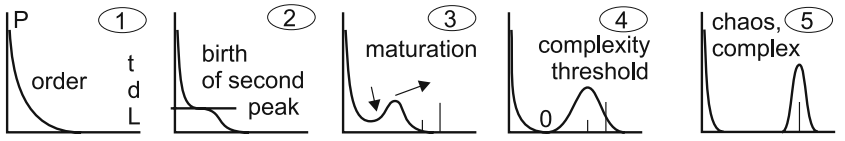


Fig. 9. Consecutive stages of shape transformation of distributions $P(d)$ and $P(L)$ during system growth.

8.2 Premise of Mechanisms and Interpretation

Before we describe simulation results, we should indicate what should be measured in these simulation and why. Cone of influence, functional order and the depth approximating it allow us to understand why the right boundary of $P(L)$ distribution for storeyed network goes to the right when depth of damage initiation grows and they allow to understand the influence of network differences on speed of this movement.

For networks without feedbacks damage is a single wave coming through network according to functional order. When this wave reaches system output, it disappears due to the missing further way. Distribution $P(L)$ is a direct state of wave front here which can be approximated by binominal distribution. It creates a peak which is equivalent to the second peak in $P(L)$ distribution. First, left peak with maximum in zero or in one is an effect of damage ability to fadeout in the first few steps. In the case of network with feedbacks let us recall an analysis of damage spreading for $s > 2$ or $K > 2$ from ch.3.1. Theoretical expectations are used there which gives equilibrium level of damage for large autonomous networks. Assumption of stable system inputs and observed system outputs introduced in ch.7 does not change damage spreading inside the system. This equilibrium level denoted above as dmx gives Lmx which when measured creates a bell- like curve of statistical dispersion. This is the right peak in $P(L)$ distribution at the end of its movement.

In the small system there is no well defined equilibrium state in damage size d. For such a well defined state feedback loops are needed and a number of such loops and network elements should be large for statistics work. When a system is small, it is similar to a system without feedbacks, however, feedback can happen. At this stage mechanism of cone of influence plays the main role but a few cases when feedbacks happen significantly influent the effect. That time a system is more 'parallel' than 'serial' and flow of signals 'cross this parallel structure' is 'difficult'. Between small and large system, during its growth, system changes its properties. Number of feedback loops increases, average way to outputs grows exceptionally fast and system becomes therefore more 'serial' and 'pseudo deep'. This process, like for lx and lw networks, results the moving peak and decreasing probability between peaks.

Let us compare what 'small' system (for $s = 4$ and $m = 64$) means. Size of system of aa type (with feedbacks) is an equivalent of 8 levels of 32 nodes on complexity threshold. Small system has less levels. For lx network, which is an ideal case for small depth - the fastest growth of cone of influence, even $P(L|D = 8)$ is far from stable position of the right peak, despite the fact that $D = 6$ is sufficient for a cone to contain all outputs.

We use a simplified algorithm (described in ch.5) for a simulation whose results we will present in next chapter. One of its useful properties is that process stops on damage equilibrium - it is 'pseudo fadeout' of damage avalanche. This stops creating our interesting right peak in $P(d)$ and in $P(L)$. Distribution $P(L)$ is an effect of $P(d)$, but this connection, as described above, in interesting period of growth is an effect of dynamically changing mechanisms and therefore it is hard for mathematical description. We can expect, that $P(d)$ and $P(L)$ should be similar and really asymptotic values are simply connected: $dmx = Lmx/m$ but such a connection is not true in period that is of our interest and L is less than we should expect. Fig.10 shows simulation results as comparison of the right maximum positions for different network types and other different parameters. It will be described in more precisely in the next chapter.

Proposed complexity thresholds, especially those based on immobilization of right peak, are connected with the system reaching particular parameters of size allowing damage to reach full equilibrium level. System ability to damage explosion is a system property named 'chaotic'. We have discussed this property in ch.2 where we estimate, that adaptive systems are chaotic, later following damage spreading we have investigated in ch.3 and 6 the degree of chaos in different network types near the phase transition to order.

Complexity, whose thresholds we have defined, is a chaotic property of system, which is present over the considered threshold and not fully expressed under this threshold. As it is visible in fig.10, different network types reach the thresholds in a different tempo connected to the network growth measured in number N of nodes which create the network. Network sf (scale-free) is the last one to reach the thresholds which can be expected basing on results shown in ch.6. It contains the biggest hubs which decrease the local coefficient of damage propagation in the remaining area. As it was shown in ch.6, such a decrease causes higher probability of damage fadeout in the first peak range.

8.3 Parameters of Complexity Thresholds for Different Networks - Simulation Results

Fig.10 shows results of multidimensional simulation series of randomly created functioning system with feedbacks. The variables of systems are: 1 - network type (aa, ak, er, ss, sf described in ch.4); 2 - network size, i.e. number N of

nodes in the network; 3 - number s of equally probable variants of signals (introduced in ch.2); 4 - number K of node inputs [21].

Each point is an effect of 3 changes of output state each of network nodes of 100 randomly created networks. It is $N * 3 * 100$ different damage initiations. Distributions $P(d)$ and $P(L)$ are collected and position of the right peak maximum is taken. In addition, the above defined two complexity thresholds are signed.

As it is visible in fig.10.1 and fig.10.3, curves for er and ss networks are placed between curves for ak and sf. In fig.5 upper right corner the degree of chaos distribution for different network types confirms this conclusion. In ch.6 (in effects of results shown in fig.5) we have discussed a conception of approximating a directed 'axis' of network types. When the results for different s ($s = 4, 16, 64$) and K ($K = 2, 3$) are shown (fig.10.2 and fig.10.4) presented curves are limited to these extreme ak and sf networks. For aa $P(L)$ distribution in fig.10.2 and fig.10.4 is identical to ak.

As it is visible, aa and ak networks quickly reach stable position of the second maximum for both distributions $P(d)$ and $P(L)$. The remain Kauffman networks (er, ss, sf) do not reach such a stable position of the second peak due to program limitation and we do not know yet whether they are able to reach the same theoretical value but there is no premise that they do not. To calculate 90% of dmx and 80% of Lmx one theoretical value of dmx and Lmx is used for all networks. For $s = 4$ the points 90% dmx and 80% Lmx are approximately the same and lie close to 'zero occurrence' (see table 1). Unfortunately, when s increases, 'zero occurrence' moves to the lower N in the fastest way, N decreases in a slower way for 90% of dmx and in the slowest way for 80% of Lmx. Therefore, these three criteria stop describing one event. Note that 'zero occurrence' is only a practical method, it not indicates a critical point and it depends, however a little bit, on number of events.

Increasing K moves faster to the lower N the boundary of 90% dmx.

The most significant differences, however, result from network type. Network sf is significantly extreme. Achievement of asymptotic values for sf network is very slow. The shape of the second peak even for higher statistics is still irregular, wider, even an assemble of a few moved peaks. The causes of these properties were discussed in ch.6. The main ones are hubs which strongly influence network dynamics. Due to the same causes, ss network has similar behaviour but much less extreme.

The separation of peaks by segment of exact zero frequency occurs in a short period of network growth and nearly simultaneously in both distributions,

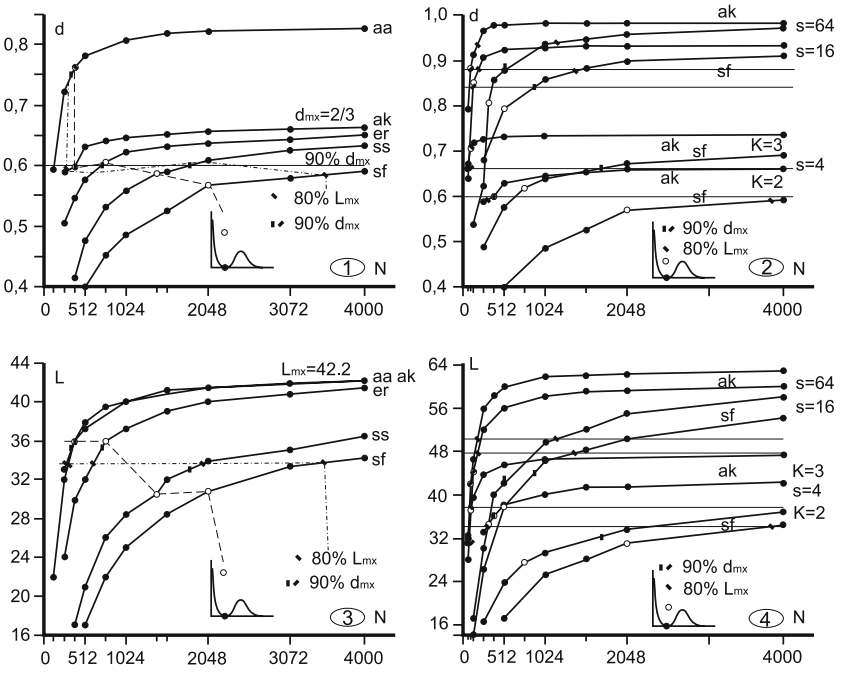


Fig. 10. The left diagrams (10.1 and 10.3) show position of the right peak maximum in $P(d)$ distribution. d - damage size is a part of all network nodes ($= N$) which are damaged. White circle is a sign of occurrence of zero between peaks. Immobilization of the right peak is signed using vertical or right sloped short line in point of 90% of asymptotic value dmx. The right diagrams (10.2 and 10.4) show analogue position of the right peak maximum in $P(L)$ distribution. L - damage size in number of damaged output signals, is shown in exactness 0.1 due to statistical average. Immobilization of the right peak is signed using left sloped short line in point of 80% of asymptotic value Lmx.

but for $P(d)$ a little bit faster than for $P(L)$. In contrast, tempo of reaching asymptotic values is significantly different in both of these distributions.

9 Conclusion

Chaos and complexity in the randomly growing functioning systems may be interpreted as the same for systems whose parameters define them as chaotic at the same time when they are big. These systems start to grow as small systems which do not exhibit chaotic behaviour and are not complex. Later they mature during growth and cross some threshold of size, above which they are both chaotic and complex. The size of networks when it becomes chaotic and complex depends on several parameters like the network type, number s of equally probable signal variants and number K of node inputs.

Table 1. Parameters of four criteria of complexity threshold for different network types and their parameters s and K. 90% Lmx is shown to argue that 80% Lmx better fit remaining criteria. Italic values are extrapolate. For zero occurrence number N is an assumed parameter for simulation experiment, therefore it has a small exactness. $\%dmx$ and $\%Lmx$ are an effect of such N.

type s,K	N	$\%dmx$	$\%Lmx$	90%Lmx	80%Lmx	90%dmx
aa 4,2	384	91.6	85.3	768	310	340
ak 4,2	384	89.8	85.3	600	280	384
er 4,2	768	91.2	85.3	1500	620	730
ss 4,2	1408	88.0	72.0		2000	1810
sf 4,2	2048	85.3	73.0		3850	*4600*
an 4,2	-	-	-		512	
lx 4,2	-	-	-	352	320	
lw 4,2	-	-	-	1920	1696	
ak 4,2	384	89.8	85.3	600	280	384
ak 16,2	128	91.2	73.6	512	200	128
ak 64,2	96	90.1	67.0	266	190	96
ak 4,3	96	70.8	78.8	240	110	56
sf 4,2	2048	85.3	73.0		3850	*4600*
sf 16,2	512	85.0	62.9	4000	1390	900
sf 64,2	320	82.0	55.0	3200	1160	512
sf 4,3	768	84.2	58.2		*5000*	1720

We have shown that the typically not used parameter s is important and may often be more adequate when $s > 2$. In such a case it defines the system as chaotic and leads to significantly higher levels of damage equilibrium. The term 'Kauffman networks' which was up till now a synonym of 'Boolean networks', should contain new cases of $s > 2$, which can not be included in the term 'Boolean networks' [35, 8, 9].

The phenomenon of reaching chaotic state by such a system during its growth appears in three ways, which can be treated as different criteria of complexity. Each of them needs arbitrarily chosen value:

1. crossing the threshold of 80% of asymptotic value by the maximum of the second (right) peak in distribution $P(L)$ (damage size in number of damaged output signals);

2. crossing the threshold of 90% of asymptotic value by maximum of the second (right) peak in distribution $P(d)$ (damage size as part of damaged system nodes);

3. occurrences of exactly zero frequency in a large segment between both peaks in distributions $P(L)$ and $P(d)$ (it occurs nearly simultaneously in both distributions, for higher statistics an arbitrary defined value has small significance).

For $s = 4$ equally probable variants of signal and $K = 2$ inputs per node all three criteria give similar results, however, when s or K is increased, they result in different number N of system nodes treated as system size. The highest dispersion of the above criteria is detected as an effect of network type diversity whose main cause are extreme node degrees k: $k < 2$ and hubs, especially for the scale-free network type which is an extreme type. Differences between network types are larger for smaller values of parameters s and K near the phase transition order - chaos and are the largest for $s = 2, K = 3$ and scale-free network whose degree of chaos is only 20% and in remaining 80% of damage initiations the network exhibits ordered behaviour.

The shown phenomenon is undoubtedly a significant threshold in basic mechanisms of growing functioning systems. The possibility to observe its characteristic measurable effects increases its potential usefulness. Usefulness of complexity threshold defined in such a way will result from applicability to different mechanisms and properties of complex functioning systems, which need further investigations. One of such directions is investigation of structural tendencies in adaptive evolution of complex functioning systems (next article in this book). This tendencies occur over the complexity threshold defined in such a way. Living objects and most of the systems created by human belong to this class of systems.

References

1. Albert, R., Barabási, A.-L.: Dynamics of Complex Systems: Scaling Laws for the Period of Boolean Networks. Phys. Rev. Lett. 84(24), 5660–5663 (2000)
2. Albert, R., Barabási, A.-L.: Statistical mechanics of complex networks. Rev. Mod. Phys. 74(1), 47–97 (2002)
3. Aldana, M.: Dynamics of Boolean Networks with Scale Free Topology. Physica D 185, 45–66 (2003)
4. Ashby, W.R.: Design for a Brain, 2nd edn. Wiley, New York (1960)
5. Ay, N., Olbrich, E., Bertschinger, N., Jost, J.: A unifying framework for complexity measures of finite systems. In: Proceedings of ECCS 2006 (2006), http://cssociety.org/tiki-index.php?page=ECCS%2706+Programme
6. Bak, P., Tang, C., Wiesenfeld, K.: Self-organized criticality. Phys. Rev. A 38, 364–374 (1988)
7. Bak, P.: How Nature Works. Springer, New York (1996)
8. Ballesteros, F.J., Luque, B.: Phase transitions in random networks: Simple analytic determination of critical points. Phys. Rev. E 71, 031104 (2005)
9. Barabási, A.-L., Albert, R., Jeong, H.: Mean-field theory for scale-free random networks. Physica A 272, 173–187 (1999)

10. Barabási, A.-L., Bonabeau, E.: Scale-Free Networks. Scientific American, pp. 50–59 (2003), www.sciam.com
11. Chatzimeletiou, K., Morrison, E.E., Prapas, N., Prapas, Y., Handyside, A.H.: Spindle abnormalities in normally developing and arrested human preimplantation embryos in vitro identified by confocal laser scanning microscopy. Hum. Reprod. 3, 672–682 (2005)
12. Crucitti, P., Latora, V., Marchiori, M., Rapisarda, A.: Rapisarda: Error and attacktolerance of complex networks. Physica A 340, 388–394 (2004)
13. Derrida, B., Pomeau, Y.: Random Networks of Automata: A Simple Annealed Approximation. Europhys. Lett. 1(2), 45–49 (1986)
14. Dorogovtsev, S.N., Mendes, J.F.F.: Evolution of Networks: From Biological Nets to the Internet and WWW. Oxford University Press, New York (2003)
15. Erdős, P., Rényi, A.: Random graphs. Publication of the Mathematical Institute of the Hungarian Academy of Science 5, 17–61 (1960)
16. Fortunato, S.: Damage spreading and opinion dynamics on scale-free networks. Physica A 348, 683–690 (2005)
17. Gallos, L.K., Argyrakis, P., Bunde, A., Cohen, R., Havlin, S.: Tolerance of scale-free networks: from friendly to intentional attack strategies. Physica A 344, 504–509 (2004)
18. Gecow, A.: A cybernetic model of improving and its application to the evolution and ontogenesis description. In: Proceedings of Fifth International Congress of Biomathematics, Paris (1975)
19. Gecow, A., Hoffman, A.: Self-improvement in a complex cybernetic system and its implication for biology. Acta Biotheoretica 32, 61–71 (1983)
20. Gecow, A., Nowostawski, M., Purvis, M.: Structural tendencies in complex systems development and their implication for software systems. Journal of Universal Computer Science 11, 327–356 (2005), http://www.jucs.org/jucs_11_2/structural_tendencies_in_complex
21. Gecow, A.: From a "Fossil" Problem of Recapitulation Existence to Computer Simulation and Answer. Neural Network World 3, 189–201 (2005), http://www.cs.cas.cz/nnw/contents2005/number3.shtml
22. Gecow, A.: Structural Tendencies - effects of adaptive evolution of complex (chaotic) systems. Int. J. Mod. Phys. C 19(4), 647–664 (2008)
23. Gell-Mann, M.: What Is Complexity? John Wiley and Sons, Inc., Chichester (1995)
24. Grabowski, A., Kosiński, R.A.: Epidemic spreading in a hierarchical social network. Phys. Rev. E 70, 031908 (2004)
25. Grabowski, A., Kosiński, R.A.: Ising-based model of opinion formation in a complex network of interpersonal interactions. Physica A 361, 651–664 (2006)
26. Holyst, J.A., Fronczak, A., Fronczak, P.: Supremacy distribution in evolving networks. Phys. Rev. E 70, 046119 (2004)
27. Iguchi, K., Kinoshita, S.I., Yamada, H.: Boolean dynamics of Kauffman models with a scale-free network. J. Theor. Biol. 247, 138–151 (2007)
28. Jacobmeier, D.: Multidimensional Consensus Model on a Barabasi-Albert Network. Int. J. Mod. Phys. C 16(4), 633–646 (2005)
29. Jan, N., de Arcangelis, L.: Computational Aspects of Damage Spreading. In: Stauffer, D. (ed.) Annual Reviews of Computational Physics I, pp. 1–16. World Scientific, Singapore (1994)

30. Jost, J.: External and internal complexity of complex adaptive systems. Theory Biosci. 123, 69–88 (2004)
31. Kauffman, S.A.: Metabolic stability and epigenesis in randomly constructed genetic nets. J. Theor. Biol. 22, 437–467 (1969)
32. Kauffman, S.A.: Gene regulation networks: a theory for their global structure and behaviour. Current topics in dev. biol. 6, 145 (1971)
33. Kauffman, S.A.: The Origins of Order: Self-Organization and Selection in Evolution. Oxford University Press, New York (1993)
34. Kauffman, S.A., Peterson, C., Samuelsson, B., Troein, C.: Genetic networks with canalyzing Boolean rules are always stable. PNAS 101(49), 17102–17107 (2004)
35. Luque, B., Solé, R.V.: Phase transitions in random networks: Simple analytic determination of critical points. Phys. Rev. E 55(1), 257–260 (1997)
36. Luque, B., Ballesteros, F.J.: Random walk networks. Physica A 342, 207–213 (2004)
37. Luttge, U., Beck, F.: Endogenous rhythms and chaos in crassulcean acid metabolism. Planta 188, 28–38 (1992)
38. Li, M., Vitányi, P.: An Introduction to Kolmogorov Complexity and Its Applications. Springer, Heidelberg (1997)
39. Nowostawski, M., Purvis, M.: Evolution and Hypercomputing in Global Distributed Evolvable Virtual Machines Environment. In: Brueckner, S.A., Hassas, S., Jelasity, M., Yamins, D. (eds.) Engineering Self-Organising Systems, pp. 176–191. Springer, Heidelberg (2007)
40. Olbrich, E., Brertschinger, N., Jost, J.: How should complexity scale with system size? In: Jost, J., Helbing, D. (eds.) Proceedings of ECCS 2007: European Conference on Complex Systems, paper #276 (2007) (CD-Rom)
41. Peliti, L., Vulpiani, A.: Measures of Complexity. Lect. Notes in Phys. 314 (1988)
42. Schuster, H.: Deterministic Chaos: An Introduction. Physik-Verlag (1984)
43. Serra, R., Villani, M., Semeria, A.: Genetic network models and statistical properties of gene expression data in knock-out experiments. J. Theor. Biol. 227, 149–157 (2004)
44. Serra, R., Villani, M., Agostini, L.: On the dynamics of random Boolean networks with scale-free outgoing connections. Physica A 339, 665–673 (2004)
45. Serra, R., Villani, M., Graudenzi, A., Kauffman, S.A.: Why a simple model of genetic regulatory networks describes the distribution of avalanches in gene expression data. J. Theor. Biol. 246, 449–460 (2007), http://dx.doi.org/10.1016/j.jtbi.2007.01.012
46. Serra, R., Villani, M., Damiani, C., Graudenzi, A., Colacci, A., Kauffamn, S.A.: Interacting random boolean networks. In: Jost, J., Helbing, D. (eds.) Proceedings of ECCS 2007: European Conference on Complex Systems, paper #165 (2007) (CD-Rom)
47. Sole, R.V., Luque, B., Kauffman, S.: Phase transitions in random networks with multiple states. Technical Report 00-02-011, Santa Fe Institute (2000)
48. Sousa, A.O.: Consensus formation on a triad scale-free network. Physica A 348, 701–710 (2005)
49. Stauffer, D., Sousa, A., Schulze, C.: Discretized Opinion Dynamics of The Deffuant Model on Scale-Free Networks. Journal of Artificial Societies and Social Simulation 7, http://jasss.soc.surrey.ac.uk/7/3/7.html

50. Stauffer, D., Moss de Oliveira, S., de Oliveira, P.M.C., Sa Martins, J.S.: Biology, Sociology, Geology by Computational Physicists, 276 + IX pages. Elsevier, Amsterdam (2006)
51. Wagner, A.: Estimating coarse gene network structure from large-scale gene perturbation data. Santa Fe Institute Working Paper, 01-09-051 (2001)
52. Yamano, T.: A statistical measure of complexity with nonextensive entropy. Physica A 340, 131–137 (2004)

Theory of Reaction-Diffusion and Emergence of the Geographical Forms

André Dauphiné

University of Nice Sophia-Antipolis
98, bld Edouard Herriot
06204 Nice Cedex 3, France
andre.dauphine@unice.fr

Summary. Geography studies the organisation of territories in their physical dimensions as well as in their social dimensions. This is thus a knowledge about emergence of forms such as the scattering of fallows and then forests over abandoned agricultural soils, the growth of a city or of a transport network, or the spreading of a pioneer front in Brazil or Siberia.

In order to account for the emergence of spreading or stationary forms, a lot of scientists from other disciplines often refer to a theory simultaneously suggested by Fisher, Kolmogorov, Petrovskii and Piskunov, and usually designated by the acronym FKPP. Numerous papers and books dedicated to this theory and to the models derived from, such as Turing activator-inhibitor model or their generalisation by Meinhardt. As the creation of a chemical, its production and growth in time, is usually obtained by reaction, this theory is generally called "theory of reaction-diffusion". We have highlighted the interest of this theory since 1985, at the European Colloquium on Theoretical and Quantitative Geography. But it was not a source of inspiration for the geographers.

We suggest then to insist on the components that should be added to make this theory effective in geography. Beyond the long distance interactions designated as convection, advection, turbulence in nature sciences or transport in the societal sphere, it is advisable to think about the initial conditions that strongly influence the emergence of the geographical forms and more about the introduction of human actors adaptability.

Keywords: theory of reaction-diffusion, growth, diffusion, interaction, segregation

1 A theory very quickly enriched

1.1 A centenary theory

The theory of reaction-diffusion resumes Heraclite's idea for who every creation of a form is the product of the struggle between two principles. In its simplest form, this theory is illustrated by an equation which is written as following in usual language:

$$Variation\ of\ a = Growth\ of\ a - Diffusion\ of\ a$$

In a given point of space, the law of reaction, production or growth integrates the creation and the disappearance of new "individuals", molecules in chemistry, cells in biology, individuals in ecology, or persons in social sciences. Frequently, this growth process is auto-catalytic: the fabrication of a product "x" depends on the presence of "x" and usually on its density. The autocatalyse is actually synonym of positive retroaction. For example, the demographic growth is function of the number of reproductive couples and, then, of the population size.

In this theory, the role of diffusion becomes anti-intuitive. As a rule, diffusion restores equilibria in a gradient field. It homogenizes this field and tends to erase forms. Coupled to a growth mechanism, diffusion has therefore a reverse effect and generates new forms. These forms are created and remained because diffusion is constantly feeded by the growth mechanism.

Following the considered discipline, this equation or system of equations, when at least one of both components is considered, has different names. In chemistry, it stands for a process of reaction-diffusion illustrated by the experiment of Belousov-Zabothinski (BZ). In ecology, this theory is named by different terms because ecologists use different growth models. They mention a KISS model (Kierstead, Slobodkin et Skellam) when growth is represented by an exponential law. The corresponding model is often solicited to explain the emergence and size of marine phytoplankton. In biology, this theory originates the activator-inhibitor model first figured out by Alan Turing in 1952. Actually, this theoretical model is well adapted to every statistical population composed by individuals that reproduce and diffuse in space.

1.2 Three types of generalisation

This theory, illustrated by an equation which integrates both mechanisms of growth and diffusion, can be generally applied. To make it applicable to the study of more realistic issues, three ways have been explored: modifying terms of growth or diffusion, taking into account several populations in interaction and finally, studying a mechanism additional to the initial equation.

A first form of generalisation consists in modifying the growth or diffusion mechanism. Indeed, there is not one but several growth laws. Thus, the observed growths in human societies are rarely exponential. The logistical Verhulst law, more realistic than the exponential growth, simulates a growth slowed down by density. This slowing down applies to most of the living populations which have a restricted food available. In ecology, the model of reaction-diffusion, built from this logistical growth, was qualified as the Fisher model because of the name of the mathematician who introduced it since 1934. This model will be then applied to represent the agriculture emergence in Europe in the Neolithic.

This first generalisation sometimes applies on the second mechanism, the diffusion. In its simple form, diffusion depends on the difference of the concentration of the considered product in space. Diffusion transfers the product from high densities to low densities. It is a slow process because it acts only gradually by contiguity. And the diffusion coefficient is often assimilated to a constant. But, it is possible to make the diffusion coefficient vary, for example in function of the density. Indeed, high densities fasten the diffusion process as testify the studies carried out on migratory fluxes: migrants rather leave very peopled spaces.

A second type of generalisation is obtained by adding new terms to the elementary equation in order to take other factors into account. Thus, a city grows up and spreads by the growth of its population and its diffusion in suburbs. But, in the same time, the centre, as well as the peripheral spaces, receives a demographic excess from the rural environment, other cities or further countries. A migratory phenomenon is superimposed on auto-growth and diffusion and is easily assimilated to an advection. To insert this new mechanism, it is sufficient to insert this advection mechanism in the initial equation, which then can be written as following:

$$\textit{Variation of } a = \textit{Growth of } a - \textit{Diffusion of } a + \textit{Advection of } a$$

This advection is symbolised, as diffusion, by a partial derivative affected of an advection coefficient which obviously does not have the same value as the diffusion coefficient. Thus, both elementary mechanisms are conserved but they are amplified or thwarted by these larger moves.

A third generalisation, already announced for the BZ model, consists on coupling two or several equations. Indeed, in the reality, a species never lives alone in its environment or its ecological niche. In physics and chemistry, molecules and products mix together. And, in geography or in social sciences, these interactions between different human groups are often more frequent. In a city, several ethnic groups or social classes cooperate or confront each other. In numerous disciplines, these interactions concern two populations. The ecologist analyses relations between preys and their predators from Volterra studies.

Biologists who are specialists of animal forms development use the same equations. But they talk about activator-inhibitor model [9]. For two populations, the Turing theory is then expressed by a model which includes both of the following equations:

$$\begin{cases} \textit{Variation of } a = \textit{Growth of } a - \textit{Diffusion of } a + \textit{Interaction of } a \textit{ and } b \\ \textit{Variation of } b = \textit{Growth of } b - \textit{Diffusion of } b + \textit{Interaction of } b \textit{ and } a \end{cases}$$

Each equation includes both mechanisms of the basis model, simulating growth and diffusion and also an interaction mechanism. In this system of two equations, it is easy, in function of the positive or negative value attributed to interactions, to represent an effect of competition or, inversely, an effect of mutualism, of cooperation between the activator and the inhibitor, between both species. Fighting is not always the rule even if it was considered as essential by Malthus, Darwin and Marx.

This form of generalisation can be applied beyond two populations. Epidemiologists also use these equations by distinguishing sane and infected populations and populations which re-cover or disappear. They build then SIR (Susceptible, Infected, Release) models which simulate the behaviour of three populations. Recently, Michaël Batty (1999) adopted this model to represent dynamics of urban growths.

This theory of reaction-diffusion generates an infinity of forms. The combination of the simple laws is sufficient to simulate the emergence of linear forms, networks, gradient forms, repetitive forms, spiral forms of cellular textures, hexagonal for example. This set of two laws succeeds in creating homogeneous or heterogeneous forms, isotropic or anisotropic, punctual or areolar. Actually, as highlighted by James Murray (1990): "The theory of reaction-diffusion predicts a very rich diversity of complex forms, from the simplest as invasion fronts and solitary waves, to the most complex as spirals and spatio-temporal chaos ". The first works were about fronts dynamics. More recent studies insist on the persistence and stable condition of some forms. Thus, strips and hexagons are more stable than other forms. This attests of their higher frequency in Nature. Biologists are the ones who best studied the laws which govern forms generated by these mechanisms. Usually, forms stabilise when the diffusion coefficient of the inhibitor is seven times higher than the activator's one. In function of the size of the considered space and conditions of production, diffusion and interaction, several similar forms can appear at more or less regular space intervals.

2 Two geographical constraints: initial conditions and role of the actors

To apply the theory of reaction-diffusion in geography, a double enrichment is imperative. It concerns the initial conditions and the role of the social actors who have an amazing capacity of adaptation.

2.1 Initial conditions often determining

The initial conditions considered by physicists are usually simple. These mechanisms act on homogeneous spaces, isotropic or random. But, in geography, the terrestrial space is always structured and heterogeneous. Agents do not interact from random or homogeneous distributions. In a river, a pollutant will spread in function of the pressure gradient (diffusion) and the associated wind (advection) but also in function of the underlying topography. In a city, people diffusion takes road networks. All preliminary structures of the terrestrial space, physic or human, channel diffusion and advection mechanisms and they also intervene on growth.

It is then imperative to take into account this ground reality by introducing this terrestrial pre-structure in the initial conditions and then by qualifying the associated effects. Works of that type, developed by physicists, can guide geographers such as, for example, the book of Ben-Avraham and Havlin [1] dedicated to the study of the mechanisms of reaction and diffusion in heterogeneous and fractal environments. They show how pre-existing structures have a direct influence on the emergence of new forms. This is a source of inspiration for geographers.

2.2 The agents of human societies are adaptive

A second specificity of geographic forms lies in the originality of the human action. Humans are not simple molecules. They pursue goals and their actions tend to reach economical, social or cultural objectives. It is then necessary to include these rules of behaviours in the theory of reaction-diffusion, rules which will depend on the scientific issue considered. These rules of behaviours are very numerous but can be reduced to some generic laws.

Humans act by comparing themselves to others but also by learning, by taking the past into account and anticipating the future. Thus, in a research about the emergence of segregation forms inside a city, Thomas Schelling showed that diffusive moves of each person are made after comparison of his cultural values with his neighbours' ones. Beyond a threshold of dissimilarity which can be modified in simulations, people leave their home and move to people sharing the same cultural values. In the course of time, people who have similar values gather and homogeneous quarters emerge from a random

initial distribution. Then, the setting up of a city with its segregation phenomena needs no chief, no villain as suggested by some theories.

But, this initial model deserves to be deepened. Originally, both populations have a fixed and constant size. It is thus necessary to introduce the real growth of these populations to obtain a simple model of reaction-diffusion. Then, it is desirable to include, not one but several types of ethnic or social classes. Geographers make then a double segregation emerge which show homogeneous quarters but also gatherings always similar to these homogeneous quarters. Segregation thus appears at two different spatial scales : green with green but also quarters which became green still stay in contact with blue quarters. This pattern corresponds to the urban organisation of great American cities.

Besides, in contact with a population which shows different value, a person can evolve, learn and adopt these values. This learning is frequent when persons in contact have different social status. It is assumed that classes said to be "underprivileged" tend to adopt quite easily values or behaviours of classes qualified as "well-off". More globally, numerous socio-cultural criteria evolve under the neighbourhood pressure whereas others, such as religious believes, are very stable.

Contrarily to gaseous molecules, humans can even determine neighbours' choice with who they compare each other and interact and they can even take into account the past interactions whether negative or positive. This is namely the case of the models built on the prisoner's dilemma. Rules of decision and interaction are then implemented in the models of reaction-diffusion. Numerous simulation models illustrating the results of these behaviours are already available (Richard Gaylord and Louis d'Andria [6]). They generally include a moving process by diffusion under the form of a brownian move.

Eventually, numerous human moves are made on long distances and are not assimilable to diffusion any more. Couples leave their city, their region or their country. An advection mechanism should be added to the diffusive moves and the brownian move should be replaced by a Levy flight.

Obviously, if this generalised cultural process of reaction-diffusion generates socio-spatial segregation, the accurate location of the quarters is often arbitrary and changes in function of the initial conditions in the simulation models. This is not the case in real cities where land-prices often determine these locations. In French Mediterranean cities, Italian quarters of the Fifties and actual North African quarters are rather located in wet glens devoid of a large view over the sea.

3 Conclusion

In order to understand the emergence of forms in geography, the theory of "Reaction-Diffusion" is an excellent starting point. However, the geographer must choose the most relevant growth (reaction) model. Moreover, he must add an advective component because only diffusive moves of human order are exceptional. Humans as albatross have daily moves which obey more generally to the formalism of a Levy flight, integrating short moves and further moves. Besides, the weight of the initial conditions is essential. Diffusion said to be normal is exceptional. Whatever the fluxes of matter, energy, people, money or information which generate territorial forms, they diffuse over a non-homogeneous space which imposes its structure. Real forms are not "free" but their emergence is channelled by these initial conditions and at the limits. In addition, human actions and interactions obey to a principle of adaptation including memory, learning and then evolutionary behaviours. This diversity does really not condemn an approach by the model of reaction-diffusion. It adds a source of complexity in the research in social sciences.

References

1. Ben-Avraham, D., Havlin, S.: Diffusion and Reactions in Fractals and Disordered Systems, 316 p. Cambridge University Press, Cambridge (2000)
2. Batty, M.: Cities and Complexity, 565 p. MIT Press, Cambridge (2005)
3. Bourgine, P., Lesne, A.: Morphogenése, L'origine des formes, Belin, Paris, 352 p. (2006)
4. Cantrell, R., Cosner, C.: Spatial Ecology via Reaction-Diffusion Equations, 4 11p. John Wiley & Sons, Chichester (2003)
5. Dauphiné, A.: Les théories de la complexité chez les géographes, Economica, Paris, 248 p. (2003)
6. Gaylord, R., d'Andria, L.: Simulating Society, TELOS, New York, 218 p. (1998)
7. Grindrod, P.: The Theory and Applications of Reaction-Diffusion Equations, 275 p. Clarendon Press, Oxford (1996)
8. Livi, R., Vulpiani, A.: L'héritage de Kolmogorov en physique, Belin, Paris, 254 p. (2003)
9. Meinhardt, H.: Models of Biology Pattern Formation, 230 p. Academic Press, London (1982)
10. Meinhardt, H.: The Algorithmic Beauty of Sea Shells, 204 p. Springer, Berlin (1995)
11. Murray, J.: Mathematical Biology tome II Spatial Models and Biomedical Applications, 811 p. Springer, New York (2003)
12. Okubo, A., Levin, S.: Diffusion and Ecological Problems, 467 p. Springer, New York (2001)
13. Rabino, G.: Processi Decisionali e Territorio nella Simulazione Multi-Agente, Esculapio, Bologna, 210 p. (2005)
14. Zwirn, H.: Les Systèmes Complexes, Odile Jacob, Paris, 219 p. (2006)

Spatial risks and complex systems : methodological perspectives

Eric Daudé[1], Damienne Provitolo[2], Edwige Dubos-Paillard[2], David Gaillard[3], Emmanuel Eliot[4], Patrice Langlois[1], Eliane Propeck Zimmermamnn[3], and Thierry Saint-Gérand[3]

[1] MTG Lab., UMR-IDEES 6228, Rouen, France
eric.daude@univ-rouen.fr
patrice.langlois@univ-rouen.fr
[2] ThéMA, U.M.R. 6049 C.N.R.S. Université de Franche-Comté, France
damienne.provitolo@univ-fcomte.fr
edwige.dubos-paillard@univ-fcomte.fr
[3] GéoSysCom Lab., UMR-IDEES 6228, Caen, France
gaillard.geo@free.fr
eliane.propeck@unicaen.fr
thierry.saint-gerand@unicaen.fr
[4] CIRTAI Lab., UMR-IDEES 6228, Le Havre, France
emmanuel.eliot@univ-lehavre.fr

Summary. Research on risk and industrial catastrophes question the complexity theories. Besides the concepts of complexity which lead us to reconsider concepts of risk, hazard and vulnerability, we propose to think about more practical aspects, for example the modelling of human behaviour in crisis situations. The link between concepts as critical self-organization, emergence, bifurcation, and the methods in the Distributed Artificial Intelligence (DAI) used to model them is however difficult.

In this paper, we present ongoing analysis on the key concepts of risk science, such as hazards and catastrophes. We propose to enrich them with complex systems theories. First, we present methodological perspectives of the DAI, for example multi-agent systems, and compare them with other simulation methods used in the context of risks. Secondly, we present the MOSAIIC model (Modelling and Simulation of Industrial Accidents by Individual-Based methods) which gives possibilities to simulate the behaviour of individuals during an industrial accident. The project and the MOSAIIC model aim to explore the effects of a major industrial accident on public health. For instance, the emission and the spread of a toxic gas in an urban environment may be a serious danger for the human health. Thus we propose to study the consequences of this type of event in order to reduce the vulnerability of the populations. In the model, we emphasize both on spatial and behavioral dimensions (ie. mobility and perception of risk).

All these questions lead us to use different methodologies of analysis. For example, concerning mobility, the daily traffic can be simulated at a meso scale: a road axis for example. In that way, we aim to simulate the global dynamics of the network from the modelling of flows on arcs of the network (modulated according to the time of day and the day of week). Yet, we plan to use classical models (for instance equilibrium models) because they give an "average image" of the flows of vehicles on the arcs. Based on this first structural mobility, it is then possible to consider "a change of level" regarding both the representation and the analysis: if a risk occurs or if a specific context disrupts the structure. As a consequence, from a management of flows on the arc, we turn to an analysis of the individual behaviours in a multi-agent system.

1 Introduction

Cities are high-risk areas. Every year, around one hundred urban disasters are observed. One third of these disasters break out in the metropolises over 500 000 inhabitants, although smaller cities are affected, too. These natural and human disasters emphasize the vulnerability of the urban areas, where more and more people live. In urban areas, risk may be linked with natural disasters (earthquakes, tsunamis, volcanic eruptions, typhoons or drought), technological (industrial outbursts; air, water or ground pollutions), social or even the result of "domino effect". "Domino effects" are caused by a set of hazards that leads consequently to the disaster event. In other words, the activated hazard (natural or technological) may cause derived hazards which are thus sources of new disasters. In the cities, the complex risks, which integrate several risks, are more frequent than the sector-related risks. Our case study is the diffusion of a toxic cloud caused by the explosion of an industrial plant and its spread in an urban area. Aim of this research is to study, through modelling, the consequences of this kind of events on the population and especially on their mobility behaviours, in order to reduce the vulnerability of networks. If some models, such as fluid dynamics, seem to be efficient to handle the structural traffic in a network, they seem to be less useful if we aim to model a disrupted situation. Agent-based models offer a solution. They allow, during the time of the event, to switch from a management of flows at the level of arc, i.e. the road of a network, to a formalization of behaviours at the level of the individual. This methodology is more appropriated to deal with different situations, and especially for the analysis of panic for which nonlinear dynamics are important. Nonlinear means that very small variation on the model parameters (for example, the number of population susceptible to panic) have significant effects on the evolution of the system (from non-panic to collective panic), which is qualified as a bifurcation. In this paper, we will first go back over the concepts of the science of risks, such as hazard and disasters, that we will reformulate and enrich through the theory of complex systems. We considered risks and hazards from the point of view of

complexity, and especially through the theory of self-organization and critical behaviour. Second, we will present briefly the MOSAIIC project designed to simulate the behaviour of individuals facing an industrial accident.

2 Risks and complexities

Hazard and vulnerability are two key concepts to define risk. After delimiting them briefly, we will question them in relation to the theories of complexity, especially that of self-organized criticality [3].

2.1 The risk, a compound of hazard and vulnerability

Natural, societal or technological risks (R) are essentially tackled through two key concepts: hazard (A) and vulnerability (V). The risk is a measure crossing hazard and vulnerability according to a function $R = f(A, V)$. In this acceptance, and as A. Dauphiné emphasizes [4] it, a disaster is a reality whereas a risk is a probability. Hazard is generally defined both as a probability to occur and as an intensity. Regarding environmental risks, hazard is a probability which results from the frequencies observed on a long time scale. On the contrary, regarding technological risks, the probability to occur is less relevant because of the infrequency of these risks and of the theoretical absence of another occurrence in the future, except if we consider a natural cycle of Human mistake. In order to put off this limit, one tries to find the conditions liable to trigger a harmful event for people and equipments. An event tree analysis can be implemented to identify these causes and effect sequences as well as to determine theoretical probabilities. Next to probability, intensity of the phenomena is the second dimension of hazard. It depends on the duration of this phenomena and of the considered area. Intensity is often employed to define the areas of protection surrounding industrial sites for example, which leads us to the concept of vulnerability. Regarding risk, most of the studies use to focus on hazards but less on vulnerabilities. Because vulnerabilities use to appear during the disasters, we need to consider the human impacts of such events. Both risk and disasters may endanger cities, regions, countries and even the world.

The probability to occur and the intensity are thus the two key elements of studies which deal with hazard, but both of them become significant if the stakes, also called here the targets, have an interest recognised by a society at a particular time.

Vulnerability generally expresses both the measure of a damage to equipments and people and the ability of a society to resist a disaster. The determination of the vulnerability at a global level thus depends on the stakes exposed to a hazard and on factors of vulnerability (sensitivity of population, resistance of houses and premises, but also the quality of risk prevention, the management

of disasters with supervision and alarm systems). So, the identification of the stakes potentially important for the system in question is needed: people for an epidemic, buildings and population for an earthquake, biological factors for an oil slick... However, the existence of "domino effects" add difficulties to analyse the stakes. For instance, an earthquake destroying buildings may also cause victims among their inhabitants and may provoke an industrial accident whose consequences may be the flow of toxic products in a nearby river. These multi-risk scenarios are feared, especially in countries like Japan. In order to take into account this complexity, we can either use synthetic indicators based on qualitative studies or model and simulate the dynamics of such systems [5].

2.2 Complexity, bifurcation and resilience

In order to integrate the theories of complexity in the study of risks, we have to face a major challenge: that of time of processes and their interactions. Following the above concepts, risk is defined as a probability of spatial interaction between sources and targets. From this definition, the probability of setting off the source is defined as the hazard and the probability of reaction by the target is defined as the vulnerability (Fig. 1). The complexity

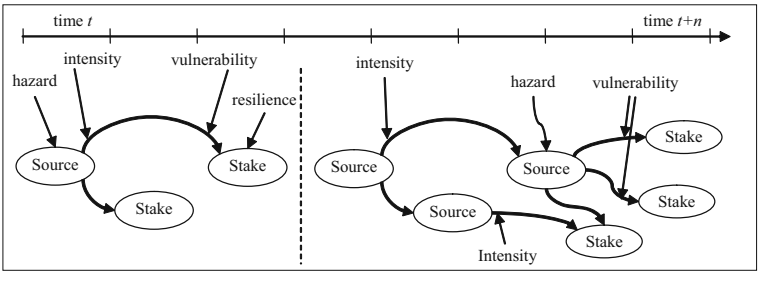


Fig. 1. Interactions between sources and stakes

of phenomena implies that risks are interlinked: from the successive interactions, each target may become the probable source of a new event. These sequences of events both occur through different temporalities, and in one or several territorial dimensions. Risk is thus a spatial and temporal concept. This approach of risks in a dynamic approach would partly explain the gap between the extent and forms of disasters (what is observed) and the extent and forms of risks (what is probable) that we generally define with concentric buffer zones. In this conception of risk as a spatio-temporal process, it is fundamental to distinguish:

- Hazard, as the probability for a source to change its state in a qualitative and eventually quantitative way, at a time t. This change of state depends

on the problem: for instance, in the case of an epidemic and at the individual level, it refers to the transition from a susceptible to an infected state;

- Intensity, as a measure of the quantity of energy released by a source (output), from its change of state and towards the outside. If the reference period of the study of intensity is relatively short, then we can comprehend it, not as a simple result, but as a process which describes the behaviour of energy in time and space. For example, the study of the diffusion of a toxic cloud after its release in the atmosphere;
- Vulnerability, as the probability for a stake to be disrupted by the energy released by a source, and thus to change its state (qualitatively or quantitatively). This probability is the result of a process which describes how the stake behaves to protect its entrances from the exits of the source, and the quantity of energy which gets into the stake. For example, the level of individual's sensitivity to panic behaviours which are displayed around him, and thus the probability for him to panic;
- Resilience, as a measure which describes the ability of the stake to adapt to change after a disturbance. For example, how individuals recover from a situation of individual, and eventually collective panic?

This last concept, richer than that of resistance already mentioned, is differently defined in the scientific literature. Coming from physics, the notion of resilience defines the ability of a material to rebound after a shock or after a continuous pressure [6]. The notion is also used in the field of psychology, of ecology and of social sciences. In ecology, Holling [7] defined it as the magnitude of disturbance that a system may absorb, without any change of its behaviour, i.e. with the ability to return to the same equilibrium. It is called "ecosystem resilience" or "ecological resilience". In the social sciences, Dovers and Handmer [8], Berkes and Folke [9] use the notion of "proactive resilience" to take into account the process of learning and anticipation of the societies towards the future. Thus, all these definitions focus on the times of return to the equilibrium state after a disturbance. However, dynamic systems may present several equilibrium states, i.e. the system may either find again its original equilibrium state or tend to a new one, after a bifurcation. Because a disaster may be considered as a dynamic system, we can evaluate the resilience from the phase spaces and the time used by the system to find again its original state or to present a new equilibrium state. In brief, the concept of resilience overcomes the boundaries of the disciplines because all the systems (either physical or social) are subject to disturbances and are thus more or less resilient to them. In addition, the concept of resilience is suitable for the disaster studies because the disasters combine social, economical and physical parameters. Some researchers have suggested, in order to reduce the damages of disasters and thus to minimize the vulnerability of exposed elements, to adopt a management strategy of risks based on this concept of resilience [10, 11, 12] in the sense of the ability of a system to return to a single steady or to a cyclic state after a perturbation. In the same

way, the UN International Strategy for Disaster Reduction (UN/ISDR) has also adopted the term resilience. It is defined by the degree to which a social system is able to organize itself and to increase its capacity for learning and adaptation.

This overview of the main concepts used to define risk enlightens the complexity of the objects considered: they may be successively sources and targets, and damages may have most of the time indirect origins with the first phenomena. In addition, according to the type of event, a same structure of interactions between a source and a stake may be a positive or negative part of vulnerability. The traffic network is a good example. If it is known to be a driving force for an epidemic hazard, it is however considered as useful for a fire hazard. Therefore, an analysis of network vulnerability in the field of risk studies is fundamental. In the case of panic phenomena, we find this same kind of complexity. In a situation of panic, the most vulnerable individuals to the intensity of the phenomenon (for example the diffusion of a toxic cloud) can quickly change their behaviours: from a state of non-panicking population to that of population in panic. If for the French school on the one hand, the terrified individuals are submitted to their gregarious instincts and irrational behaviours, for the American school on the other hand, the individuals keep forms of lucidity, abilities to analyse the situation and to take decisions: copying the neighbours, escaping... [13]. How irrational individuals may be or not, the non-linear interactions are very important in the diffusion of a panic. A few individuals may spread a panic among a whole group [14]. The crowd as a whole can then become a source for other human and technological stakes. The self-organization theory is well adapted to give an account of the emergence of such phenomena for which little local disruptions may product global and unpredicted events.

The self-organization theory identifies the principles which allow us to describe how a system creates his own behaviour at a global level, persisting in time and space, from the numerous interactions among entities displaying at one or several lower levels. These interactions are generally local ones, develop in the vicinity of each other, and such systems are characterized by the absence of planning: no global control which would pilot such structure, such behaviour, or such form. These kinds of systems are called "emergent systems" because their developments are not fully explained by the properties of entities at lower levels.

As most of natural or technical systems, self-organized systems are not systems whose equilibrium is permanent. The activity of a system, dynamic and open to the outside like all the living systems, is in evolution. Self-organized systems arrange their behaviour in relation to certain points in their environments. Order in an interconnected system of element arises in the vicinity of attractors, which create and maintain patterns within the system. Evolution

between attractors can be cyclic like prey-predator systems. Such systems are characterized by phases of intense activities: the curves of population linked to the two groups reversing more or less regularly along the time. Otherwise, the system can evolve towards a stationary state, converged on a point of attraction and absorbing progressively its activity. The activity of the system can lead it through different states through the time. This switch from a state to another is situated close to a point of bifurcation that may lead towards chaos.

In an earlier work, we explored the different phases of activity of a system from the example of the logistic function often linked with diffusion processes [15, 16].When the system evolves from a bifurcation threshold, the transition from one state to another qualitatively similar refers us to the concept of resilience. The stability of self-organized systems goes hand in hand with a possibility of change which explains that all living systems go through different phases during their activities. These phases are theorised by the criticality [1 op. cit.], which shows that all self-organized systems evolve towards a critical state and that a small and local disruption is sufficient to make the system change. This phase is characterized by a system which goes into a phase of mutual and global interaction during which the level of connections and interdependences of the elements of the system is maximal.

If they are useful in a heuristic context, such concepts are however difficult to use when one wants to apply them or to spot them in an empirical way. How knowing if a particular system comes from a decentralised context or not? How qualifying, identifying the emergent phenomena in such systems? How evaluating the intensity of relationships between elements at the same level and between the elements at different levels? How measuring resilience in a system? From measures based on particular methodologies? Systemic measures? Indicators? These uncertainties lead us to propose simple models of the complexity. This way of modelling is based on a constructivist approach for which the principle of parsimony is the crucial point for scientists who wish to tame the "artificial creatures" they build.

3 Simulating panic phenomena: methodological orientations

The aim of the MOSAIIC project is to study individual behaviours and their consequences during an industrial accident, and mainly through an analysis of the traffic network vulnerability. The hypothesis of this project is that any traffic system, made of numerous mobile entities in mutual and environmental interactions, tends to evolve towards a critical state. A small fluctuation may thus disturb the system toward a phase which considerably increases the vulnerability of the persons if the origin of the disturbance, even indirect, is a technological or natural event. The experimentation in this field is of course

either hardly conceivable or difficult to realise. Thus, the modelling methods and computer simulation offer an interesting option.

3.1 An environment of risk where hazard and vulnerability are statistically weak

According to Dauphiné's classification of risks [2 op. cit.], our case study belongs to the category of local and short time events. The analysis of technological disasters is especially relevant today in a context of urban growth leading to a proximity between residential areas and industrial plants, and after recent industrial events (AZF in Toulouse, Mède in Marseille...).

The gap between the dynamics of an observed disaster and the structure of the risk generally estimated in cities is partly explained by environmental factors. In urban areas, the variety and the number of sources and targets may grow during the event. This fact is partly the result of the proximity of elements and of the growing interactions between them. However, the main characteristics of these environments for the individuals are both the quality of buildings to confine and the traffic network in order to escape and to be rescued. As a consequence, based on the conception of risk developed in 2.2, we view global risk as a measure of all the local and contextual risks that can be observed in a situation, and for which the sequence of interactions is determining for the magnitude of the risk and of the disaster. Thus, the source is an object partly submitted to hazard and probabilities, whose outputs are a quantity of energy released (virus, toxic cloud, individual aggressiveness, physical pressure...). This energy moves in space depending on the nature of the released energy. In order to counter the flows and to prevent the target stakes from being reached, preventive measures can be taken to limit the vulnerability upstream: alarms, buffer zones, educational measures etc. Aim is to reduce the inputs to the target objects (stakes). However, if the quantity of energy coming in the stakes is important, then these last may have a high probability to become targets. These last will then determine the outputs towards new stakes (Fig.1).

In this conception of global risk as a dynamic process in which the sequences of phenomena may be numerous, the MOSAIIC project proposes to focus on individual behaviours and on vulnerabilities linked to them. Thus in this project stakes and sources are humans.

3.2 The vulnerability of traffic networks during disasters

Studies dealing with vulnerability of systems need to take into account the constraints that both time and space present. The spatial level has to be defined. If a disaster may destroy a system, this last is - most of the time - a subsystem belonging to a bigger system which may not be disturbed by this

disaster. Once the spatial level of the vulnerable system is both defined and analysed, elements and interactions of the system have to be recognized. In this paper, we will especially focus on the vulnerability of traffic networks in order to estimate its consequences on the exposed populations.

Self-organised and stable systems

We aim to explore the vulnerability of the transport networks because disasters use to disrupt the territorial organisation. The disasters may sometimes affect nodes or areas, but they often affect networks. The energy, the information and the transport networks are often the most fragile. Their vulnerability is associated with their properties of connectivity. Anisotropic networks are thus more vulnerable than polarized ones. The violent disruption of the network connectivity may thus create a lack of autonomy and may disorganize the territory. Yet, a better understanding of both the forms and the connectivity of networks contribute to improve risk prevention and risk management policies. Traffic networks may be directly or indirectly the origin of a disaster. For instance, the transport of dangerous materials in cities is a factor which tends to increase hazards. In this case, a mobile source (a truck) is not submitted to the same types of control and is not classified in the same category of risks as a static source (a factory). In addition, the nature and the quantity of stakes may vary during the shift of the object. An other aspect concerns the stakes and the vulnerability, and the need to qualify the role and to quantify the impact of traffic networks on a disaster. Urban environment is strongly restricted by its traffic network and may create use conflicts between the 'active' or 'passive' actors of the crisis. As one of the options to avoid epidemics is to isolate individuals [17], a sound management of technological disaster would propose to evacuate them out of a perimeter and then to confine them beyond a security line. However this strategy is rarely used and, in fact, is difficult to implement.

Different individual strategies and behaviours coexist during an accident. We can summarize these strategies by considering two forces: a centrifugal one (moving away) and a centripetal one (moving closer) [18]. These two forces, constrained by the reticular environment which limits the possible paths may produce three types of movement or flows within the network: a flow in the opposite direction of the source, a flow towards the source, and a parallel one (the source is a front) or perpendicular one (the source is a point) to the source. According to the connectivity of the network[1] (Fig. 2) and the area where the event is located, the different types of mobility will be more or less possible in a given perimeter. In this context, how characterizing the "normal" regime of a network and how detecting the change of this regime

[1] we use the connectivity index β [19] which is based on non oriented graphs, and can be calculated by carrying over the number of nodes (s) to the number of links (l) : $\beta = l/s$

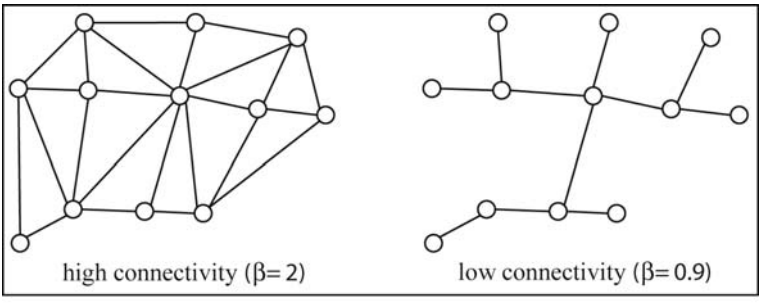


Fig. 2. Two graph models of a road network

toward an exceptional activity? This new regime could be considered as a precursory sign of dysfunctions in the system, and in our case of a possible disaster.

The average regime observed in a traffic network can be written with a function A which represents the functional dynamics of the system . It is an indicator of the level of the functioning of the system at a time t.

$$A_i = tmp_i - tp_i \tag{1}$$

where *tmpi* is the average traffic on the network at a time i and *tpi* is the instantaneous traffic on the same network. Within this framework, *tmp* correspond to a traffic modulation on different routes which are always the same in the network and at different moments in a day. The global *tmp*, as shown in the Figure 3, is an average value at different hours of the day (continuous line on the graphic). For example, the *tmp* indicates the average value of density for a week day (for instance at 8am, in the graph from the southernmost point to the northernmost one) (Fig. 3). We can select a standard day to analyse the variability of situations, and thus to observe the consequences on the dynamic of a technological accident. Doing this way, we observe three very different types of daily modulations: the working days (JO), Saturday and the days before official holidays (SVF), Sunday and days before official holidays (DJF)[2]. This typology emphasizes the intervals between different days at the same hour. In addition, it allows us to compare the average hourly values between them in order to give a qualitative level of the functioning of the network: a level of fluidity, for example (horizontal continuous line on the graphic).

At the same level of analysis, the *tpi* is the traffic observed on the network, measured 'in live' during the simulation. So if the *tmp* is an estimation

[2] This parameter (C) is equal to the hourly trafic car average on an axes (VMH) divided by the daily trafic car average (VMJ), divided by 24. Thus, $C = VMH/(VMJ/24)$, this parameter has an average equal to 1.

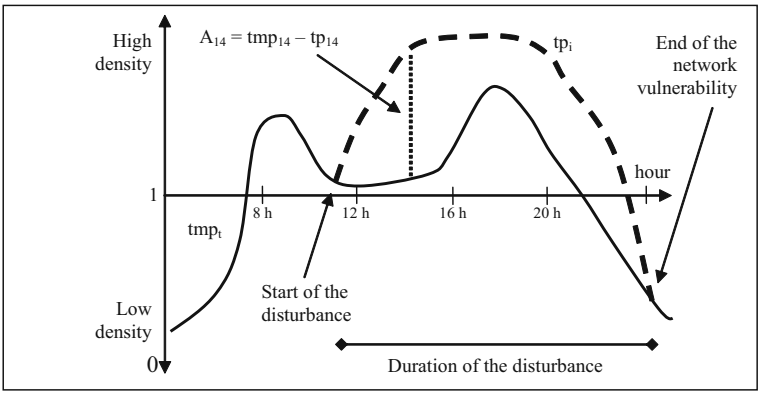


Fig. 3. Comparison between a diagram of disturbance and a series of daily average values

calculated from a counting during a period of time, then the *tpi* is the traffic in the network, or in a path, observed at a given moment. For example, the *tp*8 value indicates that at 8am on the 2nd of August, n vehicles are in the graph from the southernmost point to the northernmost one.

As a consequence, if the *At* value is positive, the traffic on the section is, on average, higher at this hour and during this type of day, than the traffic recorded at the moment i. On the contrary, if *Ai* is negative, the traffic, observed at the moment i on the network, is higher than the average traffic. The advantage of such a formalization is to show, at which moment the system enters in a disrupted phase, the duration of this phase, its intensity and at which moment the system recovers its stable regime.

Such a diagram presents our conception of the traffic network: a complex, stable and partly self-organised system whose dynamic is independent from the entities which constitute it. Although the entities (cars, pedestrians), may change, the laws which organise it remain. Yet, there is an autonomy of the traffic in relation to its constituent elements. The existence of dynamic independent patterns at a meso level of the network will lead us to model this dynamic at a meso level: for example with differential equations. But how switching from this modelling of the dynamic at a meso level to a modelling at a level where micro-changes have impacts on the global functioning?

Phases of criticalities

Why a system may move from a attraction point to another one and thus globally move away from the average observed traffic? Our hypothesis is that changes of aims and motivation within a group of individuals favour this bifurcation. These individual changes will be all the more quickly transmitted

to the whole people (a total correlation between the elements of the system [1 op. cit.]) since they will be constrained by a network and a territory whose capacities of adaptation on a so short period is nearly non-existent. Only few entities have to change their behaviour to produce feedback effects on a part or on the whole system. Yet, its properties may be changed partly, temporally or on a long term basis. This criticality of the system is all the more high since we are situated, in the space of parameters, in a zone of instabilities which are characterised for example by a level close to congestion. The possibility for the system to bifurcate is all the more high since the degree of freedom of the entities in the system is weak, and since interdependence becomes global. In our case, that means that hazard and vulnerability factors are all the more important since the number of individuals circulating in the network and in the urban area is high. The diagram in Figure 4 presents these ideas. It combines the density (number of vehicles on an x axis) and the flow (number of vehicles along an y period of time). This diagram is in general estimated and observed on small intervals of space and time, in order not to combine different states of traffic.

So the density is low if the individuals adopt the speeds of their choices only limited by statutory constraints. On the contrary, if both the density or the rate of occupation of the way rises, the traffic is more and more constrained and, beyond a certain limit - a critical density (Kc) - it reaches the congestion. We plan to model the micro level dynamics close to this threshold: for example, the variable of density will be converted into a list of agents on the relevant axes. In a next step, we will define the properties, methodologies, aims or strategies of these agents.

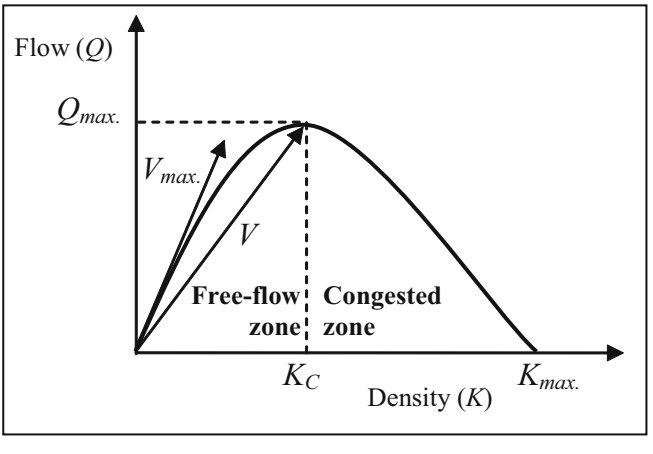


Fig. 4. Parabolic diagram: estimating congestion with flow and density

The state of traffic thus goes through different points, a possible trajectory leading to $Kmax$: points of attraction which indicate a full saturation of the road, or even of the network. The network resilience is thus the ability of the network to move to a point of attraction below the critical threshold Kc: a point that we plan to discover thanks to the analysis of the behaviour of the agents.

4 Conclusion

The analysis of the network vulnerability, measured by its trend to maintain a phase of global disturbance, proposes to stress one of the elements of complexity of disasters: the interactions between individuals and territory as well as between individuals.

Despite this paper as well as the MOSAIIC project both focus on a particular category of risk, these reflections may lead us to suggest a more general model dealing with human behaviours in situations of crises. Thus, this model may be applied to other categories of risks that imply population shifts in strongly constrained spaces. Such a model offers an opportunity to test a great variety of behaviour scenarios as well as to analyse the incidence of the network structure on behaviours. Besides these different tests, we plan to study to which extent individual behaviours are likely to lead towards a critical point, even during situations when events occur far away from a critical phase.

At last, this project plans to provide a cartography of the different strategies possible in contexts of crises to decision makers and people dealing with the problem of crisis management: evacuation strategies by getting closer to potentially disaster zones or strategies of intervention by mapping routes that allow to reach the targets.

References

1. Bak, P.: How Nature Works. Springer, New York (1996)
2. Dauphiné, A.: Risques et catastrophes: observer, spatialiser, comprendre, gérer. Armand Colin, Paris (2003)
3. Provitolo, D.: Modélisation et simulation de catastrophe urbaine: le couplage de l'aléa et de la vulnérabilité, Actes du colloque SIRNAT, La Prévention des Risques Naturels, Orléans (2003), `http://www.brgm.fr/divers/sirnatActesColl.htm`
4. Mathieu, J.P.: Dictionnaire de physique. Masson, Paris (1991)
5. Holling, C.S.: Resilience ans stability of ecological systems. Annual Review of Ecology and Systematics 4, 1–23 (1973)
6. Dovers, S., Handmer, J.: Uncertainty, sustainability and change. Global Environmental Change 2(4), 262–276 (1992)

7. Berkes, F., Folke, C.: Linking Social and Ecological Systems: Management Practices and Social Mechanisms for Building resilience. Cambridge University Press, New York (1998)
8. De Bruijn, K.M.: Resilience and flood risk management, a system approach applied to lowland rivers. Thése, Université de Delft (2005)
9. Vis, M., Klijn, F., Van Buuren, M.: Living with floods. Resilience strategies for flood risk management and multiple land use in the lower Rhine River basin. NCR-publication 10-2001. NCR, Delft, The Netherlands (2001)
10. Dauphiné, A., Provitolo, D.: La résilience: un concept pour la gestion des risques. Annales de géographie 654, 115–125 (2007)
11. Dupuis, J.P.: La panique. Les empêcheurs de tourner en rond, Paris (1991)
12. Provitolo, D.: A proposition for a classification of the catastrophe systems based on complexity criteria. In: 4th European Conference on Complex Systems (ECCS 2007), - EPNACS 2007 - Emergent Properties in Natural and Artificial Complex Systems, Dresden, Germany (2007), http://www-lih.univ-lehavre.fr/bertelle/epnacs2007-proceedings/provitolo4epnacs07.pdf
13. Langlois, P., Daudé, E.: Concepts et modélisations de la diffusion géographique. Cybergeo: Revue européenne de géographie 364 (2007), http://www.cybergeo.eu/index2898.html
14. Daudé, E.: Apports de la simulation multi-agents á l'étude des processus de diffusion. Cybergeo: Revue européenne de géographie 255 (2004), http://www.cybergeo.eu/index3835.html
15. Eubank, S., Guclu, H., Anil Kumar, V., Marathe, M., Srinivasan, A., Toroczkai, Z., Wang, N.: Modelling disease outbreaks in realistic urban social networks. Nature 429, 180–184 (2004)
16. Provitolo, D.: Un exemple d'effets de domino: la panique dans les catastrophes urbaines. Cybergeo: Revue européenne de géographie 328 (2005), http://www.cybergeo.eu/index2991.html
17. Pumain, D., Saint-Julien, T.: L'analyse spatiale: Localisation dans l'espace. Arman Colin, CURSUS Géographie, Paris (1997)

A new classification of catastrophes based on "Complexity Criteria"

Damienne Provitolo

ThéMA, U.M.R. 6049 C.N.R.S. Université de Franche-Comté
32 rue Mégevand 25030 Besançon, France
damienne.provitolo@univ-fcomte.fr
Géosciences Azur, U.M.R. 6526, Université de Nice Sophia-Antipolis, CNRS
250 rue Albert Einstein, Les Lucioles 1, Sophia Antipolis, 06560, Valbonne, France
damienne.provitolo@geoazur.unice.fr

Summary. A classification system of catastrophic events is a methodology assembling all the catastrophe groups. There are many classifications of catastrophes. The most widely known are classified into: nature, consequences of the event, duration, affected territories and areas of the destroyed zone, and at last into the needed intervention measures. It is difficult to study the complexity of catastrophes with these classifications. We propose a new system of classifications of catastrophes based on "Complexity Criteria". Within the scope of this paper, we focus first on organization, self-organization and emergence of the event, and then on spatial and temporal scales of the catastrophes. The organization is a key concept of complexity. In the catastrophe field, complexity of organization results essentially from the self-organization of the systems (the system develops its internal constitution and its behaviour because of the interactions between its various components and not because of an external strength). Phenomena as different as mantels of snow, seismic hazards, behaviours of the persons and the crowd have characteristics of self-organization allowing the emergence of new events: snow avalanches, earthquakes, collective panic. A particular attention will be given to the emergence of collective panic in a context of catastrophe. We observe that emergent properties appear after the interactions between the individuals and the crowd happened It means that we need to take into account the multi-scales aspect in order to be able to study the behaviours.
The complex systems of catastrophe have characteristics able to emerge at higher or lower levels of scales. It allows us to study the complexity of the disasters through the scales. The disasters belong to the various temporal- and space scales.

First, the disasters cannot be classified in a single category of spatial scale. Some of them appear on the scale of a territory, a region, a country or the planet. If we speak about a natural or technological disaster, none of them will be automatically associated with a spatial scale. Furthermore, a local disaster can have large-scale impacts. Various events (for example, the earthquake in

the Chinese province of Sichuan in 2008, the hurricane Katrina which destroys New-Orleans in 2005, the tsunami which ravaged the South of Asia in December 2004, the catastrophe of September 11th, 2001 in New-York) remind us that the catastrophe is not always an event within a restricted area. A catastrophe have often impacts outside this area. The complexity of catastrophes can result from the interactions between different spatial levels and from systemic relations between these levels.

The complexity also results from various temporal scales of the risks and the disasters. There are three temporal phases. The first one is relative to the temporality of the potential risk I mean what takes place before the disaster. The second phase refers to the temporality of the disaster I mean all what happens during the catastrophe. During the disaster, the temporalities of hazards, vulnerabilities and domino effects rarely happen together. The third and last phase refers to the time after the disaster and to the experience feedback for the risk management. These three temporal phases are based on two scales of time: a short time, I mean a time - action, inherent to the functioning of any dynamic system [1] and a long time. The study in each of the scale gives some information about the whole of the catastrophe, or about some of its components (hazard, vulnerability and domino effects).

Keywords: catastrophe, classification, complex systems, self-organization, emergence, panic, spatial and temporal scale

1 Introduction

Complexity have different meanings. The definitions vary according to the branches of study. We could almost say that there are as many definitions as disciplines using this notion. We can explain that by different scientific fields (the physics, the biology, the mathematics, the computing, the economy, the geography, the philosophy, the law) which are interested in this new way of studying the systems, in this new multidisciplinary approach, and by the youth of this field of study. The sciences of complexity appeared in the last decades of the twentieth century. The researchers hope to find in these sciences a general explanatory framework which would allow us to understand the ensemble of dynamic systems.

There are many definitions of complexity. Some authors contrast complex with complicated [2, 3], the others, such as Casti [4] define complexity by its opposite: by illustrating simple systems. The properties of simple systems (few constituents and interactions, centralized decision-making, the predictibility of its behaviour) are different from those of complex systems with numerous elements, varied natures, and interactions, where there is no leader, and unpredictable behaviour can appear. Others distinguish between the complexity of the unpredictable aspect, where determinist systems can

present complex behaviour, and organized complexity. The latter insists on the structure and the organization of a phenomenon. All these definitions propose a global approach to the systems. All of them, to various degrees, express a relation between the whole and the parts of a system, more exactly the fact that the knowledge of the parts is not enough to explain everything [5]. Rather than giving an exact and exhaustive definition of complexity, we shall present the characteristics and properties of complexity derived from definitions found in the literature. It is possible to distinguish between the functional and spatiotemporal characteristics [6]. The functional characteristics of complexity refer to the principles of interaction, organization, self organization and emergence. The spatiotemporal characteristics reveal the unpredictability of the behaviour of system, its bifurcation and the existence of multi-scales systems. In a complex system, these characteristics are mostly interdependent. In this article, these properties are defined and applied to catastrophe systems. In the expression, "catastrophe systems", the word "system" represents a set of elements in interaction, "an entity not reducible in its parts. (...) It [a system] implies the appearance of emergent qualities which the parts did not possess" [7]. The word "catastrophe", in the sense of social and spatial disorganization of the territorial system affected by a disturbing event, implies that these interactions concern the various components of the catastrophe, namely hazards, vulnerabilities, and domino effects meaning a chain of events activated by hazard or vulnerability. Thus the catastrophe is at the interface of nature-society relations, at the interface between the site of danger and the site of vulnerability.

In this article, we set up basis for a classification of the catastrophes based on complexity criteria. We focus first on complexity of organization, self-organization and emergence of the phenomenon, and then, on complexities resulting to the spatial and temporal scales of catastrophe. But first, we present a panorama of existing systems of classifications.

2 A System of Classifications of Catastrophes Based on a Sectorial Approach

A classification system of catastrophes is a methodology assembling all the catastrophe groups. It is possible to identify many classifications of catastrophes based on a sectorial approach. The most widely known are classified into nature, it means the hazard being at the origin of the event. Therefore the catastrophes are either: -natural, technological or technical and can be as well the result of a social behaviour. Each of the groups will be again changed into subgroups. The natural catastrophes can be the result of an action of the earth, for example in case of earthquakes, volcanic eruptions, landslides, avalanches. It can be the result of an action of the water, for example in case of floods, tsunamis or drought, an action of the air and wind in case of storms and hurricanes, an action of the fire during fires caused by the lightning or the

volcanic eruptions [8]. The technological risks have always anthropic origins. They include industrial, nuclear and technical risks or accidents of transport. These events can be again decomposed into subcategories. For example, the accidents of transport can happen: in the air, on the sea, on the road or by rail. Concerning the sociological catastrophes we can find 2 groups: the accidental ones and the ones which have been intentionally caused [9]. In the first case, they appear during crowd events, the crowd creating a disaster of appear with an external event: for instance the collapse of a building, of tier of seats. In the second case, the disasters are related to warfare or terrorism attacks. A second kind of classification is set up regarding the consequences of the event. These consequences are estimated mostly in term of losses, more rarely in terms of gains. They can be either material or human. The material consequences concern the deteriorations and the destructions of various infrastructures (houses, public establishment, industries, roads etc.). The categories of catastrophe also vary according to the number of victims. According to the amount of victims, the disaster will be considered as moderate, average or major. These data are mostly mixed with the nature of the disaster, in order to obtain the amount of injured and dead people for every catastrophe. The Emergency Disaster Data Base gives this type of information. The third classification can be realized with the needed intervention measures: local, regional, national or even international management measures. Finally, scales of classification exist in the field of seismic or nuclear disasters. The Richter magnitude scale assigns a single number (from - 2 to + 9) to quantify the amount of seismic energy released by an earthquake and the earthquake effects. On the same way, the International Nuclear Event Scale gives 8 levels of severity: from the major accident to the most simple anomaly.

These classifications are operational. They allow the risk manager to anticipate the situations, to manage them during the catastrophic event, to increase the means of intervention and to ask for more support (international assistance for example during the tsunami in Asia in December, 2004), to identify the priority zones of intervention, and to realize rescue plans. If these classifications are operational, however, they present some limits. First, there are no common methodologies of catastrophe classifications on a national or international scale. Furthermore, these classifications, based on a sectorial approach, are rarely required in their variety to categorize a catastrophic event. It's thus difficult to realize comparisons between catastrophes of diverse previous history, if not in terms of human and material losses. Furthermore these classifications exist for a single group of risk. There is no multi-risks classification. Now, the urban societies, which are the most concerned by the disaster, are multi-risks societies. Finally, with these classifications, it is difficult to analyse the complexity of the catastrophe. Thus we propose another way of reading of the catastrophes by setting up basis for a classification established on the complex characteristics of the catastrophes.

Complexity of catastrophes is based on at least 4 criteria: those inherent to the organization, the self-organization and the emergence of the phenomenon, those coming from the spatial and temporal scales, those resulting from geometrical forms of the risk and the catastrophe, and finally those resulting from the non-linearity and the unpredictable dynamics of the systems. These various types of complexity don't exclude each other but can be observe together during a disaster [10]. Within the scope of this paper, we focus first on complexity of organization and emergence of the phenomenon which result from it, and then, on complexities resulting to the spatial and temporal scales of the catastrophe.

3 Organization, self-organization and emergence of phenomenon

The organization is a key concept of complexity. We must differentiate between the structural organization and the level of emergent organization. In the first case, complex systems are composed of many and various elements interacting with each other. In the catastrophe field, these elements are from now on well identified: it is about hazards, about vulnerabilities (human, building, network) and domino effects [11]. We will not mention it in details here. In the second case, the interactions between elements generate levels of emergent organization. In the catastrophe field, the complexity of organization results essentially from the self-organization of the systems. There are many definitions of self-organization. According to J.L. Deneubourg [12], for many researchers, self-organization opposes to hierarchy and to centralized decision-making. For others, self-organization refers to the structure of the system. A self-organized system develops its internal constitution and its behaviour because of the interactions between its various components and not because of an external strength [13]. In the same way, J.L. Deneubourg [12 op.cit.] defines self-organization as "the ability of a group of units to produce a structure on the scale of the group, without marking this structure at the level of individual behaviours". This richer definition refers indirectly to the principles of complexity expressed in the introduction, namely: the spreading of authority, i.e. the system behaviour is not controlled by a leader but by the multitudes of these units, the interactions between the constituents of the system, the emergence of structures, laws or unpredictable properties. We shall use here the term of emergence in the sense of S. Kauffman [14]: emergence implies that collective phenomena cannot be explained by the properties of their constituents. These emergence of phenomena can be observed equally well in animal and human societies. These phenomena can activate, strenghten or weaken a catastrophe. In the catastrophe field, phenomena as different as mantels of snow, seismic hazards and population behaviours have characteristics of self-organization which allow the emergence of new events such as snow avalanches, earthquakes, collective panic etc [13 op.cit.] [15, 16, 17]. In this paper, we propose a simulation of the emergence of the collective panic

from individual panic. The study of panic enables to put human vulnerability in the foreground of our analysis of disasters.

3.1 A model of panic

For the American school of thought, the mechanisms of panic propagation are based on behaviour of imitation, contagion or suggestion [17 op. cit.] [18]. The crowd is then the support of phenomenon of contagion [19]. In the same way, during a collective panic, there is no coordinated action or dialogue between the persons. The collective panic would thus appear from the diffusion of individual panic, without the attendance, the domination of a leader who would call to the panic.

In this article, we have chosen the system dynamics modelling and the Stella Research software to simulate the behaviour of panic during a catastrophe [20]. The dynamic modelling of the system is found originally in the General System Theory of Ludwig Von Bertallanfy [21]. This theory is based on an analysis in which we divide a system into a number of different components which can be connected by flows of material, energy or information. System dynamics is a methodology used to understand how systems change over time. The complex systems are a particular case of dynamic systems, because we study in both cases the qualitative and global behaviour of the system. But in the case of complex systems, the researcher will consider the behaviour of non-linear systems and the representation of the system in the phase space in order to identify bifurcations, i.e. the qualitative changes of the system. The mathematical formalism of the system dynamics modelling is based on differential equations; the graphical formalism on stocks, flows, converters and connectors. The Stella software solves differential equations as difference equations, what explains that the small time steps smooth the curves. The stock or state variables are the reservoirs of the system. The flow variables correspond to the processes. Converters are rates which allow the researcher to connect flows of varied nature. Finally connectors are the relations and the feedback between the various elements of the system. Feedbacks create often unpredictable behaviours.

The model represents the dynamics of transmission of the individual panic to the collective panic in a crowd. This model is based on the epidemiological models of W. Kermack and A. McKendrick. It is based on three simple hypotheses.

Hypothesis 1: the crowd is constituted by three groups of population: the Population Susceptible to Panic (Psp), the Panicking Population (Pp), and the "Non-Panicking Population" (Npp). In a crowd, we can observe interactions between these three populations.

Hypothesis 2: Panic is a phenomenon of contagion, of collective imitation [22]. We use a transmission rate to apprehend the contagion of the panic between both human populations in contact. Indeed interactions between human populations do not necessarily lead to contagion. This transmission rate is a coefficient which varies from 0 to 1, i.e. a low to a high contamination.

Hypothesis 3: After a certain period of time, people will stop panicking and resume normal behaviour. In the model, there is an outflow which "empties" the stock of the panicking persons. This outflow is proportional to the numbers of individuals in panic and the return time to normal behaviour (Rtn).

These hypotheses simplify sometimes the real situations. For example, we do not take into account the age of the population, the social structure [23], or others factors, although such distinctions are important.

Figure 1 shows a graphic version of a situation of panic behaviour. This model includes three stocks of population: the population susceptible to panic (Psp), the panicking population (Pp) and the "non-panicking population" (Npp), i.e. people will stop panicking and resume normal behaviour. These three populations are the state variables of the system. This model includes interactions between the populations, transmission rate of panic (Tr) and return time to a normal behaviour (Rtn).
The corresponding equations are:

$$Psp(t) = Psp(t - dt) - (adoptions) * dt \tag{1}$$

$$Adoptions = InteractionPspPpNpp * TransmissionRate \tag{2}$$

$$Pp(t) = Pp(t - dt) + (adoptions - Normalbehaviour) * dt \tag{3}$$

$$Normalbehaviour = Pp/Rtn \tag{4}$$

$$Npp(t) = Npp(t - dt) + (NormalBehaviour) * dt \tag{5}$$

$$TotalPopulation = Psp + Pp + Npp \tag{6}$$

$$InteractionPspPpNpp = (FractionPsp * FractionPp * FractionNpp) * TotalPopulation \tag{7}$$

$$FractionPsp = Psp/TotalPopulation \tag{8}$$

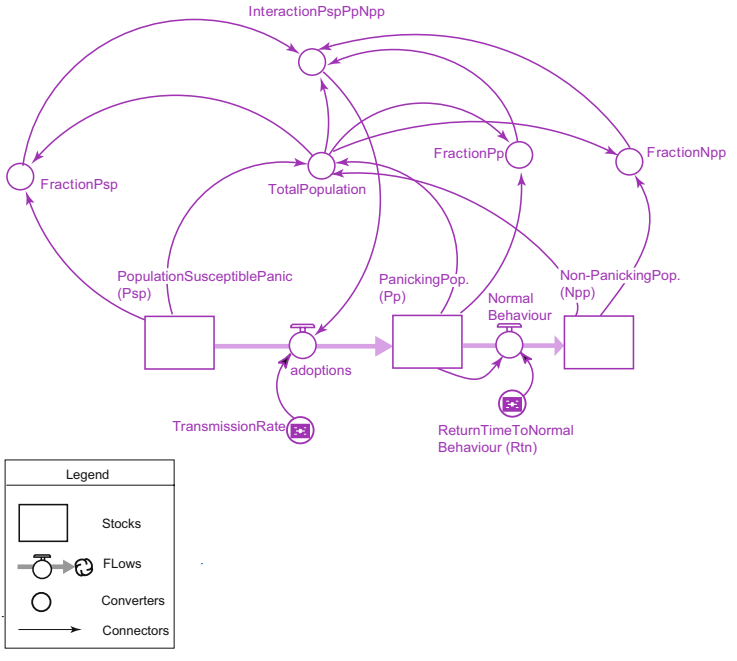


Fig. 1. A model of panic

$$FractionPp = Pp/TotalPopulation \tag{9}$$

$$FractionNpp = Npp/TotalPopulation \tag{10}$$

The focus of this paper is a formal modelling, but not the prediction results of the panic behaviour.

3.2 The emergence of collective panic is not a fate

We present results of simulation for different values of initial conditions of panicking population and of parameters related to the model. We focus more particularly our attention on the spread and the emergence of panic.The population susceptible to panic is always equal to 500 individuals. We present three cases. The cases 1 and 2 present simulation for initial different conditions of panicking population. In these two cases, the transmission rate of the panic is equal to 1 and the return time to normal behaviour to 24 units of time. The case 3 shows the evolution of the model when the transmission rate and the return time to normal behaviour vary. The analyse of the trajectories of the system is based on a phase plan method.

Case 1 :

- $Psp = 500$
- $Pp = 13, Pp = 50, Pp = 100, Pp = 200, Pp = 300, Pp = 500, Pp = 600$
- $Npp = 0$ (at the beginning of the simulation, this stock is equal to 0 because the panicking persons have not found yet their normal behaviour)
- Transmission rate $(Tr) = 1$
- Return time to a normal behaviour $(Rtn) = 24$ units of time

Figure 2 shows that for the values of the panicking population $\geq$ to 12, and for $Psp > Pp$ or $Psp < Pp$, all the trajectories of evolution aim to the equilibrium point where Psp and Pp are zero. The panicking population and the population susceptible to panic tend to disappear. For different values of Pp, we do not observe qualitative modification of the model. Trajectories have the same shape. They spread on various points on the Y axis (as soon as the curve reaches the axis Y (Pp), then the values of Psp are zero for various values of panicking population), before converging all to the same equilibrium point. All these trajectories show that for a high transmission rate of panic $(Tr = 1)$ and a return time to a normal behaviour equals to 24 units of time, there is emergence of the panic (upward slope), before reaching a equilibrium point. The emergence of the panic is particularly visible for low initial values of panicking population (13, 50 and 100). Beyond these values, the slope of the curve is smoother.

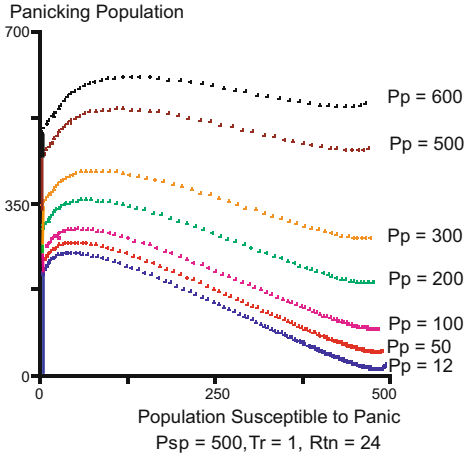


Fig. 2. From emergence of panic to the return to an equilibrium point

This equilibrium point with coordinates $(0, 0)$ is explained by the flow "normal behaviour " which tends to empty the stock "panicking population" and to feed that entitled "non-panicking population". Figure 3 shows the phase plan for "panicking population" and "non-panicking population". For every trajectory, in the similar profile, we identify two breaking points: a first point

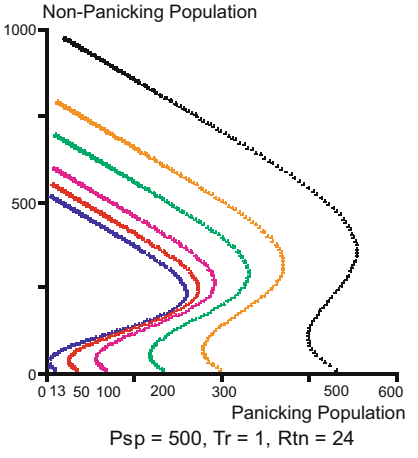


Fig. 3. Phase plan of panicking population and non-panicking population

of break indicates the transition from the decrease to the emergence of the panicking population; and a second breaking point where the trajectories show the increase of the "non-panicking population". As soon as the curve reaches the axis Y (Pnp), then Pp is zero. All the equilibrium points spread to the axis Y.

Case 2 :

- $Psp = 500$
- $Pp = 11$
- $Npp = 0$
- $Tr = 1$
- $Rtn = 24$ units of time

On the other hand, for values of $Pp < 12$, we observe a qualitative modification of trajectories (Fig. 4). The phase plan is completely different from the previous case. The trajectory converges to an equilibrium point (coordinated 0,0), but contrary to the previous case, the collective panic does not appear. The transition from 12 to 13 panicking persons with 500 persons susceptible to panic modifies qualitatively the dynamics of the system. There is a bifurcation, an effect of threshold, with $Pp = 12$, the "value threshold" beyond which there is effectively emergence of panic.

Case 3 :
Finally, we study the system evolution by making vary the transmission rate of the panic ($Tr = 0.5$) and the return time to normal behaviour. Six tests are carried out with values of Rtn equal to 14, 20 or 24 (Fig. 5) or equal to 5, 10 or 13 (Fig. 6). The population susceptible to panic is equal to 500, the panicking population is equal to 50. For a "return time to normal behaviour" (Rtn) equal to 14, 20 or 24 units of time, the phase plan is identical to the

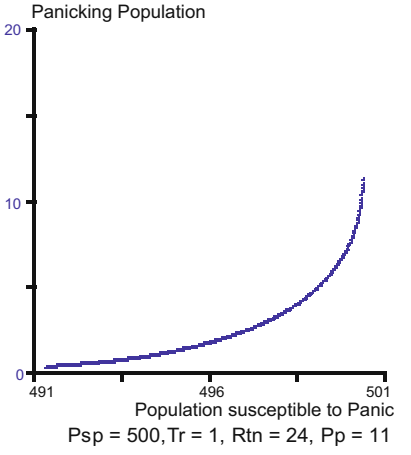


Fig. 4. Two bifurcations: from the emergence of "panicking population" to the increase of the "non-panicking population"

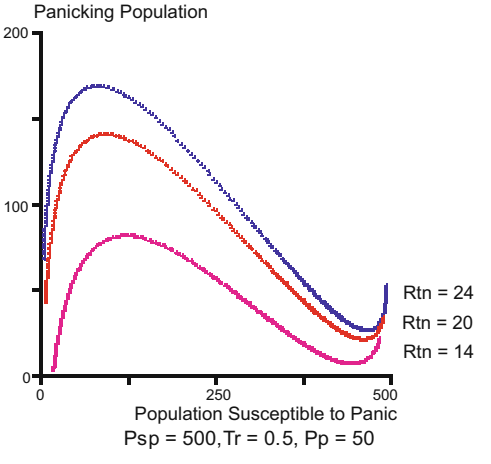


Fig. 5. An effect of threshold with $Rtn > 13$

case 1 (Fig. 2). There is a first phase corresponding to the decrease of the population susceptible to panic and the panicking population, a bifurcation then the emergence of the panic before reaching a bifurcation bringing the system to a new state of equilibrium where Psp and Pp are zero. On the other hand, trajectories are different for return time to normal behaviour, equivalent to 5, 10 or 13 units of time. The modification of Rtn influences the proportion of persons susceptible to panic and panicking people. The panic cannot spread over.

The emergence of collective panic does not appear in every situation. This emergence depends on the transmission rate, the return time to the normal

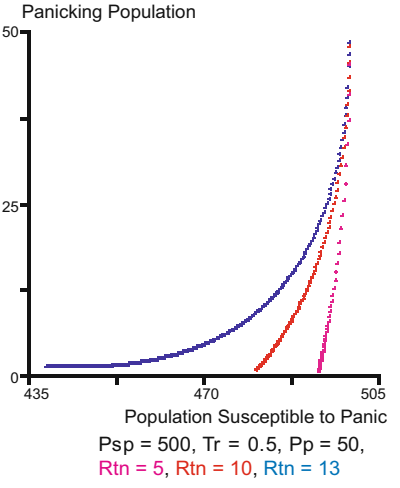


Fig. 6. The emergence of panic is not a fate

behaviour, but also on the number of panicking population at the beginning of the simulation. The complex systems of catastrophe have characteristics able to emerge at upper levels of scales. In the following section, we study the complexity of the catastrophe through the scales.

4 The spatial and temporal scales of the catastrophes

The scale notion has a main place in the study of the catastrophes. Indeed, the disasters belong to the various temporal- and space scales. First, the catastrophes can not be classified in one single category of spatial scale. Some of them appear on the scale of a territory, a region, a country or the planet. It allows us to distinct the catastrophes which are considered as localized from the more diffuse ones. But, if we speak about a natural, a technological, a social or a sanitary disaster, none of them will be automatically associated with a spatial scale. For example, terrorism, present on a world scale, can affect a confined territory, and have something of an impact on vast territories. Furthermore, we can note that metropolis play a role of "spatial switch" which enables the disturbance to spread outside the initial impact zone and on multiple scales (agglomeration, region, country, planet) [24]. Thus, a local disaster can have multiple scales impacts. We can proceed from the same assumption for more diffuse catastrophes. Various events (the tsunami which ravaged the South of Asia in December 2004, the hurricane Katrina which destroys New-Orleans in 2005, the climatic risk) remind us that the catastrophe is not always an event restricted at the affected area but can have consequences outside this area. Regarding this various spatial scales of catastrophe, we have to consider the interaction of the scales of intervention (not only the local and regional one,

but also sometimes the international one). The increase of the complexity of the catastrophe can result from the articulation between different spatial levels and from systemic relations between these levels. The question of the interaction of various spatial scales is certainly a theoretical and methodological challenge for the researchers in sciences of risks and catastrophes. Indeed, studies of risk are mostly limited to a single spatial scale. The multi-agent formalism or cellular automatons give certainly possibilities to study the various scales of a catastrophe.

The complexity results from various temporal scales of the cycle of the disasters. Three temporal phases are identified [25]. The first one is relative to the temporality of the risk as well at the level of the physical or social processes as at the level of the regional planning and prevention policies I mean what takes place before the release of the hazard. Generally these events work on a long temporality. The second phase refers to the temporality of the catastrophe I mean all what happens during the catastrophe. Mostly the temporality of the catastrophe is of short duration. But we showed that during the catastrophe, the temporalities of the hazard, the vulnerability, the domino effects and the rescue operations rarely happen together. Models of simulation showed the existence of temporal gaps between these three constituents [26]. The third and last phase refers to the time after the disaster and to the resilience of the system. It is possible to classify the systems affected by a disaster according to their resilience that is to say according to the time needed to return to the initial situation after a disturbance, for example after an environmental disorder [27]. Resilience gives us the possibility to study the catastrophe not any more from the point of view of the event but rather from the approach of the crisis. This concept of resilience requires us to apprehend the system of the catastrophe in its whole. This return time to equilibrium depends on the extent of the catastrophe and the damage, on the adaptability of the society and on the type of properties [28].

These three temporal phases are based on two scales of time: a short time, I mean a time - action, inherent to the functioning of any dynamic system [1 op.cit.] and a long time. The short time square with the rapidly developing disasters (some exceptions do exist: famine, drought are slowly developing disasters). It is also the time of the alert and of the management of the event. On the other hand the long time is the time of the prevention, the anticipation of the consequences resulting from these phenomena, the experience feedback. It can also be the time of the processes. For example, during an earthquake, the energy accumulates slowly in the fault networks (long time), before being brutally released (short time). Long phases of stability are juxtaposed to intense and brief fluctuations [29].

Furthermore, there are temporal relations not only between the various phases of the catastrophe cycle but also between the entities which constitute

every phases of the cycle. Some events depend on the occurrence of other events (for example, vulnerability of the population, buildings etc. depend on the occurrence of the hazard, prevention measures are being often established after the disaster etc.), on some states of the system (the alarm system mobilize the evacuation forces).

It is thus necessary to study the cycle of the catastrophe in its various phases: the long time of the processes, the effects of threshold at the origin of the release of the hazard, the time of the catastrophic event and that of "after catastrophe". The study in each of the scales gives some information about the whole catastrophe, or about some of its components (hazard, vulnerability and domino effects). Let us take an example: the floods of the Seine and the risk of inundation of Paris. The hazard can be studied on the scale of drainage basin (study of the precipitations on the whole basin and determination of the stream flow of the river, i.e. hazard), the catastrophe on the scale of the city (flooded surface and vulnerability of the population = vulnerability) and of the country (impact of the flood on the economy = domino effects) and the resilience on the city, national and world scale. Each of these scales is interesting. They offer different information on the catastrophe. Thus it is necessary in a modelling approach of phenomena to clarify the levels of observation and modelling [30].

5 Conclusion

In this article, two aspects of complexities were more particularly studied: the first one concerned the concepts of organization, self-organization and the emergence of phenomena; the second one is relative to the multi-scales aspects of the catastrophe. This proposal to classify the catastrophes according to criteria of complexity is certainly imperfect and must be further refined. However this classification leads to the conclusion that you have to apprehend the disaster on a different way, that is to say to exceed the disciplinary approaches and to compare catastrophes one with the other. The natural, technological or social catastrophes have similarities which can be identified with the help of the sciences of the complexity. It leads us to compare catastrophes of different origins, different impacts (in terms of human and financial losses), different duration, different required measures and different territories. It would not be in conflict with the already existing classifications, but so we can apprehend the catastrophes in another way, taking into account all the complex aspects. It enables us answering following question: is one system of catastrophe more complex than the other? It gives the possibility of more differentiated answers for the spatial and temporal management of the catastrophe.

References

1. Péguy C.-P.: Espace, temps, complexité - Vers une métagéographie, Belin, Paris (2001)
2. Morin, E.: Introduction à la Pensée Complexe. ESF Editeur, Paris (1990)
3. Le Moigne, J.L.: La modélisation des systèmes complexes, Dunod, Paris (2002)
4. Casti, J.-L.: Complexification. Explaining a Paradoxical World through the Science of Surprise. HarperCollins, New-York (1994)
5. Benkirane, R.: La complexité, vertiges et promesses. Le Pommier, Paris (2002)
6. Dauphiné, A.: Les théories de la complexité chez les géographes. Economica, Paris (2003)
7. Durand, D.: La systémique. Presses universitaires de France, Paris (1979)
8. Favre, R.: L'homme et les catastrophes. France-Sélections, Aubervilliers (1992)
9. Noto, R., Huguenard, P., Larcan, A.: Médecine de catastrophe, Masson, Paris (1994)
10. Provitolo, D.: Structural and dynamic complexities of risk and catastrophe systems: an approach by system dynamics modelling. In: Actes du colloque international The European Simulation and Modelling conference, Toulouse (2006), http://www-lih.univ-lehavre.fr/~bertelle/cossombook/complex22.pdf
11. Provitolo, D.: Vulnérabilité aux inondations méditerranéennes en milieu urbain: une nouvelle démarche géographique. Annales de Géographie 653, 23–40 (2007)
12. Deneubourg, J.-L.: Emergence et insectes sociaux. In: Benkirane, R. (ed.) La complexité, vertiges et promesses, Le Pommier, Paris (2002)
13. Bak, P.: How Nature Works. Springer, New York (1996)
14. Kauffman, S.: At home in the universe: the search for laws of self-organization and complexity. Oxford University Press, New York (1995)
15. Byrne, D.: Complexity theory and the social sciences: an introduction. Routledge, New York (1998)
16. Gribbin, J.: Deep Simplicity. Chaos, Complexity and the Emergence of Life. Allen Lane, Penguin (2004)
17. Dupuy, J.P.: La panique. Les empêcheurs de penser en rond, Paris (1991)
18. Quarantelli, E.L.: The Nature and Conditions of Panic. The American Journal of Sociology 9(3), 267–275 (1954)
19. Le Bon, G.: Psychologie des foules. Presses universitaires de France, Paris (2003)
20. Provitolo, D.: Un exemple d'effets de dominos: la panique dans les catastrophes urbaines. Cybergéo 328 (2005), http://www.cybergeo.eu/index2991.html
21. Von Bertallanfy, L.: General System Theory: Foundations, Development, Applications. Georges Braziller Inc., New York (1968)
22. Jeudy, H.P.: La panique. Editions Galilée, Paris (1981)
23. Granovetter, M.: Threshold Models of Collective Behavior. The American Journal of Sociology 83(6), 1420–1443 (1978)
24. Reghezza, M.: Réflexions sur la vulnérabilité métropolitaine: le cas de la métropole francilienne face au risque de crue centennale. Thèse de l'Université Paris X (2006)
25. Provitolo, D.: Les différentes formes de complexité des systèmes de risque et de catastrophe. In: Actes du colloque international Catastrophes, discontinuités, ruptures, limites, frontières, Rochebrune, Megève (2007)

26. Provitolo, D.: Modélisation et simulation de catastrophe urbaine: le couplage de l'aléa et de la vulnérabilité. In: Actes du colloque SIRNAT, La Prévention des Risques Naturels, Orléans (2003), http://www.brgm.fr/divers/sirnatActesColl.htm
27. Walker, B., Holling, C.S., Carpenter, R., Kinzig, A.: Resilience, Adaptability and Transformability in Social-ecological systems. Ecology and Society 9(2), 5 (2004), http://www.ecologyandsociety.org/vol9/iss2/art5
28. Dauphiné, A., Provitolo, D.: La résilience: un concept pour la gestion des risques. Annales de Géographie 654, 115–124 (2007)
29. Dauphiné, A., Provitolo, D.: Les catastrophes et la théorie des systèmes auto organisés critiques. In: Moriniaux, V. (ed.) Les risques, Éditions du Temps, Nantes (2003), http://www.edutemps.fr/extrait/EX1risques.pdf
30. Sanders, L.: Agent Models in Urban Geography. In: Agent-Based Modelling and Simulation in the Social and Human Sciences. Bardwell Press, Oxford (2007)

Community Swarm Optimization

Rawan Ghnemat[1], Cyrille Bertelle[1], and Gérard H.E. Duchamp[2]

[1] LITIS - University of Le Havre,
25 rue Philippe Lebon - BP 540
76058 Le Havre cedex, France
rawan.ghnemat@litislab.eu
cyrille.bertelle@litislab.eu
[2] LIPN - University of Paris XIII,
99 avenue Jean-Baptiste Clément
93430 Villetaneuse, France
ghed@lipn.univ-paris13.fr

Summary. The development of distributed computations and complex systems modelling [11] leads to the creation of innovative algorithms based on interacting virtual entities, specifically for optimisation purposes within complex phenomena. Particule Swarm Optimisation (PSO) and Ant Colony Optimisation (ACO) are two of these algorithms. We propose in this paper a method called Community Swarm Optimisation (CSO). This method is based on more sophisticated entities which are defined by behavioral automata. This algorithm leads to the emergence of the solution by the co-evolution of their behavioral and spatial characteristics. This method is suitable for urban management, in order to improve the understanding of the individual behaviors over the emergent urban organizations.

Keywords: swarm optimisation, community dectection, self-organization, automata, evolutive methods, geographic systems, ressource management.

1 Introduction

Artificial complex systems allow to implement distributed problems solving. Such systems are composed of interacting entities from where emergent properties appear. We focus, in this paper, on the artificial swarm or population methods allowing these emergent property computations. The artificial swarms or populations move on a representation of the environment or on a representation of the space of solutions. The emergent properties of these artificial systems are the collective meaning of the system itself, according to some objective functions. In many cases, we can express the control of the system by these objective functions as an optimization problem; the optimal configuration could be expected in advance or could be computed during the evolution of the system in order to some adaptive properties.

In the following, we will present some of these modern methods concerning artificial swarm intelligence and we will propose a new one, called Community Swarm Optimization. This method is mainly based on the concept of spatial evolutive populations of behavioral entities. The concept of community is the basic property of this method.

Definition 1. (Community operational definition)
A community is a system or an organization which is characterized by a spatial property, a behavior property and the interaction between both.

Example 1. In ecology, a community is a group of plants or animals living in a specific region and interacting with one another.

Example 2. The spatial patterns generated by Schelling's segregation models [13] are some examples of communities and these spatial patterns are linked with some elementary behavioral rules implemented for each grid case. These rules describe, for each step, the movement of each individual according to its neighborhood.

In SCO method, we need to represent an efficient way to describe the behavior of each entity and we use algebraic structures called automata with multiplicities [14]. The main advantage of these automata is to be associated with algebraic operators leading to automatic computation. With these operators, we can define behavioral distances for the entities modelled with these automata. The behavioral distance is one of the major keys of this new method. Section 3 describes the algebraic basis for the automata management used in this method. In section 4, we describe the proposed method and in section 5, we discuss some applications which can be efficiently modelled by this method, according to their own complexity.

2 Swarm Optimisation Methods

Decentralized algorithms have been implemented for many years for various purposes. In this algorithm category, multi-agent systems can be considered as generic methods [17]. Agent-based programming deals with two main categories of agent concepts: cognitive agents and reactive agents. The first category concerns sophisticated entities able to integrate, for example, knowledge basis or communications systems. Generally, efficient computations, based on these cognitive architectures, implement few agents. The second category of agents, based on reactive architecture, is expected to be used inside numerous entity-based systems. The goals of programs using such architectures, is to deal with emergent organizations using specific algorithms called emergent computing algorithms. Swarm Intelligence is the terminology used to point out such reactive agent-based methods where each entity is built with the

same basis of behavior, but reacts in autonoumous way. Swarm Optimization methods concern the problems of optimization where the computation of a function extramum is based on the concept of swarm intelligence.

Ant Colony Optimization (ACO) methods [3] is a bio-inpirated method family where the basic entities are virtual ants which cooperate to find the solution of graph-based problems, like network routing problems, for example. Using indirect communications, based on pheromon deposites over the environment (here a graph), the virtual ants react in elementary way by a probabilistic choice of path weighted with two coefficients, one comes from the problem heuristic and the other represent the pheromon rate deposit by all the ants until now. The feed-back process of the whole system over the entities is modelled by the pheromon action on the ants themselves.

Particule Swarm Optimization (PSO) is a metaheuristic method initially proposed by J. Kennedy and R. Ebenhart [10].This method is initialized with a virtual particle set which can move over the space of the solutions corresponding to a specific optimization problem. The method can be considered like an extension of a bird flocking model, like the BOIDS simulation from C.W. Reynolds [12]. In PSO algorithm, each virtual particle moves according to its current velocity, its best previous position and the best position obtained from the particles of its neighborhood. The feed-back process of the whole system over the entities is modelled by the storage of this two best positions as the result of communications between the system entities.

Other swarm optimization methods have been developped like Artificial Immune Systems [5] which is based on the metaphor of immune system as a collective intelligence process. F. Schweitzer proposes also a generic method based on distributed agents, using approaches of statistical many-particle physics [15].

The method proposed in this chapter, is called Communities Swarm Optimization (CSO) and it consists in the co-evolving of both the spatial coordinates and the behavior of each individual of a virtual population of automata. The feed-back process of the whole system over the entities is modelled by a genetic algorithm based on this co-evolving. The automata behaviors allow to define for each individual, a set of arbitrary complex transition rules. We develop the formalism needed to describe this method and the associated algorithm in the two next sections.

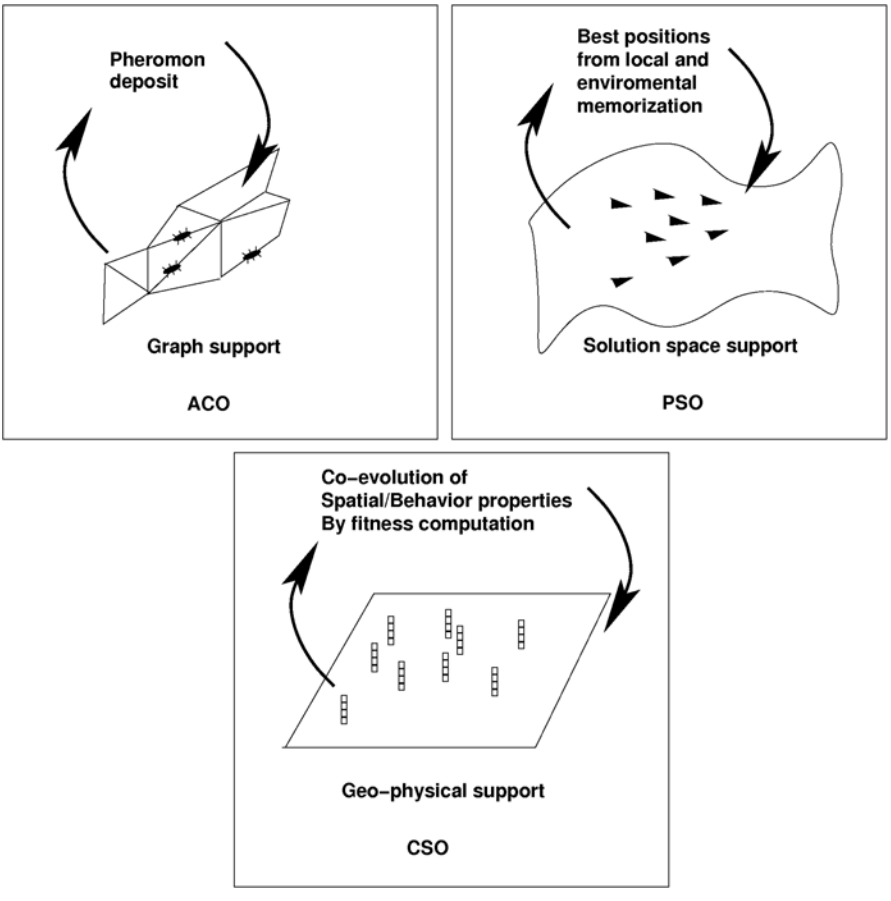


Fig. 1. Support and feed-back comparison from Ant Colony Optilization (ACO), Particule Swarm Optimization (PSO) and Community Swarm Optimization (CSO)

3 Spatial Behavioral Automata

3.1 Behavior modelling using automata

An automaton with multiplicities is an automaton with output values belonging to a specific algebraic structure, a semiring, including real, complex, probabilistic, non commutative semantic outputs (transducers) [8,16]. In this way, we are able to build effective operations on such automata, using the properties of the algebraic structures which belong the output data. We are specifically able to describe automata by means of a matrix representation with all the power of the new (i.e. with semirings) linear algebra.

Definition 2. (Automaton with multiplicities)
An automaton with multiplicities over an alphabet A and a semiring K is the 5-uple (A, Q, I, T, F) where

- $Q = \{S_1, S_2 \cdots S_n\}$ *is the (finite) set of states;*
- $I : Q \mapsto K$ *is a function on the set of states, which associates to each initial state a value in K, called entry cost, and to each non- initial state a zero value ;*
- $F : Q \mapsto K$ *is a function on the set states, which associates to each final state a value of K, called final cost, and to each non-final state a zero value;*
- T *is the transition function, that is* $T : Q \times A \times Q \mapsto K$ *which to a state* S_i*, a letter a and a state* S_j *associates a value z of K (the cost of the transition) if it exist a transition labelled with a from the state* S_i *to the state* S_j *and and zero otherwise.*

Remark 1. We have not yet, on purpose, defined what a semiring is. Roughly it is the least structure which allows the matrix "calculus" with unit (one can think of a ring without the "minus" operation). The previous automata with multiplicities can be, equivalently, expressed by a matrix representation which is a triplet

- $\lambda \in K^{1\times Q}$ which is a row-vector which coefficients are $\lambda_i = I(S_i)$,
- $\gamma \in K^{Q\times 1}$ is a column-vector which coefficients are $\gamma_i = F(S_i)$,
- $\mu : A^* \mapsto K^{Q\times Q}$ is a morphism of monoids (indeed $K^{Q\times Q}$ is endowed with the product of matrices) such that the coefficient on the q_ith row and q_jth column of $\mu(a)$ is $T(q_i, a, q_j)$

Figure 2 describes the linear representation of a probabilistic automaton which is a specific automaton where output values are probabilistic values. For these probabilistic automata, the sum of the coefficients of each matrix row is equal to 1 (being the sum of outgoing and loop probability).

Definition 3. (Automata-Based Agent Behavior)
We represent the agent behavior by automata with multiplicities (A, Q, I, T, F) *over a semiring K:*

- *The agent behavior is composed of a states set Q and of rule-based transitions between them. These transitions are represented by T; I and F represent the initial and final costs;*
- *Alphabet A corresponds to the agent perceptions set;*
- *The semiring K is the set of agent actions, possibly associated to a probabilistic value which is the action realization probability (as defined in [6]).*

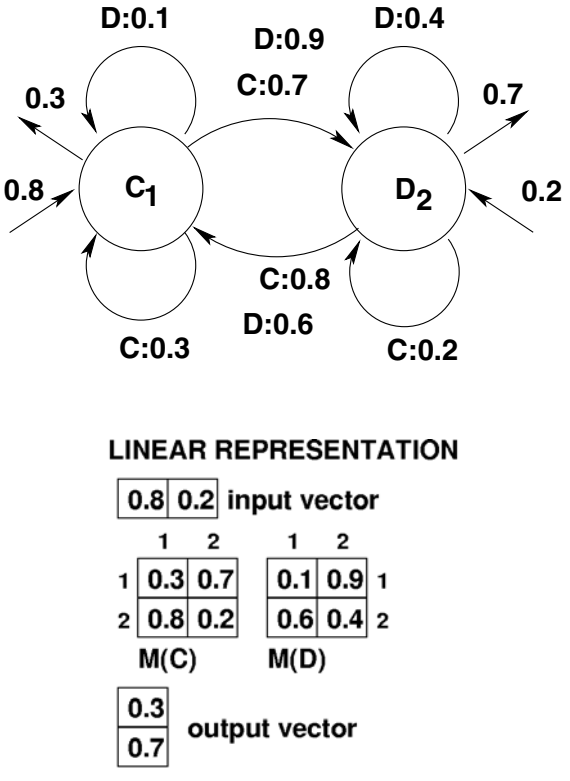


Fig. 2. Probabilistic automata over the alphabet {C,D} and its linear representation

3.2 Spatial Automata and associated spatial distance

Definition 4. (Spatial Automata-Based Agent)
A spatial automata-based agent is defined by its structural representation:

- *An automaton with multiplicities corresponding to its behavior as a whole processus managing its perceptions and its actions over its environment. They include its communication capabilities and so its social behavior;*
- *A spatial location defined on some specific metric space.*

Remark 2. According with this previous definition, we define two metrics on the spatial automata-based agent. The first one concerns a spatial distance which is directly induced by the metrics of the spatial location (from any standard Hölder norm). The second one is more innovative and concerns a behavioral distance or semi-distance. One of the major interest of the previous automata-based modelling is to be able to define such behavioral distance which leads to powerful automatic processes dealing with self-organization. We detail this definition in the next section.

3.3 Metric spaces for behavioral distances

The main advantage of automata-based agent modelling is their efficient operators. We deal is this paragraph with an innovative way to define behavioral semi-distance as the essential key of the swarm algorithm proposed later.

Definition 5. (Evaluation function for automata-based behavior) *Let x an agent whom behavior is defined by $\mathcal{A}$, an automaton with multiplicities over the semiring K. We define the evaluation function $e(x)$ by:*

$$e(x) = V(\mathcal{A})$$

where $V(\mathcal{A})$ stands for the vector of all coefficients of (λ, μ, γ), the linear representation of $\mathcal{A}$, defined in remark 1.

Definition 6. (Behavioral distance) *Let x and y two agents and $e(x)$ and $e(y)$ their respective evaluations as described in the previous definition 5. We define $d(x, y)$ a distance or metrics between the two agents x and y as*

$$d(x, y) = ||e(x) - e(y)||$$

a vector norm of the difference of their evaluations.

3.4 Genetic operators on spatial automata-based agent

We consider in the following, a population of spatial automata-based agents, each of them is represented by a chromosome, following the genetic algorithm basis. We define the chromosome for each spatial automata-based agent as a couple of two sequences:

- the sequence of all the rows of the matrices of the linear representation of the automata. The matrices, associated to each letter from the alphabet of the agent perceptions, are linearly ordered by this alphabet and we order all the rows following these matrices order [2]. The figure 3 describes how this sequence is created from a linear representation of two matrices;
- the sequence of all its spatial coordinates.

In the following, genetic algorithms are going to generate new automata containing possibly new transitions from the ones included in the initial automata.

The genetic algorithm over the population of spatial automata-based agent follows an evolutive iteration composed of two main operators, as on adaptation of the classical genetic operators [9]:

- *Reproduction (Duplication and Crossing-over)*: This operator is a combination of the standard duplication and crossing-over genetic operators. For each couple of spatial automata (called the parents), we generate two

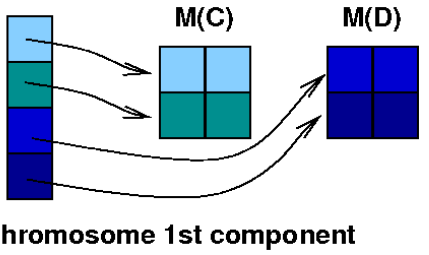


Fig. 3. Chromosome first component building from the matrix rows of the linear representation of an automaton over the alphabet {C,D}

new spatial automata (called the children) as the result of the chromosome crossings and we keep, without change, the parent spatial automata. To operate for the crossing-over operation, we have to compute

- the automata of the behaviors of the two children. For this purpose, we consider a sequence of rows for each matrix of the linear representation of one of the two parents and we make a permutation on these chosen sequences of rows between the analogue matrix rows of the other parent;
- the spatial locations of the two children. These children locations can be choosen from many ways: on the linear segment defined by the parent locations or as the nodes of the square obtained with the parent and the children as describe in the figure 4.

- *Mutation*: This operator deals only with the linear representation of the spatial automata-based agent. With a low probability, each matrix row from this linear representation is randomly chosen and a sequence of new values is given for this row (respecting some constraints if exist, like probabilistic values [2]).

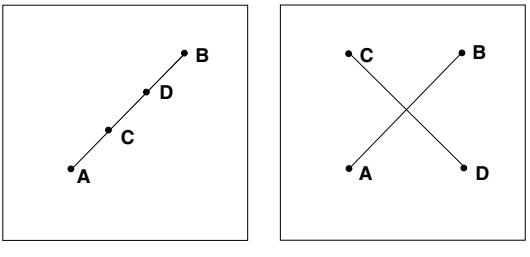


Fig. 4. Spatial locations for the children C and D from the parents A and B, after a reproduction step. Two possible locations computation are presented in the two sub-figures.

4 Community Swarm Optimization Algorithm

4.1 Adaptive objective function for community-based swarm optimization

The community swarm optimization method is based on a genetic algorithm over a population of spatial automata-based agents. The formation of the community is the result of the population evolution, crossing by a selection process computed with the fitness function defined in the following.

For this computation, we deal with two distances defined on agent sets. The first is the spatial distance associated with the agent spatial location and the second is the behavioral semi-distance defined in the definition 6.

Definition 7. Community clustering and detection fitness
Let $\mathcal{V}_x$ a neighbourhood of the agent x, relative to its spatial location. We define $f(x)$ the agent fitness of the agent x as :

$$f(x) = \begin{cases} \dfrac{card(\mathcal{V}_x)}{\sum\limits_{y_i \in \mathcal{V}_x} d(x, y_i)^2} & \text{if } \sum\limits_{y_i \in \mathcal{V}_x} d(x, y_i)^2 \neq 0 \\ \infty & \text{otherwise} \end{cases}$$

where $d(x, y)$ is the behavioral distance between the two agents x and y.

This fitness allows to implement a co-evolution process which generates an emergent set of community swarms. These community swarm formation is the result of both the emergence of the spatial location of the generated communities and the adaptive behavior of the communities as the result of the homogeneisation of the behavioral automata of all agents which compose these emergent communities.

4.2 General CSO algorithm

CSO algorithm needs a initial step description which is the major step of the modelling process. The way of going from the problem formulation to the initial spatial automata-based agents must be realized in accuracity. The formal description of the methodology to use, for this initial step, is described in Algorithm 1.
The core of the CSO algorithm is described by the iterative scheme defined in the Algorithm 2.

An example of fitness function computation output is illustrated by the figure 5 where we show, on the same population, an high level fitness individual which will be probably kept inside the population at the next iteration and a

Algorithm 1: Methodology to model the initial step of CSO

1. Problem formulation by the definition of a set of transition rules ;
2. Building of the behavioral automata based on the previous set of transition rules, describing the sequences and the context of their applications ;
3. Discretization of the spatial domain, according to its topological properties (Cellular automaton, network or graph, Geographical Information System) with the spatial location of the initial virtual population of spatial automata-based agents;

Algorithm 2: Iteratice scheme of CSO

Building the initial virtual population of the spatial automata-based agents (following the methodology of Algorithm (1)) ;
repeat
  for *Each couple of individuals in the population* **do**
    Reproduction step generating 2 new children as described in the section (3.4) ;
    Mutation step as described in the section (3.4) ;
  Selection of the half population of the individuals corresponding to the highest values of the agent fitness described in section (4.1) ;
until *(the sum of the fitness values of the whole population reachs a threshold) or (the maximum iteration number is reached)* ;

low level fitness individual which be probably removed from the population at the next iteration. The colors used describe graphically the chromosome composition, allows to appreciate the similarity of the individuals.

5 Conclusion and Perspectives

In this paper, we describe Community Swarm Optimization (CSO) method which can be described as a swarm optimization process. With the comparison of other methods from this category (ACO and PSO), CSO differs mainly on the modelling purpose. CSO deals with transition rules included in data structures (automata with multiplicities) for which algebraic operators allow to implement automatic computation for self-organizational phenomena.

Presented as a very generic method in this paper, CSO can be efficiently applied in engineering problems where the spatial characteristics are not only additional coordinates for the data but the spatial organization is the result of the self-organization process as the output of the method. Typically, urban and territorial management need this kind of modelling to improve the decision making within a sustainable development (including environmental, economic and social apects).

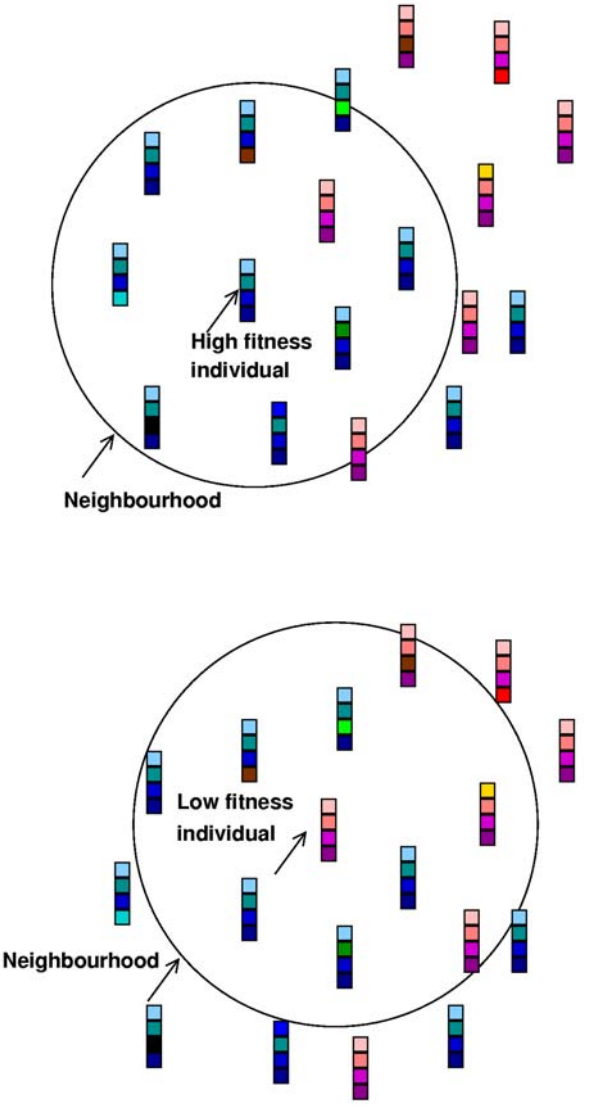


Fig. 5. Two examples of the result of selection operators within the CSO algorithm

Schelling's Segregation model [13] is one of the basic problem dealing with spatial self-organizations within urban or territorial management. In this model, the spatial domain, based on cellular automata, is constrained and the rules are elementary. Another example concerns partitionning problem for territorial management which can be computed by genetic algorithms [18], using objective functions with various parameters to adjust. Following Benenson and Torrens [1], we have to deal nowadays with the understanding of the development of urban dynamics as collective phenomena. Models of election/voting (studying the political affiliation of citizen under the influence of their environment and neighbors), diffusion of innovation in the city (studying the acceptation or rejection of innovation from citizen depending on their environment and neighbors) are typically spatial complex systems with multi-dimensional (social, economic, ...) rule-based aspects which need accurate transition machines and efficient self-organization processes that CSO is expected to solve.

References

1. Benenson, I., Torrens, P.M.: Geosimulation - Automata-based modeling of urban phenomena. Wiley, Chichester (2004)
2. Bertelle, C., Flouret, M., Jay, V., Olivier, D., Ponty, J.-L.: Genetic algorithms on automata with multiplicities for adaptive agent behaviour in emergent organizations. In: SCI 2001, Orlando, Florida, USA, July 22-25 (2001)
3. Bonabeau, E., Dorigo, M., Theraulaz, G.: Swam Intelligence, from natural to artificial systems. In: The Santa Fe Institute Studies in the Sciences of Complexity. Oxford University Press, Oxford (1999)
4. Bourbaki, N.: Elements of Mathematics: General Topology, ch. 5-10. Springer, Heidelberg (1998)
5. De Castro, L.N., Timmis, J.: Artificial immune system: a new computational approach. Springer, London (2002)
6. Ghnemat, R., Oqeili, S., Bertelle, C., Duchamp, G.H.E.: Automata-Based Adaptive Behavior for Economic Modelling Using Game Theory. In: Aziz-Alaoui, M.A., Bertelle, C. (eds.) Emergent Properties in Natural and Artificial Dynamical Systems. Springer, Heidelberg (2006)
7. Ghnemat, R., Bertelle, C., Duchamp, G.H.E.: Adaptive Automata Community Detection and Clustering, a generic methodology. In: Proceedings of World Congress on Engineering 2007, International Conference of Computational Intelligence and Intelligent Systems, London, U.K, July 2-4, pp. 25–30 (2007)
8. Golan, J.S.: Power algebras over semirings. Kluwer Academic Publishers, Dordrecht (1999)
9. Goldberg, D.E.: Genetic Algorithms. Addison-Wesley, Reading (1989)
10. Kennedy, J., Eberhart, R.C.: Particle Swarm Optimization. In: Proceedings of IEEE International Conference on Neural Networks, Perth, Australia, vol. 5(3), pp. 1942–1948. IEEE Service Center, Piscataway (1995)
11. Le Moigne, J.-L.: La modélisation des systèmes complexes, Dunod (1999)
12. Reynolds, C.W.: Flocks, Herds and Schools: a distributed behavioral model. Computer Graphics 21(4), 25–34 (1987) (SIGGRAPH 1987 Conference Proceedings)

13. Schelling, T.C.: Dynamic Models of Segregation. Journal of Mathematical Sociology 1, 143–186 (1971)
14. Schützenberger, M.P.: On the definition of a family of automata. Information and Control 4, 245–270 (1961)
15. Schweitzer, F.: Brownian Agents and Active Particles. Springer, Heidelberg (2003)
16. Stanley, R.P.: Enumerative combinatorics. Cambridge University Press, Cambridge (1999)
17. Weiss, G. (ed.): Multiagent Systems. MIT Press, Cambridge (1999)
18. Xiao, N.: Geographic optimization using evolutionary algorithms. In: 8th International Conference on GeoComputation, University of Michigan, USA (2005)

Emergence of Growth and Structural Tendencies During Adaptive Evolution of System

Andrzej Gecow*

Institute of Paleobiology Polish Academy of Science, Warsaw Poland
gecow@twarda.pan.pl

Summary. We consider (continuing earlier article in this book) functioning directed networks like Kauffman networks or aggregate of automata which have some external outputs. A measure of fitness is defined on the external output signals. We use more than two equally probable signal variants which define our systems as chaotic (and different than RBN-familly and RNS and RWN), however they become really chaotic and complex when they cross certain complexity threshold during their growth. Above this threshold adaptive condition describing Darwinian mechanism creates some structural effects which we name 'structural tendencies' and investigate in this paper. We found mechanisms connected to damage spreading which lead to these tendencies. Different network types are considered including random Erdős-Rényi and BA scale free networks. Our model describes a wide range of systems like living objects and human-designed systems because they all grow under adaptive condition and are complex. The structural tendencies are therefore commonly met, however they are known only on an intuitive level. Old classical regularities of ontogeny evolution like terminal modifications and terminal additions are such tendencies. They are contemporarily forgotten and even negated due to lack of their explanations which we propose in this paper. One of such tendencies is the growth of the system.

1 Introduction

This paper is the second and the main part in the path we develop to describe structural tendencies in adaptive evolution of functioning (dynamical) complex networks. The first part also placed in this book concerns complexity threshold, chaos, a special algorithm and other main notions which we will use here without describing them in detail for a second time. E.g. we will use the notion 'chaos' following Kauffman [21] for damage behaviour in finite networks. It is a different meaning than the commonly used one [30].

* Research financed by government grant.

The structural tendencies are a new area in complex networks. Our 'structure' is a statistical nonuniformity in variables defined basing on network construction. It is neither a lattice shape or dimension nor particular connections type as in [23]. We use fully random connections of nodes. This area contains a wide range of generally meeting phenomena, but these were not investigated on a scientific level up till now. A few of them have been discussed in classic biology more than century ago, but at that time there were no tools to solve the problem of their occurrence and now they are either forgotten or negated [15, 28, 36, 11] due to lack of their explanation. The most famous and the most controversial of them is Haeckel's 'biogenetic law' [16,17]. It is followed by Weismann's 'terminal additions' [35] and Naef's 'terminal modifications' [26]. However, the classic biology problem was only the first signal to focus our attention - most of human-designed systems, maybe even all of them, are complex, functioning networks which evolved under adaptive condition, therefore we encounter effects of structural tendencies on each step. A phenomenon does not 'exist' if it has no name, therefore it is important to name it and its environment (to establish a set of notions for the description), and this is one of a main goals of our path. The second goal is to show mechanisms leading to these phenomena.

We estimate that the typical complex, functioning and adapted system is chaotic and this is one of the basic assumptions of our research path. This is to be discussed basing on interpretation, we introduce some intuitive arguments in our previous article in this book. We also conclude there, that the number s of equally probable variants of signal in such models should be greater than two, but most of the models use $s = 2$. This can lead to very inadequate results. The Kauffman approach uses Boolean networks [19,21] and focuses attention on the boundary between order and chaos where 'homeostatic stability' plays an important role. This 'homeostatic' notion, however, has a meaning which is different than the commonly used one and which is based on negative feedbacks. Such typically homeostatic features were named 'ultrastability' by Kauffman. Although Kauffman's gene regulatory model [20, 33] works mostly in ordered regime - this model well fits the experimental data [34, 31, 32] and has become a strange exception. However, neglecting role of 'ultrastability' in living objects is a large simplification.

In the previous article we used simulation to investigate damage spreading behaviour in different types of autonomous networks and networks with external outputs and for several parameters in the context of degree of chaos. We have found there a transition to chaos (to exhibition mature chaotic features) during network growth which we interpret as complexity threshold. Above this threshold we identified some mechanisms which we are going to investigate in this paper. In that article we have also defined the depth D measured from system outputs which describes place (e.g. position of node

or connection) in the network. It is the main parameter of the found mechanisms, it approximates functional order.

For these simulations we have described and used our simplified algorithm dedicated to statistical investigation. Its specific important feature is that damage avalanche stops at an equilibrium level as 'pseudo fadeout'. This yields a single, statistically correct value of damage size (instead of infinite oscillations around the equilibrium level). This simplified algorithm will be also used in this paper as a key to the investigation of the structural tendencies.

In this paper we introduce a fitness parameter based on similarity of the system output signal vector to an arbitrarily defined 'ideal vector'. In the general case a complex functioning system with feedbacks has no particular stable output signal vector (point attractor), but a periodic attractor of such outputs. It is problematic to define useful fitness in such a case. However, in case of our special algorithm, we have a particular, statistically correct output signal vector, which allows us to use the simple fitness definition given above.

The simplest adaptive condition for evolution is that fitness should not decrease. Using it we obtain adaptive evolution which we can compare to process free of such condition. We investigate such a comparison for complex systems, means: over above mentioned complexity threshold, and there we find some strong systematic differences. These differences, named 'structural tendencies' are a direct effect of the adaptive condition treated as condition (pattern) of network growth. Note that patterns of network growth are the most important element defining network features - it is the basis of network type definition like scale-free Barabási-Albert networks or random Erdős-Rényi networks.

In this chapter we limit ourselves to describing a few of the most basic structural tendencies and their mechanisms: 'terminal modifications and conservation of early areas' tendency equivalent to Naef's regularity of ontogeny evolution in biology [26], 'terminal predominance of additions over removal' tendency and 'simplification of early areas' tendency equivalent to Weismann's 'terminal additions and compression of early stages' [35] and 'similarity of historic and functional order' connected to Haeckel's 'recapitulation of phylogeny in ontogeny' [16, 17, 11, 8] named also the 'biogenetic law'.

Especially this 'biogenetic law has always been controversial and at present seems to be definitely discredited (Richardson et al. 1997) [28]' - this sentence found in [27] is typical of contemporary biologists, however this theme is typically omitted in writing. 'It is a taboo theme' remarked Gould [15]. It is because the hopes for its exactness were too big and there was lack of other explanation when the first proposition of explanation occurred contradictory with genetics. Practically no discussion exists, new data are not referred to this theme only new extremely negative messages [28, 36] or old ideas and

opinions inadequate to current level of knowledge (especially of system science) are repeated (e.g. de Beer's repetitions conception [4]).

Our investigation of structural tendencies explains outside of biology all above mentioned biological regularities including recapitulation, moreover, it expands their applications. Now we should expect those tendencies not only in biology, but also in human-designed systems. In biology they should be expected in all of the more complex developmental processes similar to 'ontogeny' and in complex body structure and function of 'phenotype', not only of ontogeny of vertebrates or multicellural animals. E.g. in complex networks of metabolisms, even in the development of one-cell protista like Foraminifera. This last example is estimated as correct by Mikhalevitch basing on her works [25, 24]. Weismann's terminal additions regularity is negated together with Haeckel's recapitulation but Naef's terminal modification has currently in force de Beer's old, naive and false explanation [4], which has not been discussed for a long time.

We expect many more similar phenomena awaiting their description using our model or discovery. Some of them - 'covering' similar to caenogenetic modifications, 'specialization' and 'integration' we have announced in [10].

2 Fitness and Adaptive Condition

2.1 Fitness and Adaptive Condition Definitions

The most important element introduced in this article in comparison to the preceding one is fitness parameter and its application to construction of an adaptive evolution model by defining the adaptive condition. Kauffman [21] defines fitness of system using states of all system elements. These elements - nodes constitute the network. Kauffman consider autonomous random Erdős-Rényi, directed, functioning network. Signals in such a network have two variants, therefore it is named Boolean. Kauffman assumes, that these two variants are equally probable. As we discuss in our first article in this book we use more than two equally probable signal variants, therefore our networks are not Boolean networks and we cannot expect Kauffman's 'homeostatic stability' which occurs in ordered networks. Our networks are therefore potentially chaotic, however they must mature during growth and cross some complexity threshold, discussed in previous article, to be indeed chaotic.

We treat fitness as a parameter which assesses the system in the long period using its effects obtained outside, therefore we use system outputs to define fitness, not internal state of nodes like in the Kauffman model. In this chapter we assume stable system input signals as an equivalent of statistical stability of this part of environment which sends signals to object but not the one which assesses it.

In the general case a complex system with feedbacks has no single stable output signal vector which we can use to define fitness but only a periodic attractor of such outputs. However, in case of our special algorithm which we have described in previous chapter, the process stops when it reaches the damage equilibrium level and we have a particular output signal vector. Its damage size L parameter (number of damaged output signals) is statistically correct. We define fitness b of system using such an output signal vector as this vector's similarity to some arbitrarily defined 'ideal vector'. The simplest fitness b definition is a number of identical signal variants (in short - signals) in both vectors. We also use a more complex definition, but only incidentally to *aa* networks (see ch.3.4) and now we limit ourselves to this simplest one. As earlier, we fix the number $m = 64$ of system outputs in our investigation.

The simplest adaptive condition a for evolution is that fitness should not decrease. We denote adaptive condition $a \equiv b_{t+1} \geq b_t$ where t is the time sequence of changes in an adaptive evolution process which consists only of changes meeting this condition. This is different time parameter than used in previous article. We compare the adaptive evolution process to a process free of such condition which we name as the free process.

We consider damage spreading in the first article and here. We named it 'damage' assuming that any large change in any system typically destroys this system. Now we ask for the basis of this assumption or estimation. Terms 'damage' or 'destroy' are defined based on some fitness, which in the first article was not yet defined and up to this point the name 'damage' has no model valorisation as wrong or good. Now we consider changes as candidates to acceptance by adaptive condition. Such a change, if it is accepted, is not 'wrong' by definition and the term 'damage' is no longer adequate. Therefore we will use 'change size' term for L instead of 'damage size'.

In comparison to Kauffman model our fitness is very simple - it has no local extremes and this problem, a basic one in the Kauffman model, does not exist in our model together with Kauffman's complexity catastrophe [21]. It does not mean that we suggest that the problem of local extremes does not exist in modelled reality. Instead it means that we investigate other problems. We focus on the results of adaptive condition in adaptive evolution for high fitness but when the gradient of fitness is distinct, i.e. not in stasis. This is the proper state of evolution [6] when new changes are collected (i.e. accepted and this way they increase evolutionary chain of changes). Typically we fix fitness $b = 48$, more generally to 3/4 of the maximum value. We construct a stable value of b by changing the ideal of output signal in a random way. If we do not do it, the evolution will reach the maximum value of $b = m$ in a short time due to lack of local maxima. We estimate that the problem of local optima and the need for long jumps are not so statistically important

due to evolution of fitness landscape and therefore a 'greedy' algorithm is sufficient for the investigated problems.

2.2 Simple Recognition for Intuition

In the free process the fitness b converges to a point m/s (fig.1.3), for which there is the maximum of probability (fig.1.1) and entropy but in adaptive evolution fitness b should statistically grow up to m, which can be derived directly from the adaptive condition. For higher b, the probability $P(a|L,b)$ is lower (fig.2.1), and we expect the growth to become more 'expensive' (more trials before one is accepted), it is also slower growth because average accepted change is smaller but we can show this for a particular distribution of $P(L)$.

To build some intuition, we may assume a certain $P(L)$ - the most natural at this stage of model development is probably the assumption that each signal of the assessed vector has an identical and independent probability of change. This assumption gives a binomial distribution of $P(L)$ which is depicted in fig.1.1 for value $1/4$ of signal change probability. It was used for the rest of calculation depicted in fig.1. Fig.1 is only an example but it describes our intuition and expectation well. Later, from a model with network whose changes will define the assessed vector we will obtain another distribution $P(L)$ (see fig.2.3) but it will be an effect of some particular mechanism (described in previous article).

Using the assumed distribution $P(L)$ and this very simple model we have calculated an average history of fitness b in free and adaptive processes shown in fig.1.2. The speed of growth of fitness b is the average increase of b (Δb) shown in dependency of b in fig.1.3 in logarithmic scale. This is the main difference between Δb distribution for free and adaptive processes - not only the value (speed) but also the sign (direction). In fig.1.4 $P(b_{t+1}|b_t)$ for free process is shown but here, a part, which is formed by improvement cases - cases accepted by adaptive condition - is noted with grey. We can see, how small this part is for $b_t = 32$ (but still visible) and for $b_t = 48$, where it is too small even to be visible. The remaining great white part is formed by damage cases - cases of decrease of fitness, which are removed in adaptive evolution by the adaptive condition. (Remember, that this example was calculated using an assumed $P(L)$, different from the one in fig.2.3.)

2.3 Tendency as Difference in Distribution in Adaptive and Free Processes

We have two processes - a free one and an adaptive one - based on the same mechanism of random changing. In the adaptive process all changes which

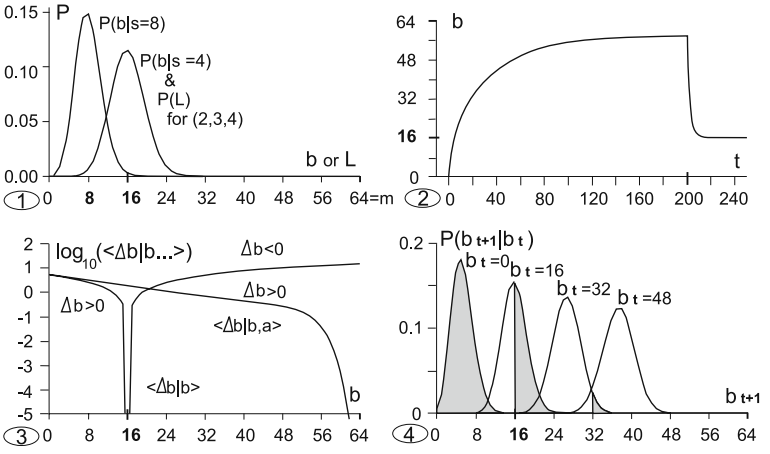


Fig. 1. Example (for $s = 4$) of average fitness history (2) and its analysis calculated for assumed $P(L)$ (1) - equal and independent probability of signal change = 1/4. Fitness grows much faster when it is low (2), dependency of average fitness increase Δb on fitness level is shown in (3) for adaptive and free processes. When in (2) adaptive condition was disconnected at $t = 200$, adaptive evolution ends and free process starts, then fitness drops down rapidly to a level m/s, here =16. It is a stable equilibrium in the free process visible in (3) and (1) where curve for $s = 8$ is also shown. Note that the scale of (3) is logarithmic to show how slow the improvement of a system with high fitness is. Other aspect is how expensive this improvement is - how many random trials are missing, this is shown in (4), where the grey part of area under the curve is formed by the improvement cases and the white area - the damage cases.

do not meet the adaptive condition are rejected and they do not form the process. We ask: what is the difference caused by adaptive condition in distribution $P(X)$ and $P(X|a)$ of change parameter X? We call this difference a tendency. In this paper we are looking for such tendencies. Note, X must describe a change; it cannot be a state parameter.

From Bayes: $P(a)P(X|a) = P(a|X)P(X)$ Because $P(a)$ is not dependent on X, i.e. is constant, therefore the tendency is shown by: $P(a|X)$. We do not have to know $P(X)$ to know a tendency. It is enough that for different X, $P(a|X)$ is different. E.g. this is the easiest way to show the obvious 'fitness b growth tendency': $P(a|\Delta b)$ depends on Δb because $a \equiv \Delta b \geq 0$. However, in the model of structure development of complex system the choice of change distribution in a free process is important due to other causes.

Let us introduce a general parameter g of process advancement connected to the t. It is the state describing parameter, e.g. it can be t or b. Similarly as

above we can find, that tendency is described by $P(a|X, g)$. We will use this form later, e.g. for small change tendency $P(a|L, b)$ in the next section.

2.4 Small Change Tendency

A simple but very important one is the tendency to collect much smaller changes in adaptive process than changes creating the free process. We named it 'small change tendency'. It is the base of mechanisms of all the structural tendencies, which we find more interesting and investigate later. This small change tendency is not a structural one because it does not need any structure of the system. It only requires the output signal vector and the ideal vector as they are necessary to define fitness b. Output signal vectors can be interpreted as a property of the whole system.

Assuming that each signal in this vector changes independently, we can calculate $P(a|L, b)$ for given s. We use here the assumption that all s variants of signal are equally probable. This $P(a|L, b)$ is the form discussed above, which indicates the tendency ($P(a|X, g)$) if it really depends on L. Fitness b describes here the advancement of process - we show above the obvious 'fitness b growth tendency'. For intuition look at fig.1.2. For higher b (we assume above $b = 48$ for $m = 64$) only very small changes are acceptable (smaller than $L = 8$, but not all of them). If we compare it to $P(L)$ distribution for networks upon complexity threshold (fig.2.3.a) only the changes forming first, left peak can form adaptive evolution process, all the changes from the right peak cannot be accepted by the adaptive condition test. From discussion in the first article we know that the first peak is created by real fade out of damage in the first few steps. Therefore anything which can help damage to fade out in these first steps will create structural tendencies.

Here we can come back to the background of the term 'damage'. Large change, i.e. in right peak in fig.2.3.a, must be destructive, i.e. not meet adaptive condition (fig.2.2). In a typical system a change, i.e. damage, grows, i.e. spreads, because a typical system is chaotic, i.e. coefficient of damage propagation $w > 1$. However, keeping in mind this aspect and that for accepted changes the damage term is not adequate, we will use this useful term only when aspect of spreading of change is important.

The 'small change tendency' is a different view on Kauffman's 'structural stability' [21] as a important condition of adaptive evolution. The structural stability of system typically leads to small and smooth changes, only a small part of the changes cross the bifurcation 'walls'.

The 'small change tendency' is trivial and known but up till now it has not been named, which we consider necessary to 'exist' in science, and it has not

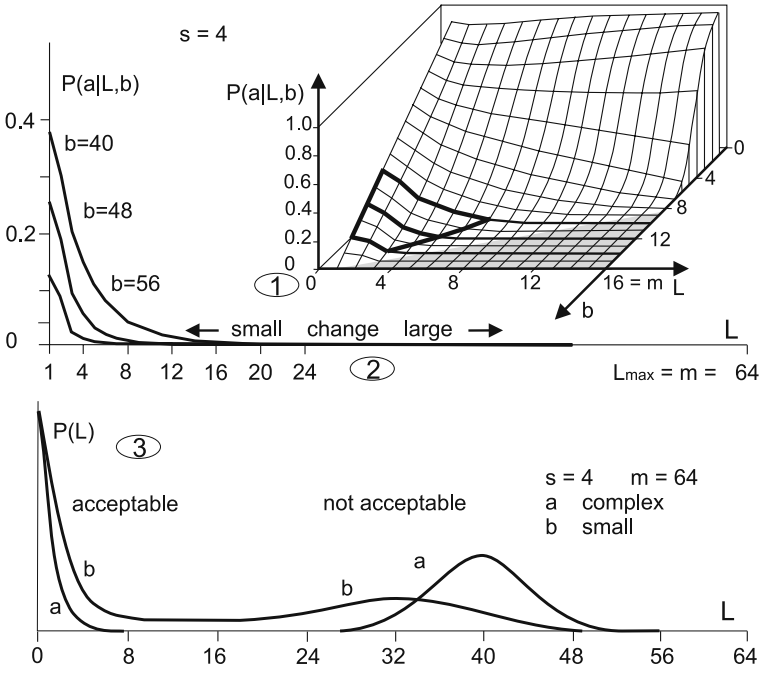


Fig. 2. Small change tendency (1,2) and complexity threshold (3). Theoretically calculated distribution of acceptance probability $P(a|L,b)$ (probability of meeting the adaptive condition) for $s = 4$ in relation to fitness b and change size L is shown in two cases. Case (1) for $m = 16$ output signals and case (2) for $m = 64$ and only for three higher values of fitness b. For interesting higher fitness the probability of acceptance is significantly different from zero only for very small change size L. Note, that for higher b and m this tendency is stronger. Comparing $P(a|L,b)$ in (2) to $P(L)$ distribution for complex networks (3.a) we see that only changes forming the left peak can form adaptive evolution process. When the network is still small (3.b) its fitness can be also smaller therefore much greater changes and much greater part of changes can form adaptive evolution.

been used adequately to its importance. It is a very basic and very important tendency. In all next investigations we will be basing on it. To explain recapitulation of phylogeny in ontogeny it has been proposed for use by Darwin [29]: 'Changes in the early stages of ontogeny cause larger phenotype changes which are typically lethal'. Now this proposition is forgotten due to lack of quantitative validation, however, in this form it only explains the terminal modifications regularity.

The 'small change tendency' also creates a natural identity criterion, which has a deeper, philosophical meaning and may be used for life definition. There is some problem, as to what does 'the same' mean for two objects in different

time and place, especially if they are not identical. It occurs when the investigated object can change. What is the subject of evolution - one object or a sequence of different objects? Our intuition agrees, that one object evolves if in a sequence of its evolutionary states there is no discontinuity i.e. no single large change. Identity of an evolving object emerges with its fitness growth in the effect of small change tendency controlled by adaptive condition.

3 Structural Tendencies - Assumptions and Expectations

3.1 Premise of Terminal Modification and Conservation of Deep Area Tendency

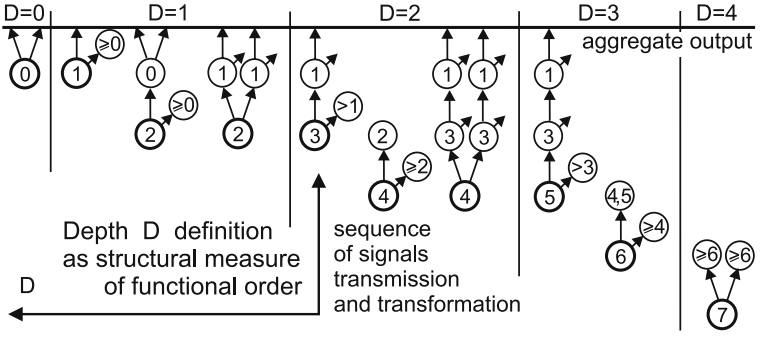


Fig. 3. Depth definition for aggregate of automata (*aa* and *an* case with $K = k = 2$). It is a structural sequential measure of functional order defined experimentally basing on similarity of acceptance probability of a change. It considers two similar ways to outputs.

In the previous article starting from the coefficient of damage propagation and the 'functional order' described as direction in directed network we developed the cone of influence structure and we discussed depth D, (i.e. height of this cone measured from system outputs down to the damage initiation point), as a structural approximation of functional order. The terms 'early', 'late' and 'terminal' were defined in connection to the functional order and depth D. Let us now consider small change tendency $P(a|L, b = 48)$ and dependency of change size $L(D)$ on depth D in the cone of influence. We can expect that changes initiated deep will give higher change size L and will be rejected by the adaptive condition. On the other hand, changes initiated shallow under outputs (late places) or closely near outputs (terminal) will create small change size L and will have a high probability of acceptance. It means we expect strong differences in $P(a|D, b = 48)$ for different D. Due

to fixed fitness $b = 48$ which well represents all cases of high fitness, we will omit the assumption of $b = 48$ in writing and use simply $P(a|D)$ (instead of $P(a|D, b = 48)$). This is a 'terminal modifications and conservation of deeper area' tendency known in classic developmental biology [26]. Depth D is an approximation of time of ontogeny stages. However, cone of influence is well defined only in system without feedbacks but if feedbacks are present, then it can only be a premise. The answer can only be given by simulation, where we should check $P(a|D)$.

In the previous article we use as depth D the shortest way to outputs measured in number of links and according to their direction i.e. functional order. This is the most convenient and general definition for D, but for many reasons it is not fully adequate. For network type aa (aggregate of automata with feedbacks) and $K = 2$ (number of inputs per node) which we have investigated first [9,10], we have used the special and more adequate definition of depth D (fig.3) in simulations presented later. It considers different ways of similar length, but it is not to be applied to various k (number of node outputs) in other networks types. In the following paragraphs, areas of the network which consist of nodes of large depth will be called 'deep areas'. We will use the term 'shallow areas' analogously.

3.2 Problem of $k < 2$ in Kauffman Networks

From famous Barabási-Albert discovery of scale-free network [3], it become clear that constant degree of node $k = 2$ used in aa is too simple a model of general real networks and we should especially investigate networks with variable number k of outgoing links per node. In such a case we must come back to the Kauffman network where we also encounter nodes of $k < 2$ e.g. $k = 1$ and $k = 0$ although the average $< k >= 2$. Distribution of node degree k is proportional to $k^{-\alpha}$ ('the power-law degree distribution') where α ranges from 2 to 3 for scale-free (sf) networks, and proportional to exponential distribution 2^{-k} for single-scale (ss) networks [1]. This means, that near half of the nodes have $k = 1$ and that damage starting from an initiation point with high probability meets nodes of $w < 1$ (coefficient of damage propagation described in detail in previous article) for any assumed s and fades out on these nodes before reaching nodes of higher k. This phenomenon does not occur in an aggregate of automata. We have discussed it more in detail in previous article. In this discussion we underline the role of hubs which decrease the average k and coefficient w for remaining node and this way increase probability of an 'real fadeout'.

Note that a change which fades out is accepted by the adaptive condition. Assuming scale-free or single-scale algorithms (for node additions) we cannot expect influence of adaptive condition on node degree distribution because it is fixed. We can only expect an unequal distribution of nodes with different

k in space of network. This mechanism in other way depends on depth (the probability of acceptance is greater in the deep areas) and may compete with the one shown above in terminal modification tendency. We name it deep fadeout tendency (it is discussed more in details in [14]).

Case $k = 0$ cannot appear as effect of addition of node in our patterns of changeability of ss and sf networks described in the first article. However, if we also use removing of node and assume that removal should be similar to addition but in opposite direction, then $k = 0$ will be created. We introduce removing of node as one of type of investigated changeability in the next chapter (see fig.4), but in order to complete the problem of $k < 2$ we should discuss the case of $k = 0$ now. The existence of $k < 2$, especially $k = 0$, in the networks creates a new dimension of problems whose significance is similar to feedback loops dimension. On the node of $k = 1$ damage can fade out with high probability but if $k = 0$, then damage must fade out.

We will call node without outputs ($k = 0$) 'a blind node' or 'a blind'. However, if certain node connects its outputs directly or indirectly only to blind nodes, then it also cannot influence the assessed system outputs. Such a node is also 'a blind'. Addition of a new node which connects its output to a blind node (and becomes a blind) is always accepted. It can be tolerated when it happens seldom, but we have found [14], that in our simulation described earlier in [12, 13] such events are two or tree times more likely than all the other additions which are interesting. Now we introduce a special 'cost of a blind' which removes the blind additions. The 'cost' introduced later (in ch.3.4) is much stronger (of course it also prevents blind additions), however, we will compare processes with 'cost' and without it. To prevent the above described distortion, even the 'case without cost' described in figures using letter 'b' instead of 'n' (no cost) - will from now on contain the cost of a blind.

In the scale-free algorithm the nodes of $k = 0$ (i.e. a part of the blind nodes) have no chance to change this state and if the algorithm allows to generate such cases, then their number can grow and grow. Such a case looks unrealistic. It needs some special assumption to be corrected. We have modified the scale-free algorithm from proportionality to k to proportionality to $k + 1$ (fig.4, ch.4). Simultaneously we change our name of such a network from sf to se. It reduces share of blind nodes which also influence on expression level of deep fadeout [14]. Antidote for deep fadeout will be a 'cost'. In ch.3.4 and ch.4.1 we come back to this problem, which has strong influence on the investigated phenomena. The removing of nodes is a second way to reduce number of a blind - each attempt is accepted. Introducing of se pattern also reduces therefore speed of accepted removing and allows to growth of larger network [14]. Now, introducing of cost of a blind reduces blinds more radically [14] and removing of nodes occurs enough for limitation of nodes with $k = 0$ with allows us to coming back to simpler pattern sf.

3.3 Balance of Additions and Removals

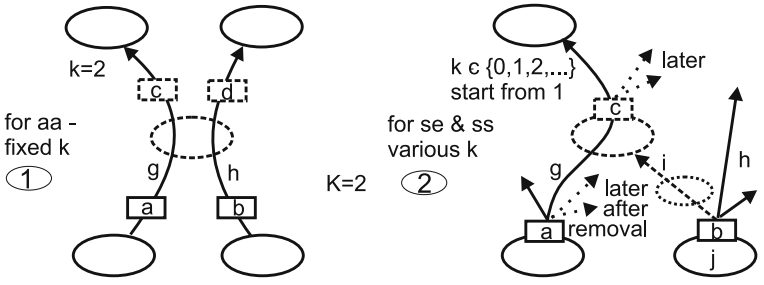


Fig. 4. Aggregate of automata *aa* (1), *sf*, *se* and *ss* networks (2) changeability (additions and removals) patterns for $K = 2$. In case of addition, links g and h and function of a node are drawn. Node j is drawn instead of link h for *ss* and moreover for *se* (change in comparison to *sf*). Removal only needs a draw of a node to remove. For node removing in *ss*, *sf* and *se* the outgoing links, which were added after the addition of this node to the network, are moved to the start node of link g. Node added on link i remains a node with $k = 0$ during the removal, therefore for adaptive growth with removing the *sf* network pattern for addition is supplemented by drawing node j additionally which gives *se*. Consequently, additions for *se* are proportional to $k + 1$ instead of k for *sf*.

To investigate the adaptive evolution as system development, i.e. growth, this system must have the ability to grow, therefore in the set of possible changes there should be the addition of a new node. For symmetry and higher adequateness to reality, removing a node should also be possible.

The patterns of network growth defining network type (e.g. preferential attachment defining scale-free type) typically only use additions. We did so in the first article but it is a simplification which cuts one dimension of the network phenomena. In this article we explore this seldom visited dimension and we find a few very important phenomena. One of them is network growth, which becomes an effect and no more just an assumption. Second one is the explanation of Weismann terminal additions regularity which occur as a phenomenon of this new dimension and is very important for biology. The great problem of $k = 0$ which we describe in ch.3.2 above is also an effect of removing of nodes.

Additions and removing should both be drawn randomly, but the sets of possibilities for such a draw cannot be the same. Removing a node can only be drawn from nodes currently present in the network. The set of possibilities

for a draw of an addition is much larger, for a single case we must draw independently function and places of all connections.

At this point we must complete changeability patterns for networks used in our investigation and add removing patterns to addition patterns showed in fig.2 in the previous article. These completions are shown in fig.4. There is also a change of addition pattern for sf used to solve the problem of $k = 0$ which defines se.

It is probable that that there is no node for an accepted removal in a network area but it is impossible to find such a situation for addition - we can always add a 'transparent' node, which produces just the same output signals for given inputs. When most of nodes have been checked for removing, new accepted addition can change the situation and create a new possibility for removal. We can introduce a parameter for area in the network structure which describes its ability to accept removal. The main part of this ability is related to the density of 'transparent' nodes. This parameter should differ in places which differ in the probability of acceptance of any change. In the places with very small probability of non-transparent modification acceptance (e.g. deep in the aggregate) between two 'non- transparent' changes, which change situation, more present nodes are checked for transparent removing, therefore, density of 'transparent' nodes is lower. But such a possibility of removing is better exploited in deep places. Transparent addition in deep and shallow areas are equal (fig.8). In deep places we can expect more of accepted removals in comparison to additions than in places of high probability of modification acceptance (e.g. shallow area). This is the second general mechanism creating structural tendencies. It creates 'terminal predominance of addition' tendency and 'simplifying of early areas' tendency as well as 'total growth of the system' tendency (see fig.5.1 distributions $P(D, +and-|a)$ or fig.6.right).

3.4 Transparent Nodes and Cost

Acceptance of addition of 'transparent' node by the adaptive condition may be challenged by interpretation similarly to the addition of a blind node. Output signals of the transparent nodes are the same as their appropriate input signals and addition or removing of such a node does not change the signals of remaining nodes. For other input signals these nodes usually have output signals different from the input signals, therefore this effect may be a little bit overestimated in our special algorithm, however it generally also exists. If the changed system is maintained by a human, then such an useless addition is an unnecessary cost. For higher s the probability of the presence of a transparent node is lower and this problem can be neglected. However, for low values of s ($s = 4$ or $s = 8$) used in the simulation of aggregate it is significant and interferes with observations of some interesting tendencies.

E.g. see fig.8.1-4. Therefore, we have introduced the 'cost' of node addition using strict inequality in the adaptive condition for additions $a \equiv b_{t+1} > b_t$ instead of the weak one. The adaptive condition for removals is still a weak inequality $a \equiv b_{t+1} \geq b_t$. Such a modification of the model had a strong influence on the dynamics of evolution process through the parameter of 'ability to removal' and on balance of additions and removals. It is described in more detail in [10] in chapters 4.4.3 and 5.1 with fig.5 (where 4- in the figure e should be and 4+ in the fig. f should be) and in [14].

In the simulation description we use the letter 'c' (which means 'with cost') and 'n' or 'b' (which means 'no cost', 'b' - 'only cost of a blind') to indicate cost implementation. These letter follow the number s. The effect of transparent node and cost in aggregate simulation will be discussed in detail in chapter 4.4 and shown in fig.8.1-4.

Another effect of cost is also cutting-out of cases of additions, whose damage avalanche fades out later than in the first step but still without reaching the network output. For the aggregate this effect was small and can be neglected but for Kauffman networks sf, se and ss nearly half of the nodes have $k < 2$ and their coefficient of damage propagation w is smaller than 1 also for very high s. It means that damage fades out easily on such nodes, what we have remarked above in ch.3.2 and in the previous article. We observe that in the deeper part of the network lots of changes fade out. We will discuss it later in more detail in chapter 4.1. We introduced cost to reject such cases but speed of producing a new ability to remove, which uses the same mechanism of fade-out, is too high and we cannot use similar probability of additions and removals because then the network would not grow.

It has been important to obtain a complex network - an area of investigation - in the most 'natural' way possible because network may be different and these differences may have an influence on the investigated phenomena. One of such differences is the above mentioned parameter of 'ability to remove' but others are not known. Therefore, we have started our simulation from a null network - signals of the system inputs have been directly connected to system outputs. We have randomly added and removed nodes and implemented adaptive condition and we expected growth. The problem is that at the first two stages of growth (in each stage the number N of nodes in the network increases by 128 in aa and 256 in current simulation of Kauffman networks), when the network is not complex yet, there are no mechanisms of growth tendency nor other normal mechanisms which make growth 'natural'. There is no problem for cases without cost but for an aggregate with cost and low $s = 4$ (with equal probability of addition and removal) we do not obtain growth nor the necessary level of complexity. Therefore we have altered the model by modification of the formula for fitness b to have s-times more states: $b = \Sigma(s-1) - |y_j - z_j|$ where y is an output signals vector and

z 'ideal' vector and signals are in range from zero to $s-1$. The presented results of aggregate simulation (aa and an networks) are obtained for such an assumption. For Kauffman networks we stay with the first simple b definition, but for the network with cost the problem of growth also occurs. This time we use different probabilities of addition and removing (99% and 1%) to obtain growth. However, this growth seems limited. This is an effect of deep fadeout (see ch.4.1) for removing.

This simple 'cost' condition appears very strong. This condition generates all the structural tendencies investigated in this paper: terminal modifications, terminal predominance of additions, simplifications of deeper parts and similarity of functional and historic orders also for ordered networks [12] where $s=2$ and $K=2$. It is easy to understand - newly connected nodes must lie near the assessed outputs to influence at least one signal of system outputs but not much more. It eliminates additions at longer distance from system outputs in both regimes - in chaotic one because change of outputs is too large and in ordered one because there the damage fades out without affecting the outputs.

4 Simulation Results

As in our previous chapter in this book, we have checked the above premise in simulation for different network types. Now we use networks: aa, an, sf, se, ss. They are investigated in the previous article, however, the scale-free network needs some change of pattern which also effects in a change of name from sf to se (see end of ch.3.2). Random Erdős-Rényi er network [7] which was used as basic by Kauffman [21,18] cannot grow under the control of adaptive condition and therefore we do not use it for these experiments. For all simulations we use a fixed number of system outputs $m=64$ and a fixed number of node inputs $K=2$. Kaufman did not use system's external outputs but he used a fixed number K and a flexible number of output links k of node in er networks. For scale-free network Iguchi et. al. [18] and earlier Aldana [2] and Kauffman [22] have also used flexible K.

The results of three main series of simulation are presented: the first series: aa - aggregate of automata with feedbacks where the described tendencies were investigated for the first time; the second: an - aggregate without feedbacks to check the role of feedbacks in the main effects; and the last one - Kauffman networks with variable node degree k and feedbacks: ss - single-scale and sf, se - scale-free networks - modern networks which appear much more adequate. This third series differs from the one described in much shorter earlier papers [12,13] with respect to the 'cost of a blind' used here. All simulations keep b on the level of 3/4 of its maximum value - it is a relatively high value of fitness b but it can still be higher.

For important scale-free network a problem appears with removing which creates $k = 0$ node. Such a node can be removed but it cannot come back into play, which creates a dummy network where most of the nodes have $k = 0$. To avoid this situation, we also add drawing of a node to the drawing of a link like in *ss* network and for this modified *se* type we obtain connection proportionality to $k+1$ instead of k. Dorogovtsev uses [5] a little similar correction. Unfortunately, this correction, especially important for the case without cost, modifies $P(k)$ which becomes more like for *ss* [14].

Investigations of the phenomena similar to transparent addition and removing need particular but random functions in our algorithm now, as opposed to simulations in the previous article. As it can be expected basing on the previous article, the networks sf, se and *ss* need much a higher s to exhibit phenomena typical for *aa* and *an* networks. The program was prepared for seven-bit s only. To create damage spreading inside the network as for very high $s \sim 1800$ we have used in [12] (where we have not yet implemented the cost of a blind) a special function (with $s = 64$). Such a case is there denoted in figures as \$. It is interesting as an example of influence of extreme values of the s parameter, however, we do not repeat this case now with 'cost of a blind'.

Aggregate *aa* and *an* grows and this growth is described by up to 8 stages W of 128 automata each. The first stage of a complex aggregate is $W = 3$ (256-384 automata). The depth definition used for *aa* and *an* networks (fig.3) shares the complex aggregate (stages from $W = 3$ to 8 for *aa* - aggregate with feedbacks or $W = 3$ and 4 for *an* - aggregate without feedbacks) as shown in fig.5.1 by $P(D)$ distribution (see also fig.8 in [10] or fig.6 in [11]). These two distributions are different mainly because the simulation of *an* case has a smaller range of stages and a smaller network.

Depth D, as the structural measure of functional order is the main parameter describing the place in the network where change is initialised. However, we have used two different definitions of it - depth in the form shown in fig.3 was used only for *aa* and *an*. For modern networks sf, se - scale-free and *ss* - single-scale with flexible node degree k we have used the shortest way to outputs as depth D. Distributions $P(D)$ for these complex Kauffman networks are shown in fig.5.2. For interpretation of simulation effects the places of 90% and 99% of network nodes are marked. The shortest way to output is not as adequate to describe functional order as depth defined for aggregate but we did not find [14] a much better solution for flexible k (see ch.3.1 here and ch.7 in the previous article).

The main result of simulation which we are going to present in this paper are three tendencies in dependency on depth D as structural measure of functional order. These are 'terminal modifications and conservation of early

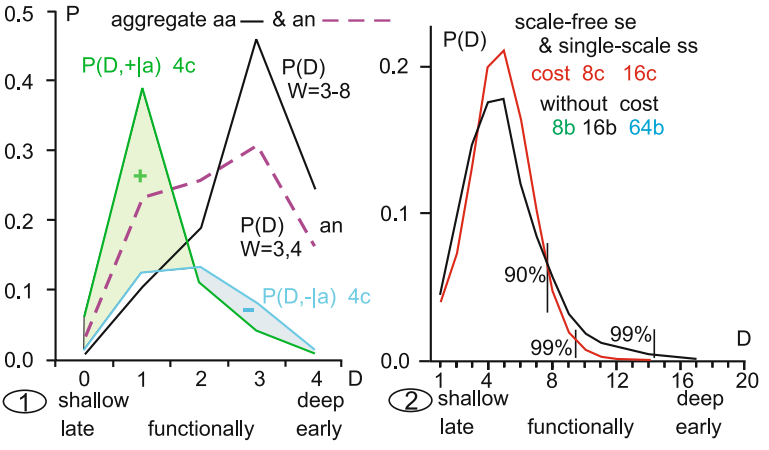


Fig. 5. Depth distribution P(D) for aggregate *aa* and *an* (1) and for Kauffman networks sf, se and ss (2). Definitions of depth D in these cases are different but the role - functional order description in sequential form based on structure - is the same. At each stage of growth W the aggregate increases by 128 nodes, $W = 3$ is the first stage of complex network. Kauffman networks have grown up to 2000 nodes and results (here $P(D)$) are collected from N=750. Additions '+' and removal '-' of nodes are drawn with the same probability (except for the cost case for Kauffman network) and with the same distribution P(D) but adaptive condition a deforms these distributions as shown in (1) for *aa* ($s = 4$, cost: 4c). This depicts the large predominance of addition '+' in the shallow area and small predominance of removal '-' in the deep area.

places' tendency shown on left column in fig.6 and 'terminal predominance of addition over removing' tendency and 'simplification of early places' tendency shown on the right column in fig.6. 'Terminal' and 'early places' in these tendencies should be understood in the sense of functional order, however, we also observe similar tendencies in historical order, which are also described and are shown in fig.8. The similarity of functional and historical orders (shown in fig.7.1-2) accomplished by the stability of node function (fig.7.3-4) and the effect of terminal predominance of additions (fig.5-6) is the theoretical expectation of the highly controversial Haeckel's biogenetic law. These tendencies are the effects of adaptive condition of network growth especially in adaptive condition's form with cost. Network growth is not assumed - it is also a tendency and, therefore, the effect of adaptive condition, except of Kauffman networks sf, se, ss with cost.

4.1 Terminal Modifications and Conservation of Early Area

A tendency known in biology and not as controversial as others is Naef's terminal modifications and conservation of early stages tendency. It has few

explanations, the one which is commonly accepted in biology is an old explanation, extremely simple but not correct, proposed by de Beer [4]. The results of our simulation of our not so simple model of this phenomenon are depicted in the left column of fig.6. In fig.6 in left column $P(a|D)$ (more exact: $P(a|D, b$ =3/4 of $bmax)$) is depicted as a sum -, and in the right column the balance of additions and removals is depicted as a difference of $P(a|+, D)$ and $P(a|-, D)$. For aggregate, where $P(+|D) = P(-|D)$ it is enough to obtain a correct form of curves but for Kauffman networks with cost $P(-|D)$ is only 1% and $P(+|D) = 99\%$. Therefore, $P(a|D)$ and the balance are calculated exactly as $P(a|+, D)P(+|D) \pm P(a|-, D)P(-|D)$. To maintain the same convention for aggregate the scale should be divided by two but the scale is not crucial here. The important thing is the form of these dependencies, therefore they are multiplied by shown constants to make the comparison of the forms more useful. In each series we show several different simulations which differ with respect to s and the cost. If there was no tendency, there should be a horizontal line. Remember that the definition of depth D is different for aggregate and Kauffman network. It is not crucial but we treat D for aggregate as a range but for Kauffman network as a value.

In the case of *aa* the shown results are obtained in five different simulations. These results refer to complex networks - aggregates at stages of growth $W = 3$ to 8. Only for $s = 4$ without cost (4n), for $P(a|D)$ there is no tendency - we consider $s = 4$ as still extremely low (see [10]). Total dependency of accepted changeability $P(a|D)$ on depth D - terminal modifications and conservation of early places tendency, concerns additions and removals together but has different power for each one of them separately. In [10] in fig.6 $P(a|+, D)$ and $P(a|-, D)$ for the aggregate are shown separately.

For *an* - aggregate without feedbacks only the stages $W = 3$ and 4 are summed up but it is enough to state that feedbacks are not necessary for tendency mechanisms. This is important especially for interpretation if system represent a process, however, it is to discuss.

The case of Kauffman networks contains sf, se - scale-free networks and ss - single-scale with feedbacks, without and with cost. Unlike in the short articles [12, 13], the 'cost of a blind' is now used for the case 'without cost' (ch.3.2). Results for sf, se and ss occur so similar that they are not distinguished in the most figures. These networks are much richer with different mechanisms than the aggregate.

Especially, there appears a strong mechanism of 'deep fade-out' on nodes with $k = 1$ and $k = 0$. These nodes group on deeper part of networks [14], deeper than the inputs of system which typically became hubs as the oldest. The deep part of the network is then significantly more ordered and therefore more resistant to the effects of changes which usually fade out. They are therefore

accepted by the adaptive condition. This mechanism gives an opposite effect than the tendency of terminal modifications and conservation of early areas but this deeper part of network is a small part as shown in fig.7.2 and fig.5.2.

The power of this mechanism depends on s but to get significant differences a big range of s is needed. We use $s = 8$, 16, 64. A special one marked in the figures as '$' was used for *se* in [12] but without the 'cost of a blind' (fig.8.5 and 6 *se*n). In this special case we have used $s = 64$ in adaptive condition but for damage propagation we force a non-random function whose output signal changed in most of the cases when the node inputs changed. Its effect for damage propagation is similar to having an s of about 1600 to 1900 in different cases. For $s = 8$ there is a slight terminal modification effect and only 'deep fade-out' is observed. For $s = 64$ these two mechanisms have similar power but 'deep fade-out' occurs in the smaller part of the network. It is for network of $N = 750 - 2000$ nodes and $m = 64$ output signals, for more nodes it is a greater part and this effects in a higher probability of real fade-out. If cost is applied then the probability of acceptance becomes higher for removing (but not for addition) and the growth of network stops. For very high s='$' for damage propagation 'deep fade-out' has a small influence. We postulate the high value of s in interpretation but we are not sure if it is indeed so high, it is one of the reasons why we do not repeat this case in the simulation with cost of a blind. We implement cost to reject the questionable cases of addition which totally fade out and have such a high frequency of acceptance. In these simulations we cannot use similar probability of additions and removals due to lack of growth. We use 1% for removals and 99% for additions. As one can expect, we obtain a strong terminal modification tendency. For the case with cost we use $s = 8$ and $s = 16$, we observe that there appears the boundary of network growth and for sf and *se* networks the planned number of nodes 2000 is not reached, only about 1500. For *ss* this boundary is higher but near 2000. Effective conservation of deeper parts for network with $k = 0$ and $k = 1$ is much weaker than for the aggregate and it is much more dependent on interpretative expectations of s value and cost problem. The backgrounds of interpretative assumption of frequency of $k = 0$ and $k = 1$ need more investigations and become a problem of similar importance as the presence of feedbacks. Its importance is illustrated by the fact that e.g for *se* network with $s = 8$ and without cost (cost of a blind applied only) there are 1200 nodes with $k < 2$ in all 2000 nodes in such network and part of remaining 800 are a blind.

4.2 Terminal Predominance of Additions Tendency, Simplification of Earlier Area Tendency, Growth Tendency

In chapter 3.3 we have discussed the balance of node additions and removals. This balance is also shown in fig.6, now on the right column, with the same

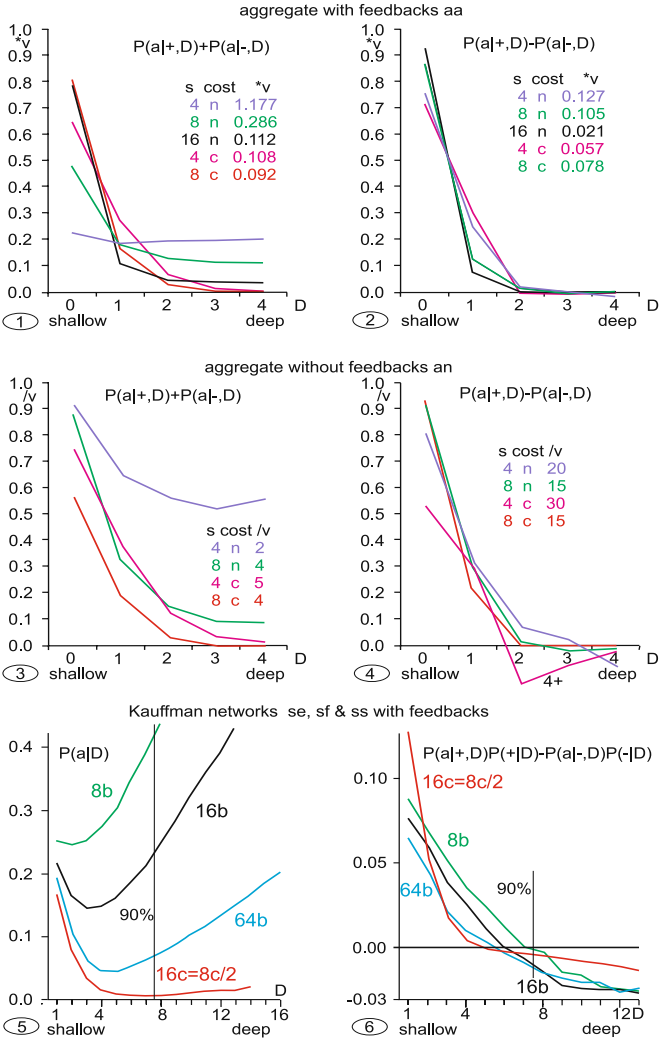


Fig. 6. The main tendencies as result of simulations. On the left (1,3,5): Terminal modifications and conservation of early area tendency. On the right (2,4,6): Balance of addition and removing: terminal predominance of additions (over removing) tendency and the tendency of simplification (predominance of removing over addition) of the early parts. On the top (1,2) for aa, in the middle (3,4) for aggregate without feedbacks 'an' and in the bottom (5,6) for Kauffman networks of variable node degree k - scale-free sf, se and single-scale ss. The results for sf, se and ss are so similar that a single curve is used for all of them. Dependency on depth D as structural measure of functional order. Definition of D for 1-4 in fig.3, for 5 and 6 it is the shortest way to outputs. For aa, an scaled by coefficient v for useful form of comparison. Cases described by s in number followed by cost: without cost: n - no cost, b - cost of a blind only; with cost - c.

assumption as on the left for terminal modifications tendency. We can see in all cases of network types and simulation parameters that there is a strong predominance of additions over removals shallow under the outputs. However, deeper in the network we encounter an opposite tendency- predominance of removals. We call it 'simplification', however, there are few accepted changes in the deep area and this effect has little importance - in *aa* it is barely statistically important. While 'terminal modifications and conservation of the early area' are two aspects of a single tendency, the 'terminal predominance (majority in [10]) of additions over removals' and the 'simplification of early area' are two independent tendencies.

They are similar to Weismann's regularity of ontogeny evolution: 'terminal additions and condensations of early stages' proposed in biology in [35]. This concept was later named 'pressing back' or 'shunting back' [17]. Weismann's condensation does not necessarily mean structural changes in the network, it could mean only faster timing but the concept of shunting back suggests the appearance of an unoccupied room to move and this is our 'simplification' (because this unoccupied room appears as an effect of removal). We regret that we cannot use the known historical name 'terminal additions' for our tendency of terminal predominance of additions, whose meaning is the same as Weismann's regularity because such a name would conflict with terminal modification tendency for additions as modifications. Weismann does not consider two types of modifications: additions and removals and their balance, he only draws conclusions from the previous observations by saying that additions are observed and they appear in the terminal area. This means that the number of transformations in ontogeny grows but the total time of ontogeny is constant and this is Weismann's condensation. Now we know the mechanism of this effect much more precisely - the observation of additions at the end of ontogeny as complex network of transformation is the effect of balance of additions and removals in this area but one of the causes of the observed time constancy is the simplification of the earlier stages of the process.

The cost has a strong influence on the balance of additions and removals, especially for networks with $k = 0$ and $k = 1$, in our simulation ss, se and sf. We will come back later to this problem for aggregate, when we discuss fig.8. The remark from the end of ch.3.4 should be recalled here.

In the clear case without cost the terminal modifications tendency creates the area differences necessary for the terminal predominance of additions tendency to occur. It also amplifies the effects of terminal predominance of additions tendency and suppresses the effects of simplifications of deeper area tendency. These two 'terminal' tendencies together produce growth of the whole network, which is also a tendency caused by adaptive condition, and produce similarity of historical and functional orders. These two effects are of great importance, therefore, the tendency of terminal predominance of

additions is considered to be their cause because it is representative of the mechanism which contains both the 'terminal' tendencies.

The cost is an independent cause of terminal predominance of additions as well as of the simplification of deeper part which in this case defines one tendency because there is one underlying mechanism. However, cost must cooperate with the adaptive condition, which remains the first main cause of the observed tendencies. Cost can be treated as a more radical form of adaptive condition.

4.3 Similarity of Historical and Functional Orders

These tendencies result in the growth of a system at its end (end of signals way through network) but if we always add new elements at one end of the linear structure, then this end becomes much younger than the opposite one. Let us introduce a parameter H as 'historical order' - it splits the current set of network nodes into 8 equal parts in the order of sequence of additions of nodes. $H = 1$ describes the oldest part of nodes and $H = 8$ - the youngest one. Parameter H is similar to the stage of growth W introduced in chapter 4, especially at the stage $W = 8$ but at the earlier stages they differ significantly - a smaller network is also split into 8 parts. This similarity for $N = 2000$ is shown in fig.7.6.

The mechanisms described above should produce similarity of functional (described by D) and historical (H) orders. We have checked this expectation - the results are shown in fig.7.1-2. For aggregate fig.7.1 (- on the left) shows this statistical condition for the same set of simulation as fig.6.1-4. We observe close similarity between the historical and the functional order. Nodes added at the earlier stages of evolution of the network occupy deeper, i.e., earlier parts of the structure, near inputs of the system. Case '4n' ($s = 4$ without cost) differs from the rest of simulations - it is extreme, as we conclude above. Due to the close similarity of results of the rest of simulations we show them together. The results for aggregate with and without feedbacks are depicted with continual (aa) and dashed (an) lines respectively.

For Kauffmamn networks (fig.7.2 - on the right) this condition is more complex. As can be expected the case $s = 8$ without cost (8b) is here an extreme one and manifests an opposite condition. There appears a similarity of functional and historical order for the case 64b in the shallow area but deeper we find an opposite dependency and these two parts are similar there. For the special case se\$n (without any cost, extremely high s for damage spreading and $s = 64$ for fitness, shown in [12] in fig.6.e and f) the main form is similar to the previous one but the part of 'typical' (like in aggregate) dependency is higher (70%) and stronger, and the remaining part is smaller and weaker.

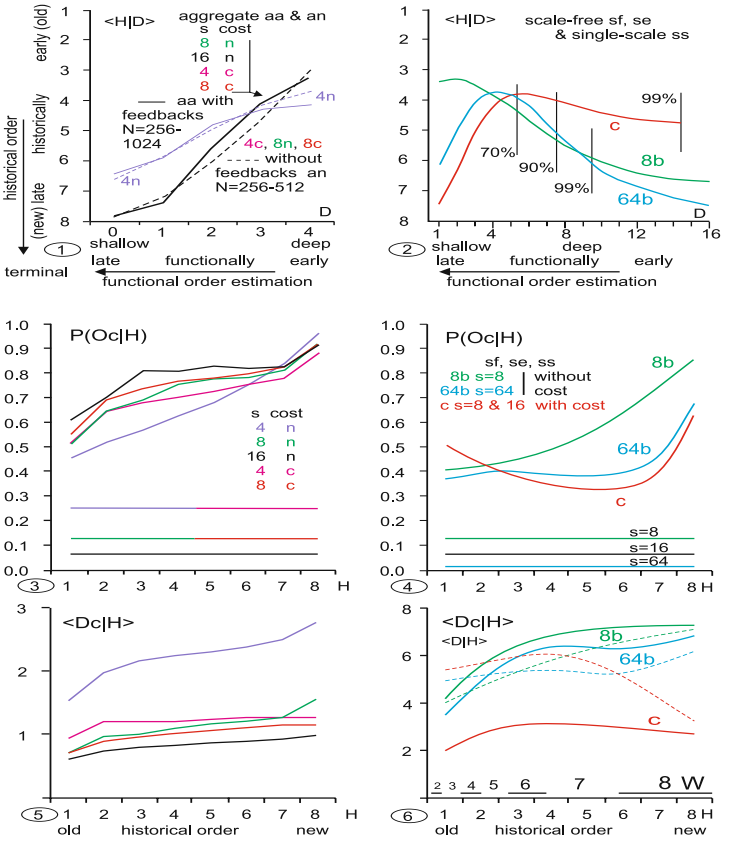


Fig. 7. Recapitulation tendency elements. (1,2) Similarity of historical H and functional D order. (3,4) Stability of nodes' functioning. Random levels are depicted. (5,6) Additions near system outputs. Compare to $P(D)$ in fig.5. H is the sequence of addition of a given node to the network. The same set of simulations as in fig.5 and fig.6. On the left (1) for aggregate with feedbacks (aa) - continuous lines, and without feedbacks (an) - dashed lines. D is defined in fig.3. Figures (3) and (5) (described in (3)) each simulation is depicted separately. Case 4n is also extreme. In (5) in all other cases (than 4n) nodes are added near the output in spite of $P(D)$ showed in fig.5.1. On the right (2,4,6) results for Kauffman networks (sf, se and ss), containing $k = 0$ and $k = 1$ are shown similarly to the left case. Here D is the shortest distance to outputs and can be large, but with small probabilities, therefore the boundaries of these probabilities are shown in (2). Here each line approximates a group of similar cases which is described in (4). All six cases with cost (sf, se and ss $s = 8$ and $s = 16$ denoted as 'c') seem very similar to each other, they exhibit in (6) shift of the placement from an area of addition of small depth $< Dc|H >$ to a deeper part ($< D|H >$ for dashed lines). For the case without cost different s cause significant differences in results. Remember, that here cost of a blind is applied, therefore 'b' is used instead 'n'. In (6) above the H axis the connection to stages W of growth is depicted.

We observe the same direction for cases with cost, which enlarges the power of a 'typical' dependency. This situation properly explains the competition of two mechanisms described above: terminal modifications and deep fade-out. Concluding - for higher s and for the case with cost we should expect similarity of functional and historical orders especially in the shallower, larger part of areas but deeper we should meet the opposite and weaker dependency.

Similarity of historical sequence of additions (of the nodes which now form the system) to the functional order (which describes the sequence of signals transformation by the system) is connected to recapitulation of phylogeny in the ontogeny. In such an interpretation the ontogeny is described as an evolving network of transformations. Recapitulation of phylogeny in the ontogeny is known as Haeckel's biogenetic law which, as mentioned above (see end of Introduction), is now disregarded because of lack of explanation (I suppose). To show the existence of recapitulation in our model we should accomplish the above similarity of historical an functional order by stability of function of old nodes and by confirmation that additions have occurred close to the outputs. Stability of function allows us to treat a particular node as the same as in the beginning. Results are shown in fig.7.3-4 as the probability of meeting the first output state at the end of evolution. Random levels are shown for comparison.

Addition near outputs should be interpreted as addition to a form similar to adults (phenotype) - extension of development. Such a mechanism cause of 'pressing back' which distinguishs recapitulation and repetition which assume constant ontogeny stage (depth) of transformation (node). It is a result of terminal predominance of additions and is shown in fig.7.5-6. These results are convincing for the *aa* network and case with cost of Kauffman networks which are lack of strong deep fadeout.

4.4 Tendencies in Functional and Historical Orders

The existence of similarity between the functional and historical orders allows us to expect that the tendencies in historical order will be similar to the observed ones in functional order in fig.6. However, it is problematic to define the parameter H for the new added node - it is new and always should be $H = 8$ but if we are going to describe the place of addition, not exactly the node, then we can use H of nodes, which receive signals from the new added node. In the aggregate there are two of them because $k = 2$ and we obtain 2D argument, which we can average by one of them. Each of these two independent 'nodes' can be an aggregate external output, its H is also not well defined. For aggregate external outputs we use the case $H = 8$ together with other automata of $H = 8$. For the case of Kauffman networks this solution appears not to be accurate and we must reserve $H = 8$ only for network outputs.

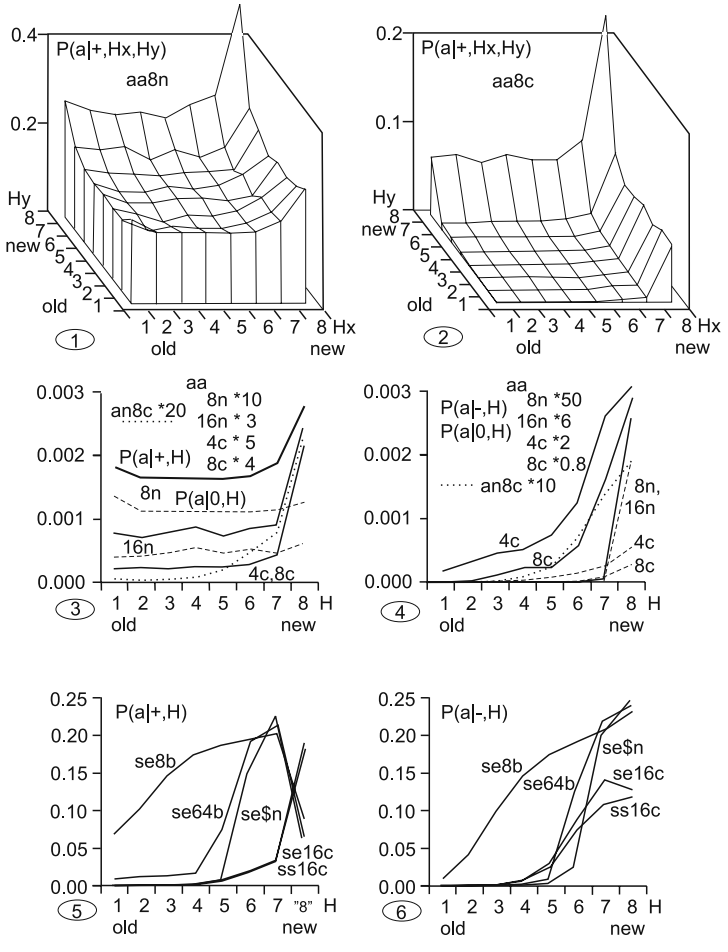


Fig. 8. Historically terminal modification tendency for additions (1-3,5) and removals (4,6) in simulations like in the case of fig.6 and fig.7 but without 4n for aggregate and with special case *se*$n. For addition in *aa* $s = 8$ the effect of cost in exact 2D distribution is shown (1,2). In (3) distribution one dimension of H is averaged. In (3 and 4) the share of transparent cases is shown by dashed lines ($P(a|0, H)$), also the case *an8c* is shown as dot lines. In (5) $H = 8$ is for connection to external outputs.

Fig.8.1-3 shows the results for additions in aggregate simulations. In fig.8.1 the case without cost is shown. We can observe flat horizontal dependency on some non-zero level, only for $H = 8$ there is a significantly higher value. When we look at fig.8.3 we get explanation of this non-zero level - these are transparent automata shown here as dashed lines and described as 0 in $P(a|0, H)$. We have discussed transparent automata in chapter 3.4 together with their questionable interpretation bases. In order to reject them we introduce cost, which radically lowers the common level practically to zero, as we can see in fig.8.2.

Definition of H for removals is clear but different than for additions, therefore, we cannot complete one historically terminal modifications distribution and we must depict additions and removals separately. In fig.8.4 distributions of $P(a|H)$ for removals are shown and the share of transparent automata in these distribution is indicated with a dashed line.

In fig.8.3 and fig.8.4 the case 8c for *an* is shown as a dot line. It looks similarly. In fig.8.5 and fig.8.6 results for *se* and *ss* are shown. In these cases when the new node was added, k was always equal to 1 and then the argument is 1D, not 2D as above but outputs of networks have not well defined H and we separate this case as $H = 8$.

In all obtained results shown in fig.8 we can state a strong tendency of historically terminal additions and removals separately. Note, that 'terminal additions' is here a part of terminal modifications but it not corresponds with Weismann's terminal additions. Due to the described problems of definition of parameter H for addition we have not checked the balance of additions and removals in H dependency (which corresponds with Weismann's terminal additions).

5 Conclusion

Adaptive condition as a condition of growth in the evolution of a wide range of networks types causes 'structural tendencies' such as: 'terminal modifications and conservation of early area', 'terminal predominance of additions over removals' and also 'simplifications of early area'. In networks with various node degree 'deep fade-out' is observed. 'Tendency' is defined as an effect of adaptive condition in some distribution of change parameter - a property of developmental changes. These change parameters are connected to the topology of network - its structure. 'Terminal' and 'early' are here understood in the sense of functional order but in the case of the 'terminal modifications and conservation of early areas' tendency they can be also understood in the historical sense (in the sense of the sequence of node additions). The growth of whole network is (except Kauffman networks with cost) not assumed, it is also a tendency, which means - it is an effect of the adaptive condition.

These 'structural tendencies' are usually strong in the range of parameters expected by interpretation.

Adaptive condition is defined basing on network's output signals in the long period of network function between consecutive changes of network. The algorithm prepared for simulation experiments uses some special simplification to omit the problem of periodic attractors, which usually appear as outputs of functioning networks with feedbacks. The role of feedbacks is checked - they are not necessary to obtain similar tendencies.

There are two different mechanisms based on the adaptive condition which leads to the tendencies discussed in this chapter.

First of these mechanisms is based on symmetrical assumptions for additions and removals of nodes. It is clearly visible for simple aa network. For sf, se and ss networks which are much richer in different phenomena it is concealed by the strong competitive mechanism of deep fadeout on nodes with $k < 2$. This first mechanism for terminal modification tendency is an effect of cone of influence and coefficient of damage propagation. For the terminal predominance of additions tendency and for simplification of deeper part it is an effect of the parameter describing the ability of a network area to accepted removal of a node, which in turn depends on the modification speed of this area and on the sets of possibilities for drawing addition and removal. This mechanism needs chaos and is stronger above the complexity threshold. This threshold of complexity (see my previous article in this book) is a phase transition to chaos (exhibition of matured chaotic features) during network growth. It is observed on the assessed outputs of the system in distribution of damage size as two peaks for very small and very large changes and near zero in-between. Feedbacks, which lead to technical and interpretative problems, help to reach faster the complexity threshold and an area where this mechanism works. However, this mechanism is 'gentle and defenceless'. Strong deep fadeout is a winner for small s. Defensive methods directed against the 'parasite' $k = 0$ nodes like cost of a blind (ch.3.2) or conversion of connection pattern for sf to se network (ch.3.3) are too weak. Most nodes in the 'node society' become 'lazy' ($k < 2$) for scale-free network and the effect disappears.

The second mechanism is simple, 'brutal, strong', but effective. It is the cost condition which we discuss as a separate condition cooperating with the adaptive condition, although it is actually only a stronger version of the adaptive condition. It is not so interesting from the philosophical point of view because to generate the most interesting 'predominance of addition over removing' tendency it is based on unsymmetrical assumptions for addition and removing. Consequently, we do not obtain the asymmetry from model as above, we simply assume it in the first place. However, this assumption is a sensible

and adequate one in the typical case. This mechanisms does not need chaos - it also works in ordered regime. In *aa* network it also helps to clean up the 'malingerers' - the transparent nodes. This 'cleaning' exposes the main working mechanism and makes it more effective. However, this strong mechanism encounters a problem with deep fadeout and $k = 0$ nodes in it, which dramatically limits the growth speed increasing the share of nodes which are available for removing.

We consider adaptive condition as one of the most ubiquitous conditions of the development of real functioning networks and, therefore, its effects should be interesting and important for the design of new networks and for understanding of the existing ones, including biological objects evolved under Darwinian mechanism. This is the topic of this paper and, therefore, we limit ourselves to a few most important tendencies but this new area is much richer with similar phenomena.

References

1. Albert, R., Barabási, A.-L.: Statistical mechanics of complex networks. Rev. Mod. Phys. 74(1), 47–97 (2002)
2. Aldana, M.: Dynamics of Boolean Networks with Scale Free Topology. Physica D 185, 45–66 (2003)
3. Barabási, A.-L., Albert, R., Jeong, H.: Mean-field theory for scale-free random networks. Physica A 272, 173–187 (1999)
4. de Beer, G.: Embryos and Ancestors. Oxford University Press, Oxford (1940)
5. Dorogovtsev, S.N., Mendes, J.F.F., Samukhin, A.N.: Structure of Growing Networks with Preferential Linking. Phys. Rev. Lett. 85, 4633 (2000)
6. Eldredge, N., Gould, S.J.: Punctuated equilibria: An alternative to Phylecic Gradualism. In: Schopf, T.J.M. (ed.) Models in Paleontology, pp. 82–115. Freeman, Cooper (1972)
7. Erdős, P., Rényi, A.: Random graphs. Publication of the Mathematical Institute of the Hungarian Academy of Science 5, 17–61 (1960)
8. Gecow, A.: A cybernetic model of improving and its application to the evolution and ontogenesis description. In: Proceedings of Fifth International Congress of Biomathematics, Paris (1975)
9. Gecow, A., Hoffman, A.: Self-improvement in a complex cybernetic system and its implication for biology. Acta Biotheoretica 32, 61–71 (1983)
10. Gecow, A., Nowostawski, M., Purvis, M.: Structural tendencies in complex systems development and their implication for software systems. Journal of Universal Computer Science 11, 327–356 (2005), `http://www.jucs.org/jucs_11_2/structural_tendencies_in_complex`
11. Gecow, A.: From a "Fossil" Problem of Recapitulation Existence to Computer Simulation and Answer. Neural Network World 3, 189–201 (2005), `http://www.cs.cas.cz/nnw/contents2005/number3.shtml`
12. Gecow, A.: Structural Tendencies - effects of adaptive evolution of complex (chaotic) systems. Int. J. Mod. Phys. C 19(4), 647–664 (2008)

13. Gecow, A.: Emergence of Growth, Complexity Threshold and Structural Tendencies During Adaptive Evolution of System. In: EPNACS 2007, Dresden (2007)
14. Gecow, A.: The conditions of structural tendencies. In: Aziz-Alaoui, M.A., Bertelle, C. (eds.) Modelling and Analysis of Complex Interacting Systems (in press)
15. Gould, S.J.: Ontogeny and phylogeny. Harvard University Press, Cambridge (1977)
16. Haeckel, E.: Generelle Morphologie der Organismen. George Reiner, Berlin (1866)
17. Holmes, S.J.: Recapitulation and Its Supposed Causes'. The Quarterly Review of Biology 19(4), 319–331 (1944)
18. Iguchi, K., Kinoshita, S.I., Yamada, H.: Boolean dynamics of Kauffman models with a scale-free network. J. Theor. Biol. 247, 138–151 (2007)
19. Kauffman, S.A.: Metabolic stability and epigenesis in randomly constructed genetic nets. J. Theor. Biol. 22, 437–467 (1969)
20. Kauffman, S.A.: Gene regulation networks: a theory for their global structure and behaviour. Current Topics in Dev. Biol. 6, 145 (1971)
21. Kauffman, S.A.: The Origins of Order: Self-Organization and Selection in Evolution. Oxford University Press, New York (1993)
22. Kauffman, S.A., Peterson, C., Samuelsson, B., Troein, C.: Genetic networks with canalyzing Boolean rules are always stable. PNAS 101(49), 17102–17107 (2004)
23. Kirillova, O.V.: Influence of a structure on system's dynamics on example of Boolean networks. Int. J. Mod. Phys. C 10(7), 1247–1260 (1999)
24. Mikhalevich, V.I.: The phylum Foraminifera d'Orbigny, 1826 - Foraminifers. In: Alimov, A.F. (ed.) Protista: Manual on Zoology, pp. 533–623. Nauka Publishers, St. Petersburg (2000) (in Russian, with English summary, 611–616)
25. Mikhalevich, V.I.: Polymerization and oligomerization in foraminiferal evolution. Studia Geologica Polonica 124, 117–141 (2005)
26. Naef, A.: Die individuelle Entwicklung organischen Formen als Urkunde ihrer Stammesgeschichte, (Jena 1917)
27. Raineri, M.: Old and new concepts in EvoDevo. In: Pontarotti, P. (ed.) Evolution Biology from Concepts to Applications. Springer, Heidelberg (in press)
28. Richardson, M.K., Hanken, J., Gooneratne, M.L., Pieau, C., Raynaud, A., Selwood, L., Wright, G.M.: There is no highly conserved embryonic stage in the vertebrates: implications for current theories of evolution and development. Anat. Embryol. 196, 91–106 (1997)
29. Schmalhausen, I.I.: Organism as a whole in individual and historical development Nauka, Moscow (in Russian) (1982); or Szmalhauzen, I.I.: Organizm jako całość w rozwoju indywidualnym i historycznym KiW Warszawa, pp. 87–113 (in Polish) (1962)
30. Schuster, H.: Deterministic Chaos: An Introduction. Physik-Verlag (1984)
31. Serra, R., Villani, M., Semeria, A.: Genetic network models and statistical properties of gene expression data in knock-out experiments. J. Theor. Biol. 227, 149–157 (2004)

32. Serra, R., Villani, M., Graudenzi, A., Kauffman, S.A.: Why a simple model of genetic regulatory networks describes the distribution of avalanches in gene expression data. J. Theor. Biol. 246, 449–460 (2007), http://dx.doi.org/10.1016/j.jtbi.2007.01.012
33. Serra, R., Villani, M., Damiani, C., Graudenzi, A., Colacci, A., Kauffamn, S.A.: Interacting random boolean networks. In: Jost, J., Helbing, D. (eds.) Proceedings of ECCS 2007: European Conference on Complex Systems, paper #165 (2007) (CD-Rom)
34. Wagner, A.: Estimating coarse gene network structure from large-scale gene perturbation data. Santa Fe Institute Working Paper, 01-09-051 (2001)
35. Weismann, A.: The Evolution Theory, 2 vols., London (1904)
36. Wilkins, A.S.: The evolution of developmental pathways, pp. 19–22. Sinauer Associates, Inc., Sunderland (2002)

Complex emergent properties in synchronized neuronal oscillations

Nathalie Corson and M.A. Aziz-Alaoui

Laboratoire de Mathématiques Appliquées du Havre
25 rue Philippe Lebon
76600 Le Havre, France
nathalie.corson@univ-lehavre.fr, aziz.alaoui@univ-lehavre.fr

Summary. This work adresses the dynamics and complexity of the Hindmarsh-Rose neuronal mathematical model. The general aim is the study of the asymptotic behaviour of neuron networks. In this paper, the analysis of these networks uses the synchronization theory via connections between neurons which give rise to emergent properties and self-organization. Our results lead to a classical law which describes many natural or artificial self-organized complex systems. This has been performed using numerical tools.

Key words : Hindmarsh-Rose model, synchronization, complexity, emergent properties.

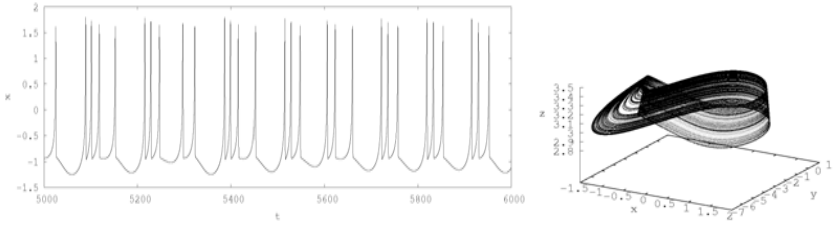
1 Introduction

Also called nerve cells, neurons are the most important cells of the nervous system. The Hindmarsh-Rose model (*HR* hereafter), see [7,8], describes the dynamics of the membrane potential in the axon of a neuron with a three dimensional system of nonlinear first order differential equations which read as,

$$\begin{cases} \dot{x} = y + ax^2 - x^3 - z + I \\ \dot{y} = 1 - dx^2 - y \\ \dot{z} = \epsilon(b(x - x_c) - z) \end{cases} \tag{1}$$

While x describes the membrane potential, y describes the exchange of ions accross the neuron membrane through fast ionic channels and z the exchange of ions through slow ionic channels.
I is the applied current while ϵ is a recovery variable, which is very small. This last parameter controls the slow motion of the neuron activity. Parameters a, b, d are constants experimentally determined and x_c is the equilibrium

x-coordinate of the two dimensional Hindmarsh-Rose system composed of the first two equations of (1) with $z = 0$ and $I = 0$.



(a) (b)

Fig. 1. System (1) for $a = 3$, $b = 4$, $d = 5$, $x_c = -\frac{1}{2}(1 + \sqrt{5})$, $I = 3.25$ and $\epsilon = 0.001$. (a) Time series (t, x), (b) (x, y, z) view.

In this paper, we mainly study synchronization motions of coupled *HR* systems which can be chaotic for some parameters choices. See for example [5, 6, 10, 16], or [9, 14, 15, 17] in which the behaviour bursting, spiking chaos or adding bifurcation are studied. For the reader convenience, let us recall that synchronization is a phenomenon characteristic of many processes in natural systems and nonlinear science. It has remained an objective of intensive research and is today considered as one of the basic nonlinear phenomena studied in mathematics, physics, engineering or life science. Several different regimes of synchronization have been investigated by many authors, see for example [1, 4, 11, 12, 13].

This word has a greek root, *syn* = *common* and *chronos* = *time*, which means to share common time or to occur at the same time, that is correlation or agreement in time of different processes.

Thus, synchronization of two dynamical systems generally means that one system somehow follows the motion of another. A lot of research has been carried out and, as a result, showed that even chaotic systems could synchronize when they are coupled. Many researchers have discussed the theory, the design or applications of synchronized motion in coupled chaotic systems. A broad variety of applications have emerged, for example to increase the power of lasers, to synchronize the output of electronic circuits, to control oscillations in chemical reactions or to encode electronic messages for secure communications. Here are some synchronization regimes,

- *Identical (or complete) Synchronization*, which is defined as the coincidence of states of interacting systems,
- *Generalized Synchronization*, which extends the identical synchronization phenomenon and implies the presence of some functional relation between two coupled systems; if this relationship is the identity we recover the identical synchronization,

- *Phase Synchronization*, which means driving of phases of chaotic oscillators, whereas their amplitudes remain uncorrelated,
- *Lag Synchronization*, which appears as a coincidence of shifted-in-time states of two systems.

Hereafter, for all numerical experiments, we use *HR* system with the following coordinates changes, see [3], $y = 1-y$, $z = 1+I+z$, $d = a+\alpha$, $c = -1-I-bx_c$. From this change of coordinates, we obtain,

$$\begin{cases} \dot{x} = ax^2 - x^3 - y - z \\ \dot{y} = (a+\alpha)x^2 - y \\ \dot{z} = \epsilon(bx + c - z) \end{cases} \tag{2}$$

Let us consider a network composed by n *HR* neurons. These neurons are coupled by their first variable x_i. A model of this network is given by,

$$\begin{cases} \dot{x_i} = ax_i^2 - x_i^3 + y_i - z_i - h(x_i, x_j), \quad i \neq j, \quad i = 1, ..., n, \quad j = 1, ..., n \\ \dot{y_i} = (a+\alpha)x_i^2 - y_i \\ \dot{z_i} = \epsilon(bx_i + c - z_i) \end{cases} \tag{3}$$

In the case of a network of n neurons linearly coupled, the coupling function h is chosen as,

$$h(x_i, x_j) = k_n \sum_{j=1}^{n} c_{ij} \Gamma(x_j) \tag{4}$$

in which the synaptic coupling Γ is a linear function, $\Gamma(x_j) = x_i - x_j$, for $i = 1, 2, ..., n$. Parameter k_n represents the coupling strength and $C_n = (c_{ij})$ is the $n \times n$ connectivity matrix,

$$\begin{cases} c_{ij} = 1 \text{ if } i \text{ and } j \text{ are connected, } i = 1, ...n, \ j = 1, ...n, \quad i \neq j \\ c_{ij} = 0 \text{ if } i \text{ and } j \text{ are not connected.} \end{cases}$$

This matrix C_n can be symmetric or not, so that unidirectional or bidirectional coupling are possible.

In the case of a network of n neurons nonlinearly coupled, the coupling function h is given by [3] and reads as,

$$h(x_i, x_j) = (x_i - V)k_n \sum_{j=1}^{n} c_{ij} \Gamma(x_j) \tag{5}$$

in which the synaptic coupling Γ is a nonlinear function with a threshold as given,

$$\Gamma(x_j) = \frac{1}{1 + exp(-\lambda(x_j - \Theta))} \tag{6}$$

where Θ is the threshold reached by every action potential for a neuron and λ is defined in [3].
As in the case of linear coupling, parameter k_n corresponds to the synaptic coupling strength. Parameter V is the reversal potential and must be larger than $x_i(t)$ for all i and all t since synapses are supposed excitatory.
We have chosen the situation in which each neuron has the same number of inputs from other neurons. Indeed, according to [3], it is a necessary condition for the synchronous solution to exist. Therefore, the synaptic connections we use have to be bidirectional. The C_n matrix should be symmetric and in our case, $C_n = 1_n - Id_n$ where $1_n = (1)_{nn}$ and $Id_n = (Id)_{nn}$ is the identical matrix.

The following parameters are fixed as follows throughout this paper,

$$a = 2.8,\ \alpha = 1.6,\ c = 5,\ b = 9,\ \epsilon = 0.001 \tag{7}$$

$$V = 2,\ \lambda = 10,\ \Theta = -0.25 \tag{8}$$

Interactions between nerve cells are possible thanks to synapses. The word *synapse* has a greek root, *syn = common* and *haptein = to touch*, which means *connection*. This is the functional contact part which exists between two neurons or between a neuron and another cell (muscular cell, sensorial receptor, ...). It operates the conversion of an action potential fired by the presynaptic neuron into a signal in the postsynaptic cell. Usually, two different types of synapses are discerned, which are,

- the chemical synapse, widely majoritary, which needs some neurotransmitters to transmit information,
- the electrical synapse, in which the signal is transmitted electrically through gap-junctions.

The size of the synaptic cleft is characteristic of one or the other kind of synapse. In the case of electrical synapses, this synaptic cleft is about two nanometers, while it can reach from ten to forty nanometers in the case of chemical synapses.

This paper is organized as follows. In the next section, the coupling function between mathematical neurons is a linear function in order to represent electrical synaptic connections. In this section, we firstly assume that neurons are identical and then, we assume that they are slightly different. In section 3, the same steps are followed with a nonlinear function in order to study chemical synaptic connections.

2 Property emerging from synchronization with a linear coupling

In this paper, the asymptotic behaviour is studied only by numerical analysis in various regions of the parameter space commonly used in the litterature. We have done the numerical computations very accurately in double precision and for different variations of parameters. In order to obtain reliable numerical results, the step size has been chosen to be equal to 10^{-4} and the first 10^6 steps are discarded to avoid the transient regime.

2.1 Coupling identical *HR* neurons

The first step is to consider the asymptotic behaviour of two neurons lineraly coupled. Therefore, we numerically study system (3) with coupling function (4) for $n = 2$. The connectivity matrix C_2 and the graph of the network are given in figure 2.

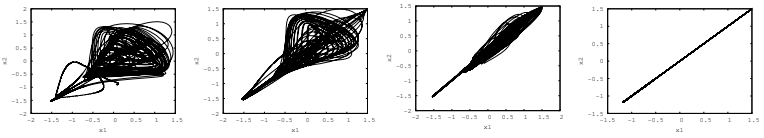
$$C_2 = 1_2 - Id_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Fig. 2. Bidirectional connection scheme for two neurons and the connectivity matrix C_2.

This gives the six-dimensional system representing two bidirectionaly coupled *HR* neurons given in (9),

$$\begin{cases} \dot{x_1} = ax_1^2 - x_1^3 + y_1 - z_1 - k_2(x_1 - x_2) \\ \dot{y_1} = (a+\alpha)x_1^2 - y_1 \\ \dot{z_1} = \epsilon(bx_1 + c - z_1) \\ \\ \dot{x_2} = ax_2^2 - x_2^3 + y_2 - z_2 - k_2(x_2 - x_1) \\ \dot{y_2} = (a+\alpha)x_2^2 - y_2 \\ \dot{z_2} = \epsilon(bx_2 + c - z_2) \end{cases} \tag{9}$$

Figure 3 shows the evolution of the (x_1, x_2) phase portrait of system (9) while making the coupling strength k_2 increase after a certain period of time. For low values of this coupling strength, the behaviour of x_1 is really different from the one of x_2, whereas when this coupling strength is large enough, x_1 and x_2 behaviours become synchronous.
We observe that synchronization motion between the first variable x_1 of one neuron and the first variable x_2 of the other, i.e : $x_1 = f(x_2)$ (here, $f(x) = Id(x)$), appears for a coupling strength $k_2 \geq 0.376$. For every coupling

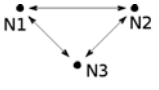


(a) (b) (c) (d)

Fig. 3. Phases portraits for system (9) showing x_2 according to x_1 for the coupling strength (a) $k_2 = 0.2$, (b) $k_2 = 0.25$, (c) $k_2 = 0.36$, (d) $k_2 = 0.376$. For two neurons linearly coupled, the synchronization threshold for the coupling strength is about $k_2 = 0.376$.

strength larger that this value of k_2, x_1 and x_2 remain synchronous. This phenomenon is also observed for variables y and z, for the same values of parameters. Therefore, this synchronization is total.

In the case of three neurons, the connectivity matric C_3 and the graph of the network are given in figure 4.



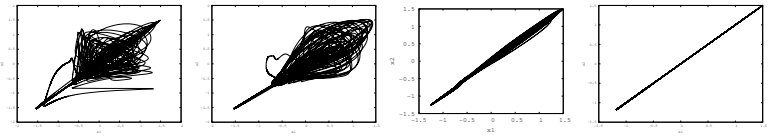
$$C_3 = 1_3 - Id_3 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

Fig. 4. Bidirectional connection scheme for three neurons and the connectivity matrix C_3.

System (3) with coupling function (4) for $n = 3$ reads as,

$$\begin{cases} \dot{x_1} = ax_1^2 - x_1^3 + y_1 - z_1 - k_3(x_1 - x_2) - k_3(x_1 - x_3) \\ \dot{y_1} = (a + \alpha)x_1^2 - y_1 \\ \dot{z_1} = \epsilon(bx_1 + c - z_1) \\ \\ \dot{x_2} = ax_2^2 - x_2^3 + y_2 - z_2 - k_3(x_2 - x_1) - k_3(x_2 - x_3) \\ \dot{y_2} = (a + \alpha)x_2^2 - y_2 \\ \dot{z_2} = \epsilon(bx_2 + c - z_2) \\ \\ \dot{x_3} = ax_3^2 - x_3^3 + y_3 - z_3 - k_3(x_3 - x_1) - k_3(x_3 - x_2) \\ \dot{y_3} = (a + \alpha)x_3^2 - y_3 \\ \dot{z_3} = \epsilon(bx_3 + c - z_3) \end{cases} \tag{10}$$

As shown in the case of two neurons, figure 5 shows the evolution of x_1 according to x_2 for system (10) while making the coupling strength increase. For low values of this coupling strength, the behaviour of x_1 is really different from the one of x_2, whereas when this coupling strength is large enough, x_1 and x_2 behaviours become identical.



(a) (b) (c) (d)

Fig. 5. Phases portraits for system (10) showing x_2 according to x_1 for the coupling strength (a) $k_3 = 0.1$, (b) $k_3 = 0.2$, (c) $k_3 = 0.24$, (d) $k_3 = 0.255$. For three linearly coupled neurons, the synchronization threshold of the coupling strength is around $k_3 = 0.255$.

If figure 5 only shows the (x_1, x_2) view of the phase portraits of system (10), a similar phenomenon is observed for between x_1 and x_3. Therefore, the minimum coupling strength k_3 to exhibit synchronization between the first variables x_1, x_2 and x_3 of three *HR* neurons is about $k_3 = 0.255$. Once again, we observe also this phenomenon for the variables y and z for the same numerical values of parameters. This is a synchronization threshold.

There are different possible ways for coupling four, five or more neurons. However, as we already said, we consider the case in which each neuron is connected to all the others. Therefore, there is a unique possible connection scheme.

Here, we consider system (3) with coupling function (4) for $n = 4$, $n = 5$, $n = 6$, $n = 7$, $n = 8$. Figures 6, 8, 10, 12 and 14 show the connection graphs of the network and the connectivity matrix C_n respectively in the case of 4, 5, 6, 7 and 8 neurons coupled all together.

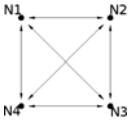
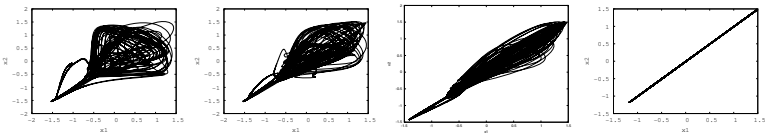


Fig. 6. Bidirectional connection scheme for four neurons. The connectivity matrix C_4 is $C_4 = 1_4 - Id_4$.

Figures 7, 9, 11, 13, 15 show the evolution of the (x_1, x_2) phase portrait while making the coupling strength increase in the case of 4, 5, 6, 7 and 8 neurons coupled all together using the coupling function given in (4). These phase portraits are plotted after a period of time. These figures show x_2 according to x_1, but the same motion is observed comparing the behaviour of the first variable of the other corresponding variables.



(a) (b) (c) (d)

Fig. 7. x_2 according to x_1 when four neurons are connected for the coupling strength (a) $k_4 = 0.1$, (b) $k_4 = 0.15$, (c) $k_4 = 0.18$, (d) $k_4 = 0.191$. Minimum coupling strength to observe synchronization between variables x_i and x_j, y_i and y_j, z_i and z_j $(i = 1, ..., 4, \quad j = 1, ..., 4, \quad i \neq j)$ when coupling four neurons is $k_4 = 0.191$.

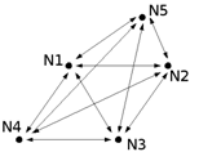
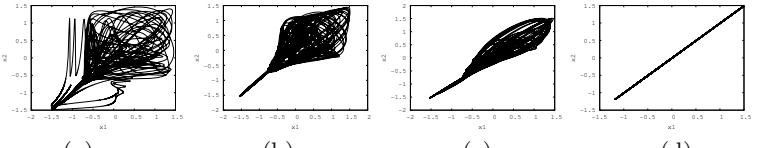


Fig. 8. Coupling five neurons. The connectivity matric C_5 is $C_5 = 1_5 - Id_5$.



(a) (b) (c) (d)

Fig. 9. x_2 according to x_1 when five neurons are connected for the coupling strength (a) $k_5 = 0.05$, (b) $k_5 = 0.11$, (c) $k_5 = 0.14$, (d) $k_5 = 0.152$. Minimum coupling strength to observe synchronization between variables x_i and x_j, y_i and y_j, z_i and z_j $(i = 1, ..., 5, \quad j = 1, ..., 5, \quad i \neq j)$ when coupling five neurons is $k_5 = 0.152$.

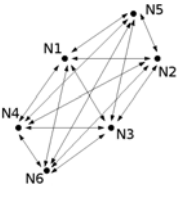
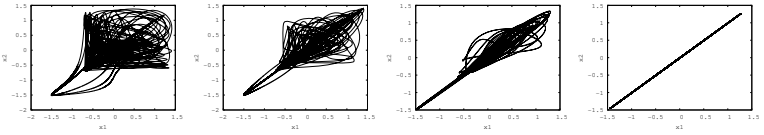


Fig. 10. Coupling six neurons. The connectivity matrix C_6 is $C_6 = 1_6 - Id_6$.



(a) (b) (c) (d)

Fig. 11. x_2 according to x_1 when six neurons are connected for the coupling strength (a) $k_6 = 0.03$, (b) $k_6 = 0.05$, (c) $k_6 = 0.1$, (d) $k_6 = 0.122$. Minimum coupling strength to observe synchronization between variables x_i and x_j, y_i and y_j, z_i and z_j $(i = 1, ..., 6, \quad j = 1, ..., 6, \quad i \neq j)$ when coupling six neurons is $k_6 = 0.122$.

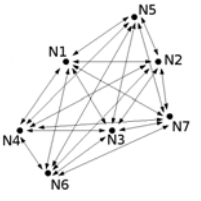
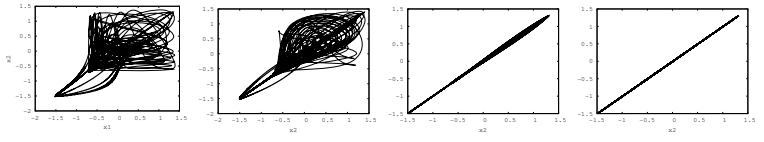


Fig. 12. Coupling seven neurons. The connectivity matrix C_7 is $C_7 = 1_7 - Id_7$.



(a) (b) (c) (d)

Fig. 13. x_2 according to x_1 when seven neurons are connected for the coupling strength (a) $k_7 = 0.025$, (b) $k_7 = 0.05$, (c) $k_7 = 0.1$, (d) $k_7 = 0.104$. Minimum coupling strength to observe synchronization between variables x_i and x_j, y_i and y_j, z_i and z_j $(i = 1, ..., 7, \quad j = 1, ..., 7, \quad i \neq j)$ when coupling seven neurons is $k_7 = 0.104$.

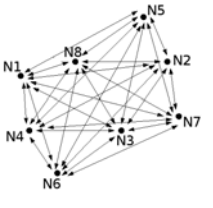
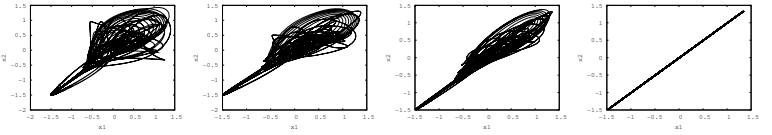


Fig. 14. Coupling eight neurons. The connectivity matrix C_8 is $C_8 = 1_8 - Id_8$.



(a) (b) (c) (d)

Fig. 15. x_2 according to x_1 when eight neurons are connected for the coupling strength (a) $k_8 = 0.05$, (b) $k_8 = 0.07$, (c) $k_8 = 0.08$, (d) $k_8 = 0.088$. Minimum coupling strength to observe synchronization between variables x_i and x_j, y_i and y_j, z_i and z_j $(i = 1, ..., 8, \quad j = 1, ..., 8, \quad i \neq j)$ when coupling eight neurons is $k_8 = 0.088$.

Synchronization thresholds of the coupling strength we obtained are summarized in table 1.

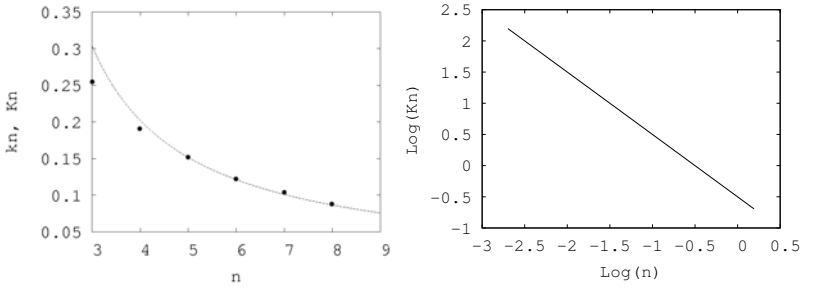
n	2	3	4	5	6	7	8
k_n	0.376	0.255	0.191	0.152	0.122	0.104	0.088
K_n		0.303	0.202	0.151	0.121	0.101	0.086

Table 1. Tabular summarizing minimal coupling strength k_n to observe synchronous motion of n neurons, with $n = 2, ..., 8$ when coupling identical *HR* neurons using the linear function (4). K_n is the theoretical value of coupling strength obtained with the heuristic law given in (11).

In this tabular, the first column shows the number n of neurons in the network. The second column shows the numerical results we obtained while the last column shows the theoretical results obtained thanks to the heuristic law we observed,

$$K_n = \frac{k_2}{0.62 * (n-1)} \tag{11}$$

Figure 16 (a) uses table 1 to show k_n versus n (dots) and K_n versus n (solid lines). Figure 16 (b) shows the curve obtained plotting $Log(K_n)$ according to $Log(n)$.



(a) (b)

Fig. 16. (a) Minimum coupling strength k_n to observe a synchronous behaviour of n neurons, according to the number of neurons in the network (dots) plotted together with $K_n = k_2/(0.62(n-1))$ (solid line). (b) $Log(K_n)$ according to $Log(n)$.

The curve of figure 16 (b) corresponds to a classical law (for n large enough) which can describe many self-organized complex systems, like earthquakes (for 1000 earthquakes of magnitude 4 on Richter scale, there are only 100 of magnitude 5 and 10 of magnitude 6), linguistics (for 1000 occurrences of "the" in an english text, there are only 100 occurrences of "I" and 10 of "say"), urban systems (big cities are rare and small ones are frequent in an exponential way).

2.2 Coupling non-identical *HR* neurons

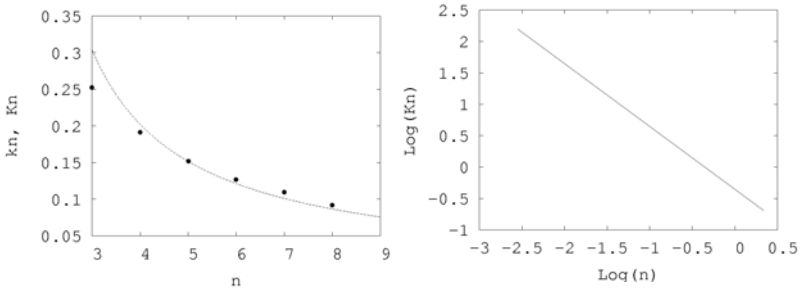
In this subsection, the same coupling function (4) is used to study the asymptotic behaviour of networks composed of *HR* neuron which parameters are slightly different from one another. Indeed, in nature, it is not realistic to assume that a neuron network is composed of exactly identical neurons. We then make all parameters vary slightly from one neuron to another. Let p_i be a generic way of writing every parameters of neuron i (i.e. a_i, α_i, b_i, c_i). The variation of parameters from one neuron to another is defined as in (12). It has been choosen around $e = 10^{-4}$,

$$\begin{cases} p_i = p_1 + (1 + \dfrac{i}{2}10^{-1}) \times e, \qquad i = 2j, \qquad j = 1, 2, 3, 4 \\ p_i = p_1 + (1 - \dfrac{i-1}{2}10^{-1}) \times e, \; i = 2j+1, \; j = 1, 2, 3 \end{cases} \tag{12}$$

With these variations of parameters from a neuron to another, we obtain the following tabular (Table 2) of minimal coupling strength to obtain synchronization, and figure 17 showing a similar law as given in figure 16.

n	2	3	4	5	6	7	8
k_n	0.376	0.252	0.192	0.152	0.127	0.11	0.092
K_n		0.303	0.202	0.151	0.121	0.101	0.086

Table 2. Tabular summarizing minimal coupling strength k_n to observe synchronous motion of n neurons, with $n = 2, ..., 8$ when coupling slightly distinct *HR* neurons using linear function (4). The difference between neurons is around 10^{-4}.



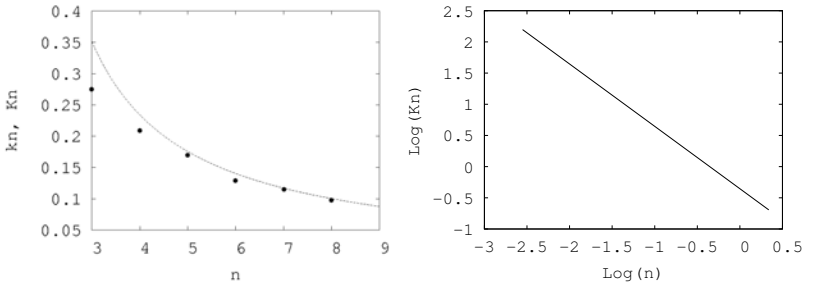
(a) (b)

Fig. 17. (a) Minimum coupling strength k_n to observe a synchronous behaviour of n neurons according to the number of neurons in the network n (dots) plotted together with $K_n = k_2/(0.62(n-1))$ (solid line). (b) $Log(K_n)$ according to $Log(n)$.

Then, in order to make the difference between neurons bigger, the variation is chosen as done in (12) with $e = 10^{-3}$. With these variations of parameters from a neuron to another, we obtain the following tabular (Table 3) of synchronization thresholds, and figure 18 showing a similar law as previously presented.

n	2	3	4	5	6	7	8
k_n	0.4	0.275	0.210	0.170	0.130	0.115	0.098
K_n		0.351	0.234	0.175	0.140	0.117	0.100

Table 3. Tabular summarizing minimal coupling strength to observe synchronous motion of n neurons, with $n = 2, ..., 8$ when coupling slightly distinct *HR* neurons using linear function (4). The difference betwenn neurons is around 10^{-3}.



(a) (b)

Fig. 18. (a) Minimum coupling strength k_n to observe a synchronous behaviour of n neurons according to the number of neurons in the network (dots) plotted together with $K_n = k_2/(0.57(n-1))$ (solid line). (b) $Log(K_n)$ according to $Log(n)$.

3 Property emerging from synchronization with a nonlinear coupling

3.1 Coupling identical *HR* neurons

For $n = 2$, system (3) with the sigmoïdal coupling function h defined as in (5) and Γ defined as in (6) reads as,

$$\begin{cases} \dot{x_1} = ax_1^2 - x_1^3 + y_1 - z_1 - (x_1 - V)k_2 \dfrac{1}{1 + exp(-\lambda(x_2 - \Theta))} \\ \dot{y_1} = (a + \alpha)x_1^2 - y_1 \\ \dot{z_1} = \epsilon(bx_1 + c - z_1) \\ \\ \dot{x_2} = ax_2^2 - x_2^3 + y_2 - z_2 - (x_2 - V)k_2 \dfrac{1}{1 + exp(-\lambda(x_1 - \Theta))} \\ \dot{y_2} = (a + \alpha)x_2^2 - y_2 \\ \dot{z_2} = \epsilon(bx_2 + c - z_2) \end{cases} \tag{13}$$

For $n = 3$, system (3) with the coupling function h defined as in (5) and Γ defined as in (6) reads as,

$$\begin{cases} \dot{x_1} = ax_1^2 - x_1^3 + y_1 - z_1 - (x_1 - V)k_3\Big(\frac{1}{1+exp(-\lambda(x_2-\Theta))} + \frac{1}{1+exp(-\lambda(x_3-\Theta))}\Big) \\ \dot{y_1} = (a+\alpha)x_1^2 - y_1 \\ \dot{z_1} = \epsilon(bx_1 + c - z_1) \\ \\ \dot{x_2} = ax_2^2 - x_2^3 + y_2 - z_2 - (x_2 - V)k_3\Big(\frac{1}{1+exp(-\lambda(x_1-\Theta))} + \frac{1}{1+exp(-\lambda(x_3-\Theta))}\Big) \\ \dot{y_2} = (a+\alpha)x_2^2 - y_2 \\ \dot{z_2} = \epsilon(bx_2 + c - z_2) \\ \\ \dot{x_3} = ax_3^2 - x_3^3 + y_3 - z_3 - (x_3 - V)k_3\Big(\frac{1}{1+exp(-\lambda(x_1-\Theta))} + \frac{1}{1+exp(-\lambda(x_2-\Theta))}\Big) \\ \dot{y_3} = (a+\alpha)x_3^2 - y_3 \\ \dot{z_3} = \epsilon(bx_3 + c - z_3) \end{cases}$$

The same numerical experiments as in the previous section lead us to table 4, in which theorical coupling strength K_n given by the heuristic law (14) is also presented,

$$K_n = \frac{k_2}{(n-1)} \tag{14}$$

n	2	3	4	5	6	7	8
k_n	1.26	0.63	0.5	0.3	0.24	0.21	0.17
K_n		0.63	0.42	0.315	0.252	0.21	0.18

Table 4. Tabular summarizing minimal coupling strength k_n to observe synchronous motion of n neurons, with $n = 2, ..., 8$ when coupling identical *HR* neurons using linear function (4) and the coupling strength obtained with the heuristic law given in (14).

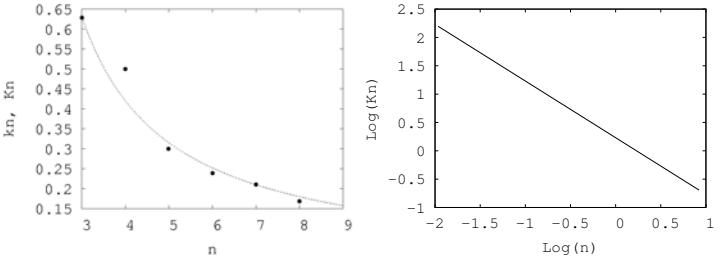


Fig. 19. (a) Minimum coupling strength k_n to observe a synchronous behaviour of n neurons according to the number of neurons in the network (dots) plotted together with $K_n = k_2/(n-1)$ (solid line). (b) $Log(K_n)$ according to $Log(n)$.

As the previous numerical results show, the synchronization threshold follows the heuristic law,

$$K_n = k_2/(n-1) \tag{15}$$

This result has been first presented in [3].

3.2 Coupling non-identical *HR* neurons

In this subsection, the coupling function (5), with Γ defined as in (6), is used to study the asymptotic behaviour of networks composed of coupled *HR* neurons, the parameters of which are slightly different from one another, as in subsection 2.2. The variation of parameters from one neuron to another is defined as in (12) with $e = 10^{-4}$. With these different neurons, we obtain the following tabular (Table 5) which gives the synchronization thresholds from two to eight coupled neurons. The obtained results are similar to those we observed in the previous section. They show the same heuristic law given by equation (15). These results plotted figure 20.

n	2	3	4	5	6	7	8
k_n	3.8	1.8	1.2	0.9	0.7	0.6	0.2
K_n		1.9	1.26	0.95	0.76	0.63	0.54

Table 5. Tabular summarizing minimal coupling strength k_n to observe synchronous motion of n neurons, with $n = 2, ..., 8$ when coupling slightly distinct *HR* neurons using the nonlinear function (5) Γ defined as in (6). The third column shows the coupling strength obtained with the heuristic law given in (15).

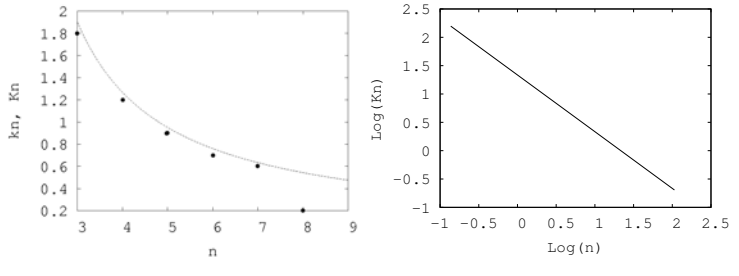


Fig. 20. (a) Minimum coupling strength k_n to observe a synchronous behaviour of n neurons according to the number of neurons in the network (dots) plotted together with $K_n = k_2/(n-1)$ (solid line). (b) $Log(K_n)$ according to $Log(n)$.

4 Conclusion

Emergence and complexity refer to the appearance of higher-level properties and behaviours of a system that obviously comes from the collective dynamics of that system's components, see [2]. These properties are not directly deductable from the lower-level motion of that system. Emergent properties are properties of the 'whole' that are not possessed by any of the individual parts making up that whole. For example, an air molecule is not a cyclone, an isolated species does not form a food chain and an isolated neuron is not conscious : emergent behaviours are typically novel and unanticipated. Thus, if a synchronization phenomenon is exhibited in a neuron network, it may develop some kind of consciousness. This is an emergent property which comes from the collective dynamics of n neurons. Moreover, as given in figure (19), as the number of neurons n increases, the synchronization threshold k_n decreases. Consciousness is more important when the number of neurons is larger.

References

1. Aziz-Alaoui, M.A.: Synchronization of chaos. Encyclopedia of mathematical physics 5, 213–226 (2006)
2. Aziz-Alaoui, M.A.: Complex emergent properties and chaos (De) synchronization. In: Emergent Properties in Natural and Artificial Dynamical Systems, Understanding complex systems, pp. 129–147. Springer, Heidelberg (2006)
3. Belykh, I., Lange, E., Hasler, M.: Synchronization of Bursting Neurons: What matters in the Network Topology. Phy. Rev. Lett. 94(18), 188101.1–188101.4 (2005)
4. Boccaletti, S., Kurths, J., Osipov, G., Valladares, D., Zhou, C.: The synchronization of chaotic systems. Physics Reports 366, 1–101 (2002)
5. Corson, N., Aziz Alaoui, M.A.: Asymptotic dynamics of the slow-fast Hindmarsh-Rose neuronal model (to be submitted) (2008)
6. Gonzalez-Miranda, J.M.: Complex bifurcation structures in the Hindmarsh-Rose neuron model. International Journal of Bifurcation and Chaos 17(9), 3071–3083 (2007)
7. Hindmarsh, J.L., Rose, R.M.: A model of the nerve impulse using two first-order differential equations. Nature 296, 162–164 (1982)
8. Hindmarsh, J.L., Rose, R.M.: A model of neuronal bursting using three coupled first order differential equations. Proc. R. Sc. Lond. B221, 87–102 (1984)
9. Mosekilde, E., Lading, B., Yanchuk, S., Maistrenko, Y.: Bifurcation structure of a model of bursting pancreatic cells. BioSystems 63, 3–13 (2001)
10. Innocenti, G., Morelli, A., Genesio, R., Torcini, A.: Dynamical phases of the Hindmarsh-Rose neuronal model: Studies of the transition from bursting to spiking chaos. Chaos 17, 043128, 1–11 (2007)
11. Izhikevich, E.M.: Dynamical systems in neuroscience - The geometry of excitability and bursting. The MIT Press, Cambridge (2007)
12. Pecora, L.M., Caroll, T.L.: Master stability functions for synchronized coupled systems. Phys. Rev. Lett. 80(10), 2109–2112 (1998)

13. Pikovsky, A., Rosenblum, M., Kurths, J.: Synchronization, A universal concept in nonlinear sciences. Cambridge Nonlinear Science Series 12 (2001)
14. Terman, D.: Chaotic spikes arising from a model of bursting in excitable membranes. SIAM J. Appl. Math. 51(5), 1418–1450 (1991)
15. Terman, D.: The transition from bursting to continuous spiking in excitable membrane models. J. Nonlinear Sc. 2, 135–182 (1992)
16. Wang, X.J.: Genesis of bursting oscillations in the Hindmarsh-Rose model and homoclinicity to a chaotic saddle. Physica D 62(1-4), 263–274 (1993)
17. Yang, Z.Q., Lu, Q.S.: Characteristics of Period-Adding Bursting Bifurcation Without Chaos in the Chay Neuron Model Chin. Phy. Let. 21(11), 2124–2128 (2004)

Validation Of A Distributed 'SmartSpace' Architecture Through Simulation

Ian Oliver

Nokia Research Center
Itämerenkatu 11-13
Helsinki
Finland

Summary. The use of formal methods in software and systems development has advantages but often suffers from difficulty in applying such methods and the visualisation of the results. We show here that the use of a light-weight formal specification language - Alloy - and the use of simulation traces similar to that of scenario explorations with object-diagrams in UML can be effective in understanding the behaviour of such systems at analysis and design-times.

Keywords: Simulation, Validation, Formal Specification, Alloy

1 Introduction

The abundance of personal, mobile devices such as mobile phones or laptop computer with significant computational power, large memories - both temporary and persistent, multi-tasking operating system and with high-speed wireless communication facilities such as UMTS, WLAN and Bluetooth provides the potential for a sophisticated, distributed computation environment which we term a 'SmartSpace' [24].
Based loosely upon the original Linda communication model, blackboard architectures and now Tuple-spaces based computation environments [27,6]have been researched and developed for an extensive period of time. Indeed some of the principles can be found in communication models such as Linda and also in some sophisticated AI work. Later blackboard architectures have been proposed as models for agent systems and suggested as bases for Semantic Web [4] architectures [18]. While there exist a number of blackboard and tuple-space based architecture implementations it is only now with the advent of high-speed data transfer mobile communications and the ubiquity of mobile [25] and embedded devices with the addition of semantic web capabilities (cf: semantic gadgets and infrastructure [19])that these types of system can start to be fully exploited.

We have developed [24] an architecture based upon an augmented blackboard system for the linking together of processes running upon various mobile devices though the notion of a 'knowledge space' (cf: [14]) in a dynamic semantic web context. The blackboard model is augmented to provide a publish-subscribe mechanism with the notion of passing *facts* (items of *knowledge*) rather than messages. This provides us with an implicit communication method. the motivation for this work was to investigate the provision of information sharing between heterogeneous clusters of devices (including mobiles).
We chose to use as the development method for this particular system a pragmatic formal approach using the Alloy specification language [23]. Alloy was chosen for the ease of use, its ability to specify dynamic systems and its suitability for producing execution traces and analyses which are more akin to how many engineers visualise a system. In addition the Alloy language has a closer semantic correspondence with defacto standard languages such as the ubiquitous UML [9,2,11]. In some previous projects have used languages such as B in isolation for the development of such systems with varying results [22] due to the strictness of the B-Method approach.
In this paper we describe our 'Smart Space' system concepts through the use of a domain modelling approach and formalisation in Alloy along with various simulations of the operations of that system. We then make discussion on the validation of the demarcation or policy structure used in that system and following this describe how these models are integrated and simulated.

2 Smart Space Domain Model

The *SmartSpace System* provides a method for concurrently running agent-like processes known as *Nodes* to interact with *SmartSpaces* which provide publish-subscribe mechanisms across a space-based communication environment in which are contained *facts* represented using the Resource Description Framework (RDF). Figure 1 shows the structure of this described using the UML class diagram notation.
Applications are constructed from dynamic combinations of Nodes at design and run-times across multiple, heterogeneous devices. Nodes interact with the SmartSpace through *Sessions* which provide facilities for the Node to add and remove facts and also subscribe and query for patterns of facts. Sessions can also provide additional pre/post-processing of the information as well as other functionalities relates to encryption and security. They also are used to hide the underlying communication mechanisms and technologies. The SmartSpace itself manages the facts and input and requested for by the Nodes. These facts however are pieces of information which means that the normal rules of message passing as one normally sees in a Linda-type system do not apply - communication here is through implicit communication as a form of publish/subscribe. Typically in a Linda system the messages are either unordered (set based) or ordered (queue based) with duplication

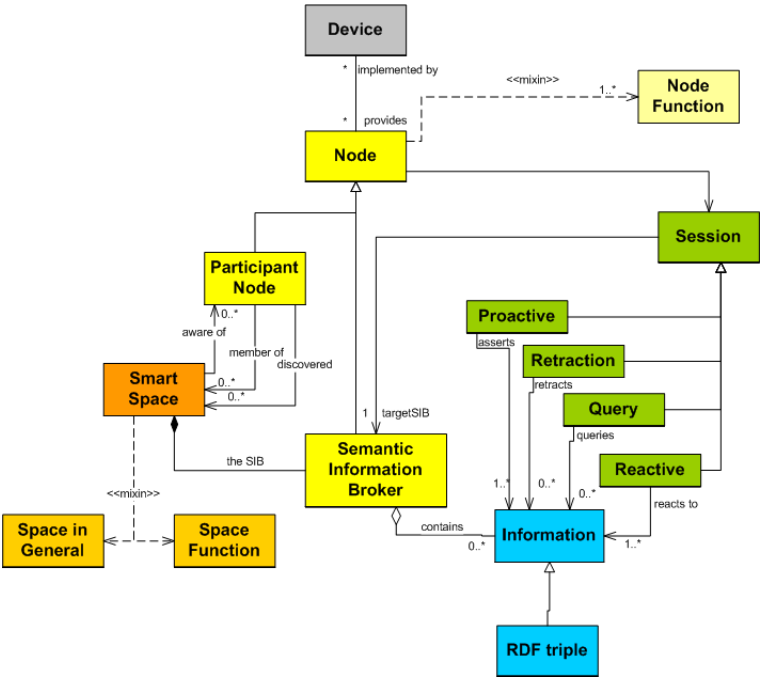


Fig. 1. SmartSpace System Domain Model

allowed and consumption of the messages explicit. We modify this to work with *information* which behaves in our current implementations according to the rules of the well-formed semantics.

In order to interact with a SmartSpace various mechanisms to associate Nodes and SmartSpaces together are utilised. A Node may actively search for SmartSpaces through a discovery mechanism and once a SmartSpace has been found a Node may then request to explicitly join that SmartSpace. Similarly a SmartSpace may actively search for Nodes and subsequently invite those Nodes into the SmartSpace.

In addition to this there are various criteria which a Node has to satisfy in order to join a SmartSpace. A SmartSpace can be demarced by any number of criteria (fig.2) such as the positioning, acceptable security and identity credentials, and so on. In order for a join operation to be successful a Node must comply with these demarcation requirements.

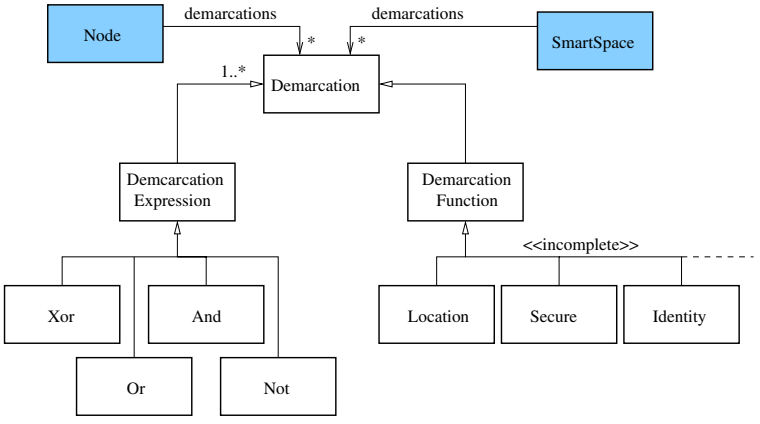


Fig. 2. Demarcation Model in UML

3 Simulation and Validation of the Domain Model

In order to validate the requirements of the system we constructed a description using the Alloy specification language [15]. This description being based upon the model in figure 1 and abstracted to concentrate on the key concepts in that model; the resultant domain model is shown in figure 3 which corresponds to the Alloy specification below (Note the relationship to Time which is explained in due course):

```
sig SmartSpace {
    aware_of :   Node set -> Time,
    facts : Fact set -> Time }

sig Node {
    discovered :   SmartSpace set-> Time,
    memberOf :   SmartSpace set-> Time,
    sessions : Session set->Time }

sig Session {
    target : one SmartSpace,
    buffer:  Fact set -> Time }

sig Fact {}
```

A number of invariants are associated with this model which control more precisely the relationships between the instances of these signatures - in UML one would use OCL for this purpose but this is rarely done in practice due to the lack of support for OCL and its analysis in the UML context.

```
fact NodeMemberOfDiscovery {
all t:Time, n : Node | n.memberOf.t in n.discovered.t }
```

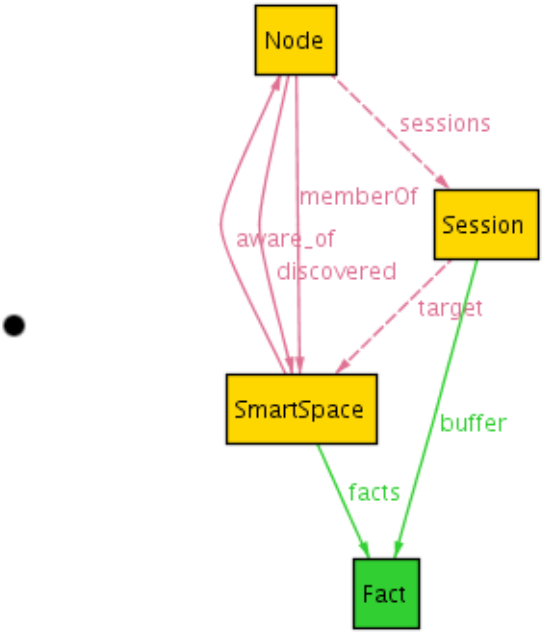


Fig. 3. Abstracted Domain Model

```
fact NodeSessionsSpaceCommutativity {
all t:Time, n: Node | (n.sessions.t).target in n.memberOf.t }

fact NoFinSe{
all se:Session,t:Time | #(sessions.t).se=0 => #buffer=0 }

fact FactsNotSharedAmonstSessions{
all f:Fact, t : Time |
 lone se:Session |
f in se.buffer.t }

fact NoSessionWithoutNode {
all n:Node, t:Time |
no se:Session |
se not in n.sessions.t }
```

This model is then augmented by the addition of the notion of time and a method for ordering events through time as a trace; it is this ability to simulate particular traces that allows us to explore particular interesting scenarios.
The traces are constructed depending upon the needs of that simulation run. We can provide a description of an empty initial state (the predicate `init...` denoting this is commented out in the specification below) which is used when we wish to simulate whether it is possible to reach a certain state, otherwise we let the Alloy

tool choose any suitable starting state which admits the events we wish to run. In this case it is not guaranteed that the said starting state is reachable from an empty initial state though we can check for this. An example trace predicate that admits any starting state and one explicit event per time step is given below. If we wish to admit more than one event per time step then we change the statement `one e:Event` to `some e:Event`.
Time is modelled as linear time or a total order of time points - Alloy provides a library for ordering which includes notions of first, next, previous and last over the given signature. Time has no meaning beyond this ordering and the syntactical use of the name 'Time' - no granularity nor duration is implied.

```
open util/ordering[Time] as TO

sig Time {}

fact Traces {
--  init [first]
   all e:Event |
      e.t' = e.t.next
      all x: Time - last |
         one e:Event | e.t = x }

pred init (t: Time) {
no aware_of.t
no discovered.t
no memberOf.t
no sessions.t
no facts.t }
```

The second modelling choice is to choose the amount of concurrency: we can choose at each step in time to simulate a single event or potentially multiple events occurring simultaneously; generally we choose the former for simplicity in reading the resulting traces. Furthermore when expressing the pre and post-conditions of the events we are free to decide on the extent of the frame-rule [5] for those events. In a sequential, atomic semantics as seen in languages such as B/EventB [1] the frame-rule is explicit and states that 'nothing else happens' however in the case of a distributed system with many points of concurrency this can prove too restrictive and mean that some interesting properties and situations may be lost; the disadvantage of this being that with concurrency it can be difficult to see what is supposed to happen with the parts of the system you are specifically interested in. Furthermore restricting the concurrency, especially in the distributed system, at specification-time has deep architectural implications later on in the development with regards to synchronisation protocols across the system.

3.1 Behaviour Simulation

The events upon the model were expressed as first class signatures in their own right leading to a structure based upon the responsibility and scope of the event. All events are expressed across two points in time: **t** and **t'**, from this are derived

specialisations with additional parameters and constraints or frame-rules if necessary as shown by the NSessFact_Event abstract signatures:

```
abstract sig Event {t,t':Time}

abstract sig NSessFact_Event extends Event{
n:Node,
se:Session,
f: one Fact }
   {    only_NseF[t,t',n,se]  }

pred only_NseF (  t,t':Time, n:Node, se:Session){
 let others = Node - n {
  others <: discovered.t' =
    others <: discovered.t
  aware_of.t' :> others =
    aware_of.t :> others
  others <: memberOf.t' =
    others <: memberOf.t  }
}
```

The two *concrete* events used in this example are given below. The first states that when a fact is sent by a node to a space via a given session that the session is 'owned' by that node, the fact is not a duplicate of a fact already in the buffers and the fact does not already exist in the system a whole. These facts appear strange but ensure that this operation selects a new fact that has not already been used in the system to this point. The result of this event is to add that fact to the buffer of the session.

```
sig SendFact extends NSessFact_Event{}
{
   se in n.sessions.t
   all sessions:Session |
      f not in sessions.buffer.t
   all s : SmartSpace |
      f not in s.facts.t
   se.buffer.t' = se.buffer.t'  + f
}
```

The following event checks that the given buffer is not empty in the before-state and in the after-state ensures that the buffer of that session is empty and that the facts formerly held in the buffer are transferred to the space.

```
abstract sig SchedulerSession_Event extends Event{
    se:  Session,
     ss: SmartSpace  }{}

sig SchedulerTakeFacts extends
   SchedulerSession_Event{}
{
   #se.buffer.t > 0
```

```
   ss.facts.t' = ss.facts.t + se.buffer.t
   no se.buffer.t'
}
```

To drive the simulation one instructs Alloy to 'run' for a number of time steps without any preference to which events should occur; this provides an idea of how the system might behave. Typically we look for particular scenarios such as when a Node sends a fact via some Session and the SmartSpace scheduler collects the waiting facts from that Session's buffer (see event declarations above) as shown in the following example.

```
run { some e:SendFact, f:SchedulerTakeFacts,
    s:Session | e.t' = f.t and e.se = f.se
       and #s.buffer.(e.t)>2
       and s.buffer.(e.t) in s.buffer.(f.t)
 } for 5 but 3 Time,2 Session
```

We also restrict this simulation run to where the session buffer at the start contains more than two facts already and that nothing else changes the state of the buffer in the meantime - at least nothing is removed and perform this for a maximum of 5 objects of each type, 3 points in time (number of events + 1 at least) and 2 sessions (for convenience). From this we obtain three models[1] one for each point in time: figures 4,5 and 6.

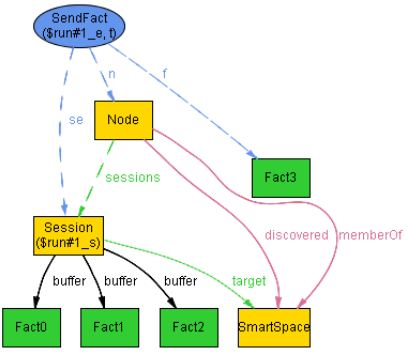


Fig. 4. Simulation (1) - Time 0

It is seen from this simulation that all the facts are indeed transferred across to the SmartSpace and nothing is left in the Node's buffer. If no trace is found then Alloy generates a counter-example which then details where the trace fails.
Note that Alloy only checks a finite state space and is used in according with the principles of the 'small scope hypothesis' [3] which states that most errors can be

[1] Actually only one model is produced but we project over the Time to obtain graphs for each of the time points individually

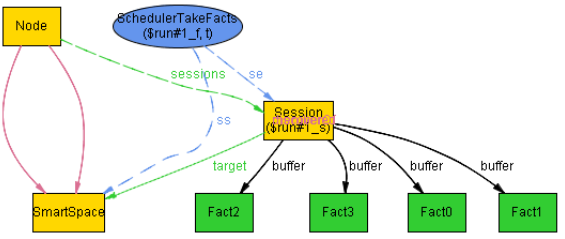


Fig. 5. Simulation (1) - Time 1

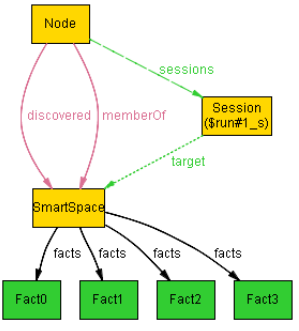


Fig. 6. Simulation (1) - Time 2

found within small examples and that it is more effectively to perform systematic validation with smaller scopes than to work with fewer tests over larger scopes. Our experience has been that given the size of events and structures that these events are operating over simulation each one in small cases provides a much larger body of easily comprehensible information rather than having to understand the results of one much larger and longer simulation.

We can explore traces related to concurrency by weakening the trace predicate to admit more than one event per time step but running the same query - an example simulation then produces the models as in figures 7,8 and 9.

In this case we have a more complex example generated with two sessions both of which have populated buffers. It is interesting to node the during the course of sending Fact4 to the buffer of Session0, Fact3 is removed. This *may* be due to an error in the specification or a concurrent action removing the fact from the buffer. In this case we are confident that another event is occurring. Note that our declarations of the event only state what must be true before-hand and what must be true afterwards, we state nothing about other changes and the atomicity or transactional properties of the event and in this case the model in figure 8 *is correct* with respect to this. We can choose to restrict the concurrency by explicitly 'locking' the buffer during this operation through expression of a frame-rule for this particular event and then rerun the simulation to check the outcome - the results of this simulation would be that the situation seen above would not occur as we would have enforced the atomicity of the event. As we are constructing a

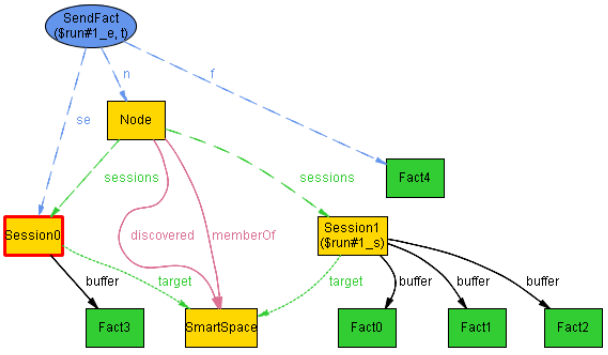


Fig. 7. Simulation (2) - Time 0

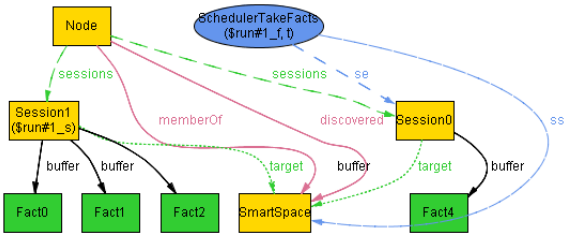


Fig. 8. Simulation (2) - Time 1

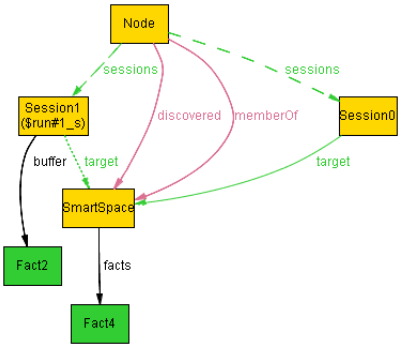


Fig. 9. Simulation (2) - Time 2

distributed system, enforcing atomicity over multiple nodes implies signalling and synchronisation protocols which may not be possible and/or desirable to write.
A similar but even more dramatic situation occurs during the course of the second event resulting in figure 9 where as before the fact is transferred from the buffer to the SmartSpace but also facts are removed from the buffer of the unrelated

Session1. Again this is due to the concurrency/frame-rule issues which may again be desirable to keep weak thus improving the concurrency properties of the system as a whole. To adequately investigate this further simulation runs under differing conditions would of course need to be run.
This is a good example of the small scale hypothesis where it is possible to see the effects of the concurrency in a restricted domain where it is easily decidable whether the concurrency and its effects are desirable. This is something that is extremely difficult or even impossible to ascertain when presented with much larger models where the difference is difficult to distinguish.

4 Demarcation Model

The demarcation model in figure 2 describes using the UML class structure the data structures involved in the policy mechanism by which Nodes are validated before admission to a given SmartSpace. Translation of this UML model into Alloy is straightforward in a similar manner to before. This model can be seen in figure 10 along with some of the invariants:

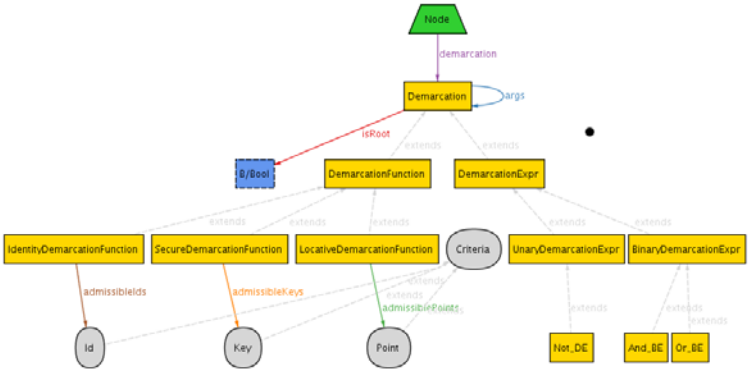


Fig. 10. Demarcation Model Structure

```
fact UnaryExpressionsHaveOneArgument{
all u : UnaryDemarcationExpr | #u.args = 1 }

fact BinaryExpressionsHaveTwoArguments{
all u : BinaryDemarcationExpr | #u.args = 2 }

fact DemarcationFunctionsAreLeafNodes
{
    all d : DemarcationFunction | #d.args = 0 }

fact ExpressionTree
{
```

```
    all u : DemarcationExpr | u not in u.^args
    all u : DemarcationExpr | u not in u.^^args
    all u : DemarcationExpr | u not in u.args }

fact AllDemarcationsContainDemarcationFunctions
{
all n : Node | Demarcation in n.demarcation.^args
}
```

Alloy presents us with the ability to generate particular instantiations of this model; an simple example of which is shown in figure 11.

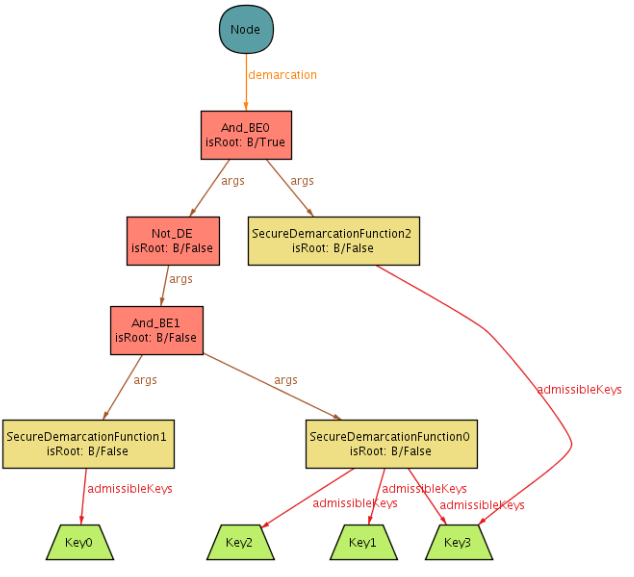


Fig. 11. Demarcation Model Instantiation Example

Through the visualisation of this model we can easily see what particular combinations are possible. By further constraining the model we can disallow certain combinations, for example, locations and identities may not appear in the same demarcation. Obviously as further subtypes of the demarcation functions (specialisation of location etc) are added then these models and their constraints become more complex. If we over-constrain a model Alloy provided us with a counter-example demonstrating that particular case.

5 Integration of Models

Our development has progressed with two separate models of differing parts of the system, though it can be seen from figure 2 that there is a relationship between

these models. Obviously some integration must be performed and the integration itself must be validated. The integration also implies a development step in the model and in this case the joining mechanism between a Node and Space must be updated to admit this policy mechanism.
The first step that must take place is changing the join operation by adding a precondition to check the correspondence between the Node's credentials and the demarcation properties of the Space. A join may only occur if the node is not a member of that particular space and the respective policies match. We also ensure that the node has also discovered the space before-hand and if all of the above are true then the node becomes a member of that space.

```
sig JoinSmartSpace extends NSS_Event{} {
    policyCheck[n,ss]
    ss not in n.memberOf.t
    ss in n.discovered.t
    n.memberOf.t' = n.memberOf.t + ss
}

policyCheck(n:Node,ss:SmartSpace) {
    n.demarcation in ss.demarcation
}
```

For the purposes of simulation and ease of specification we do make a simplification at this stage and ignore the various logical operations that could occur over the demarcation policy - hence the trivially simple policy check predicate. As the purpose of the simulation is to visualise potential situations and understand the join operation in the presence of policy information we consider this to be acceptable[2]. Indeed experience has shown that overloading the specification with precise details at this stage - especially when the simulation is to be presented to both engineers and non-engineers alike - just complicates the discussion about the system as a whole rather than concentrating on the essentials of the operation under investigation.
Alloy in the successful case will generate two points in time for the before and after states of the successful join operation as shown in figures 12 and 13. If we continue the time steps after this event we notice that the system behaves as previously. A more thorough check would involve simulating the cases previously without the demarcation criteria and comparing them more deeply.
We can also force the system to check if join is possible if the demarcations do not match. In this case as the specification suggests and the simulation run demonstrates no successful instantiation of the model in this circumstance is possible.
Of course successful integration of the models also means careful modelling at similar abstraction levels though this is an area where Alloy helps because of the simplicity and style of specification. In the case here the work required to 'bind' the two models together involved understanding that the concept of Node in both was the same, the addition off a predicate to strengthen the join event and also to add a link to a demarcation structure from the Space.

[2] for the moment - again simple (simplified) and small examples

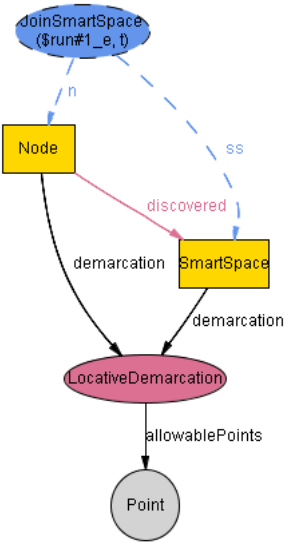


Fig. 12. Join Event with Demarcation - Time 0

6 Related Work

The use of formal methods in software engineering is not new - and Alloy has compelling examples of industrial usage[3] - however the use of more pragmatic methods which do not rely on an absolute strict method of constructing the models coupled with the ability to 'execute' or simulate models is less frequent.
As a comparison work made as part of the EU Rodin project [7] utilising models written in the B/EventB languages and coupled with the CSP [13] language to drive those models using the tool ProB [20] is very much in the same manner as explained here.
One major difference is that the use of CSP means that we are driving the model in a very specific way: we are stating that we wish to see if the model admits the trace as specified by the CSP rather than asking the model if such a trace is possible at all rather than from a particular given state. This means we can not explore individual states for their correctness - this being proven by the theorem prover before-hand - but checking now the overall progress of states occurs and whether that particular trace is admissible.

7 Conclusions and Future Work

Unlike existing solutions from formal methods tools in the software engineering domain, Alloy allows simulations and the visualisation of those simulations of the

[3] http://alloy.mit.edu/publications.php

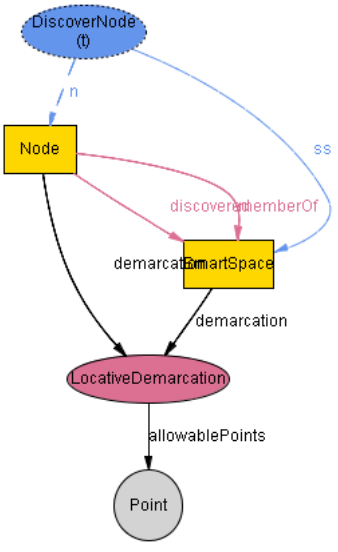


Fig. 13. Join Event with Demarcation - Time 1

behaviour of a system in a manner that is more compatible with the industrial environment . That is, systems work by single stepping through the actions over a period of time. This coupled with Alloy's much freer form of specification - one where the underlying action system is not constrained by a preconceived action semantics with additional constraints such as frame-rules etc makes the ability to explore the system even in the presence of concurrency and additionally makes the simulation even in potentially distributed environments more tractable.

Of course, other formal methods and languages have supporting tools such as Z/EVES [8], ZANS [16] etc, the most relevant of these is the aforementioned ProB tool. B/EventB however suffers from the fact that there is a strict development method and particular semantics attached that makes it unsuitable or at least more difficult and unnatural in our development environment for producing the kinds of simulations shown here.

One particular advantage of Alloy's style of specification and simulation is that it is very similar to the ubiquitous UML language extensively used within industry. The traces that Alloy produces are almost identical to the use of scenario exploration with object diagrams that are employed in a similar manner to those described in [21, 10]. UML however does not have any suitable simulation tools that scale to industrial use well, though a mention should be made of the USE tool [12] which does support in part the OCL which is used in conjunction with the UML for formally expressing actions.

The failure of UML to sufficiently express behavioural properties is known and the over reliance on state diagrams without semantics (or specific tool based semantics) by some methods is a major problem. Beyond expressing the static (class) structure and very general behavioural aspects with responsibility-based design, the

UML provides little to offer. UML's state diagram and static structure diagrams (mistaken for SDL-like block diagrams) are not sufficient. There are very few tools that are capable of analysing models written using the UML - even static analysis is not generally supported expect in some specific cases - again this is an area where formal methods (due to their very nature) have an advantage.
Regarding the overall effect of such a simulation the development of our software, as described in [26], any analysis of the system made early in the development flow has advantages - more detailed discussions of the results can be found in (specifically) [23] and [22]. In this case Alloy affords us the ability to simulate in a more explorative way than competing formal tools. For the engineer, the ability to see the effects of the operations allows much earlier feedback and subsequent modifications to the design coupled with a greater understanding of what is trying to be achieved and indeed does counteract some of the myths of formal development [17].

8 Acknowledgments

Daniel Jackson and Felix Chang for much assistance and help with the Alloy language and Jukka Honkola, Jürgen Ziegler and Antti Lappeteläinen at Nokia Research with the SmartSpace project itself.

References

1. Abrial, J.-R.: The B-Book - Assigning programs to Meanings. Cambridge University Press, Cambridge (1995)
2. Anastasakis, K., Bordbar, B., Georg, G., Ray, I.: UML2Alloy: A Challenging Model Transformation. In: ACM/IEEE 10th International Conference on Model Driven Engineering Languages and Systems (MoDELS) (2007)
3. Andoni, A., Daniliuc, D., Khurshid, S., Marinov, D.: Evaluating the small scope hypothesis (2002)
4. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American 284(5), 34–43 (2001)
5. Borgida, A., Mylopoulos, J., Reiter, R.: On the Frame Problem in Procedure Specifications. IEEE Transactions on Software Engineering 21(10), 785–797 (1995)
6. Bussler, C.: A Minimal Triple Space Computing Architecture. In: Proceedings of the 2nd WSMO Implementation Workshop (WIW 2005), Innsbruck, Austria, June 6-7 (2005)
7. Coleman, J., Jones, C., Oliver, I., Romanovsky, E.T.A.: RODIN (Rigorous Open Development Environment for Complex Systems). In: Fifth European Dependable Computing Conference EDCC-5, Budapest (2005)
8. Craigen, D., Kromodimoeljo, S., Meisels, I., Pase, B., Saaltink, M.: EVES: An Overview. In: Prehn, S., Toetenel, H. (eds.) VDM 1991. LNCS, vol. 551. Springer, Heidelberg (1991)
9. Crane, M., Dingel, J.: Runtime Conformance Checking of Objects Using Alloy. In: Third Workshop on Runtime Verification (RV 2003), Boulder, Colorado (July 2003)
10. D'Souza, D.F., Cameron Wills, A.: Object, Components and Frameworks with UML - The Catalysis Approach. Object Technology Series. Addison-Wesley, Reading (1999)

11. Georg, G., Bieman, J., France, R.: Using Alloy and UML/OCL to specify run-time configuration management: a case study. In: Proc. Workshop on Practical UML-Based Rigorous Development Methods — Countering or Integrating the eXtremists (October 2001)
12. Gogolla, M., Richters, M.: Development of UML Descriptions with USE. In: Shafazand, H., Tjoa, A.M. (eds.) EurAsia-ICT 2002. LNCS, vol. 2510, pp. 228–238. Springer, Heidelberg (2002)
13. Hoare, C.A.R.: Communicating Sequential Processes. Prentice-Hall, Englewood Cliffs (1985)
14. Hui, K.-Y., Chalmers, S., Gray, P., Preece, A.: Experience in Using RDF. In: Agent-mediated Knowledge Architectures, pp. 177–192. Springer, Germany (2004)
15. Jackson, D.: Software Abstractions: Logic, Language, and Analysis. MIT Press, Cambridge (2006)
16. Jia, X.: A Tutorial of ZANS - A Z Animation System. DePaul University, Chicago, USA (May 1995), `http://ise.cs.depaul.edu/pub/ZANS`
17. Hinchey, M.G., Bowen, J.P.: Seven more myths of formal methods. Software 12(4), 34–41 (1995)
18. Khushraj, D., Lassila, O., Finin, T.: sTuples: semantic tuple spaces. In: The First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, 2004. MOBIQUITOUS 2004, August 22-26, pp. 268–277 (2004)
19. Lassila, O., Adler, M.: Semantic Gadgets: Ubiquitous Computing Meets the Semantic Web. In: Fensel, D., Hendler, J.A., Lieberman, H., Wahlster, W. (eds.) Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential, pp. 363–376. MIT Press, Cambridge (2003)
20. Leuschel, M., Butler, M.: ProB: A Model Checker for B. In: Araki, K., Gnesi, S., Mandrioli, D. (eds.) FME 2003. LNCS, vol. 2805, pp. 855–874. Springer, Heidelberg (2003)
21. Oliver, I.: Simulation of Software Behaviour using Animation. In: Amborski, K., Meuth, H. (eds.) Proceedings of European Simulation Multiconference, SCS Europe, vol. 4, pp. 278–283 (2002) ISBN 90-77039-07-4
22. Oliver, I.: Experiences in Using B and UML in Industrial Development. In: Julliand, J., Kouchnarenko, O. (eds.) B 2007. LNCS, vol. 4355, pp. 248–251. Springer, Heidelberg (2006)
23. Oliver, I.: Experiences of Formal Methods in Conveniontal Software and Systems Design. In: BCS FACS Xmas Workshop: Formal Methods in Industry (December 2007)
24. Oliver, I., Honkola, J.: Sedvice: A Triple Space Computing Exploration Environment. In: Tripcom Workshop, Galway, Ireland, April 28-29 (2008)
25. Wade, S.: An Investigation into the use of the Tuple Space Paradigm in Mobile Computing Environments. Ph.D. thesis, Lancaster University, UK (1999)

26. Christopher Westland, J.: The cost of errors in software development: evidence from industry. The Journal of Systems and Software 62(1), 1–10 (2002)
27. Zaremba, M., Moran, M.: Enabling Execution of Semantic Web Services: WSMX Core Platform. In: Proceedings of the 1st WSMO Implementation Workshop (WIW 2004) Frankfurt, Germany (2004)

A Decentralised Approach for the Transportation On Demand Problem

Cyrille Bertelle, Michel Nabaa, Damien Olivier, and Pierrick Tranouez

LITIS, 25 rue Phillipe Lebon, 76058 Le Havre - www.litislab.eu
{Cyrille.Bertelle, Michel.Nabaa, Damien.Olivier,
Pierrick.Tranouez}@univ-lehavre.fr

Summary. Public transport systems are generally organized in a static, a priori way. In such systems, the demand must be adapted to the offer. In this paper, we propose a model based on self-organization in order to dispatch a fleet of vehicles in a purely dynamic Transportation On Demand system (TOD). Our proposal consists in a decentralized approach and a multi-agent system (MAS) to model the environment. This will tackle the problem of vehicles over-concentration or the lack of service in certain areas of the city. We demonstrate that our model addresses these problems by providing vehicle agents, for a given request, to make the final decision thanks to a negotiation process and to calculate overcosts according to an original insertion heuristic.

Keywords: transportation on demand, vehicle routing problem, collective intelligence, self-organization.

1 Introduction

Growing environmental concerns are linked to the difficulties of management of urban traffic. They lead to the creation of new solutions improving mobility in agglomeration. Current public transportation systems are determinist and based on frequencies and routes fixed in advance. They are built starting from opportunity studies and are not very adaptive to a request that can change in time: the demand has to adapt to the offer. It is thus advisable to complete the urban transportation services by flexible systems being more adapted to the individual needs. We propose a Transportation On Demand (TOD) system which must adapt to users need in real time. It must allow to generate an important reduction in the traffic and to offer a maximal quality of service to reduce the cost of exploitation. Lastly, it will be the basis for a decision support system, computing vehicle tours in real time, a service which is not offered by the traditional transportation systems. The stake of this article is to study the possibility of the installation of a TOD system to satisfy the requests of the customers at any moment, by distributing the load inside the fleet of vehicles in order to achieve the goals mentioned previously. This system will adjust dynamically to the customers demand. The scenario of the execution starts with the first customer request which appears randomly in a place of the city. It

sends a request indicating his departure point and his destination. The resolution consists in choosing the best located vehicle to satisfy the passengers already on board this vehicle as well as the new request by optimizing its rate of filling with respect to the maximal capacity, its time and cost of travel.

First of all, we will present some previous work of similar problems. Then, we will define the data of our problem. We will present our approach and the preliminary results related to the initial tests to finish by a conclusion and some perspectives.

2 Previous Works

The general problem of the construction of vehicles routes is known under the name of Vehicle Routing Problem (VRP) and represents a combinatorial problem of multi-objective optimization which was the subject of many works and many alternatives in the literature. It belongs to the NP-hard category [2,10]. In its basic version, the VRP problem (see figure 1) models a well known transportation problem which consists of pickup (and/or collect) products to serve a set of customers using a fleet of vehicles. The resolution consists in determining a set of routes which minimizes objectives as well as possible as the total traveled distance, the number of vehicles used and the sum of customers delays [8].

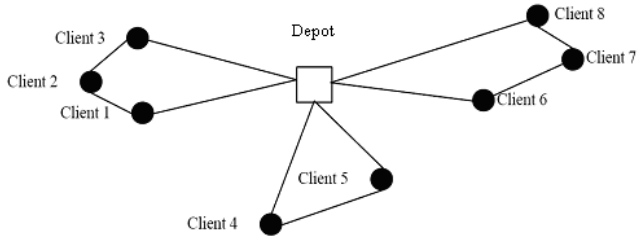


Fig. 1. Vehicle Routing Problem

A complete state of the art of the VRP problems in the static context, and in particular, the dynamic one and their applications is given by [9]. In [13], the Dynamic VRP (DVRP) problem was treated. The resolution consists in dividing time execution into slices. The DVRP problem will be a succession of VRP static problems. An ant colony algorithm was developped to solve these static problems. When it is about the problem of VRP with pickup and delivery of goods, one speaks about PDP (Pickup & Delivery Problem). The Dynamic PDP problem was studied by [6]. The authors adopted a multi agent approach. The conversation between agents was based on the Contract Net Protocol. The clients demand arrival was claculated with basis on Poisson distribution. A Dial-A-Ride Problem (DARP) is an extension of the PDP in which the transportation of goods is replaced by the transportation of people [4,7]. Since we talk about people transportation, the

DARP focuses better than the PDP on the satisfaction of these people. We can distinguish between a static or dynamic version of the DARP although the difference is not always strict. Indeed, in the static case, some reservations can be canceled at the last minute, which implies a degree of dynamism while in the dynamic case, most reservations are known a priori before planning [1]. The problem is then, in general, treated as a succession of static sub-problems [3]. An application in the urban transport related to the bus on demand was developped in [12]: a customer can give a time window in which he wishes to be served, instead of departure or arrival, but not both at the same time. Some requests are known in advance and other ones can arrive during the execution. The authors adopted a solution based on an insertion heuristic which gave interesting results with short execution time . This problem is known as the Dynamic Dial-A-Ride Problem (DDARP) with several places of pickup and delivery. In [11], the DARP problem was treated online by considering a homogeneous fleet of vehicles with unit capacities, i.e. a vehicle having a passenger on board, cannot serve another one except if it reached the first passenger destination. To our knowledge, the dynamic DARP is rarely studied and does not exist in its purely dynamic version. The problem we deal with in this paper is original, because the requests are dealt with in real time: we dispose of a purely dynamic case. This dynamicity is due to the fact that no reservation is known in advance. Moreover, traditional techniques suppose to have a central which knows vehicles positions and their planned routes (the central receives customers requests, calculates new vehicles routes and orders vehicles to service customers). Moreover, these techniques suppose a perfect knowledge by the central of the states of the vehicles (including breaks, breakdowns, communication problems) in real time, which is not realistic and can involve expensive calculations in time. That's why we adopt a decentralized approach (except for the new client request reception) to face this kind of problems; since embarked system is now standardized.

3 Proposed dynamic model

3.1 Problem Description

The model tries to arbitrating between different constraints.
Each user wishes:

- To minimize the waiting time once his demand is accepted,
- To reach his destination, respecting his desired deadline.

Each vehicle tries:

- To maximize its rate of filling by changing an already planned route to service a new request,
- To deal with the evolution of the traffic load and especially unexpected events (accidents, a road becomes blocked up, new roads) and historical events: the system must be adapted to learn about repetitive events to predict similar ones in the future,
- To negotiate with the other vehicles in order to choose the best proposal to serve each new request.

The system tends, as well as possible, to pair users and vehicles by arbitrating and adjusting the previous constraints. The system is not centralized but emerges from the fleet of vehicles.
We propose an agent-oriented approach. The system is composed of the following agents: $Vehicle$, $Interface$ and $Client$. The scenario of the execution is described as follows: a user connects to the system via a given support (service call, Web server.), it is then instantiated by a *Client agent* whose function consists in representing it in the system. The user indicates his departure point and his destination as indicated previously. Thus, the *Client agent* enters in interaction with the $Interface$ *agent* (see figure 2).
The latter broadcasts the request of the user to other $Vehicle$ *agents* located in an environment which is modelled in the following section. Our model is specific by the fact that the requests are not dealt with batch but in "real time".

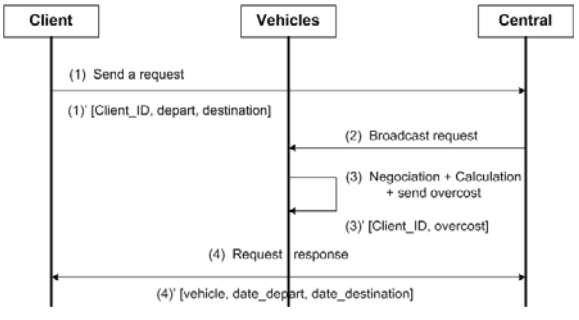


Fig. 2. UML Sequence Diagram showing interactions between the actors of the system

3.2 Environment modeling

We represent the urban network by a directed dynamic graph $G(t) = (V(t); E(t))$ where $V(t)$ is the set of nodes and $E(t)$ the set of arcs:

- The nodes represent interesting places of the network: crossroads, stations, cinemas, commercial centers
- The arcs represent the roads taken by the vehicles,
- The weight on each arc represents the needed time to cross this arc, depending on the current load of the traffic,
- Dynamic aspect relates to the weights of the arcs, which can evolve in time, according to the evolution of the fluidity of circulation. It can be related to the apparition and/or disappearance of arcs.

Customers are associated to a node or vehicles, which are themselves on nodes or arcs. The size of the population of users and vehicles is variable in time, to obtain a day/night simulation mechanism. Once a temporal model of the population of users is established, the population of vehicles must be established accordingly, in order to have a satisfying average rate of occupation for vehicles.

3.3 Offer and dynamic choice

Our model is based on two simultaneous phases, an offer phase and a choice phase. We want to establish an agreement between the proposals for a transport and the needs of the customer. This is done according to a mechanism of negotiation. A key element of the system is pairing vehicles and customers. Which vehicle is the best for servicing a given request? Who determines it and how? How a vehicle knows if it has been selected to service a new demand? These questions are not independent. The best vehicle corresponding to each user will be selected; it must minimize the additional effort $\triangle C$ to service the customers. To know this additional effort, a vehicle calculates, on the one hand, the total cost (in time) of its current route, that of the route to discharge the current passengers and charge already planned ones. On the other hand, a vehicle calculates also the cost of the new route to service actual passengers by including the new one. The difference between the two costs is the additional effort or overcost.
Each customer request is diffused to all vehicles. When receiving a new request, each vehicle calculates its overcost to service the request and diffuses it to all the fleet. Then, it compares the received answers to its overcost. It orders the received offers and broadcasts the head of list. Finally, the determined winner is the one being the most times ranked first in the received answers. Ideally, one could exempt these last phases: if the diffusion is perfect, all the fleet will obtain the correct classification directly. But, because of non perfect diffusion, we proceed to this additional phase after a possible problem (a vehicle crossing a tunnel for example), and this to be sure that all the vehicles agree on the winner vehicle wich will take the passenger.

3.4 The scheduling algorithm

The *vehicle agents* carry out the principal work of planning, and this thanks to the evaluation of the insertion of a travel (source and destination) in such a way to respect the deadlines of the existing passengers in the vehicles. The insertion heuristic is inspired from the ADARTW one [5]. For each vehicle, a scheduling block always starts with the first customer on its way and ends when the last customer reaches its destination. The following figure illustrates the insertion of a customer in a scheduling block of a vehicle having two customers on board ($C1$; $C2$) and going to servicing another client ($C3$) with an already planned route. Each one has a departure point (preceded by a + in figure 3) and a destination (preceded by -). In a block related to a vehicle already containing N clients, corresponds $K = 2N$ stops (2 stops per client) and $(K + 1)(K + 2)/2$ insertion possibilities when its pickup point must precede its delivery point.
In the example of the figure 3, the vehicle can insert the third client X with $(4 + 1)(4 + 2)/2 = 15$ possibilities. For a vehicle V with a maximal capacity Q_v, the number of possibilities $nbPossibil$ increases to $((Q_v-1)*2+1)((Q_v-1)*2+2)/2 = Q_v(2Q_v-1)$ in the worst case. The complexity of the ADARTW for a new insertion is $nbPossibil * (2Q_v - 1) * \Theta(m + nlog(n)) = Q_v(2Q_v - 1)^2 * \Theta(m + nlog(n))$ in the worst case; n and m are the number of nodes and arcs of the graph and $\Theta(m+nlog(n))$ is the complexity of the Dijkstra algorithm to calculate the shortest path in a graph. The objective function $MinZ = \triangle C$ represents the minimal overcost due to a new client insertion. $\triangle C$ depends on the following variables: additional time to service the new demand, current capacity (number of clients on

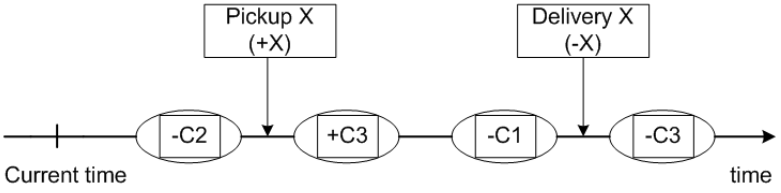


Fig. 3. Insertion heuristic

board) and proximity of the vehicle from the customer. If two or many vehicles give the same value of the overcost for the same customer, the one having the minimal capacity wins this customer. If they have also the same capacity, the winner is the nearest one (in distance).

3.5 Self-organization

The dynamic and nondeterministic aspect of the problem can lead to concentrations of demands in certain zones which are more attractive and may cause a lack of service elsewhere. Indeed, the downtown area, for example, will be a zone of strong attraction at certain hours of the day whereas certain suburban zones become badly serviced. The waiting time of clients in such a zone will then be very important. We thus have chosen to possibly degrade the performances in the attractive zones in spite of having a better service in other areas to avoid any lack of service. Several solutions are possible:

- Injecting some vehicles in the existing fleet but that can violate constraints related to environmental objectives, we could have else a maximum filling of the vehicles,
- A hierarchical centralized resolution which is opposed to the decentralized model we adopt and not very realistic,
- The use of self-organization mechanisms. We chose the last way for its distributed, local and adaptive characteristics.

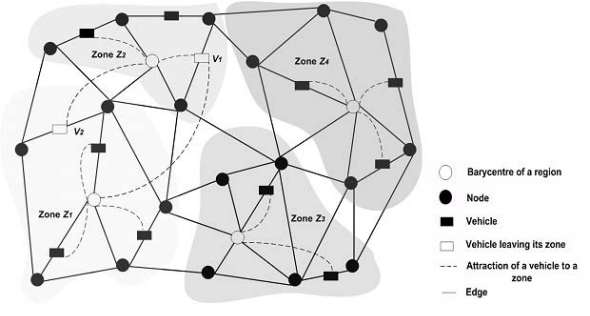


Fig. 4. Attraction of vehicles to zones

Thus, we define zones of attraction within the graph to which the vehicles are attached. These zones will evolve either geographically (the number of node they cover) or according to the number of attached vehicles. These zones have a multi-scale representation by using learning techniques since the vehicles acquire information on the road network state and are able to send the information to the graph.
When a vehicle leaves its zone, it will be penalized in its overcost function which increases while it moves away. This constitutes an exerted attraction force so we can imagine a spring fixed on the zone center by an extremity (see figure 4) and a vehicle is attached to its other extremity. If this fights against the change of a vehicle zone, it does not prohibit the vehicle from changing its zone. Indeed, the vehicle can gain the bid for a demand coming from another zone. When a vehicle leaves its zone, it can negotiate a change of zone with others. In figure 4, the vehicles $V1$ and $V2$ exchange their zones. This part of the modelisation is under developpement.

4 Simulation

Multi-agent proposed architecture was developed by using the REPAST Simphony (Recursive Porus Agent Toolkit Simulation) multiagent platform written in Java which focuses on social simulation [14]. This platform developed by the Argonne laboratory of the University of Chicago, inherits main functionalities from SWARM platform (into Objective C) and offers several advanced functionalities:

- Built-in 2D, 3D, and geographical information systems (GIS) support and tools,
- Automated connections to enterprise data sources: relational databases, GIS and to external programs for statistical analysis and visualization of model results,
- Provides information about the state of each agent,
- A scheduler which supports concurrent discrete events in a sequential or parallel way.

As mentioned previously, the calculation of an overcost related to a new request, is done by each vehicle. At a time step, if a vehicle receives a request, it collects the other vehicles answers (overcost of the other vehicles) and compares their overcost to its own, broadcasts a winner message (if he is the winner), vehicles and the concerned customer before going to service him, as described in 3.3 and detailed in algorithm 1.

5 Preliminary results

We have implemented our model on a graph with 50 nodes, 7200 time steps, with 4-passenger-seat vehicles. The customers appear at random places and hours, and give random destinations. In the table below, the optimal cost column indicates minimal total time to service all the customers from their departure until their destination nodes. The cost column represents the real time in which we serviced all customers. The client optimal itinerary column indicates the itinerary time average of the served clients. Variation per client is the percentage of difference between the two costs. Filling indicates the interval in which the capacity has oscillated.
We notice that the first results are very encouraging. The variation between real cost and optimal cost is not important if we take into account the time to park an

Algorithm 3: Taking into account of a new customer by each vehicle

```
for each vehicle v ∈ fleet do
    while not end of route do
        message = consultation of message box;
        if message! = empty then
            if message ∈ client request then
                computation of the overcost according to ADARTW for
                temporary insertion of the client in the route;
                broadcast overcost;
            else
                if message ∈ overcost then
                    sort overcost;
                    if all overcosts are arrived or timeout started then
                        winner = head (overcosts);
                    else
                        store the winner;
                        if all winners are arrived or timeout started then
                            winner = generally quoted winner;
                    if winner == v then
                        modify route;
```

Vehicle number	Clients number	Served Clients	Optimal cost	Cost	Client optimal itinerary	Variation per client	Served Clients %	Rate of filling
4	227	227	4409	4878	19.42	10.6%	100%	$0 \rightarrow 4$
4	47	47	7324	7922	155	8.2%	100%	$0 \rightarrow 4$
8	357	315	26745	29381	85	9.8%	88%	$0 \rightarrow 4$
8	200	142	16066	17243	113.14	7.31%	71%	$0 \rightarrow 4$

Table 1. Some results of simulation

individual vehicle in the real life case and the cheaper price proposed to customers in collective transportation. We remark that, for a given number of vehicles and a given period of simulation (7200 steps in the above tableau), our model gives better results with a limited number of clients having long itineraries than with a big number of clients having short itineraries. It is because of the supplementary time due to satisfy all clients which is more important in the case of a big number of clients. The simulation must be improved: probabilistic model for the population of customers, better statistics for filling ... The self-organization mechanism is under development.

6 Conclusion and perspectives

In this paper, we presented a Transportation On Demand system which is purely dynamic, in an environment in perpetual change. We have adopted a decentralized approach based on the optimization and negotiation between vehicles. To face the lack of service in certain zones or the over-concentration of vehicles in certain other zones, we have proposed a self-organizational model which can adapt to

the environmental changes. The obtained results are encouraging and the phase of self-organization is under development. We will continue our work by the complete validation of the proposed model.

References

1. Attanasio, A., Cordeau, J.F., Ghiani, G., Laporte, G.: Parallel Tabu search heuristics for the dynamic multi-vehicle dial-a-ride problem. Parallel Comput. 30, 377–387 (2004)
2. Christofides, N., Mingozzi, A., Toth, P.: The vehicle routing problem. In: Combinatorial Optimization, vol. 11, pp. 315–338. John Wiley, Chichester (1979)
3. Cordeau, J.F., Laporte, G., Potvin, J.Y., Martin, W.P.: Transportation on Demand. In: Handbooks in Operations Research and Management Science (2004)
4. Hauptmeier, D., Krumke, S.O., Rambau, J.: The online dial-a-ride problem under reasonable load. In: Bongiovanni, G., Petreschi, R., Gambosi, G. (eds.) CIAC 2000. LNCS, vol. 1767, p. 125. Springer, Heidelberg (2000)
5. Jaw, J.J., Odoni, A.R., Psaraftis, H.N., Wilson, N.H.M.: A heuristic algorithm for the multi-vehicle many-to-many advance request dial-a-ride problem with time windows. Pergamon Journals Ltd. 20B(3), 243–257 (1986)
6. Koźlak, J., Créput, J.C., Hilaire, V., Koukam, A.: Multiagent approach to dynamic pick-up and delivery problem with uncertain knowledge about future transport demands. Fundam. Inf. 71(1), 27–36 (2006)
7. Krumke, S.O., de Paepe, W.E., Poensgen, D., Lipmann, M., Marchetti-Spaccamela, A., Stougie, L.: On minimizing the maximum flow time in the online dial-a-ride problem. In: Erlebach, T., Persinao, G. (eds.) WAOA 2005. LNCS, vol. 3879, pp. 258–269. Springer, Heidelberg (2006)
8. Laporte, G., Gendreau, M., Potvin, J.Y., Semet, F.: Classical and Modern Heuristics for the Vehicle Routing Problem. International Transactions in Operational Research 7, 285–300 (1999)
9. Larsen, A.: The dynamic vehicle routing problem. PhD thesis, Technical University of Denmark (2000)
10. Lenstra, J., Rinnooy Kan, A.H.G.: Complexity of the vehicle routing and scheduling problems. Networks 11, 221–228 (1981)
11. Lipmann, M., Lu, X., De Paepe, W.E., Sitters, R.A., Stougie, L.: On-Line Diala-Ride Problems Under a Restricted Information Model. Algorithmica 40, 319–329 (2004)
12. Madsen, O.B.G., Ravn, H.F., Rygaard, J.M.: A heuristic algorithm for the a diala-ride problem with time windows, multiple capacities, and multiple objectives. Annals of Operations Research 60, 193–208 (1995)
13. Montemanni, R., Gambardella, L., Rizzoli, A., Donati, A.: A new algorithm for a dynamic vehicle routing problem based on ant colony system. In: Second International Workshop on Freight Transportation and Logistics (2003)
14. North, M.J., Howe, T.R., Collier, N.T., Vos, R.J.: The Repast Simphony Runtime System. In: Proceedings of the Agent 2005 Conference on Generative Social Processes, Models, and Mechanisms, Argonne National Laboratory, Argonne, IL USA (2005)

Modelling Complex Intermodal Freight Flows

An Caris[1], Gerrit K. Janssens[1], and Cathy Macharis[2]

[1] Transportation Research Institute, Hasselt University - campus Diepenbeek, Wetenschapspark 5 - bus 6, 3590 Diepenbeek, Belgium
{an.caris,gerrit.janssens}@uhasselt.be
[2] Department MOSI - Transport and Logistics, Vrije Universiteit Brussel - Managementschool Solvay, Pleinlaan 2, 1050 Brussel, Belgium
Cathy.Macharis@vub.ac.be

Summary. In regions with an extensive waterway network intermodal transport including inland navigation is a good alternative for unimodal road transport. Intermodal transport networks exhibit an increased complexity due to the inclusion of multiple transport modes, multiple decision makers and multiple types of load units. A discrete event simulation methodology is proposed to understand the network dynamics and analyze policy measures with the intention of stimulating intermodal barge transport. The simulation model allows to quantify a number of network properties resulting from the interaction of freight flows. The intermodal hinterland network of the port of Antwerp serves as the real-world application in this study. Various aspects in the modelling process are discussed and a first potential policy is analysed.

1 Introduction

In this paper a discrete event simulation methodology is developed to capture and analyze the interactions in intermodal freight transport networks. Macharis and Bontekoning [6] define intermodal transport as the combination of at least two modes of transport in a single transport chain, without a change of container for the goods, with most of the route travelled by rail, inland waterway or ocean-going vessel and with the shortest possible initial and final journeys by road. Intermodal transport may include various transport modes. Our aim is to analyze the performance of intermodal networks including inland navigation and road transport. In regions with an extensive waterway network, such as Western Europe, intermodal transport including inland navigation is a good alternative for unimodal road transport.
Intermodal planning problems are more complex due to the inclusion of multiple transport modes, multiple decision makers and multiple types of load units. In this paper an intermodal freight transport network is modelled with the objective to understand the system and analyze various network configurations. The complexity of the intermodal transport system makes it impossible to describe all interactions by a mathematical model. Because of this increased complexity and the required level of detail, discrete event simulation is the appropriate tool of analysis. A simulation model is created to support decisions in intermodal transport at the strategic

level. The role of complex systems models in strategic decision making is discussed in [5]. The authors indicate that simulation models are appropriate to interpret the structure of a complex system. As stated in [3], the notions of emergence and emergent properties are not well defined. Emergent properties arise at a particular level of system description by virtue of the interaction of relatively simple lower-level components. The simulation model will allow to quantify a number of network properties resulting from the interaction of freight flows, as presented in section 2.5. Another example of modelling transportation networks as complex systems can be found in [10]. The objective of the authors is to study the emergence of hierarchies in the network.
Simulation models have been widely used to optimize the design of intermodal terminals. For example, Rizzoli et al. [9] present a simulation tool for the combined rail/road transport in intermodal terminals. Parola and Sciomachen [8] describe a strategic discrete event simulation model to analyze the impact of a possible future growth in sea traffic on land infrastructure in the north-western Italian port system. We develop a simulation model that covers the hinterland waterway network of a major port in Western-Europe in order to analyze effects of future policy measures for intermodal container transport. A first policy related to the consolidation network is presented in this paper. Various studies discuss consolidation strategies for intermodal transport by rail, see for example [2] and [7]. [3] presents and evaluates a consolidation strategy for intermodal transport by barge, based on a marginal cost model. In our simulation model the operations of the inland navigation network are modelled in detail. This enables us to examine ex-ante what the effects of a certain consolidation strategy will be and to take into account interaction effects in container flows. In the future, the model will also be used to analyze other policies related to intermodal barge transport. The simulation model is part of a larger decision support system for intermodal transport policy making which will not be discussed here.

2 Modelling Methodology

In this section the methodology to model the hinterland waterway network is described. First, the main characteristics of the intermodal network under investigation are given. Next, the conceptualization of the network is developed. The inputs and outputs of the simulation model are discussed. In the final subsection special attention is given to the functioning of the locks.

2.1 Intermodal Transport Network

The intermodal hinterland network of the port of Antwerp serves as the real-world application in our study. In Belgium three regions of origin can be identified in the network. The first group of container terminals is situated along the Albert Canal towards the eastern part of Belgium. A second region of origin is located in the central part of the country, connected to the port of Antwerp by the Brussels - Scheldt Sea Canal. The third group of intermodal container flows originates in the basin of the Upper Scheldt and the river Leie. All intermodal container terminals organize shuttle services either to the port of Antwerp or to the ports of Rotterdam and Amsterdam. Two clusters of sea terminals can be identified in the port area of

Antwerp. Until recently the main center of activity was situated on the right river bank. With the construction of a new dock (Deurganckdok) in the port of Antwerp, a second cluster of sea terminals emerged on the left river bank. Barges sail through the Scheldt-Rhine connection to Rotterdam and Amsterdam. A last destination is the port of Zeebrugge, which can be reached via Antwerp and navigation on the river Scheldt. Table 1 summarizes all origins and destinations of shuttle services. Shuttle services transport containers from inland terminals to sea terminals in the port area and carry containers from sea terminals to inland destinations in a round trip.

Table 1. Origins and Destinations

Origins	**Destinations**
Albert Canal	Antwerp: right river bank
Brussels-Scheldt Sea Canal	Antwerp: left river bank
Upper Scheldt and Leie	Rotterdam
	Amsterdam
	Zeebrugge

2.2 Conceptual Model

Three major components can be identified in the intermodal hinterland network, as depicted in Fig. 1. The first component in the intermodal freight transport network is the inland waterway network. The inland waterway network is made up of terminals, waterway connections and container flows. Entities are defined as barges which originate from the different inland terminals and carry containers in round trips to the various ports. A second component is the port area of Antwerp. Barges may visit sea terminals at the left river bank and right river bank in the same round trip, go to Rotterdam or Amsterdam via the Scheldt-Rhine connection or sail to Zeebrugge via the Scheldt estuary. On the right and left river bank, barges queue for handling at the sea terminals. Barges moor as soon as enough quay length is available. The handling time at the sea terminal depends on the number of containers that need to be unloaded from or loaded into the inland vessel. In the inland waterway network as well as in the port area multiple locks are present. Therefore, the lock planning constitutes a third major component, which will be discussed hereafter.

The objective of the model is to simulate possible policy measures for intermodal barge transport. Consequences and implications can be estimated before implementation of the policy measure. Therefore, various conceptual models may be necessary to analyze the implications of proposed policies. The conceptual model of the current container flow is depicted in Fig. 2. At present all barges enter the port area and visit one or multiple sea terminals.

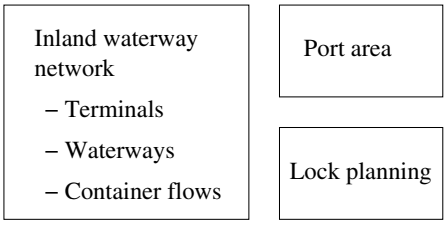


Fig. 1. Components

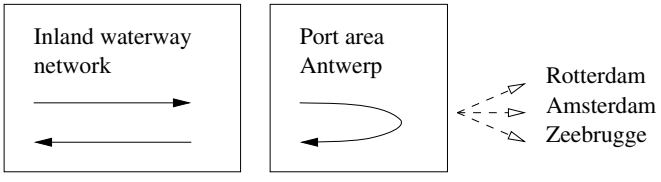


Fig. 2. Conceptual Model Current Situation

2.3 Assumptions

A number of assumptions are made to translate the actual intermodal network into a discrete event simulation model. The emphasis lays on inland waterway transport. Rail connections in the hinterland network are not taken into account. The model further assumes a homogeneous container type and equal handling time for each container. All main waterway connections between inland terminals and the port area are incorporated in the simulation model. Small waterways without inland terminals are not included in the simulation model of the current situation. Sailing times are assumed to be stochastic and follow a probability distribution. The average sailing time varies with the type of barge. A probability distribution is also used to model the stochastic lockage time.

2.4 Data Requirements

All intermodal terminals in the inland waterway network were asked for information to identify the container flows. Real data on shuttle services is used as an input for the simulation model, constructed in the simulation software Arena. For each shuttle service the following information is required: which type of barge is used, which destinations are visited and what is the average number of import and export containers for each destination. Container transport interacts with other freight flows. Therefore, the flow of non-containerized goods on the inland waterway network is introduced as an input in the simulation model. These flows affect the waiting times at locks. Information is also necessary on the network connections. The waterway administrators provided information on the number of locks on each waterway, distances between locks, average lockage times, number of lock chambers and size of the chambers. In the port area of Antwerp three clusters of locks connect the inner port area with the sea side. Data is required on the choice of locks when sailing in the port area. The average quay length available for handling inland navigation at

sea terminals gives an indication of the service capacity in the port area of Antwerp. The port authority is asked for the average mooring time and time for loading and unloading in order to model the service times of inland container barges in the port area. Finally, an enquiry is made into the turnaround times of vessels and average waiting times at locks in order to verify and validate the model.

2.5 Outputs

Table 2 gives an overview of properties measured in the simulation model. The turnaround time of shuttles is defined as the total time necessary for a barge to sail from an inland container terminal to the port area, visit all sea terminals and return to the inland terminal. The turnaround time depends on the waiting times at locks and in the port area. The outputs measured at locks are the percentage of barges that have to wait, the number of barges that have to queue and the waiting time of barges in the queue. In the port area the waiting time before handling is measured, as well as the number of vessels queueing for service. A final group of performance measures concerns the capacity utilization. In the port area this is expressed as the average percentage of quay length occupied. In the hinterland network the average and maximum number of barges on each network connection is recorded.

Table 2. Outputs

Shuttles	turnaround time
Locks	total number waiting (%) number waiting in queue waiting time in queue
Port area	waiting time in queue number waiting in queue
Capacity utilization	quay length network connections

2.6 Locks

The operations of locks strongly affect waiting times of barges for lockage. A number of decision rules are defined to make the operations of the locks in the simulation model reasonably realistic. A first decision rule relates to the size of a barge. Barges are assigned to a lock chamber only if its size is within the allowed dimensions. Secondly, barges are assigned to the smallest lock chamber that is open. This decision rule focuses on a rapid lockage process of barges. Smaller lock chambers have a shorter lockage time. On the other hand, a more intensive use of larger lock chambers may reduce waiting times because more barges can be serviced simultaneously. A third decision rule is applied when no lock chamber is open in the sailing direction of the barge. In this situation the barge is assigned to the lock chamber which

is the first available. A final decision rule concerns the closing of lock chambers. A lock chamber is closed when there is not enough remaining space for the next barge in queue or when no additional barges arrive within a predefined number of time units. From interviews with waterway administrators it appears that the operations of locks are entrusted to a lockkeeper, without fixed rules. Future research could introduce more complex decision rules in the simulation model.

3 Analysis of Container Flows

The simulation model will be used to analyze policy measures that impact container flows in the intermodal hinterland network and the port area of Antwerp. In this section a first policy measure is analyzed. A potential policy might be the introduction of an intermodal barge hub in the port of Antwerp, from which load is distributed to the different sea terminals, as described by Konings [3]. The author proposes to split existing barge services into a trunk-line operation in the hinterland and collection/distribution operations in the seaport. This leads to the conceptual model depicted in Fig. 3.

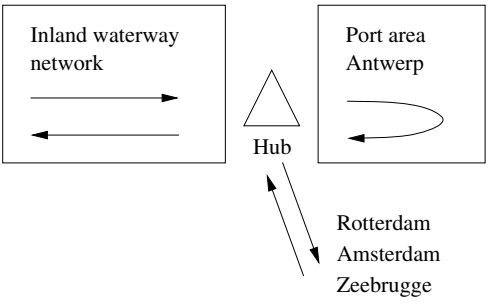


Fig. 3. Conceptual Model Future Situation

By doing so inland barges do not have to call at multiple sea terminals. They only visit the intermodal barge hub. The intermodal hub organizes shuttle services in the port area to collect containers from and distribute containers to all sea terminals. In the collection/distribution network containers with the same origin or destination can be bundled. This enables a more efficient and prompt handling of barges at sea terminals. When modeling the new situation, it is assumed that all containers are collected and distributed by barge in the port area. In reality some containers might be transferred by road when the distance between the hub and the sea terminal is small or in urgent cases. The intermodal barge hub is located in the cluster of sea terminals on the right river bank. A quay length of 500 metres is assumed at the hub. To set a service level for the hub, it is required that all inland containers are delivered within 24 hours to the sea terminals. It is further assumed that two shuttle services are organized per day in the collection/distribution network, one in the morning and one in the afternoon, each visiting terminals on the right and/or left river bank. The shuttle services are carried out with vessels of a size of 96 TEU and 196

TEU. The organization of the collection/distribution network might be optimized. However, this setting already gives an indication of possible improvements in the relevant performance measures. When comparing the current situation with the new consolidation strategy, no changes are made to the schedules of the inland terminals. It is possible that inland terminals change their departures in the new situation. Other measures to enhance the efficiency can be further simulated. A separate random-number stream is dedicated to each source of randomness in the model in order to synchronize both alternatives as much as possible.

Performance measures relevant for the comparison of the scenarios are discussed next. Ten simulation runs of 672 hours are performed. Table 3 gives the average turnaround times of all inland terminals, expressed in hours. Inland vessels may only sail to Antwerp (Antw) or they can make a combined trip to Antwerp and Rotterdam (Rdam) or Amsterdam (Adam). Standard deviations are mentioned between brackets below the average turnaround times.

Table 4 summarizes performance measures in the port area. The average and maximum waiting time before handling, expressed in hours, are given for the sea terminals on the right and left river bank and at the intermodal barge hub. Next, the average and maximum utilization of the quays on the right and left river bank and at the hub are measured.

Following [4], paired-t confidence intervals are constructed to compare the results. Table 5 presents the 95% confidence intervals for which the difference between the current situation and the intermodal barge hub is significant.

The average turnaround times of shuttles between inland terminals along the Albert Canal and the port of Antwerp are all significantly reduced. The maximum turnaround times of these inland terminals also decrease significantly due to the introduction of the hub. Shuttles originating from the Albert Canal can go directly to the intermodal barge hub without needing to pass through a lock in the port area. Shuttles from other inland terminals first have to pass through one of the locks to reach the hub. A reduction in turnaround time is also recorded for the terminal in Grimbergen. Table 4 shows that with an equal available quay length, shuttle services in the collection/distribution network of the new consolidation strategy do not have to wait for handling at sea terminals on the right and left river bank. The sea terminals can handle inland containers more efficiently because only shuttle services with consolidated load moor for service. The waiting time at the intermodal hub depends on the available quay length. A quay length of 500 metres is assumed and leads to an average waiting time of 17.82 minutes. Time windows could be negotiated with the inland terminals to reduce the waiting time at the hub at peak hours. Average capacity utilization on the right river bank decreases significantly. Finally, table 4 reveals that at peak moments the maximum capacity utilization decreases with 28.67% on the left river bank and with 19.67% on the right river bank. Less quay length is necessary to handle inland containers at peak hours. These figures demonstrate the efficiency improvements at the sea terminals in the port area. No significant influence was found of the new consolidation strategy on waiting times at locks in the port area. Inland barges constitute only a small part of total lock passages.

Table 3. Average turnaround times current situation and intermodal barge hub

Avg turnaround time	**Current**	**Hub**
Deurne - Antw	15.20	9.93
	(0.47)	(0.35)
Deurne - Antw/Rdam	22.08	22.98
	(0.89)	(0.29)
Meerhout - Antw	29.24	25.59
	(0.47)	(0.18)
Meerhout -	41.70	39.68
Antw/Rdam/Adam	(0.38)	(0.89)
Genk - Antw	38.97	35.94
	(0.62)	(0.72)
Genk - Antw/Rdam	49.89	47.24
	(0.87)	(0.48)
Luik - Antw	46.46	42.10
	(0.34)	(0.12)
Gent - Antw	20.62	19.43
	(0.49)	(0.42)
Wielsbeke - Antw	38.62	39.60
	(0.42)	(0.41)
Avelgem - Antw	41.19	40.78
	(0.88)	(2.10)
Avelgem - Antw/Rdam	62.69	61.89
	(0.48)	(0.51)
Willebroek - Antw	14.79	14.37
	(0.17)	(0.19)
Willebroek - Antw/Rdam	35.59	34.91
	(0.39)	(0.14)
Grimbergen - Antw	20.93	19.42
	(0.21)	(0.28)
Brussel - Antw	21.91	22.42
	(0.34)	(0.17)
Brussel - Antw/Rdam	40.94	40.07
	(0.29)	(0.42)
Herent - Antw	21.91	21.68
	(0.19)	(0.40)

4 Conclusions and Future Research

Intermodal freight transport networks can be seen as complex systems, showing properties that cannot be deducted from the individual components of the network. A discrete event simulation model is constructed to understand the network system and analyze network configurations. Various aspects in the modelling process are presented and network properties are discussed. Potential improvements to the model include the introduction of more complex decision rules for the operations

Table 4. Performance measures in the port area: current situation and intermodal barge hub

Port area	**Current**	**Hub**
Avg waiting time		
Right river bank	0.0629	0.0000
	(0.0306)	(0.0000)
Left river bank	0.0557	0.0000
	(0.0115)	(0.0000)
Hub	/	0.2970
		(0.0334)
Max waiting time		
Right river bank	7.6128	0.0000
Left river bank	4.3095	0.0000
Hub	/	8.4450
Avg capacity utilization		
Quay Right river bank	0.1666	0.1398
	(0.0017)	(0.0014)
Quay Left river bank	0.1741	0.1808
	(0.0017)	(0.0016)
Quay Hub	/	0.2682
		(0.0022)
Max capacity utilization		
Quay Right river bank	0.9834	0.7867
Quay Left river bank	0.9850	0.6983
Quay Hub	/	1.0000

of locks. A submodel could also be introduced to integrate the intermodal terminal planning into the simulation model.

The model will be used to make a quantitative ex-ante analysis of policy measures to stimulate intermodal barge transport. Simulation results of a first policy are presented. The introduction of an intermodal barge hub on the right river bank leads to a significant reduction in average and maximum turnaround times of shuttle services originating from the Albert Canal. This implies shorter and more stable sailing schedules for inland terminal operators along this waterway axis. However, no significant influence is found on the shuttle services of other inland terminals. Average capacity utilization on the right river bank decreases significantly. Inland containers are handled more efficiently at the sea terminals due to the consolidation of load in the collection/distribution network. Future simulation experiments will investigate the effect of a hub on the left river bank or the introduction of a multihub service model. Various scenarios may be compared to determine the most beneficial consolidation strategy.

Acknowledgement: We thank the Belgian Science Policy (BELSPO) for their support on our research project DSSITP (Decision Support Systems for Intermodal

Table 5. Confidence intervals for comparing the current and new situation

	Confidence interval hub - current
Avg turnaround time	
Deurne - Antw	-6.9143 ; -3.6159
Meerhout - Antw	-4.7188 ; -2.5713
Genk - Antw	-4.8083 ; -1.2481
Genk - Antw/Rdam	-5.1158 ; -0.1897
Luik - Antw	-5.2091 ; -3.5234
Grimbergen - Antw	-2.4003 ; -0.6144
Avg waiting time	
Left river bank	-0.0818 ; -0.0297
Avg capacity utilization	
Quay Right river bank	-0.0304 ; -0.0233

Transport Policy) in the research programme "Science for a Sustainable Development - call 2", under contract number SD/TM/08A.

References

1. Damper, R.I.: Emergence and levels of abstraction. International Journal of Systems Science 31 (2000)
2. Janic, M., Reggiani, A., Nijkamp, P.: Sustainability of the European freight transport system: evaluation of innovative bundling networks. Transportation Planning and Technology 23, 129–156 (1999)
3. Konings, R.: Smart collection and distribution of containers on barge in the port of Rotterdam. In: Proceedings of the Nectar Conference 2005, Las Palmas, Gran Canaria (2005)
4. Law, A.M.: Simulation modeling & analysis, 4th edn. McGraw-Hill, New York (2007)
5. Lyons, M.H., Adjali, I., Collings, D., Jensen, O.: Complex systems models for strategic decision making. BT Technology Journal 21(2), 11–27 (2003)
6. Macharis, C., Bontekoning, Y.M.: Opportunities for OR in intermodal freight transport research: A review 153, 400–416 (2004)
7. Newman, A.M., Yano, C.A.: Scheduling direct and indirect trains and containers in an intermodal setting. Transportation Science 34(3), 256–270 (2000)
8. Parola, F., Sciomachen, A.: Intermodal container flows in a port system network: Analysis of possible growths via simulation models. International Journal of Production Economics 97, 75–88 (2005)
9. Rizzoli, A.E., Fornara, N., Gambardella, L.M.: A simulation tool for combined rail/road transport in intermodal terminals. Mathematics and Computers in Simulation 59, 57–71 (2002)
10. Yerra, B.M., Levinson, D.M.: The emergence of hierarchy in transportation networks. The Annals of Regional Science 39, 541–553 (2005)

Modelling the Complexity of Inventory Management Systems for Intermittent Demand using a Simulation-optimisation Approach

Katrien Ramaekers and Gerrit K. Janssens

Transportation Research Institute
Hasselt University
Wetenschapspark 5 bus 6
3590 Diepenbeek, Belgium
{katrien.ramaekers;gerrit.janssens}@uhasselt.be

Summary. Inventory systems are complex systems due to the presence of several types of uncertainty. Furthermore, both an inventory management policy and a forecasting method need to be chosen in inventory management. These choices have an impact on the performance of the system and there is an interaction between the inventory management policy and the forecasting method. In this paper, the complexity of inventory systems is modeled for a special type of irregular demand using a simulation model. Like this, it is possible to predict the properties of the complete system instead of taking it to pieces and analysing its parts.

1 Introduction

Inventory systems appear with uncertainties in demand, in lead times, in transportation times, in availability of resources and in quality. Dealing with uncertainty is an important issue in inventory management. Some of the uncertainty is due to suppliers but some is also attributable to factors as customers or economic conditions. Management decisions have to take these uncertainties in consideration.
Next to these different types of uncertainty, the behaviour and performance of an inventory management system is influenced by the choice of both the inventory management policy and the demand forecasting method. Futhermore, these two factors interact with each other [8].
The impact of the uncertainties and the decisions on the performance of inventory management systems and the interaction between the uncertainties and the decisions, make the study of these systems complex. For that reason, it is necessary to predict the properties of the complete system instead of taking it to pieces and analyzing its parts. This is exactly one of the basic concepts of emergence [3].
In this paper, a special type of demand is examined: intermittent demand, i.e. demand peaks follow several periods of zero or low demand. The impact of uncertainties in demand and supply and the impact of the inventory management

policy and forecasting method are examined for inventory systems facing this type of irregular demand.
Because mathematical models cannot accurately describe the complex system, simulation models will be used. In preliminary research a simulation model is built and optimised to obtain the best strategy in combining inventory decision making and demand forecasting for intermittent demand when there is only uncertainty in demand [9]. In this paper, the same simulation model is used but the assumption of no uncertainty in supply is relaxed. First, the results of the optimal policies found for a reliable supplier are compared to results when applying the same policies for an unreliable supplier. Next, a new best strategy in combining inventory decision making and demand forecasting is determined for the situation of an unreliable supplier. The paper starts with a short overview of the literature on inventory systems.

2 Literature review

Inventory systems have to cope with uncertainty in demand. The inventory control literature mostly makes use of the Normal or Gamma distribution for describing the demand in the lead-time. The Poisson distribution has been found to provide a reasonable fit when demand is very low (only a few pieces per year). Less attention has been paid to irregular demand. This type of demand is characterised by a high level of variability, but may be also of the intermittent type, i.e. demand peaks follow several periods of zero or low demands. In practice, items with intermittent demand include service or spare parts and high-priced capital goods. A common example of such goods are spare parts for airline fleets.
Demand forecasting is one of the most crucial aspects of inventory management [13]. However, for intermittent demand, forecasting is difficult, and errors in prediction may be costly in terms of obsolescent stock or unmet demand [12]. The standard forecasting method for intermittent demand items is considered to be Croston's method [2]. This method builds estimates taking into account both demand size and the interval between demand occurrences. Despite the theoretical superiority of such an estimation procedure, empirical evidence suggests modest gains in performance when compared with simpler forecasting techniques [11]. Furthermore, the choice of the forecasting method can have an impact on the inventory management policy that is best used.
Studies in literature mainly focus on uncertainty in demand. In many inventory models the continuous availability of supply at any time in the future is an implicit assumption. However, uncertainty is also present at the supply side. This type of uncertainty occurs in delivery time, in interruption of delivery during a certain period, or in mismatches in order and delivery in terms of quality or quantity. In this paper, the focus is on uncertainty in availability. The supplier alternates randomly between an available and an unavailable state. When the supplier is available, the order is delivered after the usual lead time. When the supplier is unavailable, the order is executed when the supplier turns available again.

3 Simulation model and research approach

3.1 Simulation model

The study focuses on a single-product inventory system facing demand of the intermittent type. The simulation model is developed in Microsoft Excel spreadsheets and uses VBA. The simulation model starts by generating intermittent demand as described in the previous section. Next, the inventory system is simulated for 52 periods. At each review-time, a demand forecast and an order decision are made. The total costs of the inventory system are determined. 10 replications are made for each simulation run.

To generate intermittent demand, demand occurrence and demand size are separately generated. The demand occurrence is generated according to a first-order Markov process with transition matrix

$$\mathbf{P} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix},$$

where p_{00} is the probability of no order in the next period when there has been no order in this period and p_{10} is the probability of no order in the next period when there has been an order in the current period. Individual order sizes are generated using a Gamma distribution with shape parameter γ and scale parameter β.

The standard forecasting method for intermittent demand items is considered to be Croston's method. However, in practice, single exponential smoothing and simple moving averages are often used to deal with intermittent demand. These three forecasting methods are compared.

In this research, two periodic review models are used. The first one is the (R, s, S) system. This means that every R units of time, the inventory level is checked. If it is at or below the reorder point s, a sufficient quantity is ordered to raise it to S. The second system (R, s, Q) is similar to the (R, s, S) system but uses a fixed order quantity Q instead of an order-up-to-level S.

A deterministic lead-time L is assumed. The following costs are considered: unit holding cost per period C_h, ordering cost C_o and unit shortage cost per period C_s. The simulation starts with an initial inventory level I_0.

3.2 Experimental design

The parameters of the inventory system to optimise include both qualitative and quantitative factors. The experimental design includes two qualitative factors: the forecasting method and the inventory management policy. In addition, depending on the choice of the qualitative factors, a set of quantitative factors are part of the experimental design. If the (R, s, Q) inventory management policy is used, the safety stock SS and order quantity Q are the parameters to optimise. If the (R, s, S) inventory management policy is used, the safety stock SS and order-up-to-level S are the optimising parameters. For single exponential smoothing and Croston's method, the smoothing parameter α is optimised and for moving averages, the weights of the past values are optimised.

3.3 Research approach

Because of the dependence of the quantitative factors on the choice of the qualitative factors, we use for the optimisation the research approach described in this section. For every combination of forecasting method, inventory management policy and review period, the optimal values of the quantitative factors are determined. The total costs of the inventory system are optimised using tabu search. Tabu search is shortly described below. Once the optimal values are found, the best combination of forecasting method, inventory management policy and review period is chosen.
Tabu search uses a local or neighbourhood search procedure to iteratively move from one solution to the next in the neighbourhood of the first, until some stopping criterion has been satisfied. To explore regions in the search space that would be left unexplored by the local search procedure and escape local optimality, tabu search modifies the neighbourhood structure of each solution as the search progresses. The solutions admitted to the new neighbourhood are determined through the use of special memory structures. Tabu search uses both long-term and short-term memory, and each type of memory has its own special strategies [5, 4].
One type of short-term memory is the tabu list. This list contains solutions that have been visited in the recent past. Other tabu list structures prohibit solutions that have certain attributes or prevent certain moves. This last type of tabu list contains the moves that are not allowed at the present iteration in order to exclude backtracking moves. Subsequent to each move, the opposite move is appended to the list and the oldest move in the list is removed. To prevent this short-term memory from preventing excellent solutions from being found, aspirations levels are commonly introduced. The tabu status of a solution can be overruled if its solution quality exceeds a certain aspiration level. Long-term memory is used for both diversification and intensification of the search process. Diversification strategies are used to force the search into previously unexplored regions of the solution space. Intensification strategies are used to encourage move combinations that have worked well in the past, or to return the search to attractive regions that have been insufficiently explored [5].
Tabu search is a heuristic optimisation technique developed specifically for combinatorial problems. Very few works deal with the application to the global minimization of functions depending on continuous variables. Hu [6] is the first to adapt tabu search to continuous optimisation. However, the algorithm of Hu is rather far from original tabu search. Siarry and Berthiau [10] propose an adaptation of tabu search to the optimisation of continuous functions where the purpose is to keep as close as possible to original simple tabu search. As neighbourhood of the current solution, they perform a partition of the space around the current solution using a set of concentric balls. Inside each ball, a random neighbour is selected. The tabu list contains m balls, corresponding to the immediate neighbourhoods of the m last retained solutions. Chelouah and Siarry [1] improve the algorithm of Siarry and Berthiau [10] and propose an Enhanced Continuous Tabu Search for the global optimisation of continuous functions. They replace the balls by hyperrectangles for the partition of the current solution neighbourhood and add diversification and intensification concepts to the algorithm. The method we propose here is based on [10] and [1]. In Fig. 1, a general flowchart of the TS algorithm is shown. Two issues must be examined: the generation of current solution neighbours and the elaboration of the tabu list.

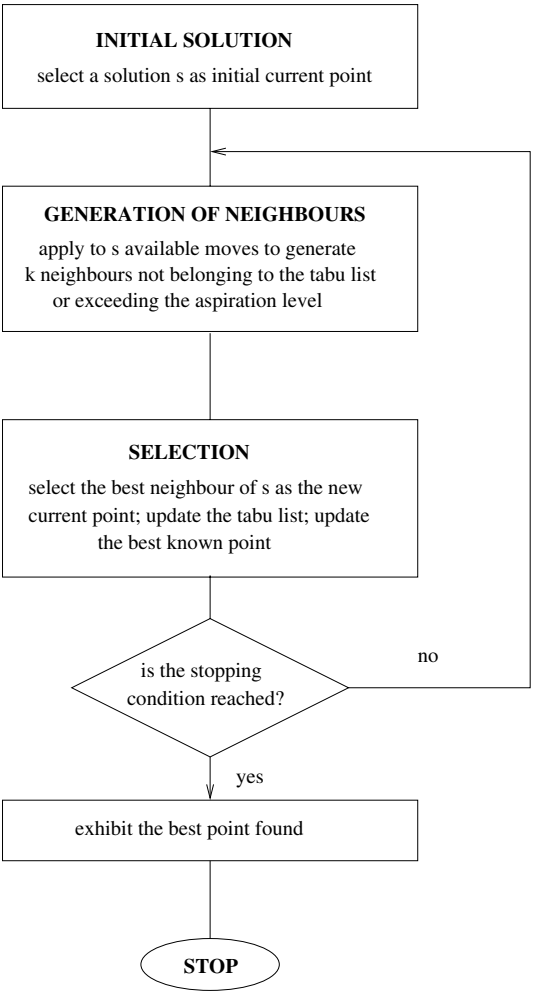


Fig. 1. General flow chart of tabu search

To define a neighbourhood of the current solution s, a set of hyperrectangles is used for the partition of the current solution neighbourhood. The k neighbours of the current solution are obtained by selecting one point at random inside each hyperrectangular zone. In Fig. 2, a two-dimensional example of such a partition for $k = 4$ neighbours of the current solution is given.

Once a new current solution is determined, the immediate neighbourhood of the previous solution is added to the tabu list. This immediate neighbourhood is also a hyperrectangle. The tabu list containts m hyperrectangles corresponding to the m last retained solutions. A solution belonging to the tabu list can loose its tabu status if its aspiration level is high enough. A solution becomes non-tabu if its objective value is better that the best value obtained at that moment.

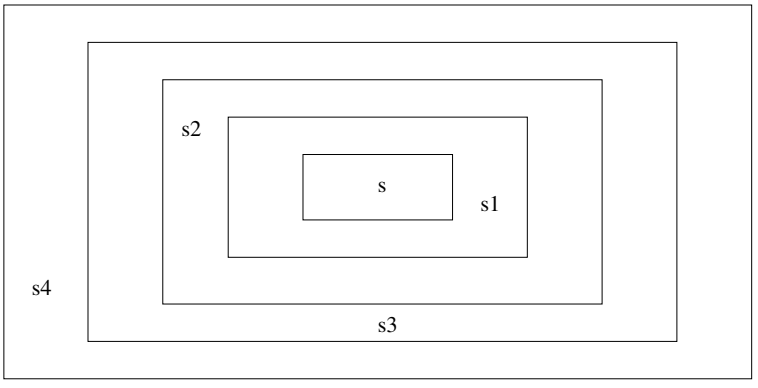


Fig. 2. Partition of current solution neighbourhood

4 Experimental environment

The experimental environment contains the uncontrollable factors of the inventory system: the costs of the inventory system and the parameters for generating intermittent demand. These factors can have an effect on the results that are obtained. The research approach described above, is executed using a single combination of the costs of the inventory system and demand. A fractional factorial design of 16 experimental points is set up for these factors and the optimisation phase is repeated for each experimental point.
Demand occurrence is generated using a first-order Markov process with transition matrices:

$$\mathbf{P_1} = \begin{pmatrix} 0.7875 & 0.2125 \\ 0.85 & 0.15 \end{pmatrix}$$

or

$$\mathbf{P_2} = \begin{pmatrix} 0.5667 & 0.4333 \\ 0.65 & 0.35 \end{pmatrix}.$$

They correspond with a probability of 20% to have demand in a certain period for the first matrix and a probability of 40% to have demand in a period for the second matrix. The *size of demand* is generated using a Gamma distribution with 4 different combinations of the scale parameter γ and the shape parameter β. These values are summarized in Table 1.

Level	γ	β
1	6	1
2	12	1
3	3	2
4	24	0.5

Table 1. Parameters of the Gamma distribution

The levels of the *costs* of the inventory system are given in Table 2. The initial inventory level I_0 equals 5.

Level	C_o	C_h	C_s
1	100	2	5
2	200	4	10

Table 2. Levels for the costs of the inventory system

The fractional factorial design is shown in Table 3. This fractional factorial design makes it possible to determine the impact of uncontrollable factors as the cost structure and the demand on the optimal strategy in inventory decision making and demand forecasting for intermittent demand.

Exp	C_o	C_h	C_s	**Freq**	γ	β
1	200	4	10	0.4	12	1
2	100	4	5	0.4	12	1
3	200	2	5	0.4	24	0.5
4	100	2	10	0.4	24	0.5
5	200	2	5	0.4	3	2
6	100	2	10	0.4	3	2
7	200	4	10	0.4	6	1
8	100	4	5	0.4	6	1
9	200	2	10	0.2	12	1
10	100	2	5	0.2	12	1
11	200	4	5	0.2	24	0.5
12	100	4	10	0.2	24	0.5
13	200	4	5	0.2	3	2
14	100	4	10	0.2	3	2
15	200	2	10	0.2	6	1
16	100	2	5	0.2	6	1

Table 3. Experimental design for uncontrollable factors

5 Results

Firstly, results are compared for a reliable and an unreliable supplier. For 16 combinations of uncontrollable factors, the best strategy in combining inventory decision making and demand forecasting for a reliable supplier is determined in preliminary research (Table 4). Based on these results, it can be concluded that the uncontrollable factors have an impact on the best strategy for combining inventory decision-making and demand forecasting for intermittent demand. Furthermore, there is interaction between these factors.

Depending on the experimental environment, two options for optimal strategies can be distinguished: an order-up-to level inventory management policy with an order-up-to level equal to 1 and a reorder point equal to 0 or an inventory management policy with a fixed order quantity $Q > 1$ or an order-up-to level $S > 1$ and a reorder point equal to 0.
To study the impact of the uncontrollable factors in more detail, a classification tree is constructed using the C4.5 algorithm, a well-known algorithm in data mining ([7]). The classification three can be found in Fig. 3. Using this tree, it can be decided which of the two strategies is best. Three factors are needed to determine the best strategy in combining inventory decision making and demand forecasting: the frequency of demand, the order cost and the inventory cost. If one of these three factors is not known, the knowledge of the stock-out cost is also sufficient to make a classification. Summarizing, it can be said that if three factors of the four just mentioned (frequency of demand, order cost, inventory cost and stock-out cost) are fixed, the best strategy is presented.

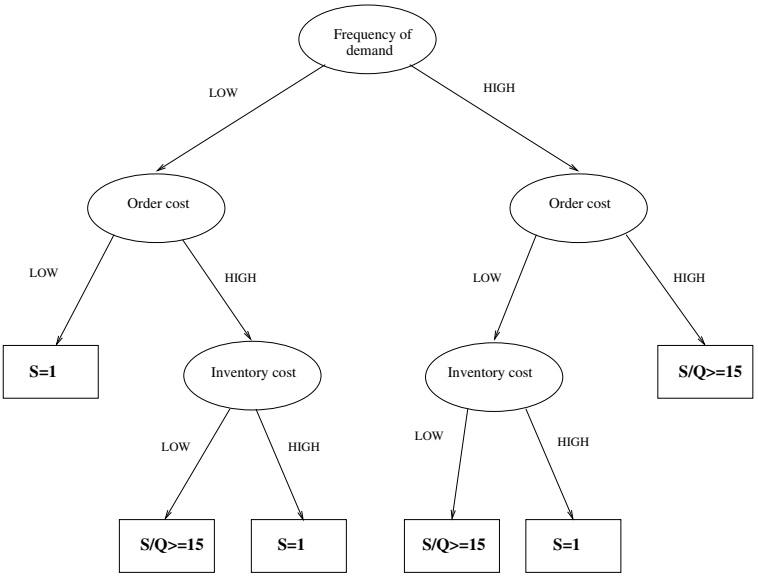


Fig. 3. Classification tree

A good classification is necessary because there is a considerable increase in the costs of the inventory system when using the other strategy.
This optimal strategies, found in Table 4 will also be used to determine output measures for the inventory system with an unreliable supplier.
Tables 5 contains 95% confidence intervals (CI) for the difference in total costs between a reliable and an unreliable supplier.
For eight of the experimental points, the difference in costs between the reliable and the unreliable case is significant. The alternative with the unreliable supplier has higher total costs than the one with the reliable supplier. The experimental

Exp	Best strategy
1	MA/FOQ; ROP=0; Q=25
2	ES/OUL; ROP=0; S=1
3	MA/OUL; ROP=0; S=30
4	MA/OUL; ROP=0; S=25
5	MA/FOQ; ROP=0; Q=20
6	MA/OUL; ROP=0; S=15
7	MA/FOQ; ROP=0; Q=15
8	ES/OUL; ROP=0; S=1
9	MA/FOQ; ROP=0; Q=20
10	ES/OUL; ROP=0; S=1
11	CR/OUL; ROP=0; S=1
12	ES/OUL; ROP=0; S=1
13	MA/OUL; ROP=0; S=1
14	MA/OUL; ROP=0; S=1
15	MA/OUL; ROP=0; S=15
16	ES/OUL; ROP=0; S=1

Table 4. Optimal results based on Tabu search for a reliable supplier

Exp	CI
1	-150.68;149.88
2	-516.98;-248.38
3	-74.65;97.89
4	-135.08;-5.64
5	-25.81;144.886
6	-65.95;74.03
7	-103.07;139.35
8	-369.92;-144.44
9	-117.4;131.08
10	-359.4;-124.36
11	-679.87;-279.44
12	-597.74;-290.24
13	-459.96;-108.36
14	-507.49;-263.19
15	-100.16;91.3
16	-232.57;-33.19

Table 5. Confidence intervals for comparing costs of a reliable and an unreliable supplier

points with a significant difference in costs are those experimental points for which the best strategy for a reliable supplier has an order-up-to-level of 1. For the other experimental points, the best strategy for a reliable supplier has an order-up-to-level or fixed order quantity of 15 or more.

A new optimal combination of forecasting method and inventory management policy is determined for the inventory management system with intermittent demand

and an unreliable supplier. Furthermore, the optimal settings for the safety stock and the fixed order quantity or order-up-to-level are determined.
Table 6 shows the best strategy for the inventory system with an unreliable supplier.

Exp	Best strategy
2	MA/FOQ; ROP=0; Q=18
8	MA/FOQ; ROP=0; Q=12
10	ES/OUL; ROP=0; S=1
11	CR/OUL; ROP=0; S=1
12	MA/OUL; ROP=0; S=1
13	MA/FOQ; ROP=0; Q=10
14	ES/OUL; ROP=0; S=1
16	MA/FOQ; ROP=0; Q=10

Table 6. Results of Tabu search for an unreliable supplier

The best strategy for the inventory system with intermittent demand and no uncertainty in supply is an order-up-to-level inventory management policy with $S = 1$. When the results in Table 6 are compared to the results in Table 4, the experimental points can be divided in two categories. For the experimental points 2, 8, 13 and 16, the best strategy is an inventory management policy with a fixed order quantity Q equal to 10 or more. When a fixed order quantity inventory management policy is used, moving averages is always the best forecasting method. For the experimental points 10, 11, 12 and 14, the best strategy is an order-up-to-level inventory management policy with S=1, which is the same strategy as found for the inventory system with a reliable supplier. When the order-up-to-level equal to 1 is the best inventory management policy, no preference for a specific forecasting method can be found.

6 Conclusions and further research

In this paper, the complexity of an inventory system for intermittent demand is modeled using a simulation model. Like this, it is possible to predict the properties of the complete system instead of taking it to pieces and analysing its parts. The simulation model includes both uncertainty in the demand as in the supply side. Intermittent demand is a special type of irregular demand. Uncertainty in supply consists of many aspects, the focus of this paper is on uncertainty in availability. The supplier alternates randomly between an available and an unavailable state. When the supplier is available, the order is delivered after the usual lead time. When the supplier is unavailable, the order is executed when the supplier turns available again.
A best strategy in combining inventory decision making and demand forecasting for intermittent demand is determined. When the best strategy for the inventory system with a reliable supplier is used for the system with an unreliable supplier, 8 of 16 experimental points show a significant difference in total costs. For these

8 experimental points, a new optimal strategy in combining demand forecasting and inventory management decision making is determined using Tabu Search. For four of these points, no better strategy could be determined. For the other four experimental points, it is better to use a fixed order quantity Q of 10 or more instead of an order-up-to-level S equal to 1.
A best strategy in combining inventory decision making and demand forecasting for intermittent demand is determined using the objective of minimising total costs. However, there are situations in which total costs are not the best decision variable. In future work, it can be interesting to determine a best strategy using a performance measure as the output measure to optimise. Multi-objective optimisation, including non-cost but service-oriented performance characteristics, can also be used to better incorporate the effect on performance measures in the optimal solution.

References

1. Chelouah, R., Siarry, P.: Tabu search applied to global optimization. European Journal of Operational Research 123, 256–270 (2000)
2. Croston, J.D.: Forecasting and stock control for intermittent demands. Operational Research Quarterly 23(3), 289–303 (1972)
3. Damper, R.I.: Emergence and levels of abstraction. International Journal of Systems Science 31(7), 811–818 (2000)
4. Dengiz, B., Alabas, C.: Simulation optimization using tabu search. In: Proceedings of the 2000 Winter Simulation Conference, pp. 805–810 (2000)
5. Glover, F.: Tabu search - part I. ORSA Journal on Computing 1(3), 190–206 (1989)
6. Hu, N.: Tabu search method with random moves for globally optimal design. International Journal for Numerical Methods in Engineering 35, 1055–1070 (1992)
7. Quinlan, J.R.: C4.5 Programs for Machine Learning. Morgan Kaufmann Publishers, San Francisco (1993)
8. Ramaekers, K., Janssens, G.K.: The interaction between forecasting methods and inventory management policies for intermittent demand. In: Bruzzone, A.G., Frydman, C., Giambiasi, N., Mosca, R. (eds.) Proceedings of the International Mediterranean Modelling Multiconference, pp. 370–375 (2004)
9. Ramaekers, K., Janssens, G.K.: Towards a best strategy in inventory decision making and demand forecasting for intermittent demand. In: Genco, A., Gentile, A., Sorce, S. (eds.) Proceedings of the Industrial Simulation Conference, pp. 409–414 (2006)
10. Siarry, P., Berthiau, G.: Fitting of tabu search to optimize functions of continuous variables. International Journal of Numerical Methods in Engineering 40, 2449–2457 (1997)
11. Syntetos, A.A., Boylan, J.E.: On the bias of intermittent demand estimates. International Journal of Production Economics 71, 457–466 (2001)
12. Syntetos, A.A., Boylan, J.E.: The accuracy of intermittent demand estimates. International Journal of Forecasting 21, 303–314 (2005)
13. Willemain, T.R., Smart, C.N., Schwarz, H.F.: A new approach to forecasting intermittent demand for service parts inventories. International Journal of Forecasting 20, 375–387 (2004)

Cooperative Medical Diagnosis Elaboration by Physicians and Artificial Agents

Barna Laszlo Iantovics

Petru Maior University of Tg. Mures,
1 Nicolae Iorga, Tg. Mures 540088, Romania
ibarna@davos.science.uttgm.ro

Summary. Cooperative medical diagnosis systems are well suited for the solving of difficult medical diagnosis problems. The solving of many medical diagnosis problems require knowledge from different medical domains, which cannot be detained by a single physician or a medical computational system. In this paper, a novel medical multiagent system called MASM (Medical Assistant Multiagent System) that can help physicians in the diagnosis processes is proposed. The proposed multiagent system is a complex system, composed from relatively simple agents that cooperatively with physicians can solve difficult medical problems. MASM multiagent system can discover autonomously emergent proprieties, like the necessary cooperation links between agents and cooperation links between physicians that emerge during the problems solving. Discovered cooperation links allows to the MASM system to self-organize depending on the necessities, in order to increases the accuracy of the diagnostics elaborated by the physicians and to reduces the complexity of the diagnosis processes realized by physicians by hiding some of tasks that must be fulfilled.

Keywords: complex systems, emergent proprieties in artificial and natural systems, self-organizing systems, intelligent agents, knowledge-based systems, diagnostic accuracy, medical diagnosis systems, multiagent systems, cooperative problem solving, medical applications, computational methods in medicine, applications to biology and medical sciences

1 Introduction

Many medical diagnosis problems solving is a difficult task, which require the cooperation of more medical human specialists and medical computational systems. Some difficult diagnostics elaboration requires a planning process to that must contribute more physicians with different medical knowledge. For example, we mention the planning of a surgery intervention for the elimination of a huge tumor. Such planning process requires the analysis of different situations that can appear during the surgery intervention (for example, the loss of a huge quantity of blood), to that

must contribute more human medical specialists. Agents can be useful during such diagnostics establishment. They can verify different medical hypothesis, established autonomously by them or required by physicians. Such analyses results may allow the detection of mistakes that may appear in the physicians' decisions. For example, we mention the suggestion from mistake of a medicine for treating a patient's illness, in the situation when the patient is allergic to that medicine.

In this paper, a novel assistant multiagent system called *MASM* (*Medical Assistant Multiagent System*) is proposed. *MASM* is a complex system, composed from relatively simple agents called *assistant expert system agents* that cooperatively help physicians in their work. The assistant expert system agents represent a novel class of agents developed in our previous researches [14, 28]. Another novelty consists in the cooperative assistance of the physicians in the medical problem solving processes offered by the system. The intelligent operation of the system allows the discovery of emergent properties (cooperation links between agents and cooperation links between physicians that appear during the problems solving), which has as effect the partially hiding the complexity of the medical diagnosis problems to the physicians (the physicians must analyze less informations and data), making easier the diagnosis processes and increasing the accuracy of the elaborated diagnostics. Discovered emergent informations that consists in cooperation links, allows the self-organization of the agents members of the system in order to improve the efficiency of the problems solving. A cooperative problem solving team composed from agents is formed when it is necessary during a problem solving. After the problem is solved, the team is dropped. The *MSAM* system can contribute to the solving of different subproblems that appear during the diagnosis problems solving by physicians. Some type of subproblems that may appear during the diagnosis problems solving are not handled by other medical computational systems. As an example of a subproblem that may appear during a diagnostic establishment by a physician, we mention a medical issue "what is the best-fitted treatment to cure a patient's illness in a particular condition". Assistant expert system agents may help the owner physicians to form cooperative problem solving teams. In addition, they can mediate the cooperative interaction of the physicians during the problems solving.

The paper is structured as follows: in Section 2 are presented agents and multiagent systems used for medical problems solving; Section 3 describes previous researches related with the *MASM* multiagent system; in Section 4 the novel *MASM* multiagent system is described; Section 5 presents the conclusions of the paper.

2 Agent-based medical diagnosis systems

The *capability* of a cognitive system consists in the problem solving *specializations* detained by the system [1, 4]. The *capacity* of a cognitive system consists in the amount of problems that can be solved in deadline by the system using the detained resources [1,4]. *Agents* represent artificial cognitive systems with properties, like [2]: increased autonomy in operation, capability of communication and cooperation with other systems. The systems composed from more agents are called *multiagent systems* [1, 2, 3, 4]. The solving of many difficult problems, require the cooperation of more agents with different [2] capabilities and capacities, which

motivate the use of cooperative multiagent systems versus the use of agents that operate in isolation.

Agents can be endowed with medical knowledge. Many medical agents that can cooperate with physicians and other agents are proposed [5,6,7,8,10,11,12,13,36,56]. The paper [9] presents the state of the art medical information systems and technologies at the beginning of the 21st century. The complexity of construction of full-scaled clinical diagnoses is also analyzed. In the medical domain many type of medical problems can be solved by human medical specialists, which uses medical computational systems, some times agent-based medical systems [30]. As examples of problems, which appear in the medical domain that can be solved by agents, we mention: medical diagnostics elaborations, medical information collection from distributed knowledge bases, medical data collection about patients from distributed databases etc.

A *medical diagnosis problem* represents the description of a patient's illness. The solution of the problem represents the identified illness and the proposed diagnostic to cure the illness. Many medical problems that must be solved in the medical domain represent subproblems of medical diagnosis problems. As an example of subproblem, we mention a medical issue whose answer is necessary in a diagnostic establishment. For example, we mention, the issue "if exists a more effective medicine to cure an illness than a medicine known by a physician" (a physician wants to know if exists a more effective medicine to cure an illness than the medicine known by himself).

Many medical diagnosis systems must have particular proprieties, depending on the types of the medical problems that they must solve. One of the main proprieties, which many times is required by a diagnosing system, consists in the adaptability realized by learning. By learning, a medical computational system may increases the detained medical knowledge accuracy. Is not always possible to create a system, which has all the necessary knowledge and the established knowledge has the required accuracy [1,10,22]. In the paper [8], an intelligent medical diagnosis system with built-in functions for knowledge discovery and data mining is described. The implementation of machine learning technology in the medical diagnosis systems seems to be well suited for medical diagnoses in specialized medical domains. By learning, a system can autonomously construct medical diagnosis rules that can be used in diagnosis processes [8].

Difficulties of some medical diagnosis problems motivate the use of medical multiagent systems for their solving versus medical systems that operate in isolation [52,14]. Cooperative medical multiagent systems may combine the members' *capabilities* and *capacities* in the problems solving. A medical diagnostic elaboration may have many difficulties that imply the cooperation of more human and/or artificial medical specialists in their solving [4]. As examples of difficult problems, we mention the combinations of illnesses where between the illnesses treatments are dependencies. In the following, are mentioned some difficulties in the establishment of a medical diagnostic. A patient may have a combination of illnesses with different symptoms. The symptoms of more illnesses may have some similarities, which make their identification difficult. Symptoms of an illness can be different at different persons who suffer from that illness. In some situations, a patient does not

exhibit the typical symptoms of a particular illness even so he suffers from it. In the case of some illnesses, the causes of the illnesses are not known. A medicine to an illness may have different effects at different persons who suffer from that illness. An illness can be in a very advanced stage that makes the diagnostic elaboration difficult. Difficult medical cases are those in which the patient's symptoms do not sufficiently match typical patterns known by physicians. An illness can be insufficiently known in medicine because it is either new or unusual. In such situations, the symptoms of an illness may or cannot be interpreted properly.

To model agent systems some methodologies were proposed, like: *Him* [49], *Gaia* [50] and *Passi* [51]. The paper [6] analyzes different aspects of the multiagent systems specialized in medical diagnosis. Understanding such systems needs a high-level visual view of how the systems operate to achieve some application related purposes. In the paper [6], a novel method of visualizing the behavior of a medical multiagent system called *Use Case Maps* is proposed. There is described a process for designing agent-based systems using a visual technique, that provide a view of the system as a whole.

In the paper [36], a medical diagnosis multiagent system that is organized according to the principles of swarm intelligence is proposed. It consists of a large number of agents that interact with each other by simple indirect communication. The proposed multiagent system real power stem from the fact that a large number of simple agents collaborates with the purpose to elaborate reliable diagnostics. Agents specialized in medical diagnosis can self-organize in order to provide viable medical diagnoses.

The main motivation that confirms the necessity to use agent-based applications in medical domains consists in the possibility to combine different technologies in the same agent body. Agents represent a very promising recent research direction in medical diagnoses elaborations [11, 12, 14, 36] and fulfilling medical tasks related with the diagnosis processes. As examples of applications of the agents for fulfilling medical tasks, we mention: *patients monitoring* [15], *patients management* [16, 17], *healthcare* [18, 19, 55], *telehealth* [52], *spreads simulation of infectious disease* [54], *web-enabled healthcare computing* [55] and *ubiquitous healthcare* [19].

3 Related works

3.1 Medical expert system agents

Medical expert systems represent relatively classical applications in the medical diagnosis. As examples of well-known medical expert systems, we mention: *Mycin* [20], *Gideon* [38], *Cardiag2* [39], *Puff* [40] and *Casnet* [41]. Expert systems had some success in particular, mainly quite narrow fields of medical expertise, but had problems to cover broader areas of expertise. Some of the problems related with the expert systems are their limited [37, 2]: flexibility, adaptability, extensibility and cooperation capability. The endowment of the expert systems with cooperation capability represents an important research direction [2, 37, 10]. In the paper [10], a system called *Feline* composed of five autonomous medical expert systems with some proprieties of the agents is proposed. These expert systems cooperate to identify the

causes of anemia at cats. There is also presented a development methodology for cooperating expert systems.

In the paper [52], a Web-centric extension to a previously developed expert system specialized in the glaucoma diagnosis is proposed. The proposed telehealth solution publishes services of the developed *Glaucoma Expert System* on the World Wide Web, allowing physicians and patients to interact with it from their own homes. The *Glaucoma Expert System* uses learning algorithms applied on patient data to update and improve its diagnosing rules.

Medical expert system agents represent a novel class of agents developed in our previous work [4, 11, 12, 36]. The medical expert system agents are medical expert systems endowed with agents' capabilities. A medical expert system agent is endowed with a medical specializations set. As examples of specializations of a medical expert system agent, we mention specializations in subdomains of: cardiology, gastroenterology, endocrinology and rheumatology. Expert system agents can perceive and interact with the environment by executing actions in the environment autonomously. Expert system agents can cooperate in the problems solving with other agents and physicians, which allows more flexible problem solving versus the expert systems. If an expert system agent cannot solve a problem (doesn't have the necessary capability and/or capacity), then he can transmit the problem for solving to another agent or physician.

In the papers [12, 48], a cooperative multiagent system specialized in medical diagnosis called *CMDS* (*Contract Net Based Medical Diagnosis System*) is proposed. *CMDS* is a complex system composed from medical expert system agents that cooperatively solve medical diagnosis problems transmitted for solving to the system. For the problems allocation for solving, in the *CMDS* system is used an adaptation of the *contract net problem allocation protocol* [1, 3]. Each agent member of the *CMDS* system can discover autonomously emergent informations and data, which allows the establishment of the agents with who must cooperate in order to solve the overtaken problems.

3.2 Medical ICMA agents

The development of *large-scale medical diagnosis systems* represents an important research direction [7,19,11]. In the paper [7], an *Internet-based holonic medical diagnosis system* for diseases is proposed. The holonic medical diagnosis system consists of a tree-like structured alliance of agents specialized in medical diagnoses, which collaborate in order to provide viable medical diagnoses, combining the advantages of holonic systems and multiagent systems.

The software mobile agents are capable to change their location in the network where they operate, which allows different problems solving using distributed resources [34, 35]. A research direction related with the development of large-scale diagnosis systems is represented by the approaches based on software mobile agents [11, 19]. *OnkoNet* mobile agents described in the literature have been used successfully for patient-centric medical problems solving [19]. In the paper [19], is introduced the notion *ubiquitous healthcare* (any-time/any-place access of health

services via mobile computing), addressing the access of health services by individual consumers using mobile agents. This access requires medical knowledge about the individual health status (relevant recent diseases). The *OnkoNet* mobile agent architecture involve architectures on the macrolevel and microlevel as well as cooperation protocols. The work presented in the paper [19], emerged from a project covering all relevant issues, from empirical process studies in cancer diagnosis/therapy, down to system implementation and validation.

The main disadvantages of the mobile agents are related with their limited [34,35]: communication capability, protection against network sources and malicious hosts, intelligence and capability to use knowledge bases in the problems solving. These disadvantages limit the use of the mobile agents for medical problems solving in insecure networks. In our previous researches, a novel mobile agent architecture called *ICMA* (*Intelligent Cooperative Mobile Agent Architecture*) was developed [24,25]. The proposed mobile agent architecture represents a combination of the mobile and static agent paradigms. The purpose of the research was the development of an architecture that allows the creation of mobile agents, which can solve intelligently medical diagnosis problems. *ICMAE* agents (*Intelligent Cooperative Mobile Agents with Evolutionary Problem Solving Capability*) represent a novel class of mobile agents developed in our previous works [31,32,33]. *ICMAE* agents are agents with the *ICMA* architecture endowed with problem solving specializations based on genetical algorithms. *ICMAE* agents can solve large numbers of problems, using efficiently problems solving resources distributed in the network [31,32,33].

Advantages of the mobile agents endowed with the *ICMA* architecture versus some of the mobile agents described in the literature consist in their increased: communication capability [25], protection against network sources and malicious hosts [26] and intelligence [53]. These advantages of the *ICMA* agents suggest their use for different problems solving in the medical domain (problems whose solving require distributed medical information collection and processing).

Medical ICMA agents represent a novel class of medical agents, based on the *ICMA* architecture. *Medical ICMA agents* can diagnoses illnesses. In the paper [11], a large-scale medical diagnosis system called *LMDS* (*Large-scale Medical Diagnosis System*) is proposed. *LMDS* is a complex system composed from medical expert system agents and medical *ICMA* agents. The agents' members of the diagnosis system cooperatively solve the overtaken diagnosis problems. A medical *ICMA* agent can cooperate with other agents in the problems solving. A medical *ICMA* agent can migrate in the network with an overtaken problem until will find an agent capable to solve the problem. The open *LMDS* system may have a large number of members.

3.3 Medical assistant expert system agents

Medical assistant expert system agents represent medical expert system agents endowed with components, which allow to them to help the physicians during their work. In (1) is illustrated a medical assistant expert system agent denoted $MAESA$. Where: ESA represents the expert system agent part of the agent; $C_1, C_2, \ldots, C_n$ represent the components that allows the assistance offering to the physicians in different problems solving.

$$MAESA =< ESA > + < C_1 : C_2 : \cdots : C_n > . \quad (1)$$

An assistant expert system agent can help intelligently physicians and other agents during the problems solving processes. This capability is implemented by the particular components of the agents. The assistant expert system agents are endowed with knowledge, about the system in which they operate and knowledge about the helped agents and/or physicians, which allows the assistance offering.

In the papers [14,28] a cooperative medical agent-based diagnosis system, proposed for difficult medical diagnosis problems solving called *BMDS* (*Blackboard-based Medical Diagnosis System*) is described. *BMDS* is a complex system, composed from: expert system agents and *assistant expert system agents*. The problems solving by the proposed diagnosis system is partially based on the *blackboard-based problems solving* [29,3].

The effectiveness of the *CMDS*, *BMDS* and *LMDS* multiagent systems demonstrate that medical expert system agents can be used successfully as members of complex medical diagnosis multiagent systems. The increased intelligence of the expert system agents is proved in [21]. *Medical assistant expert system agents* used by the *BMDS* system, proves that medical expert system agents can be adapted to help intelligently medical specialists during the medical diagnoses processes. Medical assistant expert system agents require future developments, in order to improve their limitations related with the endowment with a large quantity of problem solving knowledge, operation in a distributed environment and efficient solving of a large variety of problems.

4 The MASM multiagent system

Many medical problem solving require a cooperative multiagent system that can solve the whole problem or can cooperate during the problem solving with medical human specialists. As an example of difficult medical problem, we mention the glaucoma identification, diagnosis and therapy [52,53,58,59,60,61]. Glaucoma is a progressive eye disease that damages the optic nerve, usually associated with increased intraocular pressure. If left untreated, it can lead to blindness. The glaucoma is a challenge for the ophthalmologist is not as much the diagnosis itself but rather the evaluation of the risk for its occurrence and the prediction of disease progression. In about 70% of the cases, the diagnosis of glaucoma is evident for ophthalmologists. In preparation for a glaucoma surgery intervention, a very high amount of information about the patient has to be gathered: medical history and particular conditions that may influence the surgery outcome. Even in the simplest case, many factors have to be considered when making a decision for surgery and the necessary information is difficult to find without strong informational support, which motivate the use of an assistant multiagent system that may help the physicians in different points of decisions.

In the following, an assistant multiagent system called *MASM* (*Medical Assistant Multiagent System*), composed from a set $AS = \{As_1, As_2, \ldots, As_h\}$ of *assistant expert system agents* that can help a set $PH = \{Ph_1, Ph_2, \ldots, Ph_w\}$ of physicians in they work is proposed. In the following, the assistant expert system agents are

called shortly *assistant agents*. Each physician is the owner of an assistant agent (in the system may exists agents that are not owned by physicians). Figure 1 presents a physician (denoted Ph_i) interaction with the proposed multiagent system. The arrows used in Figure 1 present the *types* $T = \{t_a, t_c\}$ of cooperation. t_a is a cooperation link between a physician and the owned assistant agent. t_c is a cooperation link between two assistant agents.

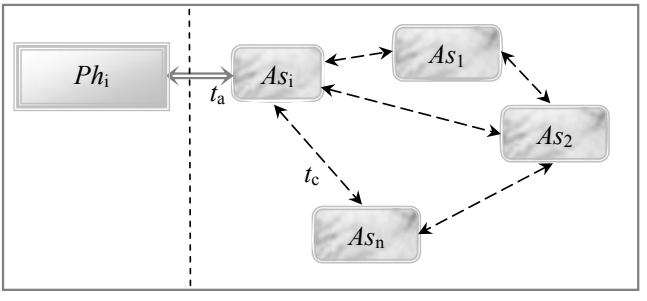


Fig. 1. A physician interaction with the MASM system

A physician denoted Ph_i ($Ph_i \in PH$) interacts directly only with its owned assistant agent As_i ($As_i \in AS$, $Link(As_i, Ph_i) = t_a$). As_i can cooperate during its operation with other assistant agents with who has a cooperation link by the type t_c (in the system may exists assistant agents between who doesn't exist a cooperation link). $As_1, As_2, \ldots, As_n$ represent the agents with who As_i can cooperate during the problems solving.

4.1 MASM system operation

The algorithm *MASM Problem Solving* describes the cooperative elaboration of a medical diagnostic (a P_z medical diagnosis problem solving) by a physician denoted Ph_i and the agents members of the *MASM* system. As_i ($As_i \in AS$) represents the agent owned by Ph_i ($Link(Ph_i, As_i) = t_a$). The solution $S_z = [Il_z, Tr_z]$ of the P_z problem, represents the patient's (denoted Pat_c) Il_z illness and the Tr_z treatment that must be applied to cure Il_z.

Algorithm - MASM Problem Solving

{IN: P_z - the diagnosis problem}

{OUT: $S_z = [Il_z, Tr_z]$ - the solution of P_z}

Step 1. Information collection about the patient.

@Pat_c personal informations (name, identification number etc) are collected and transmitted to As_i ($Link(Ph_i, As_i) = t_a$).

@Informations about the Pat_c patient's illness (the illness symptoms, the history of the symptoms etc.) are collected by Ph_i and transmitted to As_i.

@As_i will search for medical informations and data about Pat_c (Pat_c past illnesses descriptions, Pat_c known allergies to some medicines etc) detained in its medical database.

@As_i establishes the set AC of assistant agents with who have a cooperation link by the type t_c. As_i transmits to the agents AC the requirement to obtain the informations about Pat_c medical history.

@As_i selects from the informations transmitted by the agents AC, the informations considered useful for Ph_i (informations that describe Pat_c past illnesses and different medical informations about the patient). These informations are transmitted by As_i to Ph_i.

@The rest of the informations are obtained by physical examination by Ph_i. During the physical examination, Ph_i looks for signs of the illness (signs are manifestations of the illness that the physician can see or feel).

Step 2. Cooperative solving of the problem.

While (P_z solution is not obtained) *do*

@Ph_i continues P_z solving. Ph_i may require investigations such laboratory tests, X-rays, MRI, computed tomography etc.

If (Ph_i requires the help of As_i) *then*

@As_i plans how can fulfill the requirement of Ph_i cooperatively with assistant agents with who has a cooperation link by type t_c.

@As_i tries to realize the requirement of Ph_i.

If (As_i succeeded to realize the requirement of Ph_i) *then*

@As_i announces Ph_i about the obtained results.

else

@As_i announces Ph_i about its incapability to fulfill the requirement.

EndIf

EndIf

If (As_i establishes the assistance that can offer to Ph_i) *then*

@As_i plans how can help Ph_i.

@As_i helps Ph_i based on the elaborated plan.

EndIf

EndWhile

@Let $S_z = [Il_z, Tr_z]$ be the solution obtained by Ph_i.

Step 3. Formation of the problem solution.

@Ph_i transmits the identified illness and the established treatment to As_i.

If (As_i finds mistakes in S_z) *then*

@As_i announces Ph_i about the observed mistakes in S_z.

If (As_i can suggests measures that eliminate some of the mistakes) *then*

@As_i suggests to Ph_i the measures that can offer solution to the elimination of some of the mistakes.

EndIf

@Ph_i eliminates the mistakes from S_z. If is necessary Ph_i cooperates with As_i.

@Let $S_z = [Il_z, Tr_z]$ be the reviewed solution.

EndIf

Step 4. Validation of the problem solution.

@Ph_i validates S_z. During S_z validation, Ph_i may require As_i help.

EndMASMProblemSolving.

A *problem solving statement* contains a set of medical informations and data obtained during a diagnosis problem solving. The general form of a problem solving statement is illustrated in (2).

$$< [St_1 : 1 : ty_1 : nr_1]; [St_2 : 2 : ty_2 : nr_2]; \ldots ; [St_m : m : ty_m : nr_m] > . \qquad (2)$$

In $St = \{St_1, St_2, \ldots, St_m\}$ can be retained informations and data obtained during a diagnosis problem solving. $St_1, St_2, \ldots, St_m$ from (2) may have a single value or a list of values. As examples of an St_a ($St_a \in St$) parameter value (values), we mention: an Il_a illness specification, Il_a symptoms, Il_a syndromes, past illnesses of the patient, a Dl_a diagnostic proposed to cure Il_a etc. To each parameter St_a, that can contain effective medical information is associated a numerical value a (a is unique in the parameters' identifying numbers' list), ty_a specifies a type of information and/or data that can be retained in St_a, and $nr_a = \{min_a, max_a\}$ specifies how many values (minimum and maximum number) can be retained in St_a.

For example, we mention $[St_f : f : ty_f : nr_f]$ completed during a diagnosis problem solving with [*Depression*, *Increased appetite*, *Headache*, *Cough*, *Insomnia* : 5 : *symptoms* : $\{1, more\}$]. *Depression*, *Increased appetite*, *Headache*, *Cough*, *Insomnia* is a list of symptoms. $f = 5$ specifies that $[St_f : f : ty_f : nr_f]$ is the 5th parameter in the parameters list specified in the general form (2). $ty_5 = symptoms$ specifies that St_f contain symptoms. $nr_f = \{1, more\}$ specifies that in St_f can be retained a list of symptoms (at least one symptom).

Retaining the medical informations and data related with a diagnosis problem solving using the form (2), allows to the assistant agents to quickly and precisely identify the types of informations detained about the problem solving. An assistant agent based on the parameter identifier, knows with which parameter of the preconditions of its diagnosing rules must try to fit in order to establish an applicable rule. The

types of medical informations and data that can be retained in (2), are established based on the specifics of the illnesses that are diagnosed by the system (infectious disease for example). Initially, a problem solving statement contains informations that describes an illness (for example, initially are known about the diagnosed illness only the symptoms enumerated by the patient). During the solving of the problem, the problem solving statement is changing, by adding new informations and data by agents. Physicians can add, remove and modify knowledge in a problem solving statement. The physicians can make decisions that are more rational in difficult medical decisions versus the artificial agents. An artificial agent may be unable to make rational decisions when the decisions elaboration requires knowledge, that is not detained by him or there are uncertainties that are not observed or cannot be handled by the agent. A final problem solving statement contains the established diagnostic. During of a problem solving some of the parameter values may not be completed (for example, is not necessary to retain the history of the diagnosed illness symptoms).

An As_b assistant agent diagnosing knowledge used in medical diagnoses elaborations is composed from a set Rl (3) of diagnosing rules.

$$Rl = \{Rl_1, Rl_2, \ldots, Rl_g\}. \tag{3}$$

An Rl_h ($Rl_h \in Rl$) diagnosing rule has the form (4).

$$Rl_h : Prec_h \rightarrow Postc_h. \tag{4}$$

$Prec_h$ and $Postc_h$ have parameters that specify informations, which describe different aspects of a diagnosis process. In $Prec_h$ and $Postc_h$ appear a set Par (5) of parameters (some of the parameters specified in (2)).

$$Par = \{Par_1, Par_2, \ldots, Par_y\}. \tag{5}$$

A Par_e ($Par_e \in Par$) parameter in (5) has the form (6).

$$Par_e = [St_e : e : ty_e : nr_e]. \tag{6}$$

St_e contains medical informations and/or data. ty_e represents the type of medical informations retained in St_e. nr_e represents the number of values retained in St_e. e is a natural number, which specify the same type of information that specify e in (2). If Par_e appear in the precondition of a rule, then it specify conditions that must be fulfilled for the applicability of the rule in whose precondition appear. If Par_e appears in the postcondition of a rule, then it specifies informations that must be added into the problem solving statement (for example, a supposed illness or a suggested treatment), which represent the informations about the problem that is diagnosed, if the precondition of the rule is verified. When an agent overtakes a medical diagnosis problem for processing, it will identify a diagnosing rule whose postcondition specify informations that can be added into the problem solving statement. The knowledge contained in a problem solving statement is understandable to the physicians, they can modify it by adding, modifying or retracting knowledge.

An assistant agent may present a problem solving statement in an understandable form to the physicians if is required, process that may suppose the organization and

filtering of the informations. During a diagnosis process (the solving of a medical diagnosis problem by a physician), a physician can interact with the *MASM* system in different points of decisions, in order to obtain the solutions of different subproblems that have appeared or to increases the decisions accuracy. Figure 2 presents the points of interactions between a physician denoted Ph_i and the assistant multi-agent system *MASM*, in order to identify a patient's illness and establishes a proper diagnostic to cure the illness. During the diagnostic establishment, Ph_i interacts directly only with the owned assistant agent As_i ($As_i \in AS, Link(As_i, Ph_i) = t_a$). If is necessary, As_i can interact with other assistant agents with who have a cooperation link by the type t_c. As_i has a knowledge base that contains different informations about Ph_i and the members of the *MASM* system with who have a cooperation link.

About Ph_i, the owned assistant agent As_i detains informations, like:

- Ph_i specialization in medicine;
- Ph_i working program;
- the medical history of the patients of Ph_i;
- data about other physicians (specialization in medicine, working program etc.) that use the system, with who Ph_i usually cooperates in the diagnoses elaborations (require and/or offer advices from/to them).

About an As_k ($As_k \in AS$) assistant agent, the assistant agent As_i ($As_i \in AS$, $Link(As_i, As_k) = t_c$) may detains informations, like:

- As_k specialization. For example, specialization in information searches about physicians. As an example, As_k may be capable to find physicians specialized in cardiology;
- the knowledge representation language used by As_k in the detained knowledge representation. A knowledge representation language allows the representation of the used knowledge [3].

An As_i ($As_i \in AS$) assistant agent can offer autonomously assistance to its owner physician denoted Ph_i ($Link(Ph_i, As_i) = t_a$), based on the knowledge detained in its knowledge base and the informations known about the diagnosis problem that is currently solved by Ph_i. In the assistance offering, As_i can cooperate with other agents.

A physician denoted Ph_i may transmit subproblems to its As_i ($link(As_i, Ph_i) = t_a$) owned assistant agent, who if is necessary cooperates with a set $As_j, As_v, \ldots, As_q$ of agents in order to solve them (As_i have a cooperation link by the type t_c with $As_j, As_v, \ldots, As_q$). The received results by As_i from $As_j, As_v, \ldots, As_q$ will be transmitted to Ph_i. Ph_i establishes the diagnostic based on its different observations and the results obtained from As_i.

As examples of subproblems of a diagnosis problem, we mention:

- a supposed Il_c illness. Ph_i requires to As_i to analyze if a patient has the Il_c illness, based on the patient's illness description;
- the medical ontology known by As_i. A medical ontology represents a dictionary of used terms in the medical domain [42]. As examples of developed medical otologies described in the literature, we mention: *Galen* [43], *Umls* [44], *OntHoS* [19], *LinkBase* [45], *Tambis* [46] and *Gene* [47].

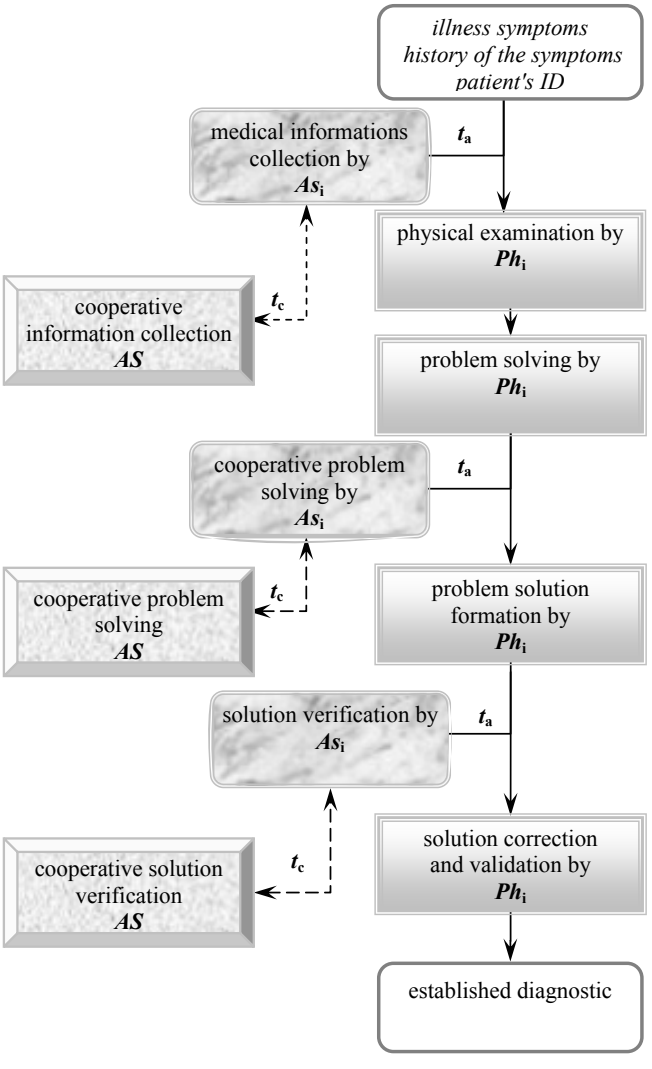


Fig. 2. Summary of a cooperative diagnosis elaboration

Figure 3 presents a difficult medical diagnosis problem solving complexity distribution between more physicians and the *MASM* system. *Problem* presents the problem that must be solved. Ph_1, Ph_2,..., Ph_j represent the physicians that contribute to the problem solving. As_1, As_2,..., As_b represent the assistant agents (owned by Ph_1, Ph_2,..., Ph_j, and may contain agents that are not owned by physicians) that contribute to the problem solving by helping the physicians during the problems solving.

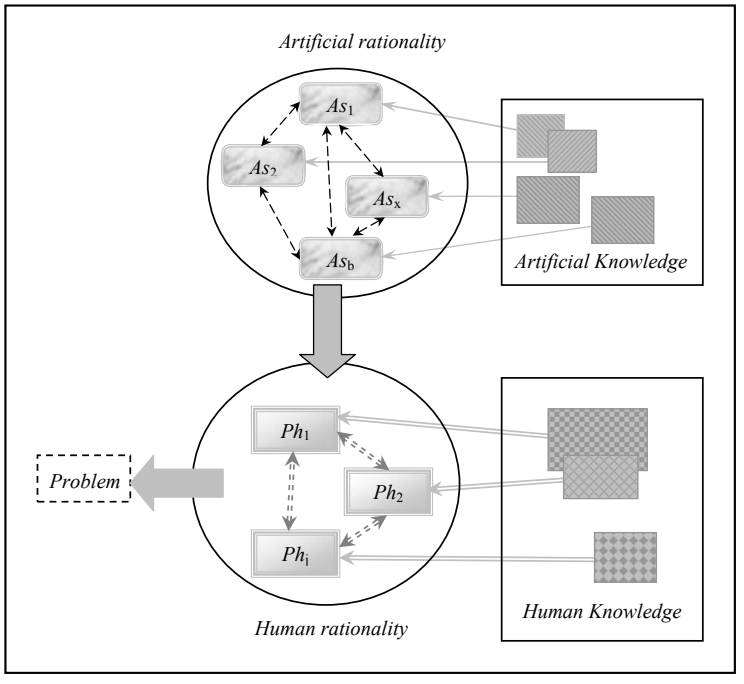


Fig. 3. Problem solving complexity distribution in the MASM system

An assistant agent can cooperates with other agents in order to collect medical knowledge and medical informations about patients. As examples of medical informations that can be collected by an As_i ($As_i \in AS$) agent about a patient, we mention the descriptions of the patient's medical history and the patient's current known illnesses. As examples of medical knowledge that can be collected by an As_k ($As_k \in AS$) agent, we mention: an illness description (the symptoms of the illness), a new medicine that can be used to cure an illness; the success of different diagnostics applied to cure an illness.

An As_i assistant agent can verify the correctness of a problem's solution obtained by its owner physician denoted Ph_i. As_i knows the problem that is solved by Ph_i (the problem description is transmitted to the assistant agent). If is necessary As_i may cooperate with other agents. For example, a physician specialized in general medicine and an agent specialized in general medicine can try to establishes simultaneously a diagnostic to cure an illness. The assistant agent can compare the obtained solutions. If the solutions differ, the assistant agent announces the physician about the uncertainty of the correctness of the obtained solution.

4.2 Cooperative diagnoses elaborations by physicians and agents

The *MSAM* system can contribute to the solving of different subproblems that appear during the diagnosis problems solving by physicians. Some types of

subproblems are not handled by other medical computational systems. In the following, is considered that the *MASM* system is used by a set $PH = \{Ph_1, Ph_2, \ldots, Ph_r\}$ of physicians. $AR = \{As_1, As_2, \ldots, As_r\}$ $(AR \subseteq AS)$ denote the assistant agents owned by $Ph_1, Ph_2, \ldots, Ph_r$. An assistant agent can manage intelligently the cooperation between the owner physician and other physicians. In the following, there are presented two cooperation scenarios. The scenario called *Cooperative Finding of the Answer to a Medical Issue* describes a medical issue cooperative solving by more physicians assisted by the *MASM* system. The scenario called *Cooperative Finding the Solution of a Diagnosis Problem* describes how the *MASM* system finds the physicians capable to solve cooperatively a medical diagnosis problem.

The Scenario of a Medical Issue Solving

Studies prove that under the same circumstances and based on the same information, judgments of the two physicians only match on 78% of the cases, which points to the necessity of further judgements or opinions [8]. In the following, is considered the situation when a physician denoted Ph_i $(Ph_i \in PH)$ wants to find the opinion of more physicians about a medical issue. As an example of medical issue, we mention a hypothesis related with a diagnostic to cure an illness, in order to establish an answer with a higher accuracy. The cooperative finding of an accurate answer to the issue, doesn't suppose self-organization in the system. Scenario - *Cooperative Finding of the Answer to a Medical Issue* describes the cooperative finding of the answer to the medical issue denoted mi emitted by Ph_i.

Scenario - Cooperative Finding of the Answer to a Medical Issue

Step 1 - The medical issue transmission to the physician's assistant agent.

$Ph_i(mi) \Rightarrow As_i$.

As_i establishes the physicians $Ph_1, Ph_2, \ldots, Ph_k$ considered capable to answer to mi.

As_i establishes the agents $As_1, As_2, \ldots, As_k$ owned by $Ph_1, Ph_2, \ldots, Ph_k$.

$As_i(mi) \Rightarrow [As_1, As_2, \ldots, As_k]$.

Step 2 - The medical issue transmission for the obtaining of the answer to it.

@mi handling by As_1.

If (As_1 establishes that Ph_1 is available) *then*

$As_1(mi) \Rightarrow Ph_1$.

$Ph_1(mi, rp_1) \Rightarrow As_1$.

else

As_1 establishes rp_1 autonomously.

EndIf

$As_1(mi, rp_1) \Rightarrow As_i$.

@mi handling by As_2.

If (As_2 establishes that Ph_2 is available) *then*

$As_2(mi) \Rightarrow Ph_2$.

$Ph_2(mi, rp_2) \Rightarrow As_2$.

else

As_2 establishes rp_2 autonomously.

EndIf

$As_2(mi, rp_2) \Rightarrow As_i$.

...

@mi handling by As_k.

If (As_k establishes that Ph_k is available) *then*

$As_k(mi) \Rightarrow Ph_k$.

$Ph_k(mi, rp_k) \Rightarrow As_k$.

else

As_k establishes rp_k autonomously.

EndIf

$As_k(mi, rp_k) \Rightarrow As_i$.

Step 3 - Establishment of the answer to the issue.

$useful = As_iFilter(rp_1, rp_2, \ldots, rp_k)$.

$As_i(mi, useful) \Rightarrow Ph_i$.

Ph_i establishes the answer to mi based on the $useful$ knowledge.

Step 4 - Learning from the received responses.

$learned = As_iLearn(rp_1, rp_2, \ldots, rp_k)$.

EndMedicalIssueSolving

" $\Rightarrow$ " denote the transmission of a message by a *source* to a *destination* (where: *source* and *destination* can be agents or humans). Ph_i transmits mi to its owned agent denoted As_i ($Link(Ph_i, As_i) = t_a$). As_i based on its detained knowledge establishes the physicians $Ph_1, Ph_2, \ldots,$ Ph_k considered capable (have the necessary medical specialization and are available) to answer to mi. As_i will transmit mi to the assistant agents $As_1, As_2, \ldots, As_k$ owned by the choused physicians $Ph_1, Ph_2, \ldots, Ph_k$ (As_i have a cooperation link by the type t_c with $As_1, As_2, \ldots,$ As_k). $As_1, As_2, \ldots,$ As_k will overtake the responsibility to transmit mi to their owner physicians and send back the answer to As_i. mi is transmitted by an assistant agent to the owner physician only if the physician is available. $rp_1, rp_2, \ldots, rp_k$ represent the responses to mi. As examples of answers that can be received by As_i from an agent As_j ($Link(As_i, As_j) = t_c$) owned by Ph_j ($Link(Ph_j, As_j) = t_a$), we mention: As_j have established that Ph_j is unavailable, Ph_j answer to the issue, specification that Ph_j cannot answer to the issue. From the received responses,

As_i filters the useful responses (contains the physicians responses to the issue). As_i can learn from the received responses. *learned* represents the learned knowledge by As_i. As examples of informations that As_i can learn, we mention: what physicians are usually unavailable, what physicians can answer fast to different issues etc. The knowledge denoted $useful$ obtained after the filtering process is transmitted to Ph_i. Ph_i will establishes the answer to mi based on the $useful$ knowledge received from As_i.

The Scenario of a Medical Diagnosis Problem Solving

In the following, is considered the situation when a physician denoted Ph_i could not solve a medical diagnosis problem denoted P_q. In order to solve the problem, Ph_i must cooperate with other physicians. Cooperation scenario - *Cooperative Finding the Solution of a Diagnosis Problem* describes the cooperative solving of P_q by a set of physicians helped by the *MASM* system. In order to solve the problem there are two types of self-organizations at different levels. The first type of self-organization is realized at the level of the assistant agents. The second type of organization is realized at the level of the assistant agents' owner physicians.

Scenario - Cooperative Finding the Solution of a Diagnosis Problem

Step 1 - The problem initial handling.

$Ph_i(P_q, D) \Rightarrow As_i$.

As_i establishes the physicians $Ph_1, Ph_2, \ldots, Ph_m$ considered capable to cooperate in order to solve P_q.

As_i establishes the $As_1, As_2, \ldots, As_m$ agents owned by $Ph_1, Ph_2, \ldots, Ph_m$.

$As_i(P_q, D) \Rightarrow [As_1, As_2, \ldots, As_m]$.

Step 2 - The analysis of the problem solving.

@Establishment if Ph_1 will cooperate in the P_q solving.

As_1 verifies if Ph_1 is capable to cooperate in the P_q solving.

If (Ph_1 is available) *then*

$As_1(P_q, D) \Rightarrow Ph_1$.

Ph_1 establishes if will cooperate in the P_q solving.

$Ph_1(P_q, rp_1) \Rightarrow As_1$.

else

As_1 establishes rp_1 autonomously.

EndIf

$As_1(P_q, rp_1) \Rightarrow As_i$.

@Establishment if Ph_2 will cooperate in the P_q solving.

As_2 verifies if Ph_2 is capable to cooperate in the P_q solving.

If (Ph_2 is available) *then*

$As_2(P_q, D) \Rightarrow Ph_2$.

Ph_2 establishes if will cooperate in the P_q solving.

$Ph_2(P_q, rp_2) \Rightarrow As_2$.

else

As_2 establishes rp_2 autonomously.

EndIf

$As_2(P_q, rp_2) \Rightarrow As_i$.

...

@Establisment if Ph_m will cooperate in the P_q solving.

As_m verifies if Ph_m is capable to cooperate in the P_q solving.

If (Ph_m is available) *then*

$As_m(P_q, D) \Rightarrow Ph_m$.

Ph_m establishes if will cooperate in the P_q solving.

$Ph_m(P_q, rp_m) \Rightarrow As_m$.

else

As_m establishes rp_m autonomously.

EndIf

$As_m(P_q, rp_m) \Rightarrow As_i$.

Step 3 - The cooperative physicians' team formation.

$useful = As_iFilter(rp_1, rp_2, \ldots, rp_m)$.

$As_i(P_q, useful) \Rightarrow Ph_i$.

Ph_i based on the informations $useful$ will establishes the set of physicians denoted Pt with who will cooperate in the P_q solving.

Ph_i(specification of the set Pt of physicians) $\Rightarrow As_i$.

As_i announces the physicians Pt about their acceptance in the cooperative problem solving team.

Step 4 - Solving of the problem by physicians helped by their owned agents.

P_q is solved cooperatively by the set $Pt \cup \{Ph_i\}$ of physicians interloped by their owned agents. S_q the solution of P_q is formed.

S_q is validated by the set $Pt \cup \{Ph_i\}$ of physicians.

EndMedicalDiagnosisProblemSolving

Ph_i requires the help of the owned As_i assistant agent ($Link(Ph_i, As_i) = t_a$) in finding physicians with who can cooperate in the P_q solving. As_i will establish the $Ph_1, Ph_2, \ldots, Ph_m$ physicians considered capable to solve cooperatively with Ph_i the P_q problem. D represents the medical domain (domains) in which is considered that the diagnosed illness is included. $rp_1, rp_2, \ldots, rp_m$ represent the responses (acceptance or rejection) of the physicians $Ph_1, Ph_2, \ldots, Ph_m$ to Ph_i requirement to cooperate in the P_q solving. After that, As_i will manage the cooperation of Ph_i with the set Pt ($Pt \subseteq PH$) of physicians. In some of the interactions, the physicians can use as interloper their owned agents. P_q solution denoted S_q is formed and validated, cooperatively by physicians from the set $Pt \cup \{Ph_i\}$ of physicians, who are specialized in the medical domain (domains) in which the identified illness (illnesses) is included.

As examples of assistance that an As_i agent can offer to its owner physician denoted Ph_i ($Link(Ph_i, As_i) = t_a$) in a P_q medical diagnosis problem cooperative solving by Ph_i with the set $Pt = \{Ph_1, Ph_2, \ldots, Ph_m\}$ of physicians, we mention:

- As_i can transmit known information to $Ph_1, Ph_2, \ldots, Ph_m$ without consulting Ph_i. For example, may transmit medical information about the patient whose illness is diagnosed cooperatively. However, is not necessary for Ph_i to specify explicitly all the details to $Ph_1, Ph_2, \ldots, Ph_m$ required in the P_q solving;
- As_i can add autonomously the ontology of the used terms in a message transmitted by Ph_i to another physician Ph_v. For example, may add alternative names of the same medicine.

4.3 Advantages of the MASM system

More physicians can use the proposed *MASM* system. Each physician interacts directly only with its owned assistant expert system agent, the rest of the system's member agents are hidden to him. Each assistant agents know how must cooperate with some other agents to offer the necessary help to its owner physician. Cooperative agents teems are formed when they become necessary during the problems solving.

More physicians using as interlopers their owned assistant agents, can form teams in order to solve cooperatively medical diagnosis problems or just establishes answers to different issues. The cooperation links between physicians are established by the assistant agents during the problems solving. The assistant agents of the physicians are responsible for the cooperative medical teams' formations. In addition, they can assist intelligently the physicians in their interaction. Cooperative physicians' teams are formed when they become necessary.

The proposed assistant multiagent system cannot substitute the physicians, but the cooperative problems solving with the physicians, may increases the accuracy of the elaborated medical diagnostics by the physicians. A final diagnostic establishment and validation in the case of an illness is the responsibility of a physician. By cooperation, the human and artificial thinking advantages in the problems solving are combined. Physicians can elaborate decisions using their medical knowledge and intuition formed in time. However, they can diagnoses difficult illnesses. The

diagnostics accuracy and elaboration time can be improved, if in some points of decisions, the physicians use the *MASM* system that can verify medical hypotheses and execute different actions autonomously, in order to make the physicians work easier.

An assistant agent's specialization contains the description of the help that can offer the agent to the physicians and other agents. An assistant agent can be endowed with new specializations; the inefficient specializations can be eliminated or improved. New assistant agents can be introduced in the multiagent system. The adaptation of a cooperative multiagent system in the efficient solving of a problem many times is easier than the adaptation of an agent that solves the same problem [1, 2, 3].

5 Conclusions

The solving of many medical diagnosis problems is a complex task that may require diagnosing knowledge from different medical domains, which cannot be detained by a single physician or an individual medical computational system. Recently developed medical agents and medical multiagent systems, prove that they represent an efficient approach for many diagnosis problems solving [27, 24, 23, 28, 11, 12, 14, 36]. Agents may solve medical tasks, related with the diagnostics elaborations and treatments fulfilling. As examples of applications of the agents for fulfilling medical tasks, different from medical diagnoses, we mention: patients monitoring and management [15, 16, 17], healthcare related problems solving [18, 19, 55, 12] and telehealth [52].

In this paper, we have proposed a novel assistant multiagent system called *MASM* (*Medical Assistant Multiagent System*) that can help more physicians during the medical diagnosis problems solving. *MASM* is composed from medical assistant expert system agents, which cooperate in order to discover difficult medical diagnosis problems solutions. Difficult medical cases are those in which is difficult the illnesses identification and the corresponding efficient treatments establishment. *Medical assistant expert system agents* represent a novel class of agents developed in our previous researches [14,28]. *Assistant expert system agents* represent an adaptation of the *medical expert system agents* [4, 11, 12, 36] developed in our previous researches, in order to help flexibly physicians and other agents in the problems solving. The application of the medical assistant expert system agents presented in [14, 28], demonstrates that they can be used as members of complex medical multiagent systems.

The proposed *MASM* multiagent system is a complex system, composed from relatively simple agents that cooperatively with physicians can solve medical diagnosis problems. *MASM* multiagent system can autonomously discover emergent proprieties, like cooperation links between agents and cooperation links between physicians, which allows to the system to self-organize, in order to help the physicians to solve more efficiently and accurately the medical diagnosis problems. The MASM system reduces the complexity of the diagnosis processes elaborated by the physicians, by hiding some tasks during the diagnosis processes. The main novelty related with the developed system consists in the particular intelligent assistance

of the physicians during the problems solving. The intelligence of an artificial computational system can be measured how "well" (flexibility, accuracy, efficiency, capability to handle uncertainties) the system can solve complex problems.

One of the purposes of some developments of complex systems consists in the efficient solving of difficult problems by them. The *MASM* multiagent system is illustrative in this sense. Represents a solution for the physicians helping during medical diagnosis problems solving by decreasing the complexity of the medical tasks that the physicians must fulfill. The system is capable to self-organize in order to handle the complexity of the problems solving.

References

1. Ferber, J.: Multi-Agent Systems: An introduction to Distributed Artificial Intelligence. Addison Wesley, London (1999)
2. Iantovics, B.L., Chira, C., Dumitrescu, D.: Principles of the Intelligent Agents. Casa Cartii de Stiinta Press, Cluj-Napoca (2007) (in Romanian)
3. Weiss, G. (ed.): Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence. MIT Press, Cambridge (2000)
4. Iantovics, B.L.: Cooperative Medical Diagnosis Systems. In: Proceedings of the 2nd International Conference Interdisciplinarity in Engineering, Tg. Mures, Romania, pp. 669–674 (2005)
5. Laita, L.M., Gonzlez-Paez, G., Roanes-Lozano, E., Maojo, V., de Ledesma, L., Laita, L.: A Methodology for Constructing Expert Systems for Medical Diagnosis. In: Crespo, J.L., Maojo, V., Martin, F. (eds.) ISMDA 2001. LNCS, vol. 2199, pp. 146–152. Springer, Heidelberg (2001)
6. Abdelaziz, T., Elammari, M., Unland, R.: Visualizing a Multiagent-Based Medical Diagnosis System Using a Methodology Based on Use Case Maps. In: Lindemann, G., Denzinger, J., Timm, I.J., Unland, R. (eds.) MATES 2004. LNCS (LNAI), vol. 3187, pp. 198–212. Springer, Heidelberg (2004)
7. Unland, R.: A Holonic Multi-agent System for Robust, Flexible, and Reliable Medical Diagnosis. In: Meersman, R., Tari, Z. (eds.) OTM-WS 2003. LNCS, vol. 2889, pp. 1017–1030. Springer, Heidelberg (2003)
8. Alves, V., Neves, J., Maia, M., Nelas, L.: A Computational Environment for Medical Diagnosis Support Systems. In: Crespo, J.L., Maojo, V., Martin, F. (eds.) ISMDA 2001. LNCS, vol. 2199, pp. 42–47. Springer, Heidelberg (2001)
9. Vesnenko, A.I., Popov, A.A., Pronenko, M.I.: Topo-typology of the structure of full-scaled clinical diagnoses in modern medical information systems and technologies. Cybernetics and Systems Analysis 38(6), 911–920 (2002)
10. Wooldridge, M., O'Hare, G.M.P., Elks, R.: Feline - A case study in the design and implementation of a co-operating expert system. In: Proceedings of the International Conference on Expert Systems and their Applications, Avignon (1991)
11. Iantovics, B.L.: Agent-Based Medical Diagnosis Systems. In: Computing and Informatics, Slovak Academy of Sciences, Bratislava (2007)
12. Iantovics, B.L.: A Novel Medical Diagnosis System. In: Bourgine, P., Kepes, F., Schoenauer, M. (eds.) Proceedings of the European Conference on Complex Systems, ECCS 2005, Paris, pp. 961–962 (2005)

13. Mitsukura, Y., Mitsukura, K., Fukumi, M., Akamatsu, N., Witold-Pedrycz, W.: Medical Diagnosis System Using the Intelligent Fuzzy Systems. In: Negoita, M.G., Howlett, R.J., Jain, L.C. (eds.) KES 2004. LNCS (LNAI), vol. 3213, pp. 807–826. Springer, Heidelberg (2004)
14. Iantovics, B.L.: A Novel Diagnosis System Specialized in Difficult Medical Diagnosis Problems Solving. Emergent Properties In Natural And Artificial Dynamical Systems. In: Aziz-Alaoui, M.A., Bertelle, C. (eds.) Understanding Complex Systems, pp. 187–197. Springer, Heidelberg (2006)
15. Hayes-Roth, B., Hewett, M., Washington, R., Hewett, R., Seiver, A.: Distributing intelligence within an individual. In: Gasser, L., Huhns, M. (eds.) Distributed Artificial Intelligence, vol. II, pp. 385–412. Pitman Publishing/Morgan Kaufmann, London/San Mateo (1989)
16. Lanzola, G., Falasconi, S., Stefanelli, M.: Cooperative Agents Implementing Distributed Patient Management. In: Proceedings of the MAAMAW 1996 Conference. Springer, Berlin (1996)
17. Lanzola, G., Falasconi, S., Stefanelli, M.: Cooperative Software Agents for Patient Management. In: Proceedings of the AIME 1995 Conference, pp. 173–184 (1995)
18. Huang, J., Jennings, N.R., Fox, J.: An agent-based approach to health care management. International Journal of Applied Artificial Intelligence 9(4), 401–420 (1995)
19. Kirn, S.: Ubiquitous Healthcare: The OnkoNet Mobile Agents Architecture. In: Aksit, M., Mezini, M., Unland, R. (eds.) NODe 2002. LNCS, vol. 2591, pp. 265–277. Springer, Heidelberg (2003)
20. Shortliffe, E.H.: Computer-Based Medical Consultations: MYCIN. Elsevier, New York (1976)
21. Iantovics, B.L.: The Intelligence of the Expert System Agents. In: Proceedings of the 2nd International Conference on Economics, Law and Management, Tg. Mures, pp. 91–101 (2006)
22. Pfeifer, R., Scheier, C.: Understanding Intelligence. MIT Press, Cambridge (1999)
23. Unland, R., Ulieru, M.: Swarm Intelligence and the Holonic Paradigm: A Promising Symbiosis for Medical Diagnostic Systems Design. In: Proceedings of the 9th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, Melbourne, Australia (2005)
24. Iantovics, B.L.: A Novel Mobile Agent Architecture. In: Proceedings of the 4-th International Conference on Theory and Application of Mathematics and Informatics, Albac county, Acta Universitatis Apulensis, vol. 11, pp. 295–306 (2005)
25. Iantovics, B.L.: A New Intelligent Mobile Multiagent System. In: Proceedings of the IEEE International Workshop on Soft Computing Applications, Szeged-Hungary and Arad-Romania, pp. 153–159. IEEE Press, Los Alamitos (2005)
26. Iantovics, B.L.: Security Issues of the Mobile Multiagent Systems. In: International Conference European Integration between Tradition and Modernity, Tg. Mures, pp. 696–705 (2005)
27. Iantovics, B.L.: Problem Solving Using Assistant Agents, vol. XVIII, pp. 173–179. Scientific Bulletin of the Petru Maior University, Tg. Mures (2006)

28. Iantovics, B.L.: A Novel Diagnosis System Specialized in Difficult Medical Diagnosis Problems Solving. In: Aziz-Alaoui, M.A., Bertelle, C. (eds.) Proceedings of the Emergent Proprieties in Natural and Artificial Dynamical Systems a workshop within European Conference on Complex Systems, pp. 107–112. Le Havre University Press, Paris (2005)
29. Jagannathan, V., Dodhiawala, R., Baum, L.S. (eds.): Blackboard Architectures and Applications. Academic Press, San Diego (1989)
30. Iantovics, B.L.: Medical Diagnosis Systems. In: Proceedings of the Symposium Colocviul Academic Clujean de Informatica, pp. 33–39. Babes-Bolyai University Press, Cluj-Napoca (2005)
31. Iantovics, B.L.: Evolutionary Mobile Agents. In: Proceedings of the International Conference European Integration between Tradition and Modernity, pp. 761–769. Petru Maior University Press, Tg. Mures (2005)
32. Iantovics, B.L.: Evolutionary Problems Solving in Networks. In: Proceedings of the Symposium, Colocviul Academic Clujean de Informatica, pp. 39–45. Babes-Bolyai University Press, Cluj-Napoca (2005)
33. Iantovics, B.L.: Problem Solving Using Evolutionary Mobile Agents. In: Megan, M. (ed.) Proceedings of 9th National Conference of the Romanian Mathematical Society. Vest University Press, Timisoara (2005)
34. Iantovics, B.L.: Mobile Agents. In: Profesor Gheorghe Farcas la varsta de 70 de ani, pp. 171–185. Petru Maior University Press, Tg. Mures (2004)
35. Iantovics, B.L.: Intelligent Mobile Agents. In: Proceedings of the Symposium, Zilele Academice Clujene, pp. 67–74. Babes-Bolyai University Press, Cluj-Napoca (2004)
36. Iantovics, B.L.: A Novel Medical Diagnosis System. In: Proceedings of the 4-th International Conference on Theory and Applications in Mathematics and Informatics, Albac county, Acta Universitatis Apulensis, vol. 10, pp. 315–330 (2005)
37. Kuhl, E.J., Graham, J.H.: ESAgent: Expert System Control of Simulated Agent-Based Mobile Robots. Intelligent Systems Research Laboratory, Technical Report TR-ISRL-04-02, University of Louisville, Louisville (2004)
38. Bravata, D.M., Sundaram, V., McDonald, K.M., Smith, W.M., Szeto, H., Schleinitz, M.D., Owens, D.K.: Evaluating Detection and Diagnostic Decision Support Systems for Bioterrorism Response Emerging Infectious Diseases. Emerging Infectious Diseases 10(1), 100–108 (2004)
39. Adlassing, K.P.: Cardiag 2 Expert System. IEEE Transactions on Systems, Man and Cybernetics SMC-16(2) (1986)
40. Aikins, J.S., Kunz, J.C., Shortliffe, E.H., Fallat, R.J.: PUFF: An Expert System for Interpretation of Pulmonary Function Data. Comput. Biomed. Res. 16(3), 199–208 (1983)
41. Kulikowski, C.A., Weiss, S.M.: Representation of Expert Knowledge for Consultation: the CASNET and EXPERT Projects. In: Szolovits, P. (ed.) Artificial Intelligence in medicine, pp. 21–56. Westview Press, Boulder (1982)
42. Nealon, J.L., Moreno, A.: The Application of Agent Technology to Healthcare. In: Proceedings of the First Agentcities Workshop: Challenges in Implementing Open Environments, Bologna, Italy (2002)
43. Zhang, S., Bodenreider, O.: Comparing Associative Relationships among Equivalent Concepts across Ontologies. Medinfo, pp. 459–463 (2004)
44. Bodenreider, O.: The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res. 32(1), 267–270 (2004)

45. Montyne, F.: The Importance of Formal Ontologies: A Case Study in Occupational Health. In: d'Atri, A., Missikoff, M. (eds.) Proceedings of the OES-SEO 2001 Rome Workshop. Luiss Publications (2001)
46. Stevens, R., Baker, P., Bechhofer, S.G., Jacoby, A., Paton, N.W., Goble, C.A., Brass, A.: TAMBIS: Transparent Access to Multiple Bioinformatics Information Sources. Bioinformatics 16(2), 184–186 (2002)
47. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al.: Gene Ontology: Tool for the Unification of Biology. The Gene Ontology Consortium. Nat. Genet. 25, 25–29 (2002)
48. Iantovics, B.L.: The CMDS Medical Diagnosis System. In: Symbolic and Numeric Algorithms for Scientific Computing. IEEE Computing Society Press, Los Alamitos (2007)
49. Elammari, M., Lalonde, W.: An agent oriented methodology: High level and intermediate models. In: Proceedings of the 1st International Workshop on Agent Oriented Information Systems, Heidelberg, Germany (1999)
50. Wooldrige, M., Jennings, N.R., Kinny, D.: The gaia methodology for agent-oriented analysis and design. In: Autonomous Agents and Multi-Agent Systems, vol. 3, pp. 285–312. Kluwer Academic Publishers, The Netherlands (2000)
51. Burrafato, P., Cossentino, M.: Designing a multi-agent solution for a bookstore with the passi methodology. In: Fourth International Bi-Conference Workshop on Agent-Oriented Information Systems, Toronto, Canada (2002)
52. Ulieru, A., Grabelkovsky, M.: Telehealth approach for glaucoma progression monitoring. International Journal: Information Theories and Applications 10, 326–329 (2005)
53. Iantovics, B.L.: The ICMA Agents Intelligence. In: International Conference European Integration between Tradition and Modernity. Petru Maior University Press, Tg. Mures (2007)
54. Yergens, D., Hiner, J., Denzinger, J., Noseworthy, T.: Multiagent simulation system for rapidly developing infectious disease models in developing countries. In: Proceedings of the 2nd International Workshop on Multi-Agent Systems for Medicine and Computational Biology, Hakodate, Japan, pp. 104–116 (2006)
55. Eder, L.: Managing Healthcare Information Systems with Web-Enabled Technologies. Idea group Publishing (2000)
56. Cabri, G., De Mola, F., Muratori, N., Quitadamo, R., Zambonelli, F.: Ubimedic: a ubiquitous computing solution for medical emergencies. In: Proceedings of the 2nd International Workshop on Multi-Agent Systems for Medicine and Computational Biology, Hakodate, Japan, pp. 42–46 (2006)
57. Nagarajappa, N.: Uploading Data from the Local to the Centralized Database for Glaucoma Diagnosis and Prediction - Internal Report, Electrical and Computer Engineering, University of Calgary (2002)
58. Ulieru, M., Geras, A.: Emergent Holarchies for e-Health Applications: A Case in Glaucoma Diagnosis. In: Proceedings of IECON 2002 - 28th Annual Conference of the IEEE Industrial Electronics Society, Sevilla, Spain (2002)

59. Ulieru, M., Pogrzeba, G.: Integrated Soft Computing Methodology for Diagnosis and Prediction with Application to Glaucoma Risk Evaluation. In: Proceedings of 6th IASTED International Conference on Artificial Intelligence and Soft Computing, Banff, Canada, July 17-19, 2002. AISC (2002)
60. Kanski, J.J., McAllister, J.A.: Glaucoma: A Coulour Manual of Diagnosis and Treatment, Butterworths, London, Boston, Singapore, Sydney, Toronto (1989)
61. Ulieru, M., Andrew, C.S., Rizzi, M., Karanicolas, C.: Using Soft Computing to Define Standards of Care in Glaucoma Monitoring. International Journal of Soft Computing: A Fusion of Foundations, Methodologies and Applications (2003)

AgentTime: A Distributed Multi-agent Software System for University's Timetabling

Eduard Babkin[1], Habib Abdulrab[2], and Tatiana Babkina[3]

[1] LITIS Laboratory, INSA de Rouen, France. State University - Higher School of Economics. Nizhny Novgorod, Russia. babkin@hse.nnov.ru
[2] LITIS Laboratory, INSA de Rouen, France. abdulrab@insa-rouen.fr
[3] State University - Higher School of Economics. Nizhny Novgorod, Russia. bts@sandy.ru

Summary. In the course of researching timetabling problems for complex distributed systems this article applies the multi-agent paradigm of computations and presents a correspondent mathematical model for university's timetabling problem solution. The model takes into account dynamic nature of this problem and individual preferences of different remote users for time and location of classes. In the framework of that model authors propose an original problem-oriented algorithm of multi-agent communication. Developed algorithm is used as a foundation for the distributed software system AgentTime. Based on multi-agent JADE platform AgentTime provides friendly graphical interface for online design of time tables for universities.

Keywords: timetabling, multi-agent algorithms, distributed systems

1 Introduction

In many modern branches of computer science, operations research and enterprise management, which deal with complex machine-human interactions and decision support we can recognize similar transitions from the paradigm of complete rationality and stable properties of systems to the new paradigm of the bounded rationality and emergent system properties [1]. Such properties are not predetermined by developers or researchers, but become visible during self-organization processes inside the system. Studies of natural self-organizing systems and development of artificial self-organizing systems exploit the principal idea of synergy in complex systems, originally proposed by H. Haken [2]. Synergy means that the complex system comprised of cooperative entities has qualitative preponderance in comparison with individual entities. We believe that among various applied tasks scheduling represents an interesting problem which should be considered within the synergic framework of self-organization of complex systems.

In modern society time scheduling plays the outstanding role because any schedule defines the well-formed obligation, which enables to carry out authentic planning of

activities for a separate person, and whole industrial systems as well. Timetabling represents an important research activity in the scheduling theory, and focuses on such problems as optimal lecture's schedules in educational institutions, week cyclic schedules of plane's flights within the framework of several airports, week or daily schedules of railway transportation, etc. For all these problems the interval of time, inside which the given set of jobs should be fulfilled, is known beforehand. Thus, the minimum of the schedule's length is not usually considered as a primary goal, – other 'system-scope' criteria are used for estimation of quality of the schedule having been built. For example, in an educational institution the timetable design process should achieve the following goals: minimization of maximal length of a working day, minimization of the number of the 'holes' in the schedules of groups and professors, maximal satisfaction of personal professor's preferences to the time and location of classes, etc. In the current situation, when many educational institutions rapidly grow in size, and distribution scale, wide application of effective software systems for distributed solution of timetabling problems becomes very important.

At present there are many various algorithms for drawing up the time tables in universities. The fundamental approaches are based on the well-known linear and integer programming paradigms [4], [5], [6], [7], [8]. However, several researches of 80s have shown that the integer programming is not equally effective from the point of view of the calculations volume. The high computational costs make integer programming poorly attractive to the large tasks of timetable design, because that method does not guarantee productivity, when the sizes and complexity of tasks grow [5], [9].

Last twenty years have shown the increased interest of the researchers to development of the approaches for the design of timetables with use of various metaheuristics [5], like simulated annealing, Tabu Search, genetic algorithms (GAs), and their hybrids [10], [11], [12], [13], [14], [15], [16], [21]. It is affirmed, that among others, GAs have larger capacity, and allow to find the greatest number of the feasible solutions [17], [18], [19]. Nevertheless, when GAs are exploited, there are difficulties in the description of controlling parameters, in definition of exact roles of crossover and mutations, as well as in analysis of convergence [20].

The majority of the considered approaches follow the paradigm of centralized predetermined systems, such approaches do not allow the authonomous users govern the process of timetable design. In complex distributed and evolving systems like modern virtual universities and peer-to-peer communities, that shortcoming makes impractical classical methods, and demands new timetabling principles. These principles should take account of real-time user's preferences, assume only partial and fuzzy knowledge of individual participants, and should adapt well to complex changing environments. A multi-agent approach represents a successful paradigm for those kinds of problems, when an optimal or quasi-optimal solution is built in the result of interaction of large number of autonomous computational entities with local partial knowledge and individual preferences. Cooperative design of the timetable in the multi-agent system represents an exemplification of the synergy in complex systems.

In general, for timetabling applications several types of multi-agent algorithms are suitable. The first type of algorithms includes economics-based models of interaction [3], [22], [23]. The second one consists of various generic algorithms for solution of Distributed Constraint Satisfaction Problems (DCSP) [24], [25]. But the most effective al-gorithms, comprising the third type, were specifically designed for a

particular sched-uling problem. Such specialized algorithms apply all domain- or problem-specific information and show unbeaten productivity. The results are known [26], [27], [28], [29], where authors propose problem-specific algorithms of agent's interaction for the meeting scheduling. Although such algorithms fit well the timetabling model, and have attractive computational efficiency, their direct application for university's time-tabling is not so straightforward and requires additional efforts.
In the given article authors propose new multi-agent algorithm for university's timetabling, and describe basic principles of the distributed system AgentTime, based on that algorithm. Presentation of the results has the following structure. Section 2 describes the mathematical model of university's timetabling, which includes user's preferences. Section 3 gives overview of the corresponding multi-agent algorithm for the design of timetables. In section 4 certain topics of AgentTime's software imple-mentation are considered. Overview of results and discussion are presented in Section 5.

2 The Proposed Mathematical Model For the University's Timetabling Problem

The exact mathematical statement of the university's timetabling problem forms the basis of our own multi-agent algorithm for design of the educational schedule. For the sake of generality we use the term 'teacher' to denote different kinds of university employees (e.g. professors, instructors, etc), the term 'stream' to denote a stream, and the term 'subject' to denote different kinds of student's subjects. Also we give the same name of user to all of the stakeholders of the schedule (e.g. the teachers and student's groups). In our mathematical model we will use the following designations.

Student's groups and Streams: $\mathbf{G}$ - the set of group's identifiers. $\mid \mathbf{G} \mid = \gamma$ - the total number of groups. $g \in \mathbf{G}$ - the unique identifier of the group. Each group belongs to one stream at least. Some streams can consist of a single group, but in most cases several groups form a stream with the following constraints:

1. All groups of the same stream exploit the same classrooms for lectures.
2. Lectures are delivered to all groups of the stream at the same time.
3. Each stream has as minimum one lesson.

$\mathbf{R}$ - the set of stream's identifiers. $\mid \mathbf{R} \mid = \rho$ - the total number of streams.$r \in \mathbf{R}$ - the unique identifier of the stream. Each single group can be treated as a separate stream, thus $\rho \geq \gamma$. $\mathbf{R}_r \subset \mathbf{G}$ - the stream. $\mathbf{C} = \{\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_\rho\}$ - the set of the streams.

Teachers: $\mathbf{P}$ - the set of unique teacher's identifiers. $p \in \mathbf{P}$ - the unique teacher's identifier.

Timetable users: Union of the group's set and the teacher's set gives us the set of the timetable users: $\mathbf{M} = \mathbf{G} \bigcup \mathbf{P}$, $m \in \mathbf{M}$ – the unique identifier of the timetable user.

Time: $\mathbf{W}$ - the set of the days of the week.$w \in \mathbf{W}_g$ - the certain day of the week. $\mathbf{W}_g \subset \mathbf{W} = \{1, 2, \ldots, 7\}$ - the set of learning days for the group g, $j \in \mathbf{J} = \{1, 2, \ldots, 8\}$ - the lesson's number.
$\mathbf{T} = \{(w, j) \mid w \in \mathbf{W}, j \in \mathbf{J}\}$ - the set of timeslots, which are the elementary units in the timetabling problem. For example, the timeslot $(1, 2)$ means the second lesson on Monday. For each timetable user m the set of free timeslots

$\mathbf{T}_m^+ \subset \mathbf{T}$ is known. The set of denied timeslots $\mathbf{T}_m^- \subset \mathbf{T}$ is known also. We assume the obvious constraints are true (i.e. $\mathbf{T}_m^+ \bigcup \mathbf{T}_m^- = \mathbf{T}$ and $\mathbf{T}_m^+ \bigcap \mathbf{T}_m^- = \oslash$).

Subjects: In our model teachers conduct lectures and manage practical exercises. Lectures are delivered to the whole stream, while practical exercises are organized for a single group only. Also some practical exercises impose restrictions on allowable classrooms, like computer or chemistry labs. To describe all these pecularities, let's introduce the following mathematical structures: $\mathbf{S}_r = \{1, 2, \ldots, \sigma_r\}$ - the set of lecture's identifiers delivered to the stream; $s_r \in \mathbf{S}_r$ - the unique lecture's identifier; $\mathbf{Q}_r = \{1, 2, \ldots, \Theta_r\}$ - the set of practical exercise's identifiers organized for the stream r; $q_r \in \mathbf{Q}_r$ - the unique exercise's identifier. Each lecture's assignment can be uniquely identified by a pair $(r, s_r) \in \mathbf{RS}$, and

$$\mathbf{RS} = \{(r, s_r) \mid r \in R, s_r \in S\} \tag{1}$$

The total number of lecture's assignments is computed as follows:

$$\mid \mathbf{RS} \mid = \sum_{r=1}^{\rho} \sigma_r \tag{2}$$

Each exercise's assignment is uniquely identified by a triple $(r, g_r, q_r) \in \mathbf{RGQ}$, and

$$\mathbf{RGQ} = \{(r, g_r, q) \mid r \in \mathbf{R}, g_r \in \mathbf{C}_r, q \in \mathbf{Q}_r\} \tag{3}$$

The total number of exercise's assignments is computed as follows:

$$\mid \mathbf{RGQ} \mid = \sum_{r=1}^{\rho} \mid \mathbf{C}_r \mid \times \Theta_r \tag{4}$$

where $\mid \mathbf{C}_r \mid$ - is the total number of groups in the stream $\mathbf{C}_r$.

For further analysis differences between lectures and practical exercises can be neglected and the united set of subjects $\mathbf{E}$ will be used:

$$\mathbf{E} = \mathbf{RF} \bigcap \mathbf{RGQ} \tag{5}$$

Curriculum consists of subjects' assignments for each of the teacher during one semester (fall) in the following form:

$$\begin{aligned} &\delta : \mathbf{E} \to \mathbf{P} \\ &\delta(e) = \begin{cases} \delta_1(e), & e \in \mathbf{RS} \\ \delta_2(e), & e \in \mathbf{RGQ} \end{cases} \\ &\delta_1 : \mathbf{RS} \to \mathbf{P} \\ &\delta_2 : \mathbf{RGQ} \to \mathbf{P} \end{aligned} \tag{6}$$

where $\mathbf{P}$ – the set of teachers; $\mathbf{RS}$ – the set of lecture's assignments; $\mathbf{RGQ}$ – the set of exercise's assignments.

For example, $\delta_1(1, 2) = 4$ means, that teacher 4 delivers lecture 2 for stream 1, and $\delta_2(1, 2, 4) = 7$ means, that teacher 7 manages practical exercise 2 for group 4, included into the stream 1. Given the curriculum δ, we can easily compute the total number of subjects $\mathbf{E}_m$ assigned to the teacher (or the group) with the identifier m:

$$\begin{aligned} \mathbf{E}_m = \{e \mid m \in \mathbf{P} \wedge \sigma(e) = m\} \bigcup \{e = (r, s) \mid r \in \mathbf{R} \wedge s \in \mathbf{S}_r\} \bigcup \\ \bigcup \{e = (r, q, m) \mid m \in \mathbf{C}_r \wedge q \in \mathbf{Q}_r\} \end{aligned} \tag{7}$$

Room's stock: this resource consists of laboratories, lecture halls and classrooms available for the subjects in the university. It is modeled by the set $\mathbf{A}$ of unique room's identifiers. For each element of the set of subjects $\mathbf{E}$, a subset of permitted rooms $\mathbf{A}_e$ is selected: $\mathbf{A}_e \in \mathbf{A}$.

The primary goal of the timetabling problem in our model is formulated as looking for the feasible mapping from the set of subjects $\mathbf{E}$ to the set of timeslots $\mathbf{T}$:

$$\tau : \mathbf{E} \to \mathbf{T} \tag{8}$$

For example, mapping $\tau(1,2) \to (4,4)$ means that subject 2 for stream 1 will be given on Thursday during the fourth lesson. Related with the mapping τ, the mapping α should assign a classroom for each subject:

$$\alpha : \mathbf{E} \to \mathbf{A} \tag{9}$$

where $\mathbf{E}$ – the set of subjects; $\mathbf{A}$–the set of classrooms.
For example, mapping $\alpha(1,2) \to 101$ means that subject 2 for stream 1 will be conducted in the room 101.
Constraints for the university's timetabling problem are defined as follows.

1. The teacher can conduct only one subject at the single timeslot.

$$\forall p \in \mathbf{P}, \forall e_1, e_2 \in \mathbf{E} : (e_1 \neq e_2 \wedge \delta(e_1) = \delta(e_2) = p) \Rightarrow \tau(e_1) \neq \tau(e_2) \tag{10}$$

2. In one classroom only one subject can be given at the single timeslot.

$$\forall a \in \mathbf{A}, \forall e_1, e_2 \in \mathbf{E} : e_1 \neq e_2 \wedge \alpha(e_1) = \alpha(e_2) = a \Rightarrow \tau(e_1) \neq \tau(e_2) \tag{11}$$

3. Each group has no more than one subject at the single timeslot.

$$\begin{array}{l} \forall g \in \mathbf{G} : \\ (e_1 = (\gamma_1, \dots), e_2 = (\gamma_2, \dots) \in \mathbf{E} \wedge g \in \mathbf{C}_{r_1} \wedge g \in \mathbf{C}_{r_2} \wedge e_1 \neq e_2) \bigvee \\ \bigvee (e_1 = (\dots, g), e_2 = (\dots, g) \in \mathbf{E} \wedge e_1 \neq e_2) \Rightarrow \tau(e_1) \neq \tau(e_2) \end{array} \tag{12}$$

Subject's priority reflects the obvious fact that not all subjects have identical importance within the framework of educational process. As such, it is necessary to set the priority order among different subjects, so subjects with higher priority will borrow the best time and location. In our model the priority is modeled as the partial order on the set of subjects $\mathbf{E}$:

$$e_1 \succ e_2 \Leftrightarrow \mathbf{U}(e_1) \geq \mathbf{U}(e_2) \tag{13}$$

where $e_1, e_2 \in \mathbf{E}$;
$\mathbf{U}(e) = \mid \mathbf{M}_e \mid + k_1(e) + k_2(e) + k_3(p_e)$ – the 'utility' of the subject e;
$\mathbf{M}_e = \{m \mid (e = (r,s) \in \mathbf{RS} \wedge m \in \mathbf{C}_r) \vee (e = (r,q,m) \in \mathbf{RGQ}\}$ – the total number of groups for those the subject e is given;
$k_1(e) \in \{0, 5, 10\}$ – the measure of the subject's importance for the stream (0–optional, 5–important in general, 10–important for the stream);
$k_2(e) \in \{0, 2\}$ – the level of the subject (0–undergraduate, 2–graduate);
$p_e = p, \sigma(e) = p$ – the teacher's identifier;
$k_3(p_e) \in \{0 \dots 5\}$ – the estimation of the novelity level of the material given by the teacher p_e.
User's Preferences comprise the important part of our model. Each preference is modeled by a numeric value from the range $[0, 1]$. Value 0 corresponds to the least desired alternative, and value 1 corresponds to the most desired alternative. The model includes two kinds of the user's preferences:

1. preferences of the user $m \in \mathbf{M}$ for the best time of the subjects:
$$f_m^1 : \mathbf{E}_m \times \mathbf{T}_m^+ \to [0,1] \tag{14}$$
2. preferences of the user $m \in \mathbf{M}$ for the best location of the subjects:
$$f_m^2 : \mathbf{EA}_m \to [0,1] \tag{15}$$
where $\mathbf{EA}_m = \{(e,a) \mid a \in \mathbf{A}_e \wedge e \in \mathbf{E}_m\}$ – the set of feasible pairs 'subject–classroom'.

We use an evident representation of the user's preferences in the form of graphics tables (Figs. 1, 2). The darker color denotes the more preferable alternative (in respect of time or location).

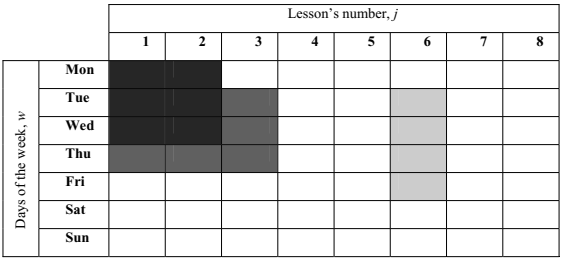


Fig. 1. Preferences of the user m for the desirable time of subjects, f_m^1.

Classroom, a	1	2	3	4	5
f_m^2					

Fig. 2. Preferences of the user m for the desirable location of subjects, f_m^2.

Criterion of timetable quality generalizes several partial criteria, and evaluates the solution found, namely the pair of mappings $\tau(e), \alpha(e)$. The first partial criterion evaluates the sum of the user's preferences for the time of the subject e:
$$F(\tau)_e^1 = \sum_{m \in \mathbf{M}} f_m^1(e, \tau(e)) \to max \tag{16}$$
The second partial criterion evaluates the sum of the user's preferences for the location of the subject e:
$$F(\alpha)_e^2 = \sum_{m \in \mathbf{M}} f_m^2(e, \alpha(e)) \to max \tag{17}$$
The generalized criterion is constructed as follows:
$$F(\tau, \alpha) = \sum_{e \in \mathbf{E}} (F(_e^1) + F(_e^2)) \to max \tag{18}$$
The solution of the described problem consists of the found mappings τ, α, assuming that all constraints are satisfied, and the generalized criterion has a maximum value.

3 The Multi-Agent Algorithm For Timetable Design

We took for the basis of our algorithm the well-known multi-agent algorithm MSRAC for meetings scheduling by A. Ben Hassine et all. [29]. Although some correspondences still remain, our algorithm is specifically designed for a quite different problem of university's timetable design, and together with the time schedule it gives also the occupancy schedule for the classrooms.
In our algorithm we recognize two roles of agents: agents-organizers and agents-participants. The agent's structure also mimics the application domain, so we classify all agents as teachers, groups and classrooms. Agents-teachers play the role of the organizer; agents-groups and agents-rooms play the role of the participant. The numbers of agents-teachers and agents-groups correspond to the real numbers of the teachers and the groups in the university. One agent-room corresponds to all classrooms in the context of the single time table. Collective search for the best time and location of the study involves communication between different agents. For each study the agent-teacher performs a set of actions, comprising the following state diagram (Fig. 3).

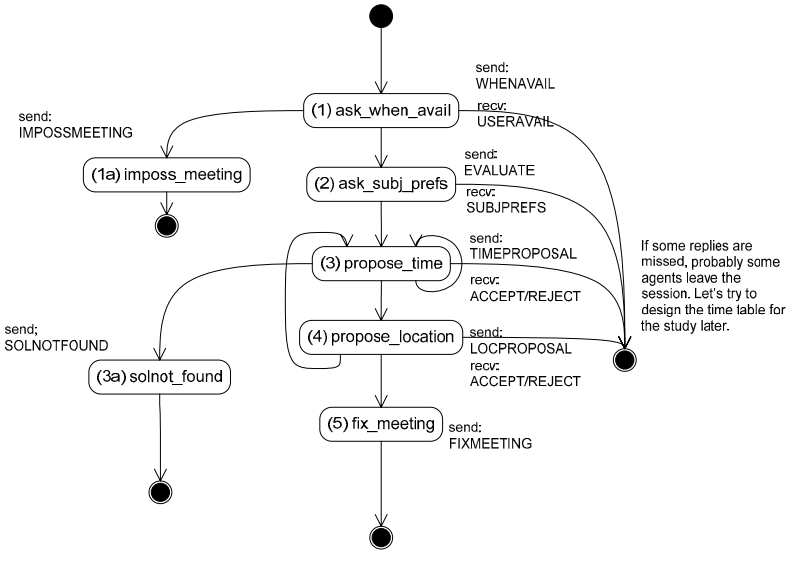


Fig. 3. The state diagram of the agent-teacher's algorithm.

The agent-teacher performs state transitions in accordance with the description given below:

1. *Ask_when_avail*. That is the first state in the algorithm. The agent-teacher sends to all agents of groups the query WHENAVAIL with the subject's identifier, requesting available time for that study. The agents of groups answer by the message USERAVAIL, in which they inform when the agent is free, and has available time for the study. If all agents have informed the answer, then the

agent-teacher finds intersection on time. If the intersection is empty, then the agent comes the final state imposs_meeting.
   a) *Imposs_meeting*. In that state the agent-teacher founds itself if intersection of available for other agents times is empty and the total solution was failed. The study is marked as 'having no solution'.
2. *Ask_subj_prefs*. In that state the agent-teacher requests preferences for time and location (the message EVALUATE). The agents of the groups reply own preferences in the message SUBJPREFS. The agent-teacher sorts received preferences for time and for location in accordance with criteria (16) and (17).
3. *Propose_time*. The agent-teacher selects the first timeslot from the ordered list of the preferences, and sends it along with the subject's identifier to the agents of groups inside the message TIMEPROPOSAL. In response the agents of groups analyze own agendas. If the proposed timeslot is free in the agent's agenda, the agent gives the positive answer, sending the message ACCEPT. Else the agent compares the priority of the subject in the agenda with the priority of the subject in the message. If the priority of the message's subject is greater, then the agent accepts new proposal and sends the message ACCEPT. In the opposite case the agent sends the message REJECT. The agent can apply the metropolis criterion [29] for decision making when the priorities are equal. In the case of total acceptance of the proposed timeslot, the agent-teacher passes to the next state Propose_location; in default the agent remains in the state propose_time, and chooses the next timeslot to negotiate. If all timeslots were rejected, it means that the decision for the currently selected subject does not exist, and the agent-teacher passes to the state (3–àa) Solnot_found.
   a) *Solnot_found*. In that state the agent-teacher founds itself if all proposed for timeslots were rejected by the agents of groups, and the total solution was failed. The study is marked as 'having no solution'.
4. *Propose_location*. The agent-teacher sends the sorted list of classrooms to the agent of classrooms inside the message LOCPROPOSAL. That message contains also the subject's identifier and the timeslot's identifier. Using own occupancy list, the agent of classrooms searches for the first classroom in the list, which is available for the timeslot given or occupied by the subject with a lower priority. If the search is successful, and the classroom is found, in reply to the agent-teacher the agent of class-rooms sends the message ACCEPT with the identifier of the room found. In the failure case the agent of classrooms sends the reply REJECT. Once the positive ACCEPT reply is received, the agent-teacher moves to the next state. In the result of REJECT receiving the agent returns to the state propose_time for selecting the available timeslot for the study.
5. *Fix_meeting*. If the agent-teacher occurs in that state, it means that both the timeslot and the classroom for the subject were successfully found. In the result the agent-teacher sends to all other agents the message FIXMEETING with the identifiers of the subject, the timeslot, and the classroom. If the agent-group does not have assignment for the received timeslot, the timeslot is fixed. By a similar way the agent of classrooms fixes the location. If the received timeslot (or the classroom) is occupied, the agent discards assignment of the subject with lower priority, and sends to the agent-teacher the message CANCEL MEETING, which is forwarded further to other agents in order to modify their agendas.

In the states (1), (2), (3) and (4), if some agents did not send the answer during a predefined time period, the agent-teacher places the subject being under

consideration into the list of the cancelled subjects, to retry attempts later. Once all agents-teachers finish state processing, the common schedule is considered to be complete. One important feature of our algorithm is that of the partial timetable is always available. The complete timetable, including all the subjects, sometimes simply does not exist. In such a case, however, the considered algorithm will build the consistent time table, with some subjects of low priority ignored.

4 Implementation Details of the Software System AgentTime

The described mathematical model and the multi-agent algorithm were applied during design and implementation of the software system for time tabling called AgentTime. AgentTime uses rich communication and agent-life cycle capabilities of Java-based JADE multi-agent platform [31], and has highly distributed software architecture (Fig. 4).

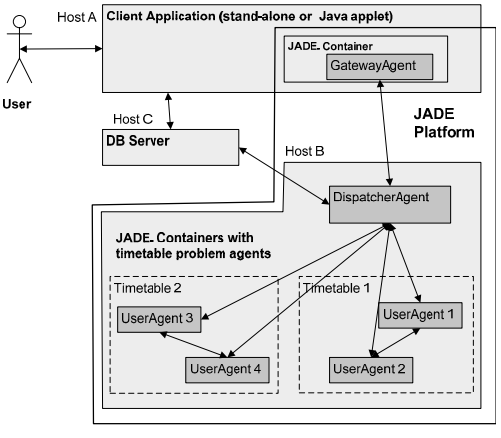


Fig. 4. The software architecture of AgentTime timetabling system.

Flexible multi-tier architecture of the system supports simultaneously multiple timetable design sessions and interaction of multiple agents. In AgentTime agents indirectly communicate with each other by passing the messages in accordance with a problem-specific ontology (Table 1).
Interaction of the agents during the design of timetable can be illustrated by the UML sequence diagram in Fig. 5. In AgentTime apart from previously mentioned types of the agents we use the dedicated ServerAgent which is responsible for communication with external data sources, logging and other technical tasks.
Different users of AgentTime can interact with the system using different end-user tools, including web-browsers and PDAs. The mostly used way of interaction assumes application of applet-based graphical interfaces (Fig. 6), but also JSP-based interfaces are available.

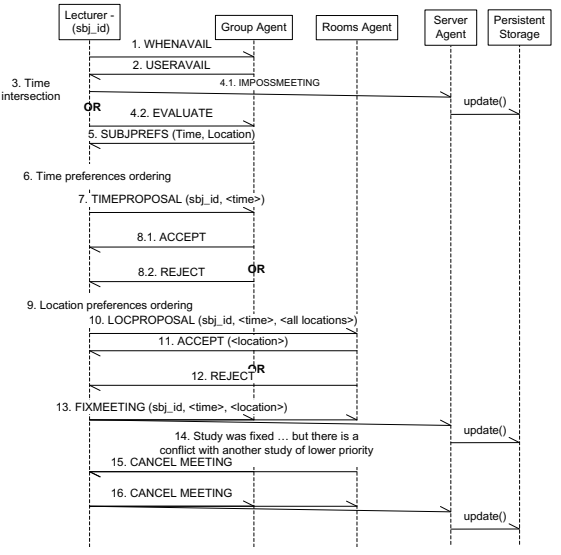


Fig. 5. The UML sequence diagram of the algorithm.

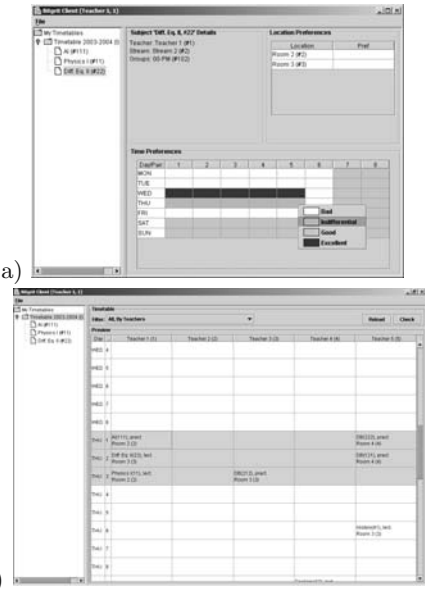


Fig. 6. Examples of AgentTime's graphics interfaces: a – assignment of preferences; b - representation of the ready timetable.

Message	Semantics
WHENAVAIL	Inquiry to the agent-group for available timeslots
USERAVAIL	Agent's response to the message WHENAVAIL. The message contains the vector with available timeslots. Format: $(a_{11}\ \ a_{12}\dots a_{18}\ \ a_{21}\ \ a_{22}\dots a_{28}\dots a_{71}\ \ a_{72}\dots a_{78})$, where $a_{ij} \in \{0,1\}$ (0–a weekend, 1– a working day).
IMPOSSMEETING	The message to the server agent about impossibility to find the time table for the subject with id sbj_id. Format: (sbj_id).
EVALUATE	Inquiry to the agent of group for the time and location preferences related with the subject sbj_id. Format: (sbj_id).
SUBJPREFS	Agent's response to the EVALUATE. Format: (sbj_id $(w_{11}w_{12}\dots w_{18}w_{21}w_{22}\dots w_{28}\dots w_{71}w_{72}\dots w_{78})$ $((L_1\ \ p_1)(L_2\ \ p_2)\dots(L_n\ \ p_n))))$, where $0 \le w_{ij} \le 1(i = \overline{1,7}, j = \overline{1,8})$ – evaluation of i-th day of week and j-th lesson; $0 \le L_k \le 1(k = \overline{1,n})$ – the number of classroom; $0 \le p_k \le 1, 0 \le k \le n$– the preference of the classroom L_k.
TIMEPROPOSAL	The agent-teacher proposes the time for the subject sbj_id. Format: (sbj_id $(d\ p)$), where d - the day of the week; p - the number of the lesson.
LOCPROPOSAL	The agent-teacher proposes the location for the subject sbj_id. Format: (sbj_id $(L_1\ L_2\dots L_k)$), where $L_k(k = \overline{1,m})$ – the identifier of the classroom. The classrooms are sorted in accordance with the preferences.
ACCEPT/REJECT	Agent's response to the message TIMEPROPOSAL (ACCEPT or REJECT). If the proposal is accepted the message contains the classroom's identifier. Format: (L), where L - is the id of classroom.
FIXMEETING	Inquiry to fix the timeslot and the location for the subject sbj_id. Format: (sbj_id $(d\ p)\ L$), where d - the day of the week; p - the id of the lesson; L - the id of the classroom.
CANCELMEETING	Notification about cancelling a conflicting subject. Format: (sbj_id $(d\ p)\ L$), where d - the day of the week; p - the id of the lesson; L - the id of the classroom.

Table 1. Multi-agent ontology for timetable design.

5 Discussion

This article focused on the important problem of time tables' design for educational institutions in presence of multiple autonomous stakeholders. To tackle this problem in the context of modern distributed and highly dynamic universities we proposed the original mathematical model and correspondent multi-agent algorithm for iterative timetabling in presence of different subjective preferences for time and location of subjects. The theoretical considerations become a foundation for development of the multi-agent software system AgentTime. That system facilitates distributed time planning allocatiing timeslots and classrooms during communication

of authonomous agents. Such approach directly corresponds to modern trends in studies and development of complex self-orginized systems.
The developed algorithm belongs to the class of domain-specific multi-agent algorithms and exposes good performance metrics. Analysis shows that in the case of single computational node computational complexity of the algorithm C for allocation of timeslots and classrooms can be estimated as follows:

$$C \leq S \times T_0(n_g, n_l) \times \log_2 S.$$

In this equation S - is the number of subjects, $T_0(n_g, n_l)$ – a constant determined by the problem's conditions.
If AgentTime is distributed among $P \leq C$ computational nodes, then estimation of processing time t_p will be $t_p \leq \frac{C}{P} = \frac{S}{P} \times T_0 \times \log_2 S$. In the extreme case, when $P = S$, t_p will not be greater then $T_0 \times \log_2 S$.
Comparing our results with other known approaches to multi-agent algorithms for timetabling like the algorithm MSRAC [29] , we can note that our system is capable of solving a more general problem, allocating not only timeslots, but classrooms also. With a few modifications proposed model and algorithm will be suitable for managing other important resources as well. At the same time we need to improve theoretical background of our algorithm to rigorously prove the optimality of the solutions found in terms of the criteria (17) and (18).
In the nearest time we are going to perform wide-area field experiments with AgentTime to test its robustness and quality of timetabling in real conditions of the complex university. We are also interested in extending the proposed mathematical model and software implementation of AgentTime with other approaches to multi-agent coordination. In this context application of the paradigm 'Controller-Variable Agent' [30] is seemed to be very promising.
This work was partially supported by Russian Fund of Basic Researches (grant # 07-07-00058).

References

1. Simon, H.A.: Administrative Behavior. Free Press, New York (1976)
2. Haken, H.: Synergetics, an Introduction: Nonequilibrium Phase Transitions and Self-Organization in Physics, Chemistry, and Biology. Springer, New York (1983)
3. Cheng, J.Q., Wellman, M.P.: The WALRAS Algorithm: A Convergent Distributed Implementation of General Equilibrium Outcomes. Journal of Computational Economics 12, 1–23 (1998)
4. Sandhu, K.S.: Automating Class Schedule Generation in the Context of a University Time-tabling Information System. PhD Thesis (2001)
5. Petrovic, S., Burke, E.: University Timetabling. University of Nottingham (2003)
6. Akkoyunly, E.A.: A Linear Algorithm for Computing the Optimum University Timetable. The Computer Journal 16(4), 347–350 (1973)
7. Shaerf, A.: Local Search Techniques for Large High School Timetabling Problems. IEEE Transactions on Systems, Man And Cybernetics 29(4), 368–377 (1999)
8. De Werra, D.: An Introduction to Timetabling. European Journal of Operational Research 19, 151–162 (1985)

9. Taha, T.R.: Numerical schemes for nonlinear evolution equations. The College Journal of Science and Technology 2, 105–116 (1987)
10. Dimopoulou, M., Miliotis, P.: Implementation of a university course and examination time-tabling system. European Journal of Operational Research 130(1), 202–213 (2001)
11. Burke, E.K., Ross, P. (eds.) PATAT 1995. LNCS, vol. 1153. Springer, Heidelberg (1996)
12. Burke, E., Erben, W. (eds.) PATAT 2000. LNCS, vol. 2079. Springer, Heidelberg (2001)
13. Burke, E.K., De Causmaecker, P. (eds.) PATAT 2002. LNCS, vol. 2740. Springer, Heidelberg (2003)
14. Back, T.: Evolutionary Algorithms in theory and practice. Oxford University Press, New-York (1995)
15. Sharma, D., Chandra, N.: An evolutionary approach to constraint-based timetabling. In: Kowalczyk, R., Loke, S.W., Reed, N.E., Graham, G. (eds.) PRICAI-WS 2000. LNCS, vol. 2112, pp. 80–92. Springer, Heidelberg (2001)
16. Sanchez, E., Shibata, T., Zadeh, L.: Genetic Algorithms and Fuzzy Logic Systems. In: Soft Computing Perspectives. World Scientific, Singapore (1997)
17. Buckles, B.P., Petry, F.E.: Genetic Algorithms. IEEE Computer Society Press, Los Alamitos (1992)
18. Goldberg, D.E.: Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley, Reading (1989)
19. Maxfield, C.: Genetic algorithms: programs that boggle the mind. EDN (1997)
20. Srinivas, M., Patnik, L.M.: Genetic Algorithms: A Survey. IEEE Computer 27(6), 17–26 (1994)
21. Abramson, D.: Constructing School Timetables Using Simulated Annealing: Sequential and Parallel Algorithms. Management-Science (1991)
22. Wellman, M.P., Walsh, W.E., Wurman, P.R., MacKie-Mason, J.K.: Auction Protocols for De-centralized Scheduling. Games and Economic Behavior 35, 271–303 (2001)
23. Walsh, W.E., Yokoo, M., Hirayama, K., Wellman, M.P.: On Market-Inspired Approaches to Propositional Satisfiability. In: Proceedings of Int. Conf. IJCAI 2000, pp. 1152–1160 (2001)
24. Yokoo, M., Hirayama, K.: Distributed Constraint Satisfaction Algorithm for Complex Local Problems. In: Proceedings of the Int. Conf. ICMAS 1998, pp. 372–379 (1998)
25. Yokoo, M., Hirayama, K.: Algorithms for Distributed Constraints Satisfaction: A Review. In: Proceedings of Int. Conf. AAMAS 2002, vol. 3(2), pp. 185–207 (2000)
26. Garrido, L., Sycara, K.: Multi-Agent Meeting Scheduling: Preliminary Experimental Results. In: Proceedings of the Int. Conf. ICMAS 1996 (1996)
27. Franzin, M.S., Freuder, E.C., Rossi, F., Wallace, R.: Multiagent Meeting Scheduling with Preferences: Efficiency, Privacy Loss, and Solution Quality. In: Proceedings of Intrl. Conf. AAAI 2002 (workshop on preference in AI and CP) (2002)
28. BenHassine, A., Ito, T., Ho, T.B.: A New Distributed Approach to Solve Meeting Scheduling Problems. In: Proceedings of IEEE/WIC Int. Conf. IAT, pp. 588–591 (2003)
29. BenHassine, A., Defago, X., Ho, T.B.: Agent-Based Approach to Dynamic Meeting Scheduling Problems. In: Proceedings of Int. Conf. AAMAS 2004, vol. 3, pp. 1132–1139 (2004)

30. Al–Maqtari, S., Abdulrab, H., Nosary, A.: Constraint Programming and Multi-Agent System mixing approach for agricultural Decision Support System. In: Proceedings of In. Conf. ECCS 2005, pp. 199–213 (2005)
31. Bellifemine, F.L., Caire, G., Greenwood, D.: Developing Multi-Agent Systems with JADE. Wiley, Chichester (2007)

Emotion: appraisal-coping model for the "Cascades" problem

Karim Mahboub[1], Evelyne Clément[2], Cyrille Bertelle[1], and Véronique Jay[1]

[1] LITIS (EA 4108) - University of Le Havre
25 rue Philippe Lebon - BP 540
76058 Le Havre Cedex, France
karim.mahboub@litislab.eu
cyrille.bertelle@litislab.eu
veronique.jay@litislab.eu
[2] Psy.NCA (EA 4306) - University of Rouen
Rue Lavoisier
76821 Mont-Saint-Aignan, France
evelyne.clement@univ-rouen.fr

Summary. Modelling emotion has become a challenge nowadays. Therefore, several models have been produced in order to express human emotional activity. However, only a few of them are currently able to express the close relationship existing between emotion and cognition. An appraisal-coping model is presented here, with the aim to simulate the emotional impact caused by the evaluation of a particular situation (appraisal), along with the consequent cognitive reaction intended to face the situation (coping). This model is applied to the "Cascades" problem, a small arithmetical exercice designed for ten-year-old pupils. The goal is to create a model corresponding to a child's behaviour when solving the problem using his own strategies.

Keywords: emotion modelling, decision making, appraisal-coping model

1 Introduction

As the study of emotion is becoming crucial today, in several fields of study such as neurology or psychology, computer science is getting more and more involved in the process of finding new models for representing emotions. Since the middle of the 19th century, psychologists, biologists, but also neurologists have tried to produce models designed to unravel the emotional processes. Scientists like Bechara and Damasio [3] have even proved that human emotional activity has an indispensable impact on decision making.

The aim of this article is to globally draw the picture of the latest models of emotion in computer science, starting from a psychologist point of view. The appraisal-coping model will be presented more precisely, as well as an existing application example. Afterwards, we will present a small toy problem illustrating a decision making problem, using the appraisal-coping strategy. Eventually, a small conclusion will introduce future aspects to be developped and other prospects.

2 State of the art

From the most recent approaches of emotion modelling, two major cognitive types of research models have been developed: hierarchical and componential models [1].

According to the hierarchical approach, emotions have their origin from early stages of development. Indeed, emotions are hierarchically organised with numerous discrete emotions at a basic level and emotional dimensions at a higher level. Fundamental emotions like thirst or fear are elements intended to build more sophisticated emotions, ending up with very complex feelings such as jealousy or pride [7]. Emotions at the basic level have an important adaptive function and are directly linked to the body stimuli and effectors.

In componential models, emotions have qualitatively different facets [23]. The so-called "emotional response triad" is composed of the three main components for the emotion production: subjective experience, peripheral physiological responses and motor expression, to which some theorists include two other components, cognitive and motivational. The componential approach deals with the relative role assigned to each of these components. Then, emotions are created by stepping through all parts of the process, from the cognitive perception, until the actual response. Lazarus [13] and Scherer [20] are usually associated with this approach.

3 Appraisal Theories

Appraisal theories suggest that emotion is the result of underlying mechanisms including the subjective evaluation of the significance of a situation and its organism circumstances (appraisal), and the coping mechanisms that guide and provide adaptive responses (Frijda, 1986 [9]; Lazarus, 1991 [14]; Scherer, 1984 [20]; Scherer, Schorr & Johnstone, 2001 [24]; Smith & Lazarus, 1990 [26], 1993 [27]). As noted by Gratch and Marsella [10]: "Appraisal theories posit that events do not have significance in of themselves, but only by virtue of their interpretation in the context of an individual's beliefs, desires, intentions and abilities". The significance of an event is supposed to be evaluated on a number of criteria such as its relevance for one's well-being, its conduciveness for one's plans and goals, and the ability to cope with such consequences.

In the framework of the Scherer's Component Process Model (Scherer 1984 [20], 2001 [24]), Sander, Grandjean & Scherer (2005 [19]) describe emotion "as an episode

of interrelated, synchronized changes in the states of all or most of the five organismic subsystems[3] in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism.", (p.318). From this point of view, rather than static and basic states of the organism (e.g. Ekman, 1984 [8]; Izard, 1971 [11]), emotions are a dynamic process whose components are the cognitive component which function is the evaluation of objects and events, the peripheral efference component which regulates the system, the motivational component which prepares and guides the actions, the motor expression component which steadies communication of reaction and behavioral intention, and the subjective feeling component which monitors the internal state and environment interaction. In other respects, this model postulates that changes in one subsystem will tend to elicit related changes in other subsystems.

3.1 Appraisal-coping example

Based on the appraisal-coping approach, several new models have been conceived (see figure 1). Gratch & Marsella [10] have produced a domain independent model using cognitive maps. This model is intended to manipulate appraisal variables to analyse the present and past situations, and to design the future decisions to make and the coping strategies to adopt. Generally speaking, the appraisal-coping approach offers a very precise model of cognitive and emotional processes in decision making (previous works have been carried out on this subject, see for example [16]).

On the one hand, the environment is appraised with respect to one's goals and beliefs. This evaluation is realised through a certain number of variables defining the different appraising dimensions to be taken into account, such as the relevance (*Does the event require attention or adaptive reaction?*) or unexpectedness (*Was the event predicted from past knowledge?*).

On the other hand, one has to cope with the situation appraised before by using coping strategies. The different strategies offer a great range of possibilities, from the perfect control of the situation until the total resignation when facing the problem.The coping outcome alters the person-environment relationship not only by modifying the environment itself related to the problem, but also by changing the interpretation and willing through emotional aspects.

In order to represent the information about the situation, Gratch & Marsella [10] use causal maps. In the following scenario (figure 2), an oncologist, Dr. Tom, is supposed to help an eleven-year-old boy, Jimmy, for his stage 4 inoperable cancer, either by giving him morphine (which relieves the pain but hastens death) or leaving him suffer (and letting him prolong his life). A causal map (see figure 2) represents the past and present situations, as well as the possible decisions to make and their expected consequences.

[3] Organismic subsystem (and their major substata are the following): Information processing (Central Nervous System-CNS-), Support (CNS, Neuro-Endocrine System, Autonomic Nervous System), Executive (CNS), Action (Somatic Nervous System), Monitor (CNS), from Sander, Grandjean & Scherer (2005, [19].

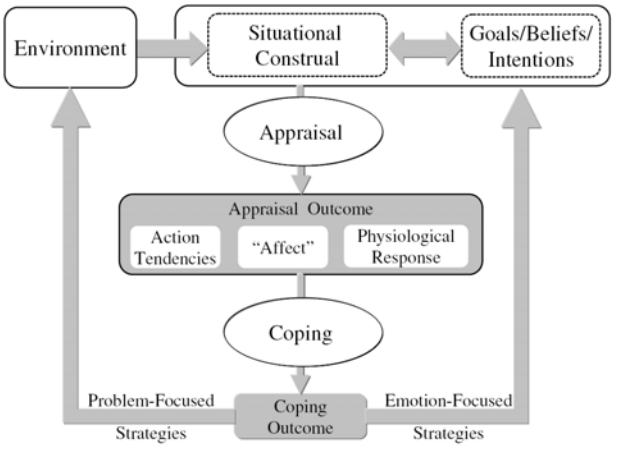


Fig. 1. The cognitive-motivational-emotive system. Adapted from Smith an Lazarus (1990, [26]).

This causal map can evolve with the modification of the parameters. For example, U_{Jim} is how Dr. Tom appreciates Jim's desirability for the corresponding event. This value can be updated after the coping process, dealing with emotional aspects. Once Dr. Tom re-appraises the consequences of his decision (by decreasing the probability parameter of the "Death hastened" event), he accepts to give Jim morphine, following Jim's mother request.

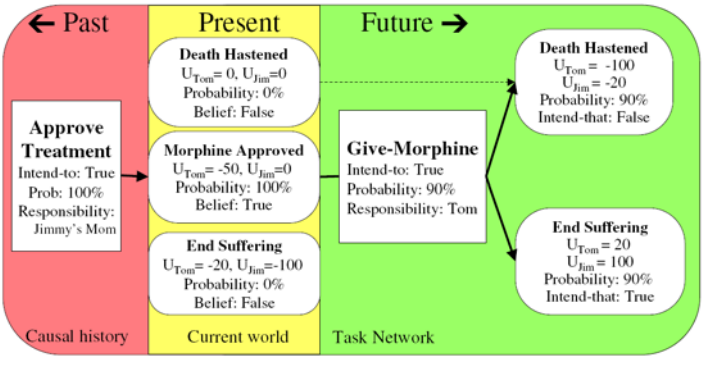


Fig. 2. Dr. Tom's causal interpretation at the end of the scenario.

However, the major problem when using this technique happens to be the storage of the information taken from the environment. In fact, we have two options.

The first option is to put all the information we need explicitly on the object to be evaluated. For instance, if an individual is watching a photograph of his wife and kids, the variables indicating pleasantness or pride are to be described on the photograph. With this method, each object in the environment is clearly identified as pleasant or mysterious or annoying, etc. The implementation is therefore easier, and the possibility of interacting with the emotional representation of the object is hence trivial. The main problem is a lack of flexibility, especially if more than one individual has to evaluate the same object.

The second option is the internal storage of the objects data, directly into the brain, using a memory strategy (Tulving et al., 1972 [29]). With this technique, memory is divided into several categories, each of which stores different kind of information. On the one hand, the long-term memory system is composed of the episodic memory (i.e. it refers to knowledge of episodes and facts that can be consciously recalled and related) and the semantic memory (underlying absolute knowledge and language; semantic memory is context-independent). On the other hand, the short-term memory, also known as working memory stores the current context-related data.

For each of the above options, emotion is supposed to be triggered from the appraisal processes. Indeed, the evaluation activity requires knowledge, and emotion is part of the memory processes of encoding, storage and retrieval (Tulving & Thomson, 1973 [30]). For modelling purpose, we assume that memory is split into three categories :

- The semantic or factual memory which stores the global knowledge of the world and the information considered to be facts, like *"Paris is the capital of France"*.
- The episodic or autobiographical memory which contains the personal events that happened in an individual's life. This type of memory is strongly linked with a spatio-temporal context.
- The working memory is the current dynamic representation that an individual has in mind when solving a particular problem.

Each time a memory element is encoded (i.e. added into the brain), it is stored along with the current emotional learning context e_{t_0}. Later on, when the information is required and must be found in the memory, the current emotional context e_t will be compared to the former emotional learning context related to the required memory item. If the old emotional context equals the current one (i.e. if $e_{t_0} = e_t$) the retrieval mechanism will be facilitated. On the contrary, if the two values are different ($e_{t_0} \neq e_t$), the retrieval process will be made more difficult.

4 Application example

4.1 The "Cascade" problem

The "Cascades" problem is a puzzle-like situation in which the goal is to fill up the grid with numbers according to the following instruction: "Each box contains the sum of the numbers situated above it. Look for the missing numbers in the grid". The initial state and the first step are presented hereafter (figure 3).
Ten-year-old children will be invited to solve the problem. In order to study how the emotions are elicited over the course of the problem, emotional manifestations will

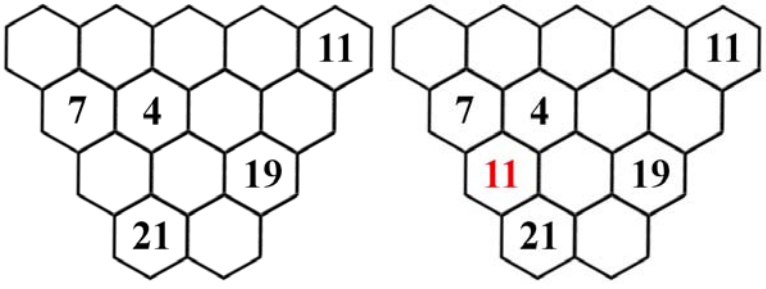


Fig. 3. The "*Cascades*" initial grid and the first solving step.

be recorded without interruption during the problem-solving activity. In line with the work of Clément & Duvallet (2007 [6]), we will focus on two kinds of response components: the physiological – electrodermal activity – and the expressive – facial expressions – components.

Although the definition of emotions remains controversial, some researchers distinguish emotions and other related notions as mood or personnality traits on the basis of their behavioral time course and intensity: emotions are defined as short-lived behavioral dispositions, moods are of longer duration and lower intensity, while personality traits reflect relatively stable behavioral tendencies.

Nevertheless, it is generally assumed that emotions may be evaluated by three kinds of responses: the physiological responses which the electrodermal activity and the heart rate are the most widely used, the expressive responses including facial, vocal, gestural and postural expressions and the subjective responses based in part on verbal report (see Bauer, 1998 [2]; Boehner, DePaula, Dourish, & Sengers, 2007 [4], for critical reviews).

Concerning the physiological component, and in particular the electrodermal activity, the study of Pecchinenda and Smith (1996) provides psychological significance of spontaneous skin conductance activity. In that work, participants were given to solve a set of anagrams which difficulty was manipulated by both the objective difficulty (easy, moderately, difficult, extremely difficult) and the amount of time available to solve the problems (30 vs. 120 seconds). The authors demonstrate that the skin conductance activity during problem solving is correlated to the appraisals of coping potential: in a difficult problem, appraisals of coping potential based on self-report are especially low and produce selective disengagement of the task, yielding reduced skin conductance activity. The spontaneous electrodermal activity is interpreted as reflecting task engagement (Pecchinenda, 2001 [17]).

Moreover, the results of Clément and Duvallet (2007 [6]) support the idea that skin conductance activity is a convergent measure of appraisal-related processes and that facial expressions reflect the appraisals of the events according to their conduciveness for the goal (Kaiser & Wehrle, 2001 [12]; Scherer, 1999 [21]; Smith, 1991 [25]; Smith & Scott, 1997 [28]).

4.2 The appraisal-coping model

The appraisal-coping model adaptation to the "Cascades" problem (see figure 4) allows us to analyse in a more accurate way the children's cognitive emotional activities. The appraisal step deals with the evaluation and the prediction of the plan selected by the child in order to solve the problem. According to this appraisal step, the child will fill in a hexagon, following his selected plan, or correct a previous result, or if the appraisal step leads to a bad evaluation, he will change his plan (we observe that the exercice instructions are interpreted in many different ways, especially for children with school difficulties).

The coping strategy is the actual decision a child will take with respect to his previous choices. The coping process is usually accompanied by an emotional reaction which depends on the appraisal consequences. For instance, if the child decides to reinforce his strategy, his emotional state will be positive. On the contrary, if he continually changes his plans, the situation will end up with a progressive disengagement for the task.

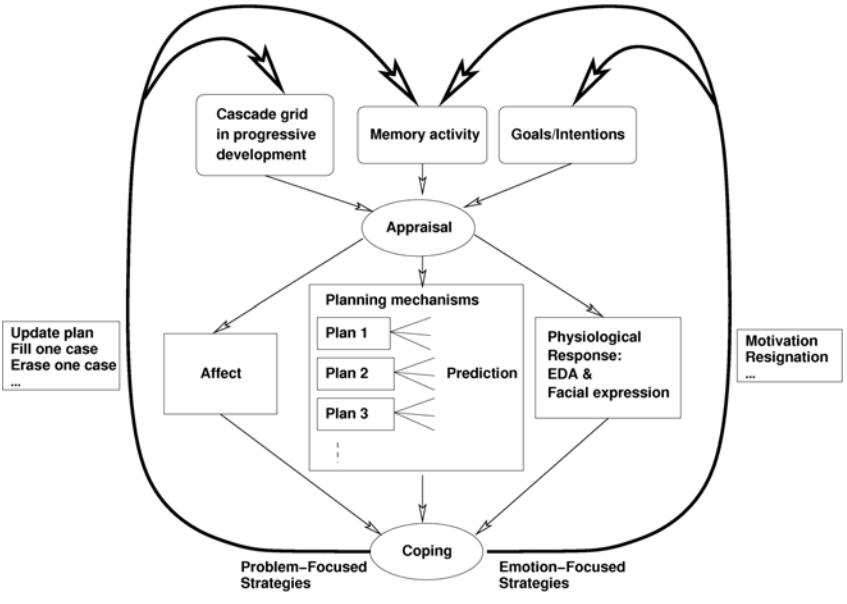


Fig. 4. The appraisal-coping model applied to the "Cascades" problem.

5 Conclusion and prospects

Contrary to the classical models which study cognition in a "cold" way, independently from any emotional process, we are now trying to unravel the mysteries underlying the emotion-cognition interaction. This new challenge makes researchers produce a new generation of cognitive models, based on more accurate systems, such as the appraisal-coping approach. The multidisciplinary work carried out so far allows us a better understanding of emotion mechanisms, by bringing out two complementary approaches:

- The produced model aims to analyse the experimental data.
- The experimental data come to strengthen or question the existing model.

As a future work, the strategies selected in the coping process are to be examined more precisely, in order to extract cognitive-emotional individual profiles, and especially when it comes to help pupils with school difficulties.

References

1. Baudic, S., Duchamp, G.H.E.: Emotions: theoretical models and clinical implications. In: European Simulation and Modelling (ESM) Conference 2006 (2006)
2. Bauer, R.M.: Physiologic measures of emotion. Journal of Clinical Neurophysiology 15, 388–396 (1998)
3. Bechara, A., Damasio, H., Damasio, A.R.: Emotion, decision making and the orbitofrontal cortex. Cerebral Cortex 10, 295–307 (2000), http://cercor.oxfordjournals.org/cgi/content/full/10/3/295
4. Boehner, K., De Paula, R., Dourish, P., Sengers, P.: How emotion is made and measured. International Journal of Human-Computer Studies 65, 275–291 (2007)
5. Clément, E.: La flexibilité représentationnelle, Mémoire de HDR - Université de Rouen, France (2007)
6. Clément, E., Duvallet, D.: Emotions elicited by impasses and subgoal achievements in problem-solving. In: CEM 2007, Cognition, Emotion and Motivation International Congress, Hammamet, Tunisia (2007)
7. Denton, D.: The Primordial Emotions: The Dawning of Consciousness. Oxford University Press, USA (2006)
8. Ekman, P.: Expression and the nature of emotion. In: Scherer, K.R., Ekman, P. (eds.) Approaches to emotion. Erlbaum, Hillsdale (1984)
9. Frijda, N.: The Emotions. Cambridge University Press, Cambridge (1986)
10. Gratch, J., Marsella, S.: A domain-independant framework for modeling emotion. Cognitive Systems Research 5, 269–306 (2004)
11. Izard, C.E.: The face of emotion. Appleton (Century/Crofts), New York (1977)
12. Kaiser, S., Wehrle, T.: Facial expressions as indicators of appraisal processes. In: Scherer, K.R., Schorr, A., Johnstone, T. (eds.) Appraisal processes in emotion: Theory, methods, research. Oxford University Press, New York (2001)
13. Lazarus, R.S.: Thoughts on relations between emotion and cognition. The American Psychologist 37, 1014–1019 (1982)
14. Lazarus, R.S.: Emotion and Adaptation. Oxford University Press, New York (1991)

15. Mahboub, K.: Modélisation des processus émotionnels dans la prise de décision, Master's thesis, Université du Havre, France (2006)
16. Mahboub, K.: Simulation of emotional processes in decision making. In: European Simulation and Modelling (ESM) Conference 2006 (2006)
17. Pecchinenda, A.: The psychophysiology of appraisals. In: Scherer, K.R., Schorr, A., Johnstone, T. (eds.) Appraisal processes in emotion: Theory, methods, research. Oxford University Press, New York (2001)
18. Pecchinenda, A., Smith, C.A.: The affective significance of skin conductance activity during a difficult problem-solving task. Cognition and Emotion 10, 481–503 (1996)
19. Sander, D., Grandjean, D., Scherer, K.R.: A systems approach to appraisal mechanisms in emotion. Neural Networks 18, 317–352 (2005)
20. Scherer, K.R.: On the nature and function of emotion. a component process. In: Scherer, K.R., Ekman, P. (eds.) Approaches to Emotion, pp. 293–317. Erlbaum, Hillsdale (1984)
21. Scherer, K.R.: Appraisal theories. In: Dalgeish, T., Power, M. (eds.) Handbook of cognition and emotion. Wiley, Chichester (1999)
22. Scherer, K.R.: Appraisal considered as a process of multi-level sequential checking. In: Scherer, K.R., Schorr, A., Johnstone, T. (eds.) Appraisal processes in emotion: Theory, methods, research. Oxford University Press, New York (2001)
23. Scherer, K.R., Peper, M.: Psychological theories of emotion and neuropsychology research. In: Boller, F., Grafman, J. (eds.) Handbook of Neuropsychology. Emotional behavior and its disorders, vol. 5, pp. 17–48. Elsevier, Amsterdam (2001)
24. Scherer, K.R., Schorr, A., Johnstone, T.: Appraisal processes in emotion: Theory, methods, research. Oxford University Press, New York (2001)
25. Smith, C.A.: The self, appraisal, and coping. In: Snyder, C.R., Forsyth, D.R. (eds.) Handbook of social and clinical psychology: The health perspective. Pergamon, New York (1991)
26. Smith, C.A., Lazarus, R.S.: Emotion and adaptation. In: Pervin, L.A. (ed.) Handbook of personality: Theory and research, pp. 609–637. Guilford, New York (1990)
27. Smith, C.A., Lazarus, R.S.: Appraisal components, core relational themes, and the emotions. Cognition and emotion 7, 233–269 (1993)
28. Smith, C.A., Scott, H.S.: A componential approach to the meaning of facial expressions. In: Russel, J., Fernandez-Dols, J.M. (eds.) The psychology of facial expression. Cambridge University Press, Cambridge (1997)
29. Tulving, E.: Episodic and semantic memory. In: Tulving, E., Donaldson, W. (eds.) Organisation of memory, pp. 381–403. Academic Press, New York (1972)
30. Tulving, E., Thomson, D.M.: Encoding specificity and retrieval processes in episodic memory. Psychological Review 80, 352–373 (1973)

Index