LAWS OF NATURE AND HUMAN CONDUCT



SYMPOSIUM

SOLVAY INSTITUTE · HONDA FOUNDATION

BRUSSELS, OCTOBER 7 - 9 1985

La page de couverture est une œuvre originale d'Octave Landuyt, spécialement conçue à l'occasion de ce symposium.

LAWS OF NATURE AND HUMAN CONDUCT

Specificities and Unifying Themes

Discoveries 1985 Symposium,
Hosted by « the International Institutes for Physics and Chemistry
Founded by E. Solvay »
and Supported by « the Honda Foundation »

Palais des Académies • Brussels 7-9 October 1985

Under the auspices of the Classe des Sciences de l'Académie Royale de Belgique.

Edited by Professor Ilya Prigogine, Nobel Prize and Doctor Michèle Sanglier

Brussels, Task Force of Research
Information and Study on Science

ISBN 2-87081-020-2

Preface

Laws of Nature and Human Conduct Specificities and Unifying Themes

Ilya Prigogine
Université Libre de Bruxelles
The University of Texas at Austin

To what extent do human activities stand outside the natural world ? Are the conceptions we have of the « Laws of Nature » applicable to the description of human conduct ?

These questions have dominated the history of ideas in the western world for two centuries ; they are at the root of the work of Kant and Hegel, Whitehead and Heidegger, to quote only a few names.

Classical science, which started with the « Newtonian paradigm », was centered around the idea of deterministic and time-reversible laws. In this context, the world was seen as a vast automaton ; man appeared as being outside nature, as a free agent in a mechanical universe, able to manipulate and exploit his environment. As expressed concisely by W.A. Weisskopf[1].

« *The Newtonian paradigm underlying classical and neo-classical economics interpreted the economy according to the pattern developped in classical physics and mechanics, in analogy to the planetary system, to a machine and to a clockwork : a closed autonomous system ruled by endogenous factors of a highly selective nature, self-regulating and moving to a determinate, predictable point of equilibrium.* »

It is clear that in this description, essential features related to the existence of multiple choices, of imaginative, structured responses are neglected.

Arnold Toynbee had devoted his entire life to the understanding of the forces which shape human history. At the end of his long career, he became worried about the relevance of this Newtonian paradigm[2] :

Have I not erred in applying to historical thought, which is a study of living creatures, a scientific method of thought which has been devised for thinking about inanimate Nature ? And have I not also erred further in treating the outcomes of encounters between persons as cases of the operation of cause- and effect ? The effect of a cause is inevitable, invariable and predictable. But the initiative that is taken to an encounter is not a cause ; it is a challenge.

Still the classic image of « science » continues to be propagated with considerable authority. It permeates many aspects of human sciences, in which rationality is identified with timelessness and equilibrium.

However, at present, our conception of nature is undergoing a radical change toward the multiple, the temporal, the complex. A new paradigm is taking shape. It is now understood that the behaviour of matter under non-equilibrium conditions can be radically different from its behaviour at, or near equilibrium ; and it is precisely this difference that introduces multiple choices, self-organization, and complex dynamics.

There is a close relation between non-linearity and distance from equilibrium. Close to equilibrium, the description of the temporal evolution can be expressed by linear equations. However, far from equilibrium, we deal with non-linear equations, which may lead to bifurcations and to the spontaneous appearance and evolution of organized states of matter, — « dissipative structures ».

Since the advent of quantum mechanics, we know that probabilistic concepts play an essential role in physics. What we observe is that they now play a basic role not only at the microscopic quantum level, but also at the macroscopic level. We are progressing towards the « open universe », with its multiple facets, to quote the title of a recent book by Sir Karl Popper[3].

It is the aim of this conference to discuss the recent reconceptualization of science as well as its possible role in the modelling of human activities.

The **first** session will be devoted mainly to the physical sciences, in order to identify the points which may be of interest in bridging the gap between the « hard » and the « soft » sciences. According to the classical view, there was a sharp distinction between simple systems, such as studied by physics or chemistry, and complex systems, such as studied in biology and human science. Indeed, one could not imagine greater contrast than the one which exists between the simple models of classical dynamics, or the simple behaviour of a gas or a liquid, and the complex processes we discover in the evolution of life or in the history of the human societies. This gap is now being filled. Over the last decade, we have learned that, in non-equilibrium conditions, simple materials such as a gas or a liquid, or simple chemical reactions, can acquire complex behaviour. This opens the way to new channels of transfer of knowledge from physics and mathematics to a variety of other fields, like the modelling of biological and human activity which will be discussed in the perspective of complex systems.

We come now to the **second** session. It is clear that the concepts involved in the dynamics of complex systems offer new possibilities for cognitive sciences. Let us quote some of the most challenging themes :
— the creation of new information as a result of bifurcation, stabilization and selection ;
— the amplification and selection of information. This theme is closely related to the description of neural networks as well as to the problem of artificial intelligence.
In this connection, it is also of interest to discuss the social behaviour of insect societies, which are immensely successfull from the ecological point of view. It has been recently established that in a changing environment, both deterministic and stochastic elements are necessary components of essential activities such as food collecting and recruiting. This leads naturally to the problem of learning, closely related to exploratory activities, and therefore to the stochastic elements apparent in collective hebaviour.
The **third** part of the symposium will be devoted to the impact of these recent developments on the description of human behaviour, with special emphasis on economics. As already mentionned, classical rationality was rooted in a static conception of the Universe, as well as in ideas such as complete information and global optimization.

The present reconceptualization of physics leads towards a revision of this concept of rationality, according to the lines of Herbert Simon[4] :

The capacity of the human mind for formulating and solving complex problems is very small compared to the size of the problems whose solution is required for objectively rational behaviour in the real world — or even for a reasonable approximation to such objective rationality.

Are we at present on the verge of a new rationality, and can we already describe some of the aspects this new rationality may take ?

It is of special interest to discuss models of human activity which make a minimum of a priori assumptions. It has often been stated that the choice of the right variables in modelling is a step of an importance equal to the formulation of the model itself. It is therefore interesting to notice that the recent progress in chaotic dynamics leads to exciting new possibilities in « recreating » the dynamic structure of the past from available time series, such as stock market indices in economics.

A new element involved in the description of human activity, as opposed to the modelling of physical processes and of simple biological activities, is the role of mental representation. Human activity is delocalized in time as it involves the memory of the past and an anticipation of the future. This new element is of crucial importance in the modelling of numerous fields of human activity (such as the so-called « rational anticipation » and the risk assessment in economics).

In short, this conference has provided an opportunity to discuss the possibilities of a new transfer of knowledge from the mathematical and physical study of complex systems to problems involving human activity. It may also be the place to discuss some interesting projects which have

been proposed recently, concerning problems such as : the « Science and praxis of Complexity » (E. Ploman) ; the new meaning of risk assessment (O. Giarini) ; or structural changes in societies destabilized by rapidly changing economic and political conditions (E. Laszlo). It is hoped that it can also serve as a natural forum for formulating other lines of investigation, taking into account the new possibilities coming from recent progress in the dynamics of complex system (such as the modelling and the dynamics of the learning process).

We hope that in this way the conference will bring out some of the unifying themes between human sciences and the new image of the natural world.

J. Salk coined the expression « the survival of the wisest ». Clearly, the growth of human population requires a new dialogue between man and nature as well as new relations among men[5].

It is burden of our generation to spell out more precisely the form that this new dialogue has to take if we are to overcome feelings of alienation and frustration, which may lead to an emotional rejection of science and technology.

NOTES

(1) Walter A. Weisskopf — « *Reflections on Uncertainty in Economics* », *The Geneva Papers on Risk and Insurance, 9 : 33, pp. 335-360.*

(2) Arnold Toynbee — *A Study of History*, Oxford University Press, 1972, p. 97.

(3) Karl R. Popper — *Postcript to The Logic of Scientific Discovery, Vol. 2, The Open Universe : An Argument for Indeterminism*, Totowa, New Jersey, Rowman and Littlefield, 1983.

(4) Herbert A. Simon — *Models of Man*, Wiley 1957, p. 198.

(5) Jonas Salk — *The Survival of the Wisest*, NY, Harper and Row, 1973. See also in this perspective A. Toffler, *The Third Wave*, NY, William Morrow 1980 ; and I. Prigogine and I. Stengers, *La Nouvelle Alliance*, Paris, Gallimard 1979, tr. as *Order out of Chaos*, NY, Bantam Books, 1984.

Opening Adresses

Opening Address

Jacques Solvay
President of the Solvay Institutes

I am very pleased and gratified by this collaboration between the Instituts internationaux de Physique et de Chimie and the Honda Foundation. I am very happy to welcome the Ambassador of Japan, who will read a message from the Prime Minister of Japan ; Mr. Honda, who is kind enough to be present in person ; Mr. Shimoda, the President of the Honda Foundation ; the Ministers Geens, and Hatry and the members of the Classe des Sciences de l'Académie Royale de Belgique for contributing to this meeting.

None could be better prepared to organize this meeting than Ilya Prigogine, who was able to gather around him a most remarkable team of researchers. As you know, he was the first belgian to receive a Nobel Prize in science. He was awarded also the prestigious Honda Prize. The ideas put forward by his research team are likely to play an important role in the discussions of this conference.

Mr. Honda and I share a somewhat similar position : we both have to deal with established enterprises, with all it implies in responsabilities. Let me take this opportunity to explain the relation I see between an enterprise and dissipative structure as defined by Prigogine.

A dissipative structure is an assembly of interacting components. The mathematical relations which describe it may have in certain cases a unique, deterministic solution ; but in other cases they may lead to a bifurcation giving rise to multiple possible solutions.

Any dissipative structure is a mesh of structures forming ensembles at different levels. A living organism is made up of cells, which have their

own structure and purpose ; cells are grouped together to constitute different organs : a leaf, a heart for example ; and the different organs interact to form a whole. Each substructure must satisfy the following conditions : they must reach a threshold value of free energy ; and they are each at its own level responsible for their own choices : nature is indeterminate, and not ruled by central planning.

These ideas may be applied to the individual, and also to the enterprise. Different members of an enterprise collaborate towards different projects ; the substructures have their own goals. In a tree for example, the roots seek to absorb water and mineral salts ; in an animal, the stomach participates in the digestion, and so on.

A project may be controlled through a variable ; for example, a chemical messenger may determine the degree of activity of the different elements of the ensemble. In animals, the control variable is often nervous impulses, or hormones circulating in the blood, responding themselves in some way to internal or external stimuli.

Whatever this may be, we are witnessing today a great arousal of interest in far-from-equilibrium conditions, which are a prerequisite for this excess of free energy.

Goal-directed dissipative structures have another property that of stability. Depending on their stability, a structure may become reinforced or weakened. A tree branch which is well exposed to sunlight will bud and develop ; it will become atrophied if maintained in the shade. One may find an exponential relation between the productive ability of a structure and the quantity of excess free energy it can afford ; this relation may determine the individual progress along the so-called learning curve, whose exponential character depends on excess free energy production, which depends in turn on the message or the messenger that indicates what the level of activity must be.

For simple organisms, the message may be a chemical variable. For insects, it may be pheromonal ; for human enterprises, the messages are the motivating values which lead an individual to produce excess free energy — which is a necessary condition for increasing productivity.

Between Mr. Honda's japanese society and our european one, we are all aware of great cultural differences and of the fact that motivating values in our enterprises are different.

I am convinced that a close study of these phenomena and a deeper analysis would lead to many most interesting findings.

We are at a turning point, be it in the field of economics, of science or of technology. Is this the transition between the second and the third wave, as suggested by Alvin Toffler ? We surely need new insights to overcome the coming decade of risk and uncertainty.

In spite of these revolutionary changes, our present is still deeply rooted in the past. Japanese enterprises on one side, European on the other are embedded in different cultural climates. Motivations, and accordingly

the reasons for success or failure are widely different. This leads us to the main question to be discussed here : are we on the verge of a new rationality ? I am looking forward to the answers that no doubt so many talents will present.

Opening Address

Takeso Shimoda
President, Honda Foundation

Mr. Chairman (Mr. Jacques Solvay, President of the Solvay Institutes), distinguished guests, ladies and gentlemen :

On behalf of all the participants of this symposium,

I would like, first of all, to express our sincere gratitude to his excellency Mr. Wilfried Martens, Prime-Minister of Belgium and to his excellency, Mr. Shizuhiko Yamamoto, Ambassador of Japan, for honouring by their presence the opening ceremony of the 8th Discoveries International Symposium

I would like also to express our deep appreciation to Prof. Ilya Prigogine, Chairman of the organizing committee of the symposium as well as to all committee members for their valuable cooperation given to the preparation of this international symposium.

It has been our belief that the greatest desire of manking at the present time should be the creation of a civilization in which utmost respect is paid to the human beings themselves. We believe also that this goal can only be achieved through inter-disciplinary cooperation of the many intellectuals belonging to various fields, especially those engaged in the field of science and technology. Based on this concept, the Honda Foundation has sponsored the « Discoveries International Symposia », which were first held in Tokyo in 1976, and then subsequently in Rome, Paris, Stockholm, Columbus in Ohio of the United States, London and last year in Melbourne, Australia.

In the previous symposia, we have had discussions on the concept of « catastrophe » which was considered to be inherent in the modern civi-

lization, and we have deepened our recognition of the status of « mega-crisis », a situation that mankind is unfortunately going to be confronted with, sooner or later.

In order to cope with such crisis, we conducted a comprehensive review on the concepts of « information » and « communication », the most fundamental aspects of human activity in the modern society.

And in Columbus, we had discussions on the theme of the « social impact of advanced technology ».

In London, on the theme of the « social and cultural challenge of modern technology » and in Melbourne, on the theme of « the use of resources and technology in the interest of mankind ».

And now, at this symposium here in Brussels, at the very center of Western Europe, father land of modern civilization, we are looking forward today to lively discussions among the numerous distinguished participants and observers from various countries of the world under the main theme of « laws of nature and human conduct ».

Among the participants of this symposium, I am very pleased to meet here once again, the distinguished leaders of past discoveries symposia, including among others, Sir Henry Chilver, Dr Chestnut, Prof. Sheridan, Dr Curran, Prof. Simon, Prof. Caianiello, Prof. Hambraeus, Dr. Margulies, Sir David Zeidler.

On the basis of a number of opinions originating from active discussions among the participants to the present symposium, we sincerely hope to make a contribution toward the creation of civilization which has high respect for humanity, the basic concept of our foundation.

In closing, I would like to express our deep appreciation to you, Sir, Mr. Chairman for taking personally the chair at the symposium and to other members of the Solvay Institutes for taking the lead in the realization of this important symposium.

Thank you for your kind attention.

I

The Reconceptualisation of the Physical Sciences in the Perspective of Complex Systems.

A New Rationality ?

Professor Ilya Prigogine
Service de Chimie Physique II — Université Libre de Bruxelles
and Center for Studies in Statistical Mechanics — The University of Texas at Austin

1. INTRODUCTION

As underlined in the preface the aim of this conference to discuss the recent reconceptualization of science as well as its possible role in the modelling of human activities. We have to identify the points which may be of interest in bridging the gap between the « hard » and the « soft » sciences. According to the classical view, there was a sharp distinction between simple systems, such as studied by physics or chemistry, and complex systems, such as studied in biology and human science. Indeed, one could not imagine greater contrast than the one which exists between the simple models of classical dynamics or the simple behavior of a gas or a liquid, and the complex processes we discover in the evolution of life or in the history of human societies. This gap is now being filled. Over the last decades, we have learned that, in nonequilibrium conditions, simple materials such as a gas or a liquid, or simple chemical reactions, can acquire complex behavior. This opens the way to new channels for transferring of knowledge from physics and mathematics to a variety of other fields.

It is clear that the concepts involved in the dynamics of complex systems offer new possibilities for cognitive sciences. Let us quote some of the most challenging themes :

— the creation of new information as a result of bifurcation, stabilization and selection ;

— the amplification and selection of information. This theme is closely related to the description of neural networks as well as to the problem of artificial intelligence.

In this connection, it is also of interest to discuss the social behavior of insect societies, which are immensely successful from the ecological point of view. It has been recently established that in a changing environment, both deterministic and stochastic elements are necessary components in essential activities such as food collecting and recruiting. This leads naturally to the problem of learning, closely related to exploratory activities, and therefore to the stochastic elements apparent in collective behavior.

We shall first summarize some relevant developments in nonequilibrium physics and in dynamical systems. How, in this perspective, can we see life and cosmology ? This will be the subject of the third section.

In all fields of interest here, we see the growing importance of irreversibility. Basic aspects of life and cosmology can only be understood with the transcription of time into space and matter. For this reason, we discuss in section 4 the microscopic meaning of irreversibility. This question, which has been so controversial for more than one hundred years, can now be presented in a simple and physically transparent way.

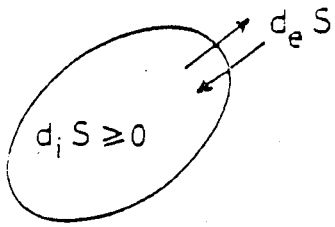
2. NONEQUILIBRIUM THERMODYNAMICS AND COMPLEX SYSTEMS

Some one hundred and twenty years ago (1865), Clausius formulated the second law of thermodynamics, introducing a new quantity : entropy. The fundamental importance of entropy is that as a result of irreversible, time-oriented processes, the entropy of our universe (considered as an isolated system) is increasing. This was an amazing statement at a time dominated by classical mechanics.

Since Clausius, physics has dealt with two concepts of time : time as repetition and time as degradation. But it is obvious that we must overcome this duality. Neither repetition (the negation of time), nor decay (time seen as a degradation), can do justice to the complexity of the physical world. We must therefore reach a third concept of time, which contains positive, constructive aspects as well.

Let us first recall that a distinction may be made between two additive terms in the variation of entropy dS (figure 2.1) : d_eS and d_iS, respectively the exchanges (positive or negative) between a system and its environment, and the internal entropy production (always positive). For an isolated system, $dS = d_iS$ is positive or zero (figure 2.2). But what happens if the system is submitted to external constraints ?

The first steps of thermodynamics were associated with the consideration of equilibrium states. There is no arrow of time for an isolated system at equilibrium. Later (1931), Onsager[5] gave general relations for nonequilibrium thermodynamics in the near-to-equilibrium region : systems submitted to weak external constraints give linear responses to these constraints. These linear, reciprocal relations between fluxes and forces designated nonequilibrium thermodynamics as a worthwhile subject of study, since they extended the usual description of equilibrium systems — in terms of thermodynamic potentials whose extrema corre-



$$dS = d_e S + d_i S$$

Figure 2.1
Second law of thermodynamics

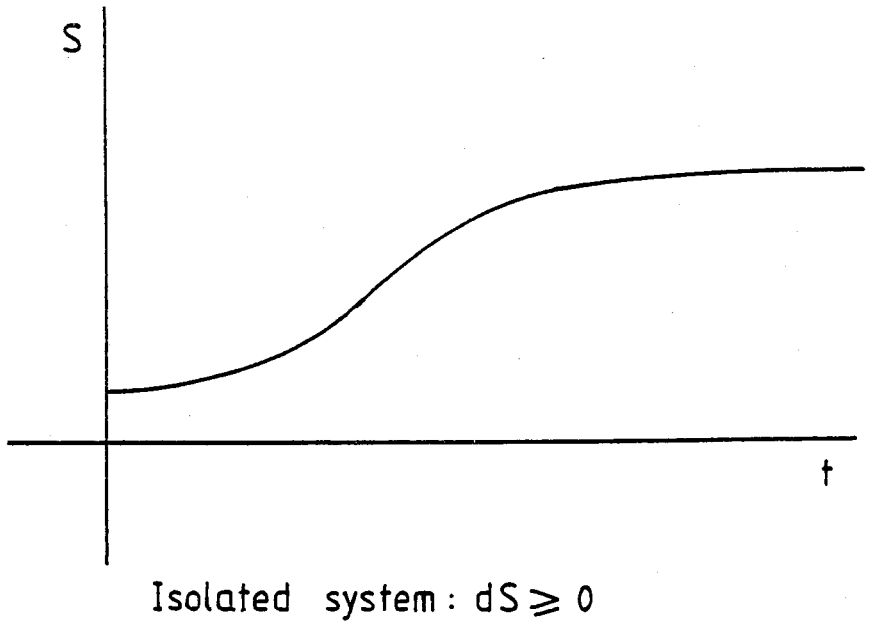


Figure 2.2
Isolated system : ds ≥ 0. Monotonous increase of entropy for an isolated system

spond to the final states — to a new description of nonequilibrium systems in terms of a new potential function : the entropy production.

Let us consider an example : A two-component system containing hydrogen and nitrogen is submitted to an external heat flow, maintaining a thermal gradient. As a result, one of the components, say H_2, accumulates in compartment I, the other in compartment II. In the frame of

macroscopic thermodynamics, the entropy production per unit time can be written as the sum of two terms,

$$\frac{d_i S}{dt} = \underset{>0}{\text{heat flow}} + \underset{<0}{\text{diffusion}} \geqslant 0$$

The important point is that even in this simple case, we have a coupled phenomenon, heat flow is producing entropy but diffusion goes *against* the concentration gradients (that is the so-called « thermal diffusion »), and would lead to a negative entropy production *if going on alone.*

This shows that irreversibility has generally a *dual* aspect : it corresponds both to dissipation (here the heat flow) and to the formation of *order* (the thermal diffusion). That is why I have put on the cover of one of my books a photograph of a well-known Indian symbol, the dancing Shiva, wo holds in one hand the fire which destroys and in another, the drum which creates !

Let us now turn to nonequilibrium thermodynamics valid for systems in far-from-equilibrium conditions [5]. Systems submitted to strong constraints present responses that are no longer linear ; accordingly, their behavior displays a multiplicity of stationary states and a lot of new related dynamical characteristics such as bifurcations, hysteresis, and so on. Let us give two examples of this newly understood behavior of nonequilibrium systems.

The so-called Bénard instability is a striking example of instability in a stationary state giving rise to a phenomenon of spontaneous self-organization ; the instability is due to a vertical temperature gradient set up in a horizontal liquid layer. The lower face is maintained at a given temperature, higher than that of the upper face. As a result of these boundary conditions, a permanent heat flux is set up, moving from bottom to top. For small temperature differences, heat can be conveyed by conduction, without any convection ; but when the imposed temperature gradient reaches a threshold value, the stationary state (the fluid's state of « rest ») becomes unstable : convection arises, corresponding to the coherent motion of a huge number of molecules, increasing the rate of heat transfer. In appropriate conditions, the convection produces a complex spatial organization in the system (figure 2.3).

There is another way of looking at this phenomenon. Two elements are involved : heat flow and gravitation ; under equilibrium conditions, the force of gravitation has hardly any effect on a thin layer of the order of 10 mm. In contrast, far from equilibrium, gravitation gives rise to macroscopic structures. Non-equilibrium matter becomes much more sensitive to the outer world conditions than matter at equilibrium. It like to say that at equilibrium, matter is blind ; far from equilibrium it may begin to « see ».

Consider secondly the example of chemical oscillations. Ideally speaking, we have a chemical reaction whose state we control through the appropriate injection of chemical products and the elimination of waste products. Suppose that two of the components are formed respectively

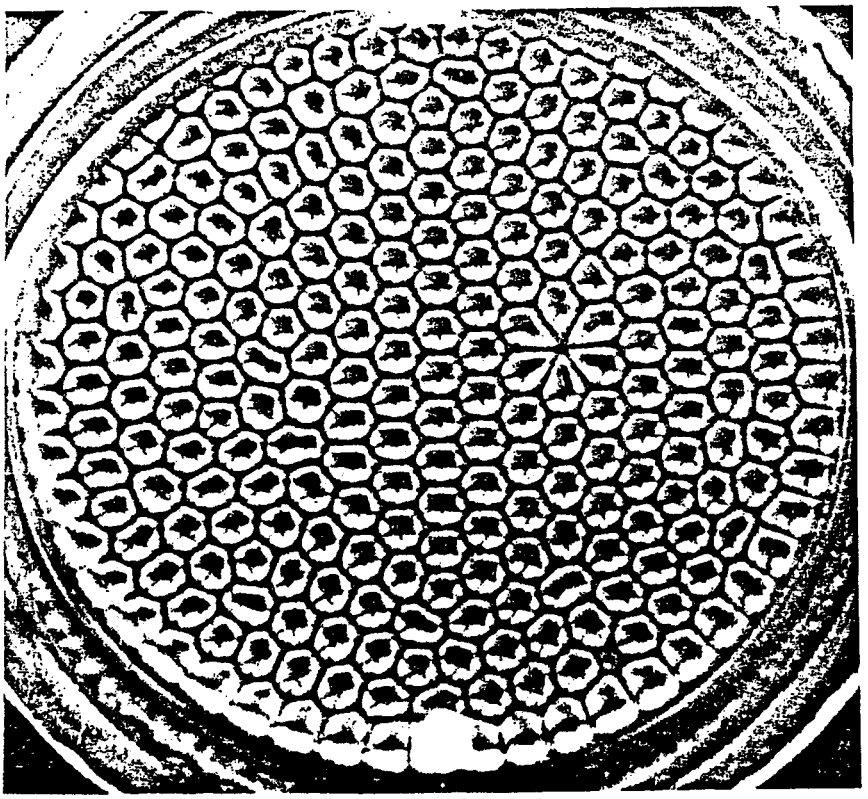


Figure 2.3

The Bénard hydrodynamical instability (as seem from above).
Liquid heated from below

by red and blue molecules in comparable quantities. We would expect to observe some kind of blurred color with perhaps occasionally some flash of red or blue spots. This is, however, not what actually happens. For a whole class of such chemical reactions, we see in sequence the whole vessel become red, then blue, then red again : we have a « chemical clock ». In a sense, this violates all our intuitions about chemical reactions.

We used to speak of chemical reactions as being produced by molecules moving in a disordered fashion and colliding at random. But, in order to synchronize their periodic change, the molecules must be able to « communicate » in a sense. In other words, we are dealing here with new supermolecular scales — in both time and space — produced by chemical activity.

The basic conditions to be satisfied for such chemical oscillations to occur are auto- or cross-catalytic relations, leading to « nonlinear » behavior, such as described in numerous studies of modern biochemistry. Remember that nucleic acids produces proteins, which in turn lead

to the formation of nucleic acids. There is a cross-catalytic loop involving proteins and nucleic acids.

Nonlinearity and far from equilibrium situations are closely related ; their effect is that they lead to a multiplicity of stable states (in contrast to near-equilibrium situations, where we find only one stable state). This multiplicity is to be seen on a « bifurcation diagram » (figure 2.4) : we have plotted the solution of the problem, X, against some bifurcation parameter λ (X would be for example the concentration in some chemical component, and λ could be related to the time the molecules spend in the chemical reactor). For some critical value of control parameter, say λ_c, new solutions emerge. Moreover, near the bifurcation point, the system has a « choice » between the branches : we could therefore expect a stochastic behavior : near a bifurcation point, fluctuations play an important role.

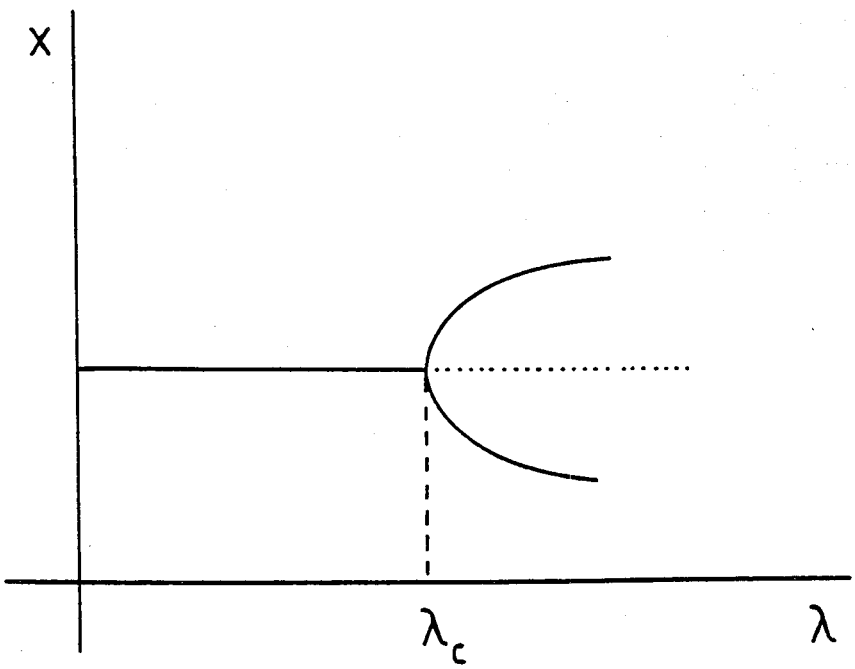


Figure 2.4

Bifurcation between stationary branches for a control parameter

The fact that dissipative systems present a behavior which can be described in terms of attractors implies another basic difference with mechanics. Suppose we have some foreign celestial body approaching the earth. This would lead to a deformation of the earth's trajectory, which would remain for astronomical times. Dynamical systems have no way

to *forget* perturbations. This is no longer the case when we include dissipation. A damped pendulum will reach a position of equilibrium, whatever the initial perturbation from it.

We can also understand in quite general terms what happens when we drive a system far from equilibrium. The « attractor » which dominated the behavior of the system near equilibrium may become unstable, as a result of the flow of matter and energy which we direct at the system. Nonequilibrium becomes a source of order : new types of attractors, more complicated ones, may appear, and give to the system remarkable new spacetime properties.

We have seen that dissipative systems may forget perturbations ; these systems are characterized by attractors. The most elementary attractors are points or lines, as represented in figures 2.5 and 2.6. In figure 2.5, we have a point attractor P in a two-dimensional space (X_1 and X_2 may be concentrations of some species) : whatever the initial conditions, the system will evolve necessarily towards P. In figure 2.6, we have a line-attractor : whatever the initial conditions, the system will eventually evolve along this line, called a limit-cycle.

But attractors may present a more complex structure : they may be formed from a set of points (figure 2.7), whose distribution may be dense enough to permit us to ascribe them an effective (nonzero) dimensionality. For example, the dimension of the attractor on figure 2.3 may be any real number between 2 and 3. Following the terminology of Benoit Mandelbrot, one may call this a « fractal » attractor. The behavior of systems in the presence of such an attractor can be described using the new qualitative nonlinear dynamics. For example, one can see that trajectories on the chaotic attractor in figure 2.8 (corresponding to a model given by Rössler) are converging along one direction, diverging along the other ; these characteristics are described by the respective Lyapunov exponents attached to these directions.

Such systems have unique properties, reminiscent of, for instance, turbulence which we encounter in everyday experience. They combine both fluctuations and stability. The system is driven to the attractor ; still, as the latter is formed by so « many » points, we may expect large fluctuations. One often speaks of « attracting chaos ». These large fluctuations are connected to a great sensitivity with respect to initial conditions. The distance between neighboring trajectories grows exponentially in time (this growth is characterized by the Lyapunov exponents). Attracting chaos has now been observed in a series of situations including chemical systems or hydrodynamics ; but the importance of these new concepts goes far beyond physics and chemistry properly. Let us indicate some examples recently studied.

We know that climate has fluctuated violently over the past. Climatic conditions that prevailed during the last two or three hundred million years were extremely different from what they are at present. During these periods, with the exception of the quaternary era (which began about two million years ago) there was practically no ice on the conti-

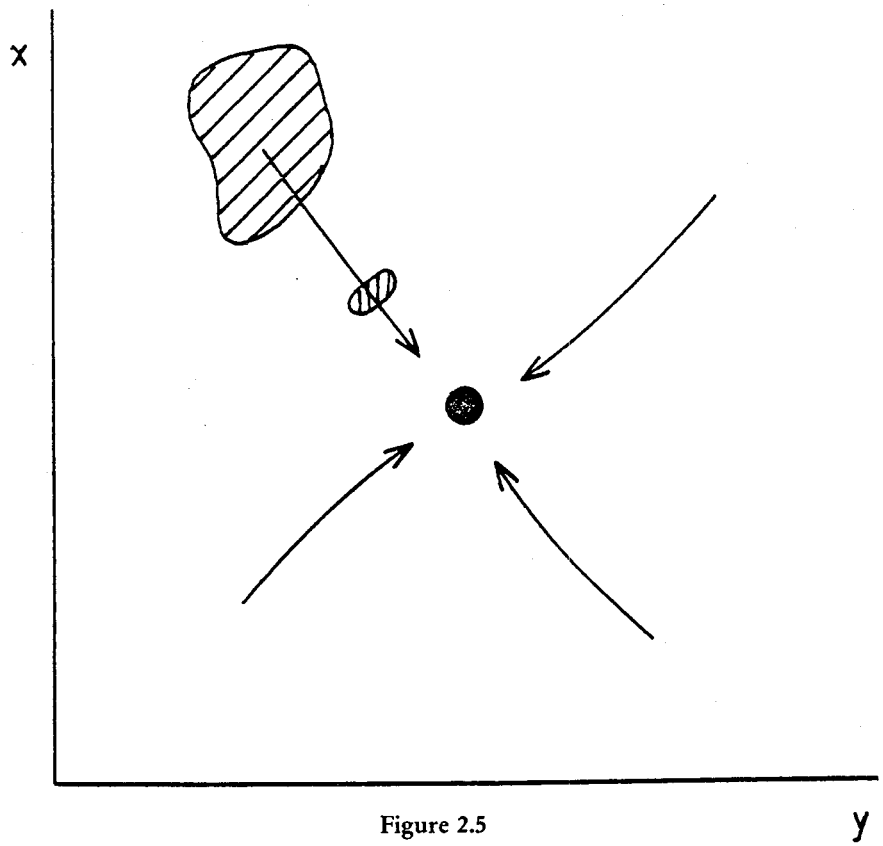


Figure 2.5

Point attractor in phase space (x,y)

nents, and the sea level was higher than its present value by about 80 meters. A striking feature of the quaternary era is the appearance of a series of glaciations, with an average periodicity of one hundred thousand years, to which is superposed an important amount of « noise ». What is the source of these violent fluctuations (figure 2.9) which have obviously played an important role in our history ? There is no indication that variation of the solar radiation may be responsible for them.

A recent analysis by C. and G. Nicolis[7] has shown that these fluctuations can be modelled in terms of four independent variables, which form a nonlinear dynamical system leading to a chaotic attractor of dimension 3.1 embedded in a phase space of dimension 4. The variability of climate could have been thought of as resulting from the interplay of a large number of variables, acting in a deterministic fashion ; the situation would then be very similar to the outcome of the law of large numbers. The new insight is that it is not so. The temporal complexity can be ascribed to only four independent variables. We may therefore speak of an intrinsic complexity or unpredictability of climate.

In a quite different field recent work[8] has shown that the electrical activity of the brain during deep sleep, as monitored by electroencephalo-

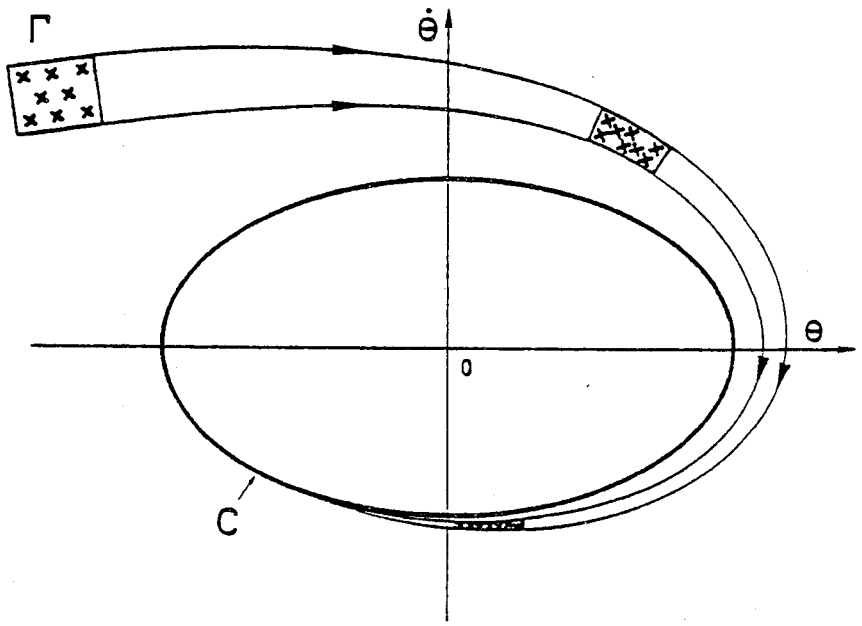


Figure 2.6

Limit-cycle attractor.
One sees a contraction of the area associated to the transition to the limit cycle

gram (EEG), may be modelled by a fractal attractor. Deep sleep EEG may be described by a dynamics involving five variables ; again, this is very remarkable as it shows the brain acting like a system possessing intrinsic complexity and unpredictability.

It is this instability which permits the amplification of inputs related to sensory impression in the waking state. Obviously, the dynamical complexity of the human brain cannot be an accident. It must have been selected for its very instability. Is biological evolution the history of dynamical instability, which would be the basic ingredient of the creativity characteristic of human existence ?

3. IRREVERSIBILITY AND INFORMATION IN BIOLOGY AND COSMOLOGY

Let us now present some (necessarily rather speculative) remarks about the two problems which have most excited the imagination of men : the origin of life and the origin of the universe.

We want to emphasize that irreversible processes have likely played an esssentiel role in inscribing, so to speak, time into matter at the early stage of the universe, and in producing the information carried by basic biochemical compounds such as DNA.

Indeed, a basic feature of biomolecules is that they are carriers of information. This information is quite similar to a text, which has to be read

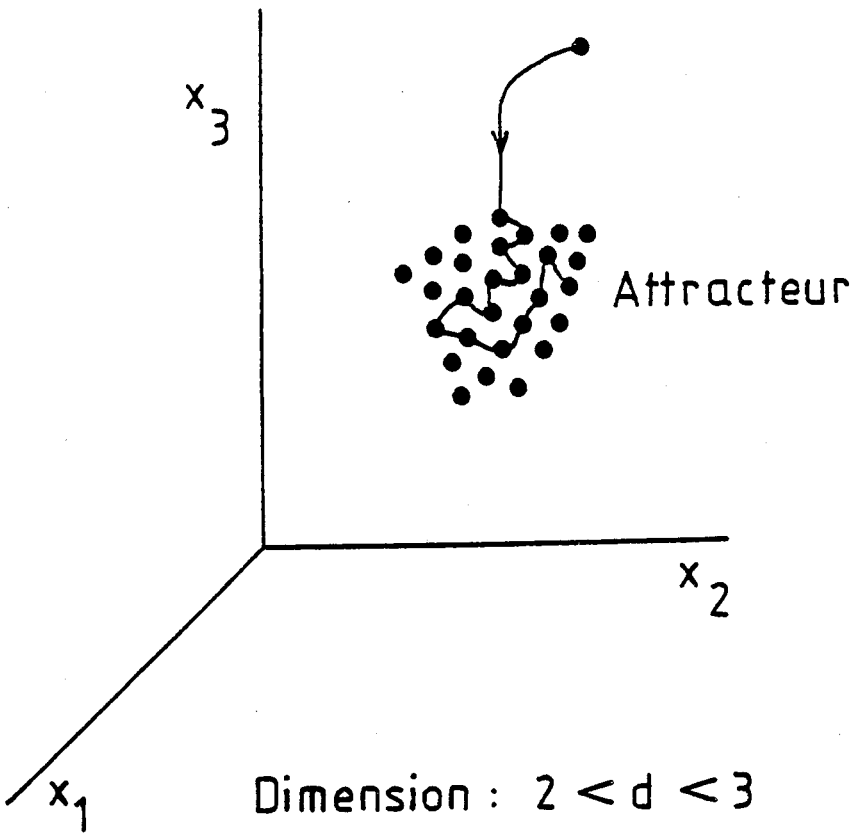


Figure 2.7

« Fractal attractor » formed by a multiplicity of points with a « dimension » between 2 and 3

in one direction, exactly like a text in human language. Biomolecules present in this sense a broken symmetry. It is natural to relate this broken symmetry to the broken time symmetry which is expressed in the second law of thermodynamics. Already in the simple case of the Bénard instability, irreversibility is transformed into pattern. However, the Bénard flow pattern persists only as long as heat is crossing the liquid layer. In contrast, life has an extraordinary degree of persistence, as it originated 3.4 billion years ago.

One of the basic questions in the origin of life is a better understanding of the selection of nucleotide sequences. Suppose we take a protein 100 amino acids long. It is well known that in nature there exist $N = 20$ kinds of amino acids. This would lead to 20^{100} possible sequences. None-quilibrium conditions are essential to reduce this number (see ref. 9, chapter IV). We have to produce sequences to which we can associate some « *information* ». This excludes purely random sequences, as well

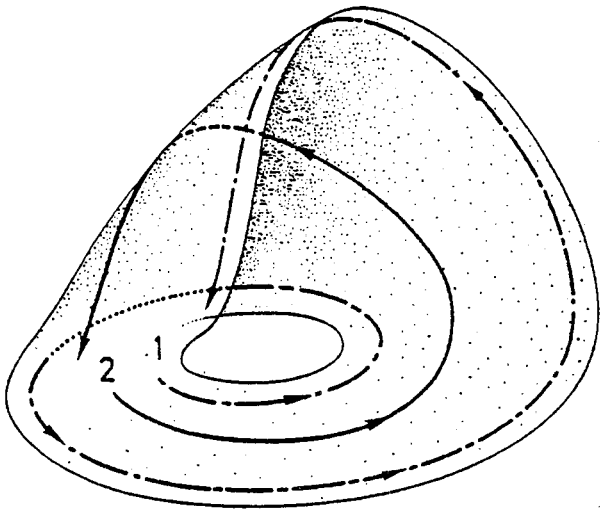


Figure 2.8

Chaotic attractor corresponding to a Rössler model

as periodic sequences which could be compressed into a few symbols. To borrow a term from communication theory, we need an *incompressible* information.

It has recently been shown that chaotic chemical reactions provide us with a very simple model for this inscription of irreversibility into matter. As suggested by G. Nicolis and G. and J. Subba Rao[10], we may consider a chemical reaction involving three components. X_1, X_2 and X_3.At each crossing of some threshold by one of these concentrations, we assume that the molecule of the corresponding component is added to some polymer (figure 3.1). As a result, we obtain a sequence which indeed carries an incompressible information.

Of course, with a limit-cycle reaction, the analogous sequence would carry a compressible information, as in this case we could only expect periodic strings like xyzxyzxyz. It has been shown by the authors that the sequence obtained from the chaotic reaction, when read in one direction, corresponds to a Markov chain. Irreversibility leads here to a « text ». This is a simplified model, but one may hope that real-world experiments can be performed to produce polymer chains in far-from-equilibrium conditions, separated from the thermodynamical branch by bifurcation points. Here, we may expect to find information-carrying molecules which could indeed be plausible ancestors for the biomolecules of today. The important point we want to emphasize is that the condition of non-equilibrium processes and chaotic dynamics can indeed generate « information ».

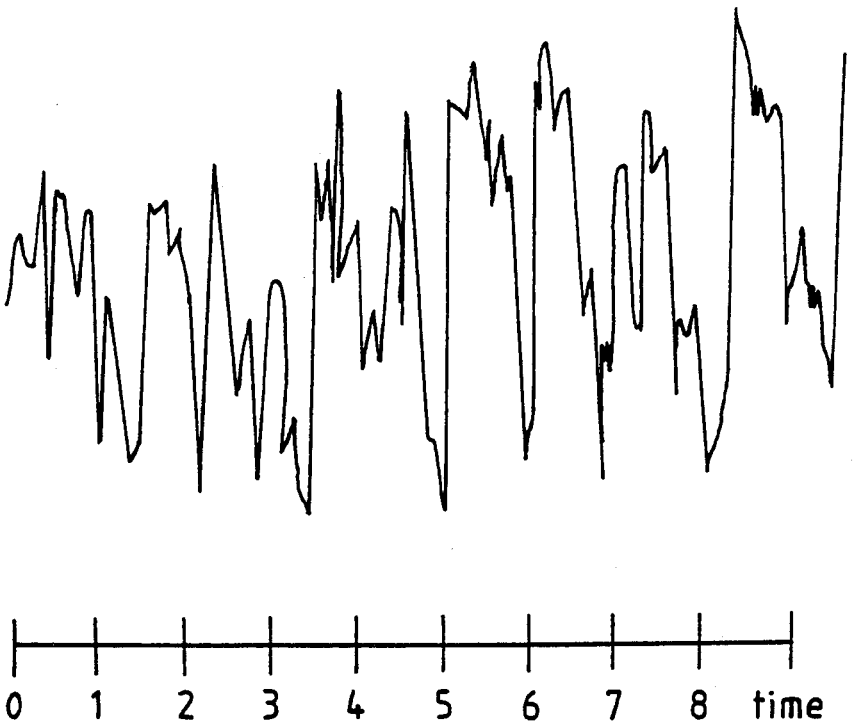
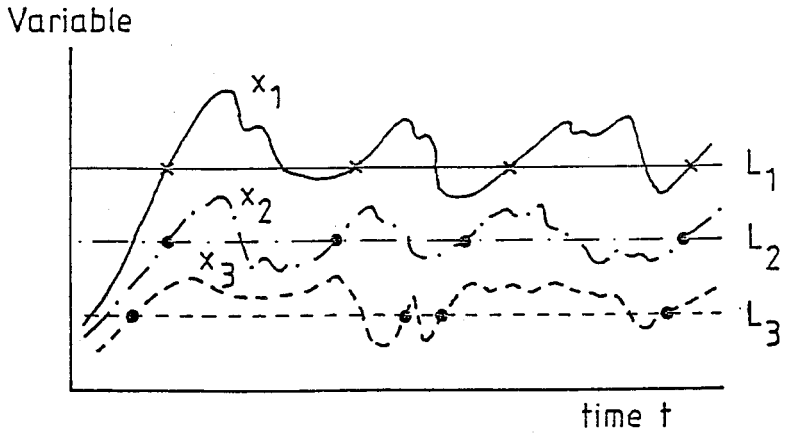


Figure 2.9

Full scale = 968 x 10^3 years.
Series of temperatures for the earth's climate over a million years

We have just made some remarks on the problem of the origin of life, but obviously the most striking of all such problems is the one of the origin of our universe. The modern history of cosmology has been a dramatic one. As is well known, Einstein proposed en 1917 a static model of the universe. General relativity seemed to him to be the ultimate achievement, as it gave the unification of gravitation with space-time. But this static image was to be rapidly given up, as in 1922 Friedmann showed Einstein's equations were unstable, and the observations of Hubble pointed in the direction of an evolving, expanding universe. Later, the celebrated residual black-body radiation suggested that beyond this geometrical evolution, there is a more basic, thermal evolution.

We come therefore to the modern « standard » cosmology (figure 3.2). It presents two fundamental aspects : the temperature of the universe increases monotonously when we come close to the « big bang », while,



zxyx zxyx zxzyx zxyx zxyx zxyx zxyx zxyx

zxyx zxyx zxyx zyx zxzyx . . . .

α = zyx , β = zxyx , γ = zx

. . . $\beta\beta\gamma\alpha\beta\beta\beta\beta\beta\alpha\gamma\alpha\beta$ . . .

Figure 3.1

Coding of a chaotic behavior leading to an « incompressible » information. (For the origin of the data, see ref. 10.)

according to the adiabatic type of evolution, the entropy of the universe remains constant. We cannot go into details here, but two obvious questions are implied : how could our universe start from this high temperature ? and how could entropy reach this level at once ?

It is very tempting to reconsider these questions from the viewpoint of irreversible processes[11]. We find in the universe a remarkable duality : it is made principally of photons and baryons. A rermarkable insight into the structure of the universe is given by the so-called « specific entropy » S, defined as the ratio of the number densities of photons to baryons. The usually accepted value is

$$S \simeq 10^8 \sim 10^9$$

Today baryons are considered to be nonequilibrium objects which are either in a metastable state or decaying, while the photons are « waste

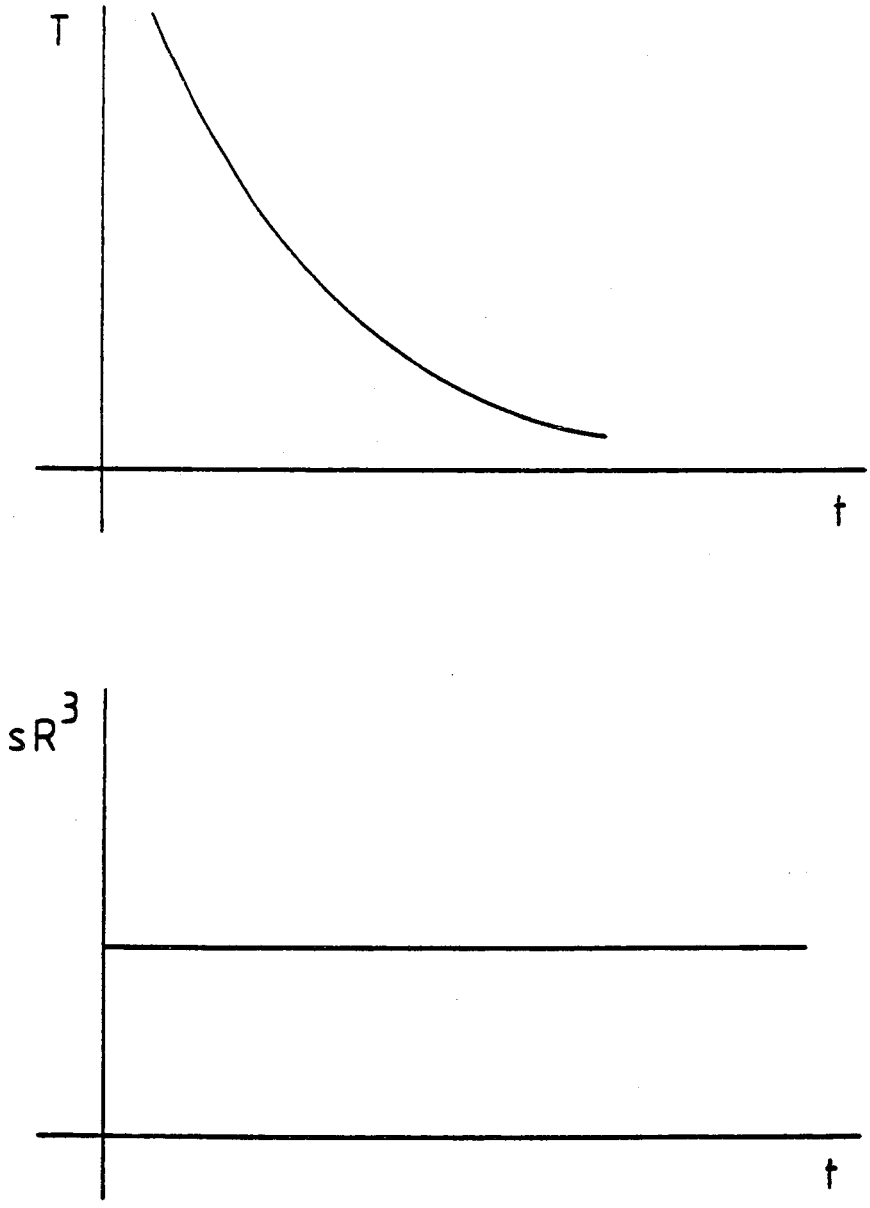


Figure 3.2

Initial conditions assumed in the standard cosmology.
Temperature goes to ∞ for $t \to 0$ and entropy remains constant

products » which can no longer decay into other known forms of matter. So, most of our universe is at thermal equilibrium from this « particle » point of view ; but (fortunately for us) there is a small nonequilibrium residuum. Such a duality is typical for coupled nonequilibrium processes. On earth, solar radiation gives rise to positive entropy pro-

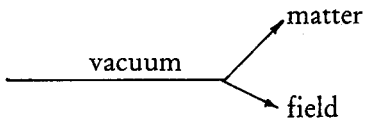
duction, most of which leads to dissipation in the form of photons. In a schematic way,

$$\frac{d_i S}{dt} = \underset{>0}{\text{production of photons}} + \underset{<0}{\text{production of biomolecules}} \geqslant 0$$

This reminds us of the situation, mentioned in section 2 (effect of thermal diffusion), where the entropy production is also the sum of a positive and a negative term. If in the example of thermal diffusion the heat flow is discontinued, the concentration of the components in the two compartments will slowly equalize.

The important point is that the part of the system which is slowly evolving to equilibrium has been itself brought out of equilibrium by a *nonequilibrium process.* Could it be that matter (the baryons and leptons) is the nonequilibrium product of some cosmic nonequilibrium process which has also produced the black body radiation ?

There is also a basic duality which appears in general relativity : on one side, spacetime ; on the other, matter. These two aspects are connected through Einstein's field equations. In recent years, describing the creation of the universe as a « free lunch » has become quite popular[12]. The energy of the vacuum would be transformed into a *positive* energy (the « matter »), and a *negative* energy (the gravitational field energy)



However, if matter is indeed created from the vacuum, this should be an *irreversible, entropy-producing process.* In this view, the entropy of the vacuum would be zero. Matter would be the « contamination » of spacetime, carrying the entropy, generalized by this creation process.

A possible mechanism for such an irreversible creation of matter related to the geometrical spacetime evolution of the universe has been discussed elsewhere[11]. The intriguing new possibility is that the universe may now start cold and empty (see fig. 3.3), it will then heat up through the decomposition of unstable particles, reach a maximum temperature and then cool down through adiabatic expansion. Similarly, the entropy would start from zero and would grow with the creation of matter.

It is interesting to notice how far we are here from the concept of « elementary » particles. On the contrary, particles can now be characterized by the entropy which was necessary to produce them. Perhaps it could always be possible to associate a high information content (which means a negative entropy) to some particles. That would mean that they could have not been created *alone*, exactly as a biomolecule which can only be found in conjunction with other molecules which provide the necessary free energy.

Let us summarize our main findings. Irreversibility plays a constructive role, not only a destructive one, in our universe. Questions such as the

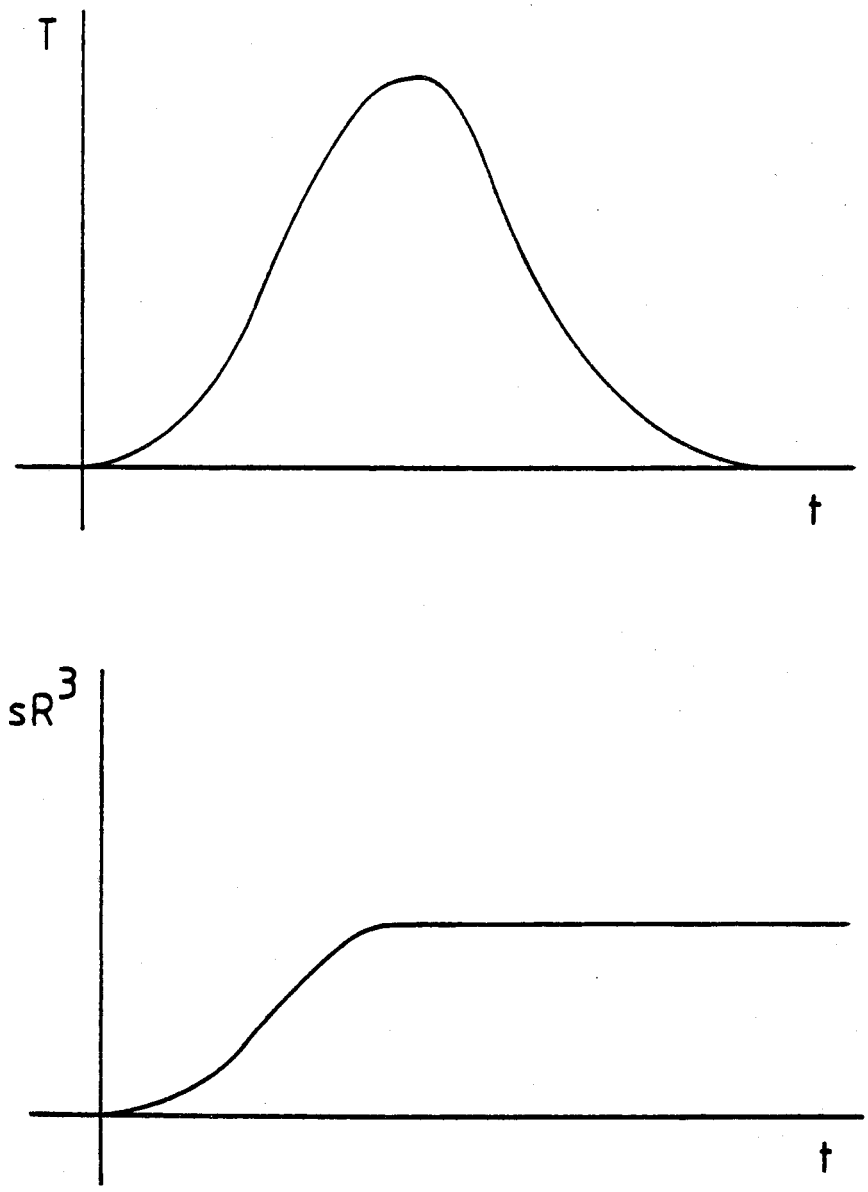


Figure 3.3

Possible initial conditions for the temperature of the universe and its entropy

origin of life, the origin of the universe, or the origin of matter, can no longer be discussed without recourse to irreversibility. The search for a fundamental time-reversible level in nature seems to have come to an end. Indeed, irreversibility has to exist on all levels or on none. It cannot emerge in the transition from on level of description to another. How then to introduce irreversibility on the fundamental dynamical level ? That is the question to which we turn now.

4. FROM INSTABILITY TO IRREVERSIBILITY

What do we mean by laws of Nature ? Let us take the example of coin tossing, already discussed by Poincaré. The familiar outcome is traditionally represented by equal probabilities : $P_0 = 0.5$ and $P_1 = 0.5$. Does this outcome correspond to a fundamental law of Nature ? If we assume that a deterministic representation of coin tossing (Newton's laws) is possible, should not the outcome be 0 or 1, according to the initial conditions ?

We can now analyze this situation in more detail. Suppose we restrict the initial conditions to smaller and smaller intervals. If, starting with some finite precision, the outcome would correspond to the deterministic 0/1 prediction, the basic law of coin tossing would indeed be deterministic. However, if for arbitrary initial conditions, whatever the finite precision, the outcome remains 0.5/0.5, then the law of coin tossing is a statistical law ; we have then to accept the outcome of a single coin toss is unpredictable — but that we can make statistical predictions if we repeat the game.

What is the difference between the two situations ? Let us represent the initial conditions on a segment of a line. Two things can happen :

— Either there are parts of this segment on which all points (with some appropriate measure) lead to 0, or to 1. Then we come to a deterministic situation. The dynamical system is then *stable* (sufficiently small perturbations will not change the outcome).

— In contrast, if each point leading to 0 is surrounded by points leading to 1 (and vice-versa), we can only make statistical predictions. The system is then *unstable*. Of course, this issue would change if we accepted the idea of infinite precision.

It is reported that Bohr stated to Planck that in quantum mechanics, coordinates and momenta cannot be determined simultaneously. Planck answered : « But God knows both ! », to which Bohr allegedly remarked that physics deals only with what *man* can know.

The illusion of a complete (infinite) knowledge comes, as it would seem, from the historical fact that classical science started with the study of periodic motions. The return of the sun and the regularity of celestial phenomena have deeply influenced man's thought since paleolithic times. It led to the prototype of knowledge as expressed in classical physics. But this regularity of periodic processes is not the general case. The message conveyed by the second principle of thermodynamics is that we are not living in a world that can be described in terms of periodic motions. It is an unstable world, which we know through a « *finite window* ».

Let us consider a highly unstable dynamical system and show how this instability leads to irreversibility. The system we shall consider corresponds to the so-called « baker transformation ». Here the phase space is a square, on which we apply the operations well known by bakers :

flatten, cut, fold (figure 4.1). In this transformation the surface area is conserved, but points are scattered over the whole phase space.

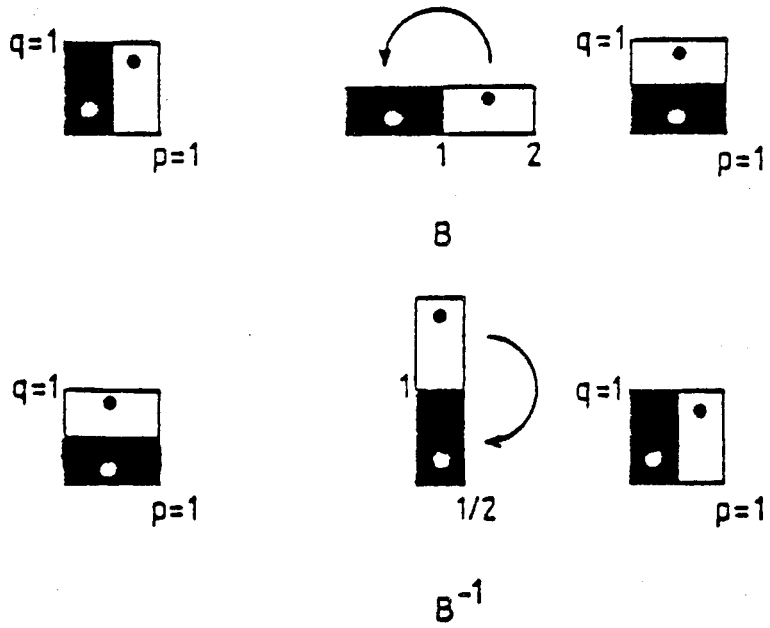


Figure 4.1

The baker transformation in phase space (p,q)

The baker transformation induces a particular geometric structure on the square. It is clear that points separated by a vertical (horizontal) interval will reduce (magnify) this distance by a factor 2 at each iteration. This determines two foliations of the phase space, one into horizontal (dilating) fibers, and one into vertical (contracting) fibers.

Each point is at the intersection of a contracting and a dilating fiber. All systems on the same contracting fiber tend to the same phase point for $t \to +\infty$. They have the same (asymptotic) *future*. All systems on the same dilating fiber have the same past (for $t \to -\infty$). The essential point is that fibers characteristic for unstable dynamical systems have a *broken time symmetry*. The dilating fiber is a time-reversal conjugate of the contracting fiber.

For this reason, every finite initial information — whatever its precision — leads to a nonlocal situation with broken time symmetry. Whatever the precision, we shall never be able to distinguish points sufficiently close on the *contracting* fiber. We have therefore to replace a point by an ensemble of points on the same contracting fiber. This leads to the consideration of bundles of trajectories (a nonlocal concept) with broken symmetry as this construction is based essentially on the distinction

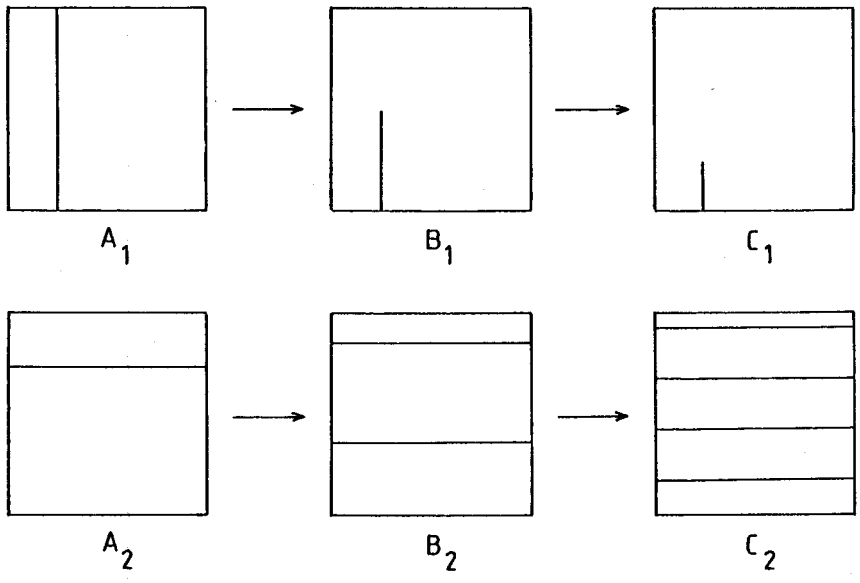


Figure 4.2

Contracting and dilating fibers

between contracting and dilating fibers. We cannot go here into the details, which are of a rather technical nature[13,14,15]. In order to avoid misunderstandings, let us however emphasize that the limitation to finite precision does not introduce any subjectivistic future. It would indeed be difficult to accept that our measurements could be based on other mechanisms than other natural processes : any interaction between dynamical systems also involves a limited « window » formed by their limited sensitivity, exactly as in our observations. That makes the description of our world inherently irreversible and stochastic, even in the field of classical dynamics.

5. CONCLUDING REMARKS

In the approach advocated here, rationality is no longer to be identified with « certainty », nor probability with ignorance. At all levels, probability plays an essential role. It was of course always understood that exact (that is, infinitely precise) initial conditions correspond to an idealization, but we find it rather unexpected that giving up this idealization leads to such momentous consequences.

It is of course not the first time that such a situation has appeared in the history of physics. The history of science includes both the progressive accumulation of klnowledge and the succession of conflicting theories. For example, it is well known that due to the presence of Planck's constant , the structure of quantum theory differs drastically from that of classical theory. In short, we witness here the inadequacy of the ideal of « complete knowledge » which has haunted western science for three

centuries. As already mentioned, the message of the second law is that we live in a highly unstable universe, in which the classical idealization do not apply.

This new rationality of science leads us to reconsider the relations between man and man as well as the relations between man and nature[16]. We have already mentioned that we live at the intersection of at least two systems of values. Clearly, a social system is by definition a nonlinear one, as interactions between the members of the society may have a catalytic effect. At each moment, fluctuations are generated, which may be damped or amplified by society. An excellent example of a huge amplification is the acquisition of knowledge which in a few decades led from the work of a few pioneers in solid state physics to the information revolution that we are witnessing today.

Scientific and technological progress « probes » the stability of the social system. In this view, there can be no question of the « axiological neutrality » of science. Problems that may arise at the science/society interface can only be solved by understanding the actual complexity of societal processes. If these are not understood, the response of the system may be ultimately a negative one.

J. Salk coined the expression « the survival of the wisest ». Clearly, the growth of human population requires a new dialogue between man and nature as well as new relations among men[17,18].

It is the burden of our generation to spell out more precisely the form this new dialogue has to take if we are to overcome feelings of alienation and frustration, which may lead to an emotional rejection of science and technology.

ACKNOWLEDGEMENTS

I would like to mention the helpful comments and suggestions of J. Géhéniau, B. Misra and G. Nicolis. During the redaction of this paper, I received the support of the Robert A. Welch Foundation (Houston, Texas), and the Institute for Constructive Capitalism (Austin, Texas).

REFERENCES

(1) Koyré, A. *From the Closed World to the Infinite Universe*, Baltimore, Johns Hopkins University Press (1957).

(2) Weisskopf, W.A. « Reflections on Uncertainty in Economics », *The Geneva Papers on Risk and Insurance, 9,* 33, 335-360.

(3) Toynbee, A. *A Study of History*, Oxford University Press, 1972.

(4) Popper, K. *Postscripts to the Logic of Scientific Discovery*, vol. II, *The Open Universe : An Argument for Undeterminism*, London, Hutchinson, 1982.

(5) See, e.g., Nicolis, G., and Prigogine, I. *Self-Organization in Nonequilibrium Systems*, New York, Wiley, 1977.

(6) Rössler, O., quoted in Nicolis G., and Prigogine, I. *Exploring Complexity*, Freeman, 1986.

(7) Nicolis, C. and G. *Nature 311* (1983) ; Nicolis, C. and G., *Proc. Natl. Acad. Sci.*, to appear.

(8) Babloyantz, A., Salazar, J.M., and Nicolis, C. submitted to *Physics letters.*
(9) Nicolis, G. and Prigogine, I. *Exploring Complexity*, Freeman, 1986.
(10) Nicolis, G. and Subba Rao, G. and J., personal communication.
(11) Géhéniau, J. and Prigogine, I., to appear in *Fund. of Physics* ; Géhéniau, J. and Prigogine, I., to appear in *Proc. Natl. Acad. Sci.* (1986)
(12) See, e.g., Gibbin, J. *In Search of the Big Bang*, Bantam, New York, 1986.
(13) Misra, B. and Prigogine, I., *Letters Math. Phys. 7* (1983) 421-429.
(14) Prigogine, I. *From Being to Becoming*, San Francisco, Freeman, 1979.
(15) Elskens, Y. and Prigogine, I., to appear in *Proc. Natl. Acad. Sci.* (1986).
(16) Prigogine, I. and Stengers, I. *Order out of Chaos*, New York, Bantam, 1984.
(17) Salk, J. *The Survival of the Wisest*, New York, Harper and Row, 1973.
(18) Toffler, A. *The Third Wave*, New York, William Morrow, 1980.

On the Concept of Chaos

Hao Bai-lin
Institute of Theoretical Physics, Academia Sinica, Beijing

I would like to make a few comments on the notion of chaos which has been mentioned repeatedly throughout this Symposium.

Firstly, the better understood part of chaotic dynamics is essentially classical. Quantum mechanics, in spite of its probabilistic interpretation, happens to be more deterministic. I do not think one needs quantum mechanics in the macroscopic description of most biological, ecological or economic systems. However, even within classical chaotic dynamics trivial or strange attractors appear only in the infinite time limit, i.e., in asymptotic or stationary regimes. Since all biological, ecological or economic systems exist only in finite time range, it seems necessary to emphasize the importance to study transient behaviour in chaotic dynamics. In the « phase space » of a complex system, besides asymptotic attractors there may exist many unstable and metastable objects. Transitions among these objects before the system settles down at some asymptotic attractor may give rise to a very rich variety of transient phenomena. Anyway, we ourselves are only transients with respect to the history, so do not ignore transient behaviour in chaotic dynamics.

Secondly, chaos is not simply disorder. It is more appropriate to consider chaos as a new kind of order without periodicity. Within generally chaotic regimes one can discover patterns of order interspersed with chaos at smaller and smaller scales, provided sufficiently high resolving power has been reached in numerical or laboratory experiments. In idealized models there are often infinite hierarchies of self-similar structures and a fractal geometry behind them. Although it may sound strange, « chaotic order » or « chaotic structure » must be considered as equal

members of the great family of dissipative structures. The mathematical manifestation of chaotic order consists just in the fractal geometry and the scaling invariance, ushering in the renormalization group ideas which have been proved so successful in studying critical phenomena in physics. However, speaking about fractals and their dimensionality, one should remember that here one deals with metric properties. While most orthodox mathematicians prefer topological dimensions, metric dimensions are much more flexible to reflect the physical world around us.

The last, but not the least comment concerns the role of chaotic dynamics to the reconceptualization of physics theory in general. Newtonian mechanics has long been established as the Plank constant h=O and the reciprocal light speed l/c =O limits of quantum and relativistic mechanics respectively. However, Newtonian mechanics may well be the Boltzmann constant k=O limit of the mechanics of complex systems. Unlike quantum and relativistic mechanics which were created by a few distinguished individuals, the formulation of a consistent mechanics of complex systems may require the joint effort of scientists from many disciplines, as represented by the participants of this Symposium. Professor I. Prigogine and his team have contributed so much towards this goal, but much more remain to be done. Perhaps, this is where to get yet another Nobel prize. It is a challenge for all of us.

Natural Order, Natural Laws and Correlated Events

Giorgio Careri
Dipartimento di Fisica, Università di Roma

1. INTRODUCTION

The aim of this conference is to find a common ground between the area governed by naturals law and the realm of those principles that regulate human behavior. To this end, my brief contribution will draw attention to a point I find interesting as well as pertinent, and that is, the way the very concept of natural law has changed. Determinism is giving way to probability ; mathematical relationships between predictable quantities are replaced by *statistical correlations between natural events.* Whereas this process of transformation has almost reached its conclusion in modern physics[1], it is only beginning in other fields of science, such as molecular biology. I am convinced that this broader outlook is the most suitable one to lead us to the common ground between the natural sciences and the humanities. Indeed, as we shall seen, several important aspects of modern Western philosophy provide significant and encouraging indications of progress in this very direction.

2. THE ORIGIN OF THE IDEA OF NATURAL LAW

Nowadays we take it for granted that experimental science began with Galileo. In the same way, it is agreed that natural order can be expressed only on the basis of analytical relationships between measured quantities ; it is these relationships that are usually called « natural laws » ». Actually, in many of his letters and finally in his *Dialogue*, Galileo — the first man to observe the heavens using a telescope — thanked God for the order he observed (1632 : « The world is highly ordered »). But Galileo's religious sense of the order in all things created never brought him to use the term « natural law », even in relation to his « law of falling

bodies ». The modern concept of physical order, meaning the presence of relationships between observed quantities, was clearly in his thoughts but the word « law » never appeared in his writing[2].

Kepler also steered clear of this term when describing his famous planetary laws, but then his interpretation of astronomical data was still quite close to the astrological outlook. Later, Stevin and Pascal proceeded in much the same way as Galileo : although both were entirely familiar with the concept, neither ever used the expression. The term « natural law » became explicit only in 1637, when Descartes published his *Discours de la Méthode*. For Descartes, the regularity observed in natural law could be traced back to the orders of the supreme architect ; accordingly, he called them « laws given by God to nature ». In any case, this notion of natural law was not widespread among seventeenth-century experimentalists, although one member of the Royal Society, Robert Hooke, did call his 1678 law on elastic bodies a « Rule or Law of Nature ».

The first extensive use in a scientific context of the notion that a law is an analytical expression of natural order did not appear until Newton. However, as Koyré observes[3], Newton's God was still the God of the Bible. So the same religious spirit was shared by Galileo, Descartes and Newton, the three seventeenth-century scientists who most contributed to the birth of the concept that natural order is grounded in « physical laws »[4].

Later, scientists began to move away from the religious feeling that linked nature as a divine creation to human behavior as a phenomenon conforming to the laws of Christianity. Scientific thought gradually came to be based only upon the presence of analytically expressed deterministic principles called « natural laws ». The fact that time was reversible in the equations of Newton's dynamics seems not to have disturbed this philosophical outlook. In the same way, the certain presence of chance in many natural occurrences never undermined deterministic mechanics until the end of the nineteenth century.

3. *PHYSICAL LAWS AS A CORRELATION OF OBSERVABLE EVENTS*

Let us consider modern physics, where this transformation from a deterministic to a probabilistic view linking observed events has occurred prevalently during the course of our century.

Physics attempts to provide a quantitative, self-contained description of observed events, based upon increasingly simple and general relationships. Thus physics does not claim to explain natural events ; it deals only with the regularities found in those natural events that can be experimentally observed. These regularities are the expression of an order that forms along with the progressive synthesis of these events, so that the initial observation of a specific regularity among a restricted group of measurements leads to the formulation of more general regularities which pertain to a broad class of physical events. Originally determinis-

tic, this system of relationships observed in nature may be called physical order. The objective, communicable quality of this learning process led to the common belief that such order exists in nature outside and independently of man, and thus is completely unrelated to human behavior.

After the turn of the century, physicists have come to realize that a natural law is essentially the expression of a correlation among a group of natural events observed in a system that has been correctly isolated from the rest of the universe. For instance, the law of falling bodies can be newly considered as the correlation of the successive positions of the falling body and the fractions of time in which those positions are measured. This regularity in the correlations among natural events expresses order from the view-point of classic mechanics. Diversely, quantum mechanics makes it possible to formulate the fundamental laws directly as correlations among successive observations of a single object. *Thus physical order can always be expressed in terms of correlations among events measured in quasi-isolated systems, and of successive relationships among these initial correlations.* According to Wigner[5], we may broaden this concept by saying that just as observed events are the raw materials for specific physical laws, these same laws are in turn the raw materials for general invariance principles. Today we know that an invariance principle equals a symmetry principle, just as it implies a general conservation law.

In all probability, the spontaneous natural event became the primary element of physics only when radioactivity was discovered at the end of the 1800s. Radioactive decay was quickly identified as a spontaneous natural event that could not be broken down to other more elementary causes and for which statistical laws could be accurately determined. But before long, at the turn of the century, Planck's quantum appeared on the scene, making it possible to define an observable elementary event in terms of energy. Almost simultaneously, the theory of relativity identified the event as the site change in space-time. Later, quantum mechanics focused upon « observables », proposing that the elementary event was the measurement itself.

Thus it is certain that both theoretical thought and laboratory experience simultaneously and progressively led to the acknowledgement of the event's leading role in modern physics.

The physical events considered above are statistically independent from one another. But there are also physical processes in which the single events are not statistically independent. One instance is the laser, where the events of light emission exert a determining influence upon one another, thus creating the coherence of the electromagnetic wave they produce. In such situations a new state of affairs is formed and *the new scale event that appears owes its origin exclusively to the correlation of elementary events*[6]. The laser is certainly not unique ; we now know many situations in which an open system can be driven by means of a cooperative transition among elementary events[7]. Because the mathematical relationships that describe these transitions are non-linear, it is still not clear how the system evolves through them. Other lecturers have treated the subject in greater depth at this conference.

4. *FUNCTIONAL ORDER AS A CORRELATION OF BIOCHEMICAL EVENTS*

At this point I would like to show how the concept of correlation among events, mentioned above in relation to the physical sciences, can be used to express the idea of order in the natural sciences as well. Of course this kind of generalization is feasible only when the event can be clearly identified as the measurable elementary change of the variable used to describe the natural process in question. In this case, *the regularities made clear by these correlations among events are the laws* that govern nature *and they are no longer deterministic.* Unfortunately, while this outlook is always correct in theory, at times practice may prove it sterile, thanks to multiple correlations among primary events. This leads to the conviction that biology is « too complex » to be described in physical terms. In camparison, molecular biology is less complicated ; I have made it the object of several studies, which I would like to outline here[8]. In keeping with the nature of this conference, I will only mention the most general aspects at this time.

In molecular biology, the living organism appears to be a system of structures that are involved in interconnected processes ; this system safeguards its internal coherence despite the continual exchange of products with the sourrounding environment. Obviously, when isolated, neither structrures nor processes are typical of the world of biology ; they certainly pertain to the realm of physics. But the unique feature of the biological organism is that this self-correlated system can generate other systems like itself by means of a *functional order made up of a long series of elementary events*, which together maintain the permanent nature of the biological species, over and above the constant variability of each individual within the species.

Let us take a closer look at functional order, to highlight a circumstance that even molecular biologists admit, which is that this type of order implies the presence of a group of relations among the events in the series[6]. Every event considered here is first and foremost a biochemical event which gives it biological meaning. At the same time, it is also a physical event because the natural transformation of matter is necessarily subject to physical laws. And *it is the ensemble of theser correlations among significant biochemical events that characterizes functional order*, in much the same way we observed above with regard to physical order. Nonetheless, it is still difficult to analytically express the correlations these events must meet in order to maintain the features of the species, inasmuch as these events must certainly be of a statistical nature at the molecular level, given the environmental thermal bath. In other words, while the general program of these events is pre-established in the material of the DNA, it can be carried out only randomly because at the molecular level it must adapt to the environment at a finite temperature.

Of course, organismic biological order is different from physical order, because in some physical systems it is actually possible to provide a detailed description for the entire hierarchy of correlations among events, whereas this is impossible in biology because the cell's complexity seems truly irreducible.

5. ORDER AND EVENTS IN MODERN WESTERN THOUGHT

At the turn of the century, Bergson firmly rejected deterministic outlook and all its implications. Alongside the mathematical order of inanimate matter, Bergson proposed a second, distinct kind of order that pertains to living matter. Clearly, this kind of order cannot be subject to Newton's concept of time, with its connotations of reversibility and predictability. Thus chance becomes an indispensable ingredient, intrinsically linked to the order of living matter, because the latter order is unpredictable, given the complex phenomena involved. As a result, according to this outlook, the evolution of living matter gives rise to an irreversible series of correlated events that characterize an authentic inner « time » which is real and experienced. The highest example of this « inner time » is the human consciousness.

Whitehead immediately adopted Bergson's outlook for his own organistic philosophy. Whitehead's point of departure is the fact that the natural event is the primary element of perception, and that man experiences nature as a correlated sequence of perceived events. Thus space and time are characterized only as permanent relationships between the events perceived ; material objects are no more than our own way of expressing certain reciprocal correlations among the physical features of the events perceived. Thus all of nature, whether inside or outside man, is seen as an initial series of events, and it is their flow which gives birth to permanent entities, not vice-versa. Thus it is clear that deterministic order pertains only to certain events which fall into a fairly isolated framework. With Whitehead, the event reaches its primary role in philosophic thought. So it is that Wittgenstein begins his *Tractatus*, « The world is all that occurs. The world is the totality of events, not things. »

Nevertheless, it is not until recently, with Michel Foucault, that we find a systematic philosophic analysis based upon *the concepts of event, series of events, the characteristics of regularity in these series and finally, the possibilities of these characteristics in environmental conditions*, i.e., the same concepts we used to describe functional order in molecular biology. Foucault decries the scarce attention Western philosophy has paid to these four concepts, upon which all analysis of philosophic reasoning must be based. He chose the project for his own work, which was tragically interrupted as a result of his premature death. Foucault may have been the first to be truly aware of the order present in all that is visible and invisible, that is characterized by the presence of correlations among things, and that acquires meaning only inasmuch as the observer mentally organizes it into a series of correlated perceived events. Thus in acknowledgement, as well as to further explain all that I have tried to illustrate with this brief contribution, I refer you to Foucault's original thoughts on order, as stated in the Introduction to his *Les mots et les choses* : « Order is, at one and the same time, that which is given in things as their inner law, the hidden network that determines the way they confront one another, and also that wich has no existence except in the grid created by a glance, an examination, a language ; and it is only in the blank spaces of this grid that order mani-

fests itself in depth as though already there, waiting in silence for the moment of its expression. »

BIBLIOGRAPHY

(1) *Stochastic Behavior in Classical and Quantum Hamiltonian Systems*, Volta Memorial Conference, Como 1977, Springer Verlag, Berlin, 1979.
(2) E. ZILSEL, *The Phil. Rev., 60*, 245 (1942).
(3) A. KOYRÉ, *From the Closed World to the Infinite Universe*, Johns Hopkins Press, Baltimore, 1957.
(4) G. CARERI, The Lincei Academy and the Self-Organization of the Early Scientific Societies. The Parallel Genesis of the Concept of Natural Order, in S. Fox, ed., *Self-Organization, Principles in Science, Events in Society*, Adenine Press, NY, 1985.
(5) E.P. WIGNER, *Symmetries and Reflections*, MIT Press, Cambridge, Mass., 1970.
(6) G. CARERI, *Order and Disorder in Matter*, Benjamin/Cummings Pub. Co., Menlo Park, Cal., 1984.
(7) I. PRIGOGINE and L. NICOLIS, *Self Organization in Non-Equilibrium Systems*, Wyley, New York, 1977.
(8) G. CARERI, P. FASELLA and E. GRATTON, Enzyme Dynamics : The Statistical Physics Approach, *Ann. Rev. of Biophysics and Bio-engineering*, 1979.

Modelling Complexity : the Dynamics of Discovery and Exploitation in a Fisheries Example

Peter M. Allen
Université Libre de Bruxelles

Jacqueline M. McGlade
Marine Fish Division, Bedford Institute of Oceanography, Dartmouth

1. MODELLING COMPLEX HUMAN SYSTEMS : A FISHERIES EXAMPLE

Fishery management at present is based on a mixture of theoretical concepts and practical considerations which in fact do not result in a clear basis for policy, nor in any self- evident criteria for management decisions. The theoretical approaches have two fundamental aspects : the biology of fish populations, and the theories of economics. The former is concerned with calculating the growth characteristics of individual fish, assessing how fast weight is put on, when spawing occurs etc. and also with establishing the relevant parameters of a particular species populations dynamics. In practice, however, instead of modelling the interactive processes of the different species inhabiting a given ecosystem, the equations discussed are for each species separately, with the effects of other species, of food abundance, and indeed of fishermen, merely represented by exogenous parameters. In addition, instead of exploring the « dynamics » of such systems, the emphasis has been on the *equilibrium or static* solutions of such equations. Initially, this sprang from the laudable desire to look at « long term yield » of a fishery, rather than the immediate « profit ». However, by studying the equilibrium solution of deterministic equations, we immediately rule out the possibility that the long term situation may be *dynamic* — that a fish population may always change and never attain a stationary state. These

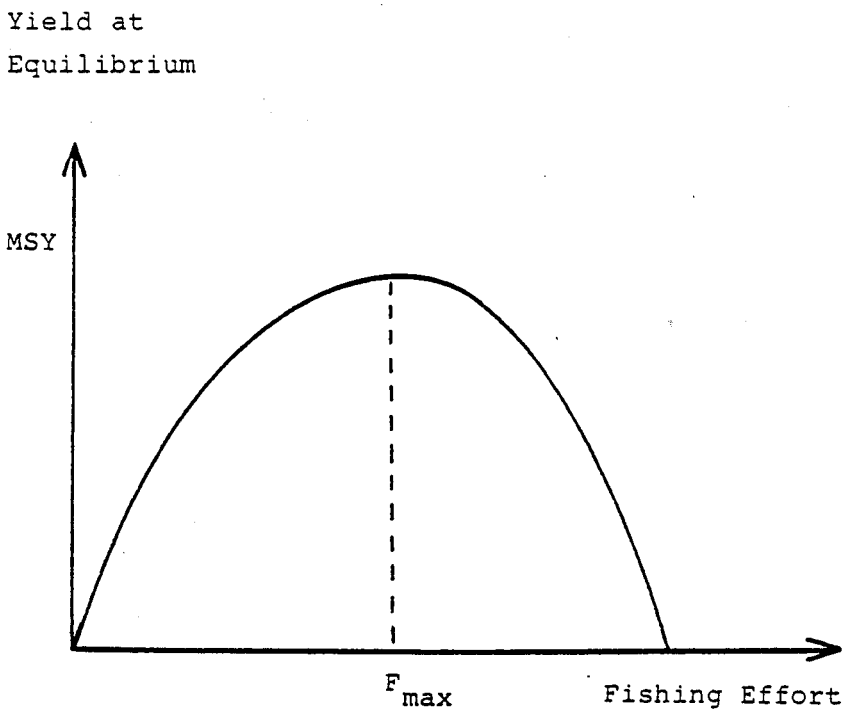


Figure 1

This type of graph showing the equilibrium yield or catch for a given level of fishing effort has been the basis of much of fishery management. Maximum Sustainable Yield (MSY, which occurs at F_{max}), gives the fishing effort needed for the maximum possible yield of the fishery.

notions of determinism and of equilibrium underly the concept of « maximum sustainable yield » that has dominated much of fishery management thinking for many years (Figure 1). Indeed, so deeply ingrained were the ideas of equilibrium, that even when a « dynamic population » model was developed by Beverton and Holt, in 1955, it still retained within it concepts which only have a clear interpretation at equilibrium. For example, the idea of calculating the « yield per recruit » which is fundamental to their model, means that we study the average weight of catch which will be made from each young fish entering the system. Clearly, this only has a sense if the system is in a static, equilibrium state for at least the average life time of a fish, and this can often be 8 or 12 years. In reality, fish populations, fishing effort, technology and even markets all change over such a time scale, and in addition, if the model is used to make management decisions, then clearly these must also disturb the system and invalidate the model on which they are based.

Some of these shortcomings were highlighted by various authors (Beddington, 1979 ; Beddington and May, 1977 ; May et al., 1979) who examined the effect of fluctuations of the parameters in the basic equations.

They showed that exploitation at or near the « maximum sustainable yield » would in fact lead to collapse of the fish population, and of the fishery. These ideas, together with the practical experience of decimated fisheries, led to exploitation which aimed at a level somewhat below that of the « maximum sustainable yield ». Nevertheless the fundamental idea of basing the biological evaluation on the equilibrium state of a single species equation, still persists.

However, if there are criticisms to be made of the biological basis for fishery assessment, the economic calculations seem even less well founded. The basic ideas underlying both the biological and the economic calculations are illustrated in figure 2 for a simple case (the logistic equation). Essentially, the economic calculation supposes not only the equilibrium state of the system, but also perfect information on the part of fishermen as to the possible « return » on fishing effort of a particular kind. In addition it supposes that the relative advantages of fishing can be weighed by each fisherman, or firm, as opposed to the advantages of investing in other economic activities. Under these assumptions, a line can be drawn as in figure 2 showing the cost to the « fishing industry » of a certain scale of effort. If it is assumed to be linear as a first approximation, then we see that the yield on a given effort by the « industry » goes through a maximum, and then decreases. When it crosses the yield curve it is assumed that the profit being made by *each* firm or fisherman, is « normal », that is, is equal to that obtainable from other activities. In a system in which firms and individuals are free to join the system, or to increase their efforts (a « free- entry » regime) then it is assumed that the system must be operating at the point P.

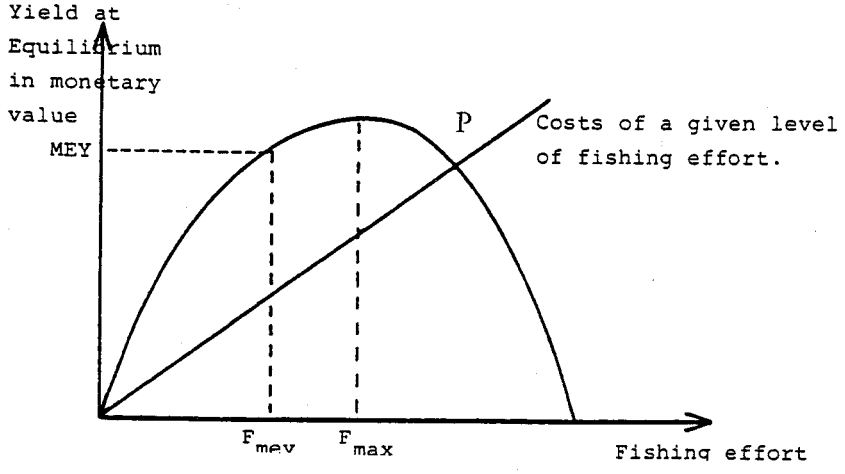


Figure 2

Here, economic theory then becomes somewhat schizophrenic because on one hand the *general theory of equilibrium* would suggest that maximum economic benefit for the whole economy would occur if the system operated at this point, P, but, equally, it would maintain that maximum economic benefit would occur at the point of *maximum eco-*

nomic yield for the fishing industry. That is the point MEY. Management policy would be different in the two cases. In the first, a free entry system would be maintained, but some limit on total catch would be imposed in order to stay below « maximum sustainable yield ». In the second case, however, there would be a system of « licences » which would be used to restrict entry to the fishery, and attempt to set the total effort, or catch at around MEY.

Unfortunately for either of these, the whole calculation and theory is based on quite untestable, and probably untenable assumptions. Not only that, it denies the fact that each action taken at one time will in fact influence the « yield » at later times, and therefore that the relation between « investment » and « profit » not only has a strong dynamic evolution, but also depends on what each of the other actors chooses to do. In other words it is inherently impossible to have « perfect information », and therefore to make perfectly « rational » decisions.

Another feature which we shall discuss further later is in the assumptions about the « price » of fish. In fact, the price will be a function of the demand and supply of fish, and hence will be affected by which policies are adopted for the industry. As we shall see, it may perfectly well be possible to have two distinct equilibria — one corresponding to fairly abundant, normally priced fish, and the other to very low catches of very rare, and highly priced, fish. It may even be true that the employment and profit made by the industry is greater in the second case, but common sense tells us that this is not necessarily the « best » manner in which the natural marine resources could be exploited for the benefit of the community as a whole. Of course, one then falls upon the totally impossible task of defining ones aim for management. This is generally not done for the simple reason that there is no indisputable definition. Management goals therefore tend to arise by tradition, and by reacting to crises as they occur.

In the case of the Canadian Fisheries which we have studied there is in fact a policy document that is an honest attempt to state government policy. This is more than exists for most other examples, but it serves in reality to show how unclear the aims of management are. In that document we find phrases like « ... the *best use* of fisheries resources... to obtain *optimal* socio-economic benefits. » « ... to maintain fishery resources at levels which will generate *maximum continuing* economic and social benefits... » « ... to create viable and stable commercial fisheries and *improved incomes* through the regulation of harvest levels... » Many questions spring to mind : what does « optimal » mean ? If we knew, then for whom should we try to attain « optimality » ? For Canada ? For the province immediately concerned ? For consumers ? For the local industry ? For employment ? Also, even if we define these aims, since each action affects the future yield, then over what time period should we consider the particular « optimality » chosen ? Now, the models which we shall discuss below do not answer these questions. However, by modelling the sytem more globally, and by rejecting the simplistic and unrealistic assumptions of economic theory, our models allow a clearer discussion of what aims are possible, and which are com-

patible with each other. In short, they highlight the fact that a compromise solution is inevitable, and enable a clearer discussion of possible compromises than can be afforded by traditional models.

2. A SIMPLE, DYNAMIC NON-SPATIAL FISHING MODEL

In this section we shall briefly describe a single species, non-spatial dynamic model which allows us to show more clearly the effects of natural fluctuations on both the fish populations and on the fishing effort exerted by fishermen. Also, we shall take into account the fact that price is not determined exogenously, but instead depends on supply and demand, and on the cultural identity of the consumer population and their response to changed prices.

In order to construct our model we have studied the data which is available for the Nova Scotia groundfish fisheries, which are mainly concerned with the catch of three species : haddock, cod and pollock. If we look at the simple direct catch data for the period 1965 to 1981, then we are forced to admit that « equilibrium » is not a word that immediately springs to mind. If we smooth out the yearly cycles of seasonal fishing then we perceive very large underlying trends. The question of course is : can we model this behaviour ? What this means is that we must attempt to find some explanatory mechanism which will « generate » these curves as an output. The most obvious mechanism is that of the interaction between fishermen and fish, an interaction which essentially is that of a predator and its prey. Now, these types of equations are very well known, and are called Volterra Lotka equations. These predator prey equations are characterized by a causal relationship between prey population density and the growth of predator density. In the case of fishing, that would correspond to a mechanism which would relate an increase in fishing effort (number of vessels and time spent fishing) to high fish density, which would be signalled by a high « catch rate ».

Now if we study the evolution over time of catch rate and fishing effort for the three fish species, then we see that only in the case of haddock is there a logical relationship. For the 17 years of data, only for haddock did effort increase in response to high « return » or « catch rate », and decrease when the catch rate fell. For cod and pollock however, there is no discernable relation of this type, with effort changing in an apparently unaccountable fashion. However, this particular fishery (the 4x zone) serves the consumers of Nova Scotia, and they have a cultural preference for haddock. Hence, haddock is more demanded, more fished, and has a higher price. In fact the relative prices of haddock, cod and pollock are 3 : 2 : 1. When we take into account the fixed costs involved in any fishing expedition therefore, we see that it is very much more attractive to fishermen to fill their boats with haddock than with the other species. In this way, there is a causal relationship of the predator prey type between haddock and local fishermen, but the other species are only fished if haddock is difficult to obtain. In other words the causality in the system includes price. But price, as we shall see, is determined partly by the endogenous dynamical system.

Because of this lack of mechanism for cod and pollock, we have therefore modelled the haddock fishery of 4x as a first step. Our equations describe the changes in five variables. The first three are the populations of young fish (1 and 2 year olds), of medium fish (3 and 4 year olds) and of old fish (5 and over). The fourth equation describes how fishing effort responds to the revenue and profit obtainable by fishing haddock in 4x, and the fifth equation takes into account the change in price of fish as supply, demand and price change.

Our equations are :

$$\frac{dx_1}{dt} = bx_3\left(1 - \frac{TX}{N}\right) - mx_1 - s_1x_1y - \frac{x_1}{T_1}$$

$$\frac{dx_2}{dt} = \frac{x_1}{T_1} - mx_2 - \frac{x_2}{T_2} - s_2x_2y \qquad (1)$$

$$\frac{dx_3}{dt} = \frac{x_2}{T_2} - mx_3 - s_3x_3y$$

$$\frac{dy}{dt} = ry\left(\frac{p(w_1s_1x_1 + w_2s_2x_2 + w_3s_3x_3)}{C + pr} - 1\right)$$

$$\frac{dp}{dt} = rrp\left(\frac{MA}{(P + \varepsilon P^4)^e\,(w_1s_1x_1 + w_2s_2x_2 + w_3s_3x_3)\,y} - 1\right)$$

were, w1, w2, w3 = weight at age

b = rate of production of x1, per adult.
TX = total fish biomass
m = natural mortality
T1, T2 = time span of groups 1 and 2
y = number of boat/effort units
c = cost of fishing for each boat
pr = profit desired in order to maintain effort
r = rate of response of effort
rr = rate of response of price
MA = market per unit time at unit price
p = price
e = elasticity of demand.
s_{123} = catchability of x_1, x_2, x_3.

If we study the behaviour of these equations for the simplest possible situation where there is no price mechanism, then we find that there is a single stationary state, and that the factor (1 — TX/N) ensures that it is stable. Figure 3 shows a typical deterministic trajectory of the system in a damped oscillatory motion towards the stable stationary state. However, it should be noted that the characteristic time in which the system goes to equilibrium is greater than 40 years for any reasonable values of the parameters.

However, the really interesting effect occurs when we add to these equations a term which takes into account the fact that *in reality* the birth rate b, of haddock is extremely variable from year to year. In order to represent this we use a random number each year to decide whether it

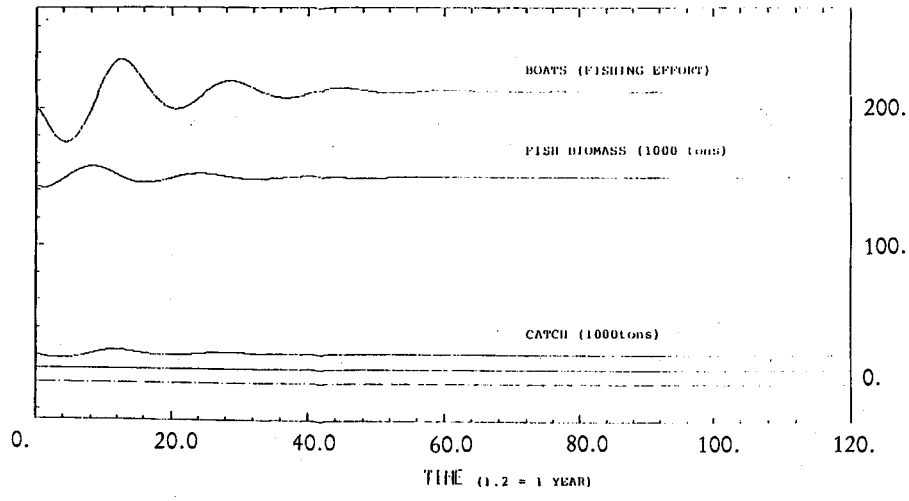


Figure 3

The model of Equations (1) run deterministically

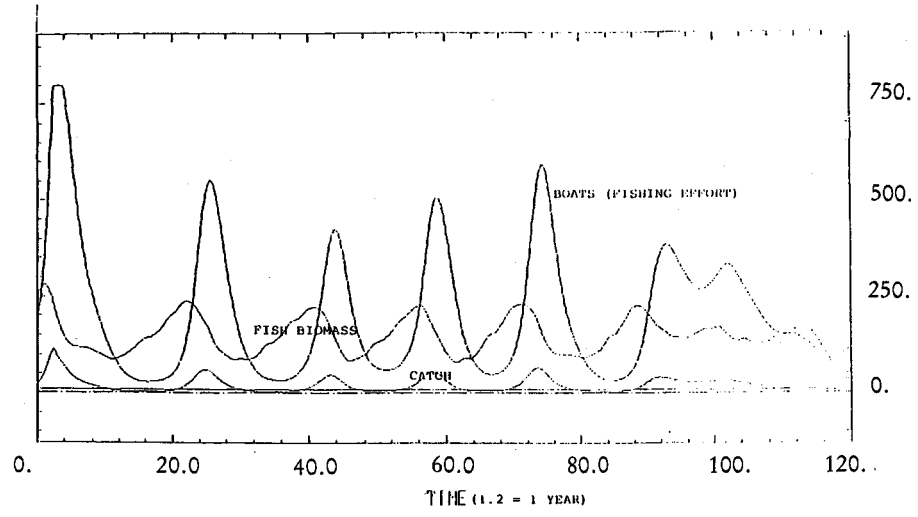


Figure 4

Exactly the same model and parameter values, but with « b » fluctuating

will be a « good » year — b=2b, a « medium » year, b=b, or a « bad » year, b=0. This ensures that on average, over a long period, the average value of b is maintained, but that in each particular year we may find either high, medium or very low rates.

The result is dramatic. In figure 4 we see how the system *amplifies* these random fluctuations in b and sets itself into relatively violent, irregular oscillations. Let us underline here that the fluctuations which we have proposed for the haddock fishery are considered by fishery experts to be conservative, if anything, compared to reality. Turning off the fluctuations would give rise to a stationary state after some 60 or 70 years, but clearly, the answer to the question, « Are cyclic crises in fisheries of natural or man- made origin ? » we may suggest a new answer. Both. The human responses can amplify relatively rapid fluctuations and exite a Volterra-Lotka type oscillation. However, without this natural « background » noise, there would be nothing to amplify, and the fishery would be stable.

We can study the evolution of the system in terms of the « probability » of finding it whith a particular value of fish and boat population. In order to do this, we therefore plot each time step on the phase space (the space of the possible boat and fish populations) of the system, and then accumulate the number of phase points in each small region, giving rise to a « mountain » in the third dimension which shows us the regions of phase space in which the system spent most time. For example, if we run our system for 500 years of fishing, with different values or « r », the rate of response of the fishermen, we find the figures 5, 6 and 7. These show us that increasing « r », makes the system less and less stable, and leads to more violent oscillations in the size of the fishing fleet and the fish stocks.

However, if we ask what kind of *behaviour* can *invade* the system, then we find that fleets with larger values of r will always replace those with smaller values. Similarly, higher values of « s », corresponding to better technology, will also « invade », as will lower values of « c ». All of these are « economically rational » in the short term, and yet all of them lead in the long term to lower yields, and to a much more unstable system.

In any real system, though, this model without a price mechanism will be unrealistic. If, therefore, we now study our equations with the parameter rr, of price response, not equal to 0, then our model will simulate the situation where fish stocks, fishing effort, profits, demand and price are all determined dynamically within the system.

One very critical factor, however, turns out to be the « elasticity of demand ». If demand should be somewhat inelastic at around the normal price, 1, then suddenly a completely new effect can appear. This is shown in figures 8 and 9, where we see that the system has « discovered » a second « basin » of attraction. It correponds to a regime of very low fish stocks, very large oscillations of fishing effort, and to very high priced fish. Depending on the parameter values, the two basins of attraction can vary in their depth, and in their nearness to one another. Howe-

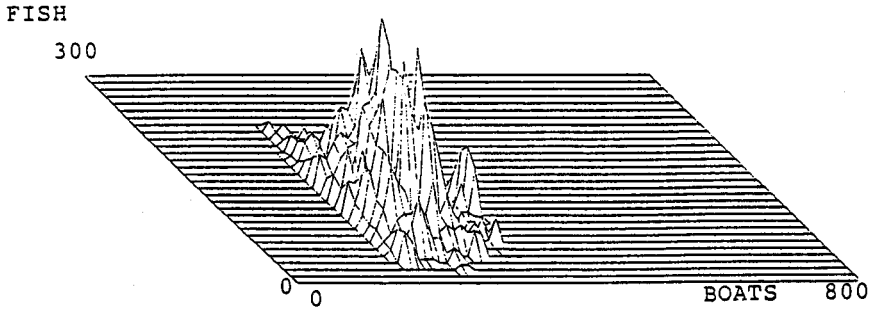


Figure 5

The probability of finding values of x and y if the model is run for 1000 years. Here r = .02, a slow rate of response.

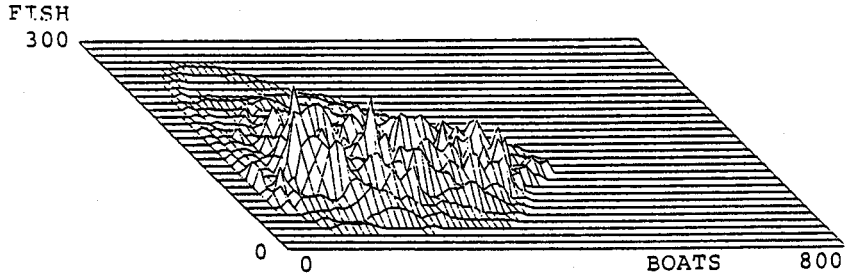


Figure 6

For a faster rate of response of effort to return, r = .5, we observe a much greater spread of values for x and y over 100 year period.

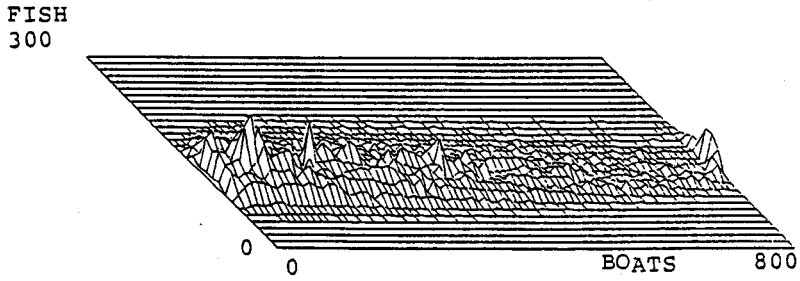


Figure 7

For a fast rate of response, r = 2, corresponding to the ability to double effort in 6 months, we see a very flat distribution. The deterministic solution will almost never be observed, and enormous oscillations will occur.

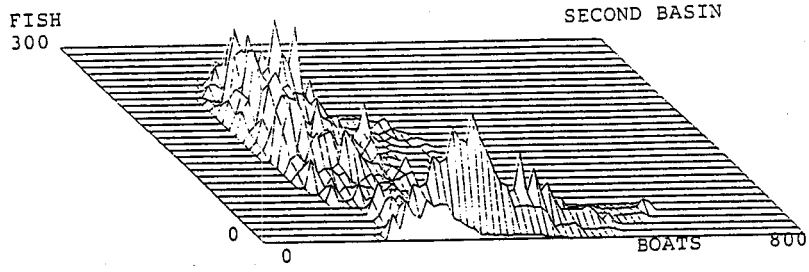


Figure 8

For an elasticity of demand of .5, a second basin of oscillation can appear corresponding to very low catches of very high priced fish. The size of the fishing industry and the « value added » may be larger than in the previous cases of « normal » fishing.

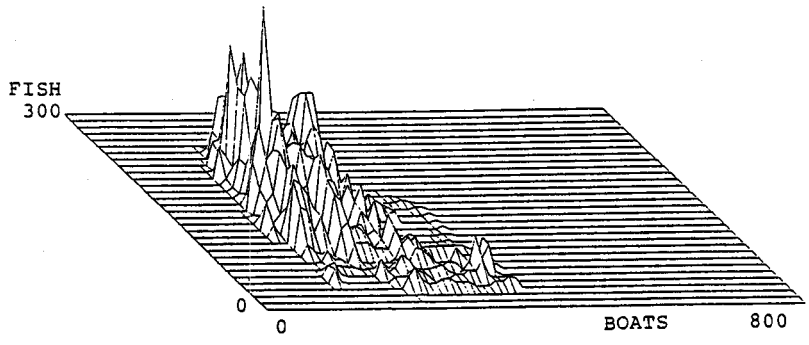


Figure 9

For identical parameter values as in figure 8 it is possible that in 1000 years the system, by chance, does not « discover » the second basin. In this case the model may suggest possibilities which have never been observed even for very long time series of data. Also, we must face the possibility that a fishery could oscillate around a quite « false » equilibrium for a very long period.

ver, the important point of principle remains that a system can oscillate in one, or the other for times of the order of 500 years, giving no indication of the existence of the « other » regime. It can also flip flop in and out of the basins in an unpredicable manner and over a relatively short time. What it shows us then is that for the same fish, the same fishermen, the same price mechanism and the same demand for fish, the *system* can operate in two qualitatively different modes, which offer quite different satisfactions to the different actors of the system.

In fact, running the model without the environmental fluctuations shows us that deterministically, in most cases, there are two possible equilibrium states. Depending on the precise initial conditions, the system will fall into one or the other basins and remain there. Clearly, the whole idea that « economic equilibrium » represents a « natural optimum » of some kind, falls by the wayside, when we see how two quite different

possible states exist. The point is that one basin corresponds to a balance between the fishing effort and the revenue from the catch at a fairly normal price. The other one, corresponds to that of fishing effort with the demand for very high priced, and rare fish. It is the precise manner in which the demand falls off with price that affects the characteristics of this basin.

However, we see the fundamental paradox that the « value » of fish can increase faster than they become scarce, and that therefore, as in our examples, fishing employment and boat numbers can be larger for the system with very few fish on average, simply because their is still some demand at the much higher price. That is, since « value » is not an « objective » quantity, but depends on the needs of people, the scarcity of the product and the existence or not of substitutes, we can arrive at the conclusion that the optimum exploitation of the system, the one that creates most « value », is this regime of « scarcity ». This is clearly in contradiction with any common sense interpretation of the « best possible » exploitation of natural resources.

A further point arise in addition to this. It seems from our simulations that there are cases, such as in figure 10, when the system under fluctuating conditions still explores the two basins, but where there are not in fact two equilibrium solutions to the deterministic problem. This could occur if there were a region of « slow » dynamics. Then everytime the system entered that region it would spend a lot of time there. This would mean that the probability of finding the system there, in a sampling, was large. However, there may not be a deterministic *stationary* steady state which corresponds to it. In that case, the deterministic equations would give no hint, from their solutions that the system would in general be found far from that state.

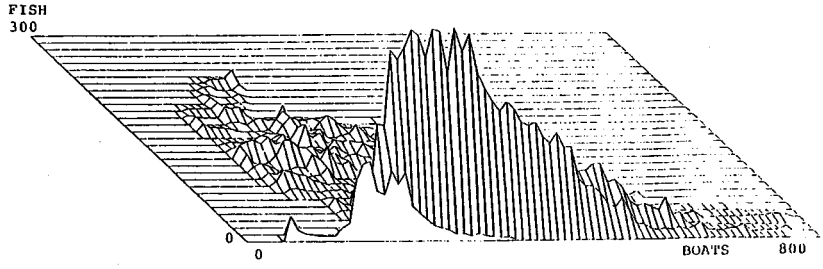


Figure 10

In this case we see that for 1000 years ofrunning the model it was predominantly to be found in the « second basin ». However, a carreful search with the deterministic model failed to reveal a second stationary solution of the equations (1). In this case the second basin seems to correspond to a region of « slow dynamics ».

In fact, we can use our model to run through the « evolution » of a typical fishery. It will start off with low technology, subsistence fishing for just the local community. At some point a price mechanism or an equivalent will appear in order to adjust the demand and supply. Once this appears, then in a « young » fishery which is hardly exploited the « price » of fish is very low. This leads naturally to an expansion of the

« market », and to greater demand for fish. Competition then leads to higher technology, lower costs, faster reactions and more powerful boats. Although at first these tend to hold down cost increases, as fish become more difficult to catch, so price increases. At some point, if free competition persists, the system will suffer collapse either of the entire industry, or into the « high price » basin where fish are a rare « luxury », supporting perhaps a thriving industry. Clearly, our model allows a much better picture of how the fisheries have evolved, and also of the real affects of different policy options. The important point being that it is not necessarily the behaviour observed in the system at present which will be important after the implementation of some policy. Perhaps other, subtle and selfish behaviours may invade the system, and thwart the management plans. As we stressed in the first section, it is very important to take into account not just what is *in* the system, but what is not (yet) there.

An important point resulting from this discussion is that our model, which is giving such complex behaviour is in fact only a non-spatial model for a single species of fish. As we said, if we wanted to apply it to a species other than Haddock, in this region, then we would not be able to, because the « causality » relating fishing effort of a species to the species density only exists for the highest price fish. In the next section, therefore, we shall describe a multi- species spatial model where fishing behaviour can be modelled much more correctly, and where the fisherman's knowledge and his information play a vital role.

3. A MULTI- SPECIES, SPATIAL DYNAMIC FISHING MODEL

In this section we shall describe a more complete model which attempts to take into account considerably more of the complexity of the real world than was possible in our previous model. In particular, this model will focus on the behaviour of fishermen, and the manner in which they make decisions about where, and what to fish. It must take into account the information which fishermen have concerning fish abundances, and the « expected revenue » that could be obtained from fishing a particular zone. This is by no means a trivial matter because essentially this information comes from the fishing activity of other boats, and therefore there is a tendency for the pattern of fishing to structure spatially because high catches result in a concentration of effort and, in consequence, high catches in that zone. Areas where fishing boats are absent send no information about potential catch and revenue. This provides a spatial positive feedback mechanism which structures fishing patterns.

Because our main focus of interest here is the behaviour of fishermen, we have used a very simple equation to describe the population dynamics of the fish populations. Of course, in the event of a real application of the model to a fishery, it would be necessary to put greater detail concerning the growth of each fish, and also the interactions between species and with the marine environment. However, here we have simply used a « logistic » equation for each species, and we have assumed that there is competition between them. This is based on the fact that the total mass of fish stocks remains remarkable constant, while the mass of each

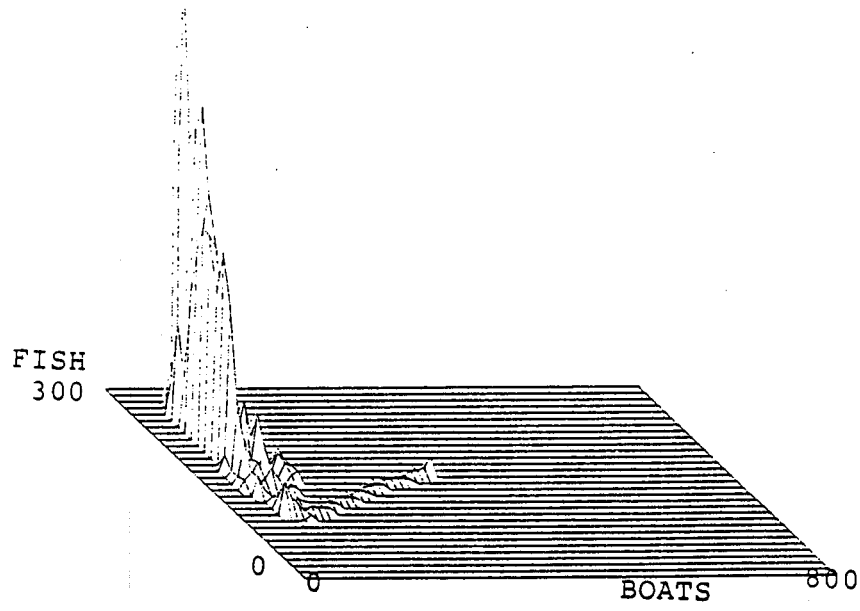


Figure 11

A « young » fishery. There is only a small market demand (10ktons at price 1) and technology is low. The average price of fish is .61, and this will attract more population.

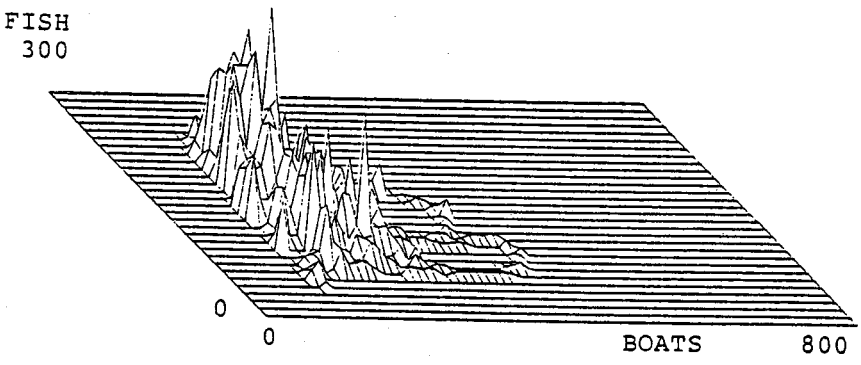


Figure 12

The fishery is « maturing ». Market demand has risen to 15ktons at price 1, and technology has increased. The system is less stable and there are less fish in the sea. Average price = .8.

species may in fact vary considerably. However, for haddock, cod and pollock the interaction is not as strong as we have supposed.

The second equation is that governing the spatial distribution of fishing boats, or effort. It has two essential terms. The first takes into account the increase or decrease of fishing effort in zone i, by fleet L, according to how profitable it is. If the catch rate (fish abundance) is high for a species of high value, then revenue is greatly exceeding the costs incurred

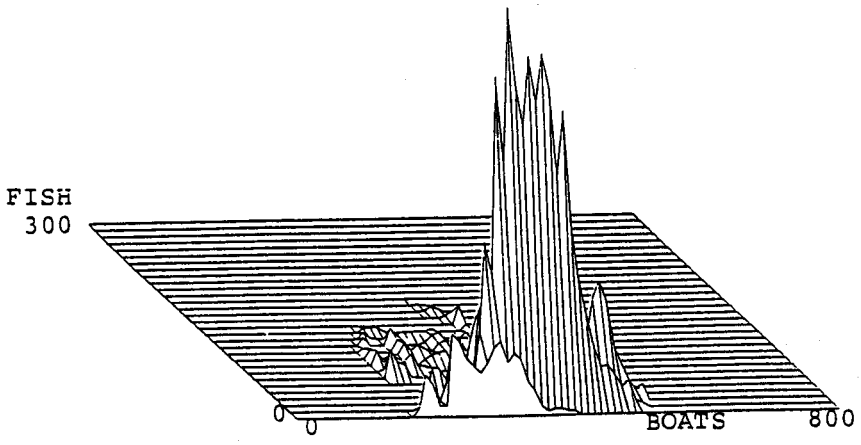


Figure 13

A « mature » fishery. The system has collapsed into the second basin. Catch has fallen from 17ktons of figure 12 to 5ktons. But prices have risen to 7.2, for a market of 29ktons at price 1. Greater « need » leads to less supply.

in fishing there, provided that the zone is not too distant from the port. Then effort will increase. If the opposite is true then effort will decrease. The second term takes into account the fact that due to information flows in the system (radio communication, conversations in port bars, spying etc.) to a certain extent each fleet is aware of the catches being made by others. Of course, boats withing the same fleet communicate freely the best locations, but even between fleets there is always some « leakage » of information. This results in the spatial migration of boats according to the « expected profit » that may be made at the different locations. Let us first write down the equations and then continue discussing how it is possible to capture these effects.

Our equations are :

$$\frac{dx_i^k}{dt} = b^k x_i^k (1 - t x_i^k / N_i) - \sum_L \frac{{}^L s x_i^k \, y_i^L}{\left(1 + s^L \tau \sum_k x_i^k\right)} - m^k x_i^k$$

$$\frac{dy_i^L}{dt} = \xi \left(\sum_i \frac{y_j^L A_{ij}^L}{\sum_{i'} A_{i'j}^L} - y_i^L \right) + r y_i^L \left(1 - (C_{i+pr}) \Big/ \sum_k \frac{{}^L s x_i^k}{\left(1 + s^L \tau \sum_k x_i^k\right)} \right) \quad (2)$$

Where b^k = the birth rate of fish type k
tx_j^k = the total fish biomass of each zone
N_j = the total carrying capacity of each zone
m^k = the natural mortality rate
s^L = the encounter rate of fish and boats
$1/\tau$ = the maximum catch rate
r = the rate of response of effort
p^k = price per kg of fish k
c_i = cost per unit time of fishing at i
pr = desired profit level
A_{ij}^L d the attractivity of i viewed from j, by L
ξ = the mobility of the spatial migration.

Now, the really new feature in this equation as compared to our previous model ( ), is the spatial dimension and the way that the pattern of fishing shifts according to the « attractivity » of the different zones. What factors are taken into account in the « attractivity » ?

But how precisely is this « attractivity » constructed, and what factors are taken into account ?

For these terms we use the idea of « boundedly rational » decision makers, who do not have « perfect » information or absolutely « rational » decision making capacity. In other words, each individual has a *probability* of being attracted to zone i, say, given by,

$$P(i) = A_i/\Sigma A_i$$

where A_i is the attractivity of zone i. Since probability must vary between 0 and l, we see that A_i must always be defined as positive, and a convenient form is :

$$A_i = e^{IU_i} \qquad (3)$$

where U_i is a « utility function ». U_i constitutes the « expected profit » of the zone i, taking into account the revenue from the « expected catch » and the costs of crew, boat and fuel etc. that must be expended to get it.

The particular form used for U_i is :

$$U_i^L = \alpha_i \sum_{L'} \varepsilon\,(L,L') \sum_k \left[\frac{s^{L'} x_i^k y^{L'}_i}{(1 + \sum_k s^{L'} x_{k'} \tau)} \cdot \frac{}{(1 + 0.3\, y_I^{L'})} \right] - \alpha_2 \phi\, d_{ij} - \alpha_3 \phi\, d_{ip}$$

where E (L,L') = the information exchange matrix between fleets L and L^1

ϕ = the fuel cost per unit distance

d_{ij} = the distance from i to j

d_{ip} = the distance from i to port.

In this way we see that boats are attracted to zones where high catches and catch rates are occurring, but the information only passes if there is communication between the boats in i and in j. This will depend on the « information exchange » matrix, which will express whether there is cooperation, spying or indifference between the different fleets. However, responses in general will be tempered by the distances involved, and the cost of fuel.

In addition to these effects, however, there is another very important factor which our equation takes into account. That is the factor I in equation (3). This factor expresses how « rationally », how « homogeneously » or with what probability a particular skipper will respond to the information he is receiving. For example, if I is small, then whatever the « real » attraction of a zone i, as expressed in U, the probability of going to any zone is roughly the same. In other words, « information » is largely disregarded, and movement is « random ». We have called this type of skipper a *stochast*. Alternatively, if I is large, then it means that even the smallest difference in the attraction of several zones will result in each skipper going, with probability 1, to the most attractive zone.

In other words, such deciders put complete faith in the information they have, and do not take the « risk » of moving outside of what they *know*. These « ultra rationalists » we have called *cartesians*.

The movement of the boats around the system is generated by the difference at a given time between the number of boats that would like to be in each zone, compared to the number that actually are there. There is only a finite rate at which these transfers can occur, however, and as the boats congregate in particular locations of high catch, so they fish out the fish population that originally attracted them. They then move on the next zone which attracts them, and in this way there is a continuing dynamic evolution of fish populations and of the pattern of fishing effort. In addition, the total effort being deployed in the system depends on the profits being made there, and this in turn requires that a fleet locates the fish aggregates and concentrates its effort in the best places. In this way, as the fish stocks change in size, and as the knowledge that the fishermen have concerning the location of fish aggregates changes, so the profit and fleet sizes change.

Let us describe briefly some of the preliminary computer runs that we have made for this model.

In the first figure we show the evolution of the spatial distribution of fish and fishing effort when we have a two species of fish, and a single fleet, characterized by an I of 1.

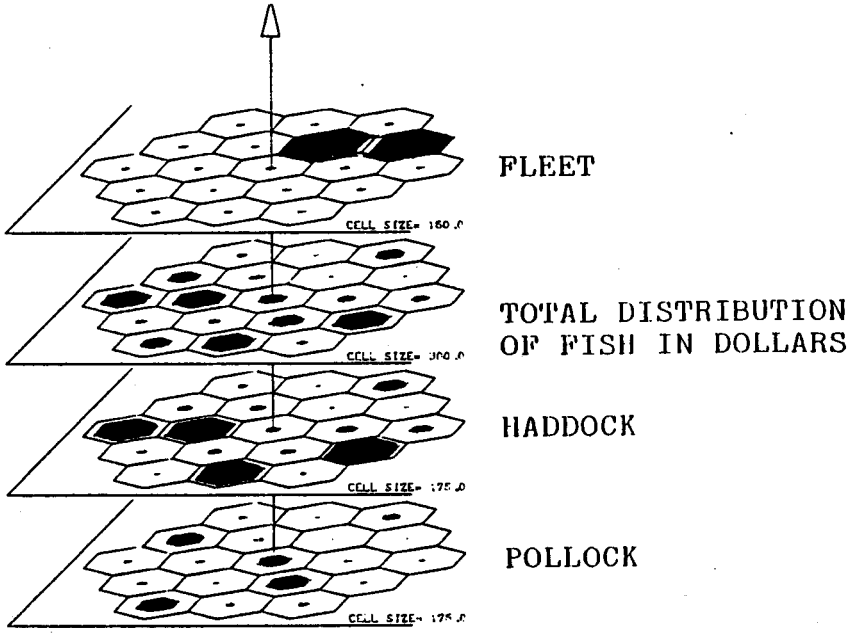


Figure 14

The initial distribution for our simulations of the evolution of a single fleet (cartesians or stochasts) and 2 fish species (Pollock and Haddock).

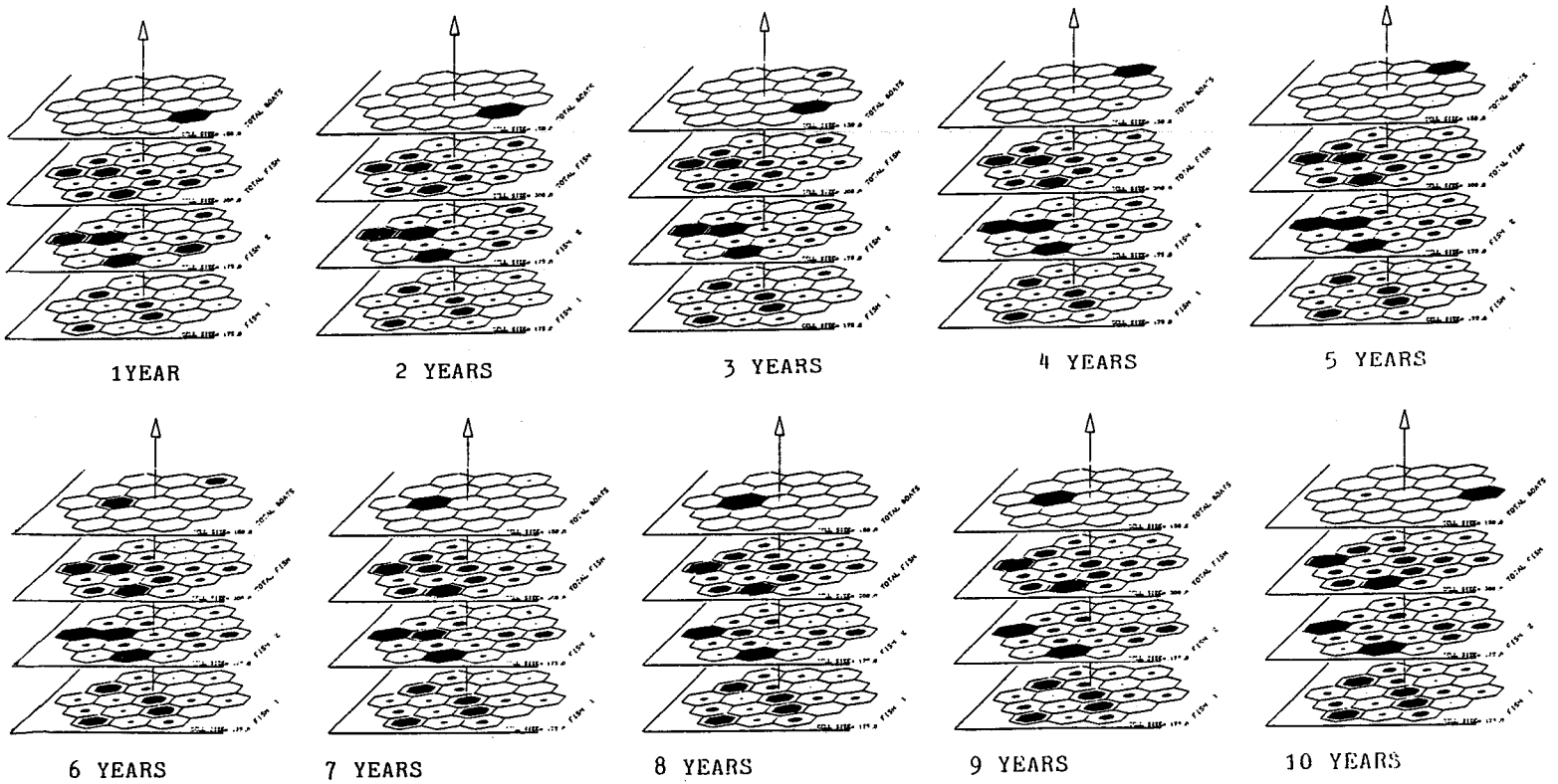


Figure (15). The spatial evolution of the distributions of boats, of Haddock and Pollock over the 19 spatial zones of our model. The Cartesians seen here fish very ineffectively because they fail to discover the best zones to fish. They remain 'locked onto' existing information.

This corresponds to a fleet of « cartesians » and as we can see after leaving the port and locating a first fish aggregate, they « lock on » to this zone and stay fishing there for 3 years. In fact, after only 4 months there were in reality better alternatives, but because the only information available to the fleet came from its own fishing activities, and because this value of I corresponds to a « non- risk taker », who tends to go only to zones where the *knows* there are fish, the pattern of fishing effort concentrates on this zone until it has been almost totally fished out. Then, after 3 years the fleet finally moves to another zone, where once again it remains for another 3 years, before moving to another zone.

When we look at the performance of this fleet then, we observe that its spatial strategy was rather inefficient, because although it concentrated very strongly on the best location which it knew, its lack of « exploratory » voyages meant that most of the time it was not fishing the most advantageous spot. In other words, it was strong on « exploiting » what it knew, but it knew far too little.

We may compare this simulation with one strating from identical initial conditions, but where the fleet is made up of « stochasts », which a value of I = 0.1. Here we see that although the initial effort is mainly at the same point as the cartesians, it is in fact spread out over several points. After only 9 months the fleet is fishing the four best zones. Because of this, profits are high at each of these points, and growth is broadly based and much more rapid than in the previous simulation. After 5 years the fleet size has reached 265 units, whereas, for the cartesians it had only attained 159 at this time. This is because the more « random » movement of the stochasts allow it to explore the system much more fully than the cartesians, and to « map » the resource more quickly and more completely. However, because of their « anarchistic » tendencies, stochasts do not necessarily exploit the knowledge they have most efficiently. Because they do not obey « information » closely, they therefore maintain a dispersed effort, even though they do in fact « know » where it is « better » to be.

This paradoxical situation results from the fact that in order to fish effectively, there are two distinct phases that must be accomplished. First, the fish must be « discovered ». This requires risk takers who will go out into the « unknown » and explore, whatever present knowledge is. The second phase, however, requires that when a concentration of fish have been discovered, then the fleet will move in massively to exploit this, the most profitable location. These two facets are both necessary, but call on different qualities.

We may gain further insight by running the simulation again, but with 2 fleets — a fleet of stochasts, and a fleet of cartesians. This is shown in figure (18). However, we must also specify what flows of information we shall allow between the two fleets. In this first case let us examine what happens if there is no information exchange.

In that case, as can be seen, the pattern of the previous two simulations is repeated except that the competition between the two fleets gradually leads to the elimination of the less effective « cartesians ». The « sto-

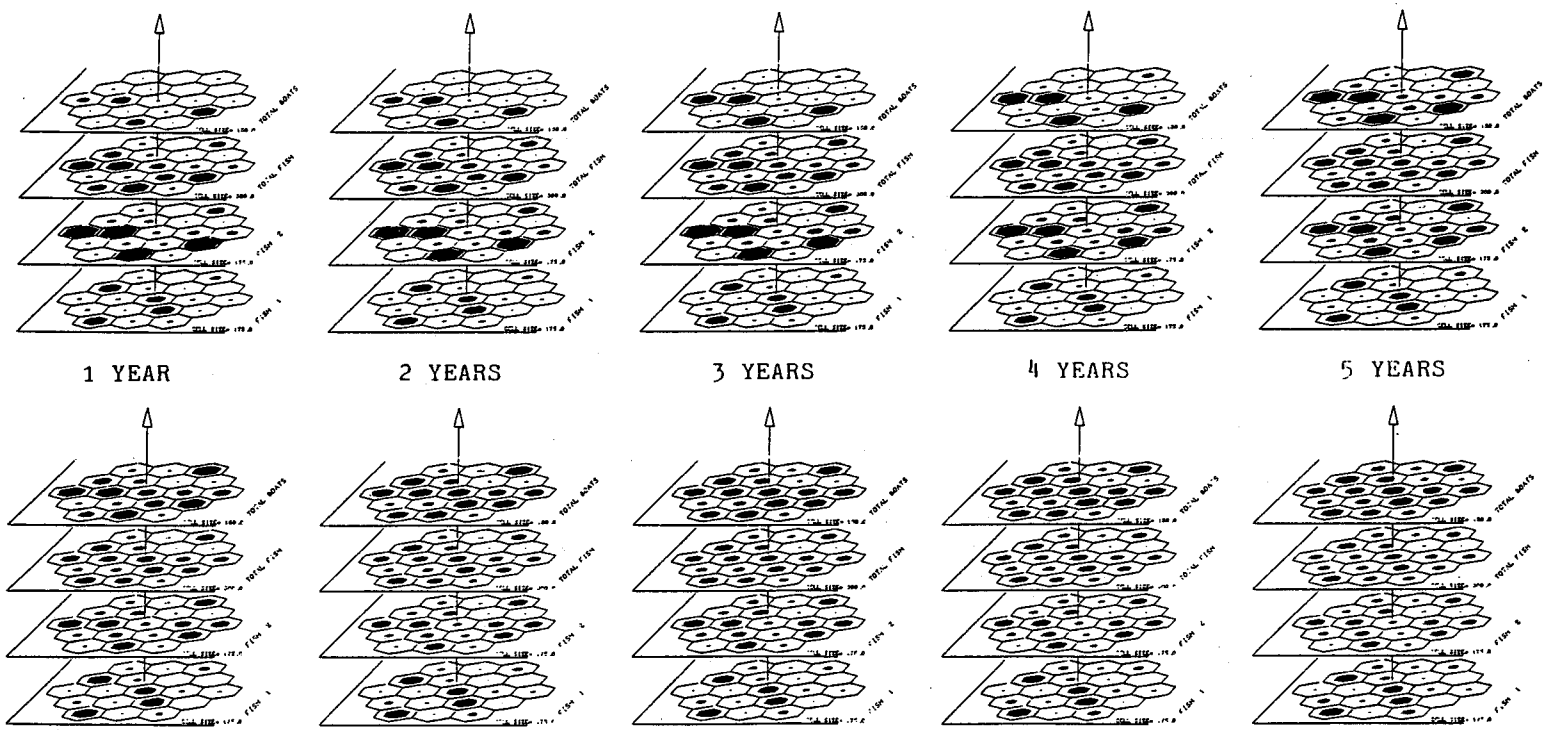


Figure (16). The evolution of boats and fish with a fleet of 'Stochasts'. They are much more effective than the 'Cartesians', but still do not muster their effort on the 'best' zones.

chast » strategy is the only one to survive, and we may surmise therefore that in a natural evolution the cartesians would be eliminated.

But, of course, we must then consider the fact that instead of allowing itself to be eliminated, the cartesians may change their behaviour in order to survive. As we stressed in the first section, the important thing may well be all the behaviours which are *not* yet in the system. Can a new behaviour invade the system ?

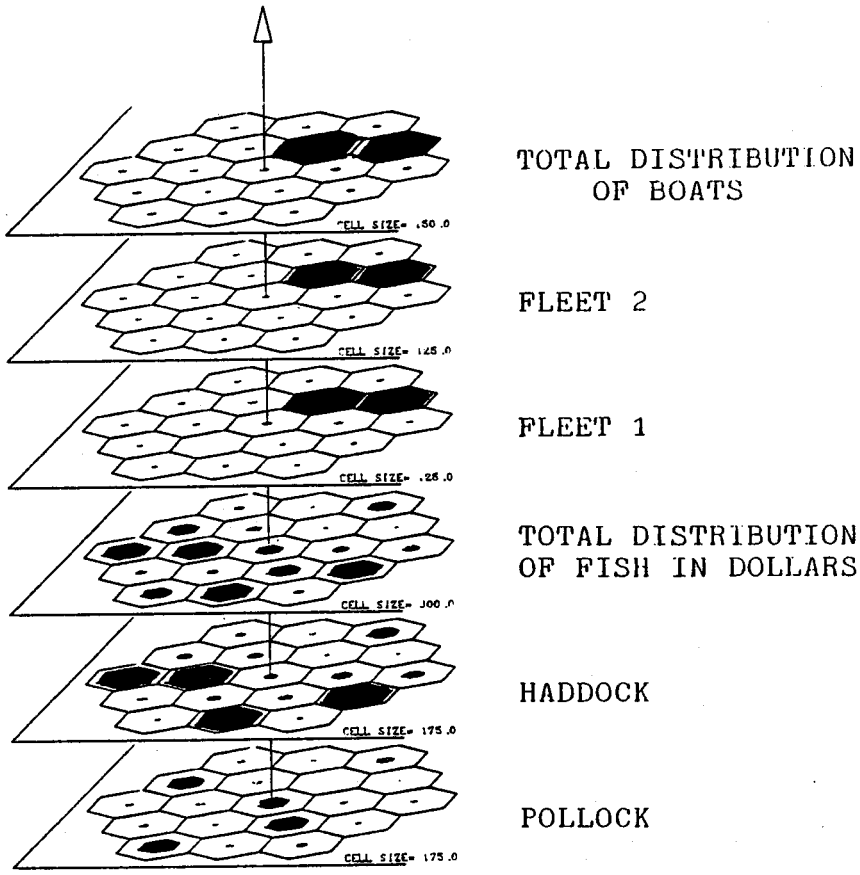


Figure 17

The initial distribution for our simulations of the evolution of 2 fleets, (Cartesians and Stochasts) and 2 fish species (Pollock and Haddock).

The answer to this is, yes. Indeed several possibilities arise. Firstly, the « cartesians » could become « stochasts ». Successive battles of this kind might lead to the selection of the « best compromise » for I. However, another possibility exists. The cartesians could *spy* on the stochasts ! If they can obtain the information which the stochasts have as to the « map » of the resource, then they could use their greater efficiency of « extraction » to good effect.

Figure (18). Stochasts out compete Cartesians. No information is shared.

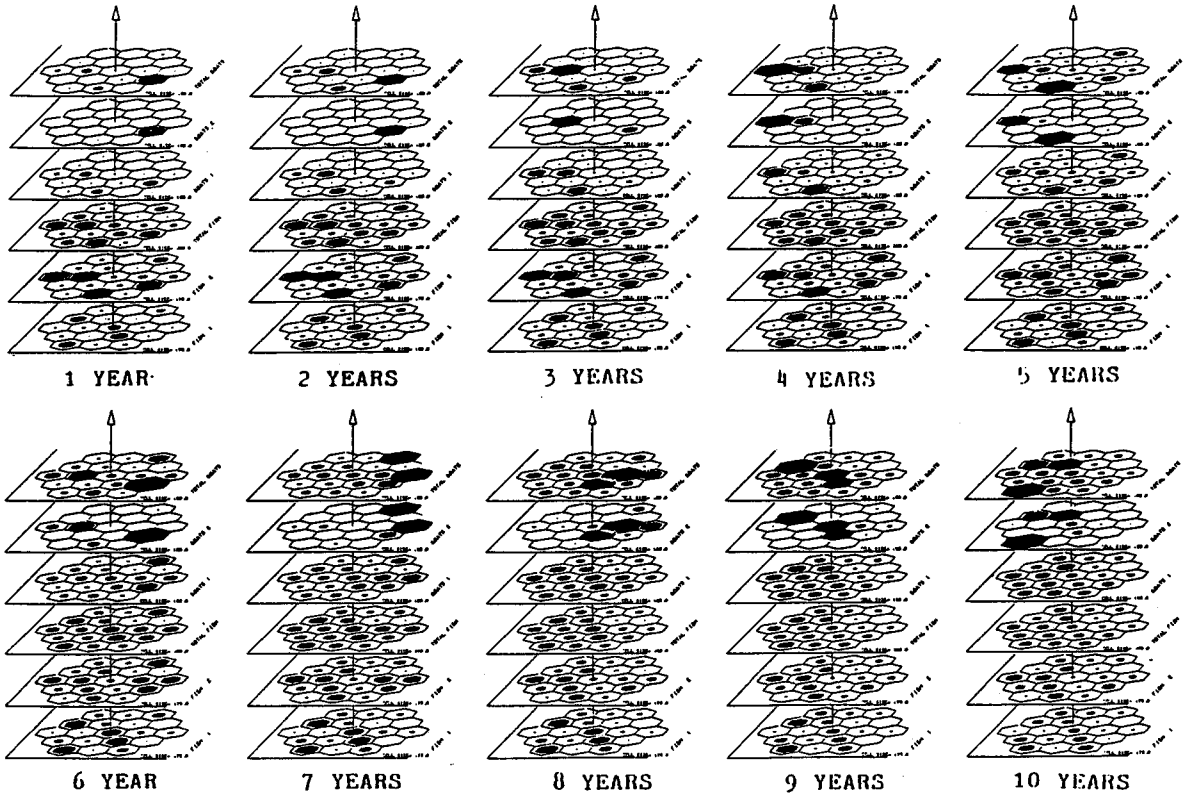


Figure (19). In this simulation the Cartesians spy on the stochasts, receiving .1 of their information. This is much more effective and the system is fished out much faster

Figure (20). Here there is a complete exchange of information between Stochasts and Cartesians. It favours the Cartesians very strongly and is only conceivable behaviour for the Stochasts if their is a profit sharing alliance between them.

In the simulation shown in figure (19), we see the evolution of the system when the « information exchange » matrix has a term 1 for the passage of information from the stochasts to the cartesians. Now we see that the pattern is different. After 5 years there are 253 cartesians and 162 stochasts. The strategy is *globally* more efficient because in all we have 415 boats already. This results from greater catches and profits than any other strategy up till now. However, the « stochasts » are now less well off than before, and will resent the cartesians profiting from their discoveries. At this point, they will attempt to reduce the information leakage by talking in codes on the radio, and by avoiding being sighted as much as possible. Similarly, the cartesians will step up their surveillance, installing several radios in order to listen on several channels simultaneously. They may have spies which literally follow the other fleet and report back. In this way we see how our model allows us to generate much of the complexity of the real world by allowing us to ask the question, what behaviour can invade the system ?

The important point is that it does not necessarily need to lead to better global performance of the system. It simply has to be able to invade. Indeed nearly all the innovations that can and will invade the fishing system lead to lower overall performance, and indeed to collapse. Thus, higher technology, more powerful boats, faster reactions etc. all lead to a decrease in output of fish.

Indeed, the lesson of our spatial simulations is similar. In reality, the strategies which invade the system are ones that pay off for a particular actor *in the short term*. But, globally, and over a long period, the effect may be quite negative. For example, in the case we have discussed above, if the cartesians can obtain information too readily from the stochasts, and act on it with great speed, then they will « out compete » the stochasts and drive them from the system. Without them, fresh discoveries will not be made so easily. However, in a situation of competition, unless there is either collusion between fleets, or management by some authority, then actors cannot afford not to implement an « innovation » which pays in the short term, for fear of being eliminated themselves. In this way the system becomes more fragile, less stable and can develop an incredible pyramid of complexity. For example, the stochasts can attempt to pass « mis- information », the cartesians can cease to listen, fleets could buy spotter planes, or rent satellite information. Alliances could form, where information was shared and profits divided. Exploring, spying, lying, and of course fishing. All of this may evolve, and much of it will require costly apparatus — radios, coders, scramblers, night manœuvres... An innovation does not have to raise the global « efficiency », it just has to be able to invade the system — that is offer a local, short term « gain » to someone.

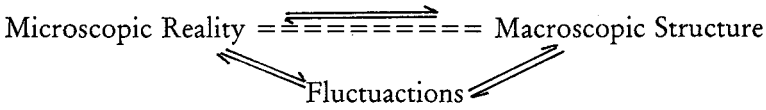
In the next section we shall briefly discuss the wider implications of these ideas for human systems in general.

GENERAL DISCUSSION

In the preceding sections we have tried to show that new concepts emerging from the study of dissipative structures and non-linear systems give

rise to models which can be usefully applied to the understanding and management of fisheries. But, we also feel that some very general lessons can be drawn from these ideas.

The first point concerns the fact that we must recognize that the evolution of a complex system is not simply given by some smooth set of deterministic equations, some System Dynamics, but by a dialogue between this « average » description of the moment, and the fluctuations and perturbations which have been excluded from it. We have :



There are three basic types of fluctuation, those simply of density which test the « stability » of the structure, those of the parameters which represent the environment, and also those of individual characteristics which have been lost in the aggregation process. Clearly, Darwinian evolution is based on the latter, where individual « mutants », or non-average individuals, can appear in the system and grow to become a population. But this is just a special case of the general principle whereby innovation and individual eccentricity can be copied by others and invade the system. Evolution concerns therefore the dialogue between the average and the non- average. Fluctuations of all kinds are the motor of evolution and the underlying « reason » for much of its structure. In the case of our fisheries example above, the feedback response of the human system were such as to amplify the yearly random fluctuations of the environment, and generate large, fairly regular cycles of boom and bust, despit the fact that the deterministic system gave rise to a *stable* equilibrium solution.

In this way fluctuations, instead of being just a minor perturbation to « average » results, actually have a major effect on the local economy. The cyclic nature of the haddock fishery is a very important feature for fishermen with small boats, and few choices, and hence affects the economic and political evolution of the region to an important degree. A model such as the one presented above can be a useful « learning » tool, even if it merely served to show that a fluctuating resource can probably never give rise to a constant and satisfactory economic return. Also, it shows us that the « learning processes » in the system, whereby for example fishermen with faster response rates (r) will replace others, lead to larger swings in catch and profits not smaller ones, and in general increase the instability of the system.

Indeed, our model shows rather clearly the opposition between short and long term decision making. The evolution of the behaviour of fishermen, if based on the short term requirements of immediate profit and competitive advantage, will lead the system to collective disaster, and this is the fundamental reason why fisheries management exists. However, management without taking into account the natural fluctuations in the system, and the dynamic mechanisms which may amplify them, will be perceived as incompetent as successive, unexpected « crises » occur.

The final point raised by this paper is, we feel, of very general interest. It concerns the behaviours which underly the processes of « discovery » and of « exploitation » of « resources », and ideas, technologies and products. The vital point is that « discovery », like « invention » and « creation », can only be achieved by going *beyond* present knowledge. In our fishery example, this showed up because « stochasts » were necessary for the discovery of new fish aggregates, and the exploitation phase concerned how this new information was channelled, and how it led to switching in the « investment » of fishing effort. And the important point about « stochasts » is that they are people who, for whatever reason, do not respond simply to the information which exists about the present returns on effort. That is, they either have some personal information, or some belief which leads to behaviour which is much less conditioned by the « average » view and assessment than is the case for « cartesians ». Broadly speaking we could perceive them as *risk takers*, and in a sense as optimists, while « cartesians » are the solid backbone of the system, who only bet on certainties, and prefer low, but safe, returns to the « promise » of some « Eldorado ».

But all of us, in our various roles in our family, firm, university, community and even Nation, are faced with the problem of deciding how to divide our time and effort between the performance of tasks characteristic of our present role, with known values and pay offs, and the continued « openness » to the possibility of new roles, new pay offs and values for the future. We can either deliberately allow ourselves to explore new paths and connections, or instead, we could organize our lives so that such diversions are minimized, and clearly this second alternative is very attractive since at the time when they occur, diversions have no obvious purpose.

There are two extremes — « stochasts » and « cartesians ». But, as we have seen for our fishing example *both* behaviours are necessary to the system. We must avoid both the total *dispersion* of randomness, and the complete enslavement by *information*. It is however the discoveries made by these risk taking « stochasts » which nourish society in the long run, and allow its expansion and greater sustenance, but of course the efficient and directed behaviour of the « cartesians » constitutes the *concensus*, and gives stability and form to society.

The key conclusion from all this must be that in order to *learn* about the world it is the « fluctuations » which are important and not the average, rational behaviour. Risk takers who go beyond the present limits and *logic* of the system can discover new « attractors » and send back the information. Today, the spread of computers and of information gathering and treatment presents us with a considerable new threat. That is the gradual supression of apparantly unmotivated, abnormal behaviour and the freedom for it to occur which was afforded by our partial ignorance of the exact state of the system. But this would be just as if we suppressed the natural fluctuations present in the system, which would tend to make it « fossilize » at each stage. « Progress » would then occur by catastrophic collapse and replacement rather than by a

more smooth, adaptive process. In other words, the *adaptive capacity* of the system comes from the fluctuations which are present and which always explore new possibilities.

Whenever a new technology or product is discovered, there follow the same phases as we have seen in our fishing model. First, the discover tries to profit from his costly exploration by developing production. However, soon he is joined by « spying cartesians » who want their slice of the new « cake ». There follows a period of expanding production as growing quantities of the product are consumed and economies of scale emerge. However, after some time either the market begins to saturate, or the resource required for the product start to become rare, and to increase in price. Either way competition intensifies. This leads to a period of « rationalization » when investment is used not to expand production but to make existing capacity more cost effective, usually decreasing employment in the sector. The less « dedicated » firms, the « stochasts », are driven out of the market and competition increases further among the remaining « cartesians ». Less and less employment is offered by the sector and its exploitation divides between either extremely efficient, mechanized production in developed economies, and less sophisticated production in developing countries. Clearly, the production of agricultural and industrial commodities and products show this trend, and now manufacturing and even the « lower » end of high technology products are displaying it already.

In order to escape from the destructive circle of increasing competition, we see the importance of management principles which would maintain diversity, and explore alternatives beyond the present « rationale ». It becomes important not to simply try to *maximize* one's strategies on the rational basis of *today's* information. This is an important point, because « cartesians » are so plausible ! After all, they can *show rationally* why their particular choice is best, given present information and conditions. But, of course, as we have tried to explain, what matters in the long term is not what we *know*, but what we *don't know*. Evolution results from the development of new things, and these are what we do not really know about. If we are to avoid a future of ferocious and ever growing competition, in a shrinking world with a single perspective and the common value of a single culture, then we must encourage « stochasts », and the diversity and expansion which only they can bring.

ACKNOWLEDGEMENTS

This work was supported by the Department of Fisheries and Oceans, Canada under contract 095C.FP101— 4— 0580 and also by the Global Learning Division of the United Nations University, Tokyo. The authors wish to thank scientists from the Marine Fish Division, Bedford Institute of Oceanography and from Chimie-Physique II at the Université Libre de Bruxelles for many helpful discussions. In particular we wish to thank G. Engelen for help with the computer program.

REFERENCES

Beddington J.R. 1979. Harvesting and population dynamics, pp. 307-319. *In* : Anderson R.M., Turner B.D. and Taylor R.L. (eds.) Population Dynamics. The 20th Symposium of the British Ecological Society, London 1978. Blackwell Scientific Publications, Oxford.

Beddington J.R. and May R.M. 1977. Harvesting natural populations in a randomly fluctuating environment. Science 197 : 463-465.

Beverton R.J.H. and Holt S.J. 1957. On the dynamics of exploited fish populations. U.K. Min. Agric. Fish. Invest. (Series 2), 19 : 1-533.

Holling C.S. (ed.) 1978. Adaptive environmental assessment and management. International Series on Applied Systems Analysis, Wiley Interscience, John Wiley and Sons, New York, N.Y.

Horsthemke W. and Lefever R. 1984. Noise-Induced Transitions. Theory and Applications in Physics, Chemistry, and Biology. Springer Series in Synergetics, Vol. 15. Springer-Verlag, New York, N.Y. 318 p.

May R.M., Beddington J.R., Clark C.W., Holt S.J. and Laws R.M. 1979. Management of multispecies fisheries. Science 205 : pp. 267-277.

McGlade J.M. and Allen P.M. 1985. The Fishery Industry as a Complex System, pp. 209-216. *In* : Mahon R. (ed.) Towards the Inclusion of Fishery Interactions in Management Advice. Can. Tech. Fish. Aquat. Sci. N° 1347.

Nicolis G. and Prigogine I. 1977. Self-Organization in Nonequilibrium systems from Dissipative Structure to Order Through Fluctuations. Wiley Interscience, John Wiley and Sons, New York, N.Y. 491 p.

Prigogine I. and Allen P.M. 1982. The challenge of complexity, pp. 3-39. *In* : W.C. Schieve and P.M. Allen (eds.) Self-Organization and Dissipative Structures. Applications in the Physical and Social Sciences. Univ. of Texas Press, Austin TX.

Stucki J.W. and Horsthemke W. 1982. Creation of order by environmental noise in the Volterra-Lotka model, pp. 187-192. *In* : W.C. Schieve and P.M. Allen (eds.) Self-Organization and Dissipative Structures. Applications in the Physical and Social Sciences. Univ. of Texas Press, Austin, TX.

On Stochastic Models for Human Conduct in the Evolution of Science and Technology

Werner Ebeling
Humboldt Universitat Berlin, GDR

Contribution to the discussion of the lecture by Peter Allen

Investigating processes of human conduct as e.g. the choice of certain field by a scientist or of a certain technology by an engineer, one is confronted with problems of much higher complexity than in models for biological or ecological communities (Allen and Ebeling, 1983). However there are some basic processes as selfreproduction, decay and transitions which are also of great importance for human conduct and can be modelled easily. Let us consider for example a scientific discipline like physics consisting of many subfields (research areas) as statistical physics, high energy physics etc.. The number of scientists in a given subfield i (i = 1, 2, 3, ...) is denoted by $N_i(t)$ and the probability of a distribution over the fields by $P(N_1,N_2, ...,N_i ...,t)$. Assuming Markovian character, the time evolution is described by a Master equation $\partial_t P = WP$ where the evolution operator W is determined by the transition probabilities. We assume the following elementary processes (Ebeling and Scharnhorst, 1986) :

(i) Selfreproduction processes : Students and young scientists make a choice of the field where they want to start their work. The transition probability is represented by linear and quadratic terms

$$W(N_1,...N_i+1... N_j.../N_1...N_i..N_j...) = A_i^* N_i + B_{ij}^* N_i N_j$$

(ii) Decay processes : Scientists are active a limited number of years, then they go to pension or simply stop to work in science. The probability of this transition is described by

$$W(N_1... N_i-1... N_j.../N_1...N_i...N_j...) = A_i^\dagger N_i + B_{ij}^\dagger N_i N_j$$

(iii) Discoveries and induced transitions : Scientists eventually make discoveries what means that they move to new nonoccupied fields or they move to other already occupied fields for various reasons (e.g. imitation). The transition probability is assumed to be

$$W(N_1 \ldots N_i+1, \ldots N_j-1 \ldots/N_1 \ldots N_i \ldots N_j \ldots) = A_{ij}N_j + B_{ij}N_iN_j$$

The social forces and motivations which lead to the elementary processes are described in this simple model by the parameter vectors and matrices. These values may be estimated from empirical material for the field mobility of scientists (Vlachy, 1981). By an appropriate choice of the coefficients not only linear processes of growth and decay but also nonlinear effects as selfinhibition saturation, advice, imitation and sponsoring may be described at least formally. Of special interest are discoveries, i.e. the transitions to new nonoccupied fields i.e. from $N_i = O$ to $N_i = 1$. In the history of science these processes are connected with the names of outstanding scientists. Another process which seems to be typical for human conduct is imitation, i.e. in our case the transition of scientists to fields where already many other scientists are working ; these are the so-called modern fields. Imitation is a nonlinear process. Further we mention other interaction processes as stimulation and learning which also are modelled by quadratic effects i.e., by the transition matrices B^{+}_{ij}, B^{*}_{ij}, B_{ij}. We want to underline that imitation, stimulation and learning are typical aspects of human conduct. The model described so far can also be applied for the formal modelling of the evolution of technologies (Jimenez Montano and Ebeling, 1980). In that case the numbers $i = 1,2, \ldots$ correspond to technologies which are used to cover certain social needs as e.g., communication, shelter etc... Further the N_i correspond to the number of workers (or plants) which are using the technology i. The processes which are modelled include innovations, selection, imitation, substitution etc.. Summarizing let us repeat again the main idea of this approach : Biological evolution and the evolution of human conduct show deep differencies but they have also several common aspects. As such common aspects we consider the fundamental role of (i) selfreproduction (in a wide sense) of (ii) finite lifetime and permanent change of generations and of (iii) mutability, variability and discoveries.

One may ask whether it makes sense to start the modelling of evolutionary aspects of human conduct before reaching full clarity about biological evolution. In this connection let us repeat an argument given by I. Prigogine : Biological evolution started several billions of years ago and we do not have reliable records and fossils, on the other hand we have a quite good knowledge about the history of science and technology. This gives us a real chance for theoretical modelling based on the observed material. As shown in the contribution by Peter Allen, the models for complex processes should include the stochastic effects, since deterministic models are in most cases inadequate and inappropriate.

BIBLIOGRAPHIE

P. ALLEN, W. EBELING, BioSystems 16 (1983) 113.
W. EBELING, A. SCHARNHORST, Czech. J. Phys. B36 (1986) 43.
M.A. JIMENEZ MONTANO, W. EBELING, Coll. Phenomena 3 (1980) 107.
J. VLACHY, Czech. J. Phys. B31 (1981) 669.

Transitions between Biochemical Oscillations and their Relevance to Biology

Benno Hess
M. Markus
Max-Planck-Institut für Ernährungsphysiologie

INTRODUCTION

Periodic reactions are a widespread feature in biochemical processes (for reviews, see 1-4). Recently, more complex dynamic states, namely aperiodic (chaotic) oscillations were discovered in biochemistry. Experimental evidence for chaotic oscillations was presented for a single-enzyme system (the peroxidase-oxidase reaction, 5-6) and for glycolyzing yeast extracts under periodic substrate input (7-9). It was found that oscillating glycolysis acts as a time pattern converter, transforming the sinusoidal input flux into a large variety of periodic, quasiperiodic and chaotic oscillations. These results may have model character of general importance in view of the ubiquitous appearance of coupling between oscillators in biochemical and biological systems. In particular, oscillatory coupling between metabolic processes has been demonstrated in yeast cells (10,11) and in the slime mould D. discoideum (12).

Besides having the function of a time pattern converter, it has been shown that the glycolytic extract system can display different time patterns (up to four) under the same set of external control parameters, depending solely on the previous history of the system (hysteresis) (11,13,14). In such a case, different attractors coexist in phase space (i.e. in a space of variables describing the system dynamics), each attractor having its own basis of attraction (set of starting values in phase space leading to the particular attractor).

In general, there are two ways of causing a transition between time patterns from outside : 1. By changing the external control parameters, or

2. By perturbating the phase variables of the system so that it may jump from one attractor to the basin of attraction of a different attractor, at constant control parameter values. Beside these exogenous transitions, the system may abruptly undergo an autonomous transition at a socalled « crisis », as has recently been reported for simple mathematical mappings (15,16). In the present contribution, we illustrate these transitions by numerical calculations of glycolytic time patterns.

THE MODEL

It has been shown that the information dimension (17) of glycolytic attractors obtained from experiments lies below the value 3 (18). Therefore, three phase variables are enough to describe the dynamics of glycolysis, in spite of the much larger number of metabolites involved. As phase variables, we choose here [ADP] and [PEP] (PEP : phosphoenolpyruvate) and the phase φ of the input flux. We have obtained satisfactory agreement between theoretical and experimental oscillations with a simplified model containing only the enzymes phosphofructokinase (PFK) and pyruvate kinase (PK) (7-9). The model is described in detail in references (13,19). It contains rate laws for PFK and PK that had been determined from extensive kinetic measurements. Unless otherwise stated, the conditions for the calculations are those given in reference (13). The input flux rate consists of a constant flux, plus a flux of the form $A \sin\omega_e t$.

TRANSITIONS THROUGH CHANGES IN THE EXTERNAL CONTROL PARAMETERS

In order to determine the effects of the external control parameters on the system, these parameters were slowly varied while integrating the differential equations (« slow time » method). We have described this method in detail in reference (13). In Fig. 1, we exemplify a result using this procedure. Fig. 1 shows hysteresis loops obtained upon variation of the input flux amplitude A. In Fig. 1a we see the influence of the amplitude A on the period T of the system oscillation, relative to the input T_e. Fig 1b shows the effect of the amplitude A on the maximum Liapunov exponent per unit time (λ_{max}) (see 19,20). λ_{max} is a measure of the average stability of an orbit in phase space. For chaotic oscillations, λ_{max} is larger than 0 and is equal to the reciprocal of the shortest average time in which a perturbation doubles its size. For periodic orbits, λ_{max} is smaller than 0 and is equal to the reciprocal of the longest average time in which a perturbation decreases to one half. Quasiperiodic oscillations (mixing of periodic oscillations having incomensurate frequencies) are characterized by $\lambda_{max} = 0$. At the left of Fig. 1 ($A<0.85$ mM/min) a quasiperiodic orbit occurs. Farther to the right on this figure ($0.88<A<1.01$ mM/min) two periodic orbits coexist in phase space (periods $T=3T_e$ and $T=10T_e$). At even larger values of A ($1.01<A<1.03$ mM/min) a chaotic attractor coexists with a periodic orbit having period $T=3T_e$. At the extreme right of Fig. 1 ($A>1.03$ mM/min) there exists only a periodic orbit with $T=3T_e$. In Fig. 1b, one sees that λ_{max} is in fact close to zero for quasiperiodicity, and greater than zero for chaos.

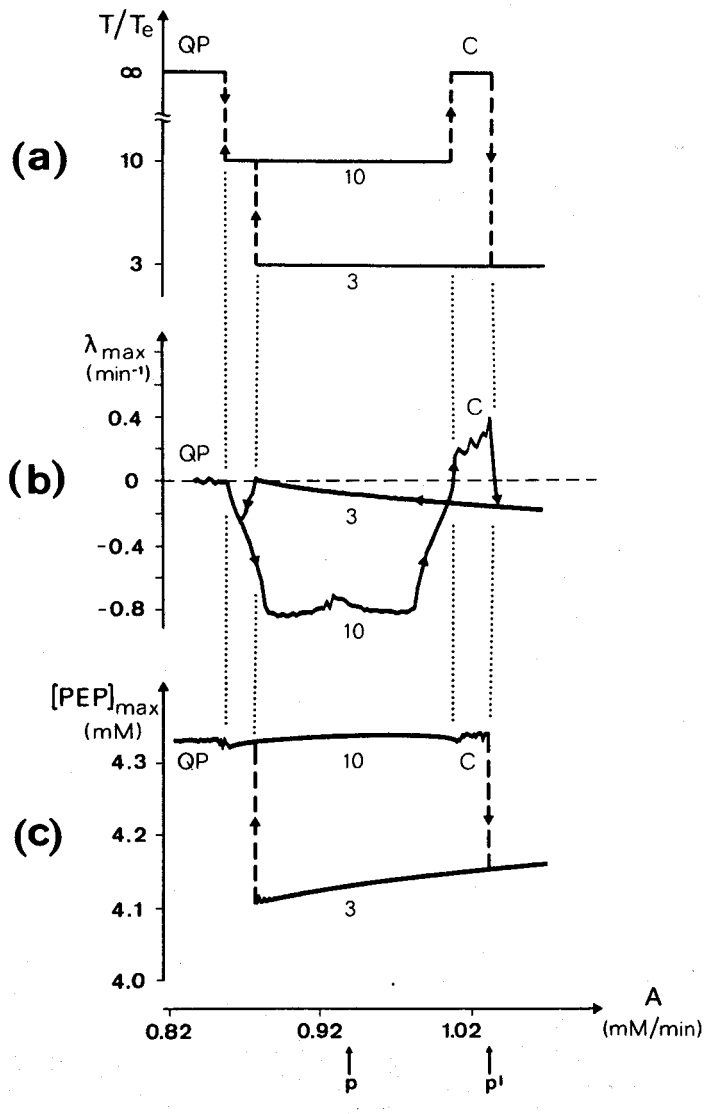


Figure 1

In the middle portion of Fig. 1b, it is note-worthy that the oscillations with period $10T_e$ have a higher stability (lower λ_{max}) than those with period $3T_e$. We have called systems displaying deep minima in λ_{max}, as exemplified here for $T=10T_e$, « hyperstable » systems (19).

Fig. 1c shows the effect of the input flux amplitude A on the maximum amplitude of the [PEP] oscillation. Clearly, only the oscillation with period $3T_e$ differs in amplitude from the other oscillations. Thus, any clipping process (for example, through an enzyme with a sigmoidal rate law) aimed to recognize the oscillation mode by cutting off lower values of [PEP] would not discriminate between period $10T_e$ and chaos, but

would discriminate between period $3T_e$ and chaos, or period $3T_e$ and period $10T_e$ (see also Fig. 2). Since for the conditions of Fig. 1, only one, resp. two ordinate values appear for each abscissa value, only one or two attractors coexist in phase space for a given set of control parameter values. The presence of two coexisting attractors is called « birhythmicity ». Also, hysteresis loops of high complexity are obtained, involving three or four coexisting attractors (« tri »- and « tetrarhythmicity » (11,13,14).

TRANSITIONS THROUGH EXTERNALLY INDUCED JUMPS IN PHASE SPACE

Oscillations corresponding to attractors in phase space can be switched from one to the other by suitable jumps in the phase variables (13,14,19). In our case, this can be accomplished by metabolite pulses. Fig. 2 shows such a switching process for the conditions indicated by arrow p pointing to the abscissa of Fig. 1. In Fig. 2, first a transition from $3T_e$ to $10T_e$ is obtained by addition of [PEP]. Through proper timing and dimensioning of the pulse, the system is driven directly from one attractor to the other, so that no transients occur. The second switching process (from left to right) shows the reverse transition by substraction of [PEP]. The third and fourth switching process in Fig. 2 show that both transitions can be accomplished by pulses of PEP in sequence of the reverse sign. However, the pulses in these last two situations do not drive the system directly to the final attractor, so that transients occur.

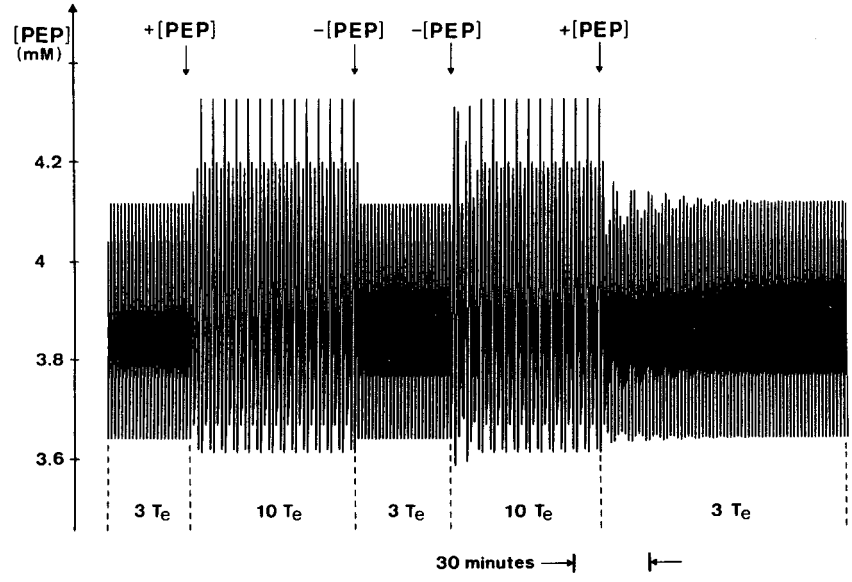


Figure 2

In order to determine the size of the pulses needed to cause a switching process, it is necessary to know the « watershed », also called « separatrix », which is the boundary separating the basins of attraction of the orbits in phase space. The watershed is given by the stable manifold of a

saddle-type solution (see 15,21,22). In some cases, especially when three or more attractors coexist, the watersheds may be complicately interleaved in phase space and very close to the attractors. Noise is then strongly amplified by the system (14). This is explained by the fact that the noise keeps switching the system from one basin of attraction to the other. The system thus displays ever changing transient behaviour, never settling on an attractor.

AUTONOMOUS TRANSITIONS (CRISES)

Recently, special types of transients, called « crises », have been reported (15,16). These transients have two typical features : 1. They have the characteristics of a chaotic attractor before they die away, and 2. they can be extremely long-lived.

Crises are explained as follows. A chaotic attractor coexists with some other solution (periodic, quasiperiodic, or chaotic) at slightly different values of the control parameters than those where a crisis occurs. At the crisis, the chaotic attractor collides with its own basin boundary. Under these conditions, it may take a long time until the system reaches the point of collision, which is the point where the chaotic oscillation ceases to exist, as it is attracted beyond its basin. In our model, such a collision takes place at the arrow p' pointing to the abscissa of Fig. 1. This collision causes the disappearance of a chaotic time pattern in favour of a periodic pattern with period $3T_e$. The corresponding crisis is illustrated in Fig. 3. The point of collision is reached here after about 4.5 hours. The collision itself is illustrated in Fig. 4. In this figure, we show the phase variables [ADP] and [PEP] in a stroboscopic representation, in which these variables are plotted at times $(\varphi/360° + n)T_e$, $n=1,2,3,\ldots$ φ is set to 0°. A periodic solution with period mT_e leads to m points in this representation. The crosses are the stroboscopic points corresponding to period $3T_e$. The dark circles are the stroboscopic points corresponding to the unstable saddle-type solution with period $3T_e$. The watershed between the basins of attraction of chaos and of the stable solution with period $3T_e$ is given by the stable manifolds of the saddle points. These manifolds are shown in the figure with arrows pointing in the direction in which they proceed in time. They were obtained by integrating the differential equations backwards in time starting from the saddle points. The chaotic attractor corresponding to Fig. 4 has a Liapunov dimension (17,19) very close to 2. Thus, the attractor is almost a surface (« thin » attractor) and the stroboscopic plot of Fig. 4 is almost a curve. The figure shows a tangential collision of the chaotic attractor with its basin boundary. At smaller values of A (see Fig. 1), where chaos coexists with period $3T_e$, the stroboscopic plot looks similar to Fig. 4, except that the chaotic attractor lies further outwards, not touching its basin boundary.

Fig. 4 correponds to one particular value of the phase variable φ. A variation of φ can be represented graphically by rotating the [PEP]-[ADP]-plane around the [PEP]-axis with angular velocity ω_e (11,14). Thus, one rotation is completed after the time T_e. This procedure allows presentation of Fig. 5. Fig. 5a shows the « thin », almost toroidal chaotic attractor. Fig. 5b shows its basin boundary under the conditions of col-

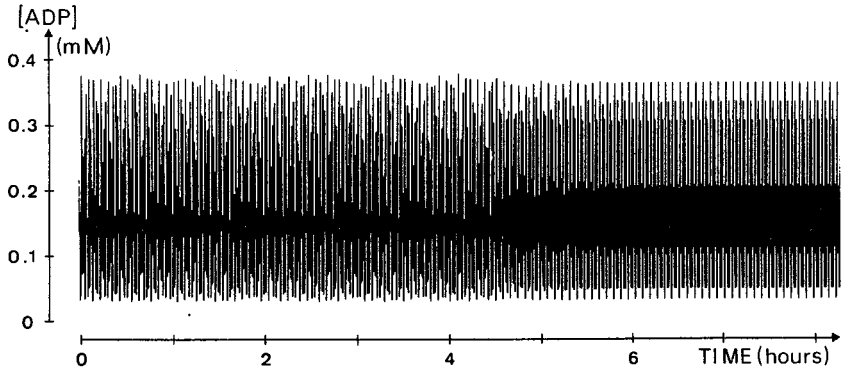


Figure 3

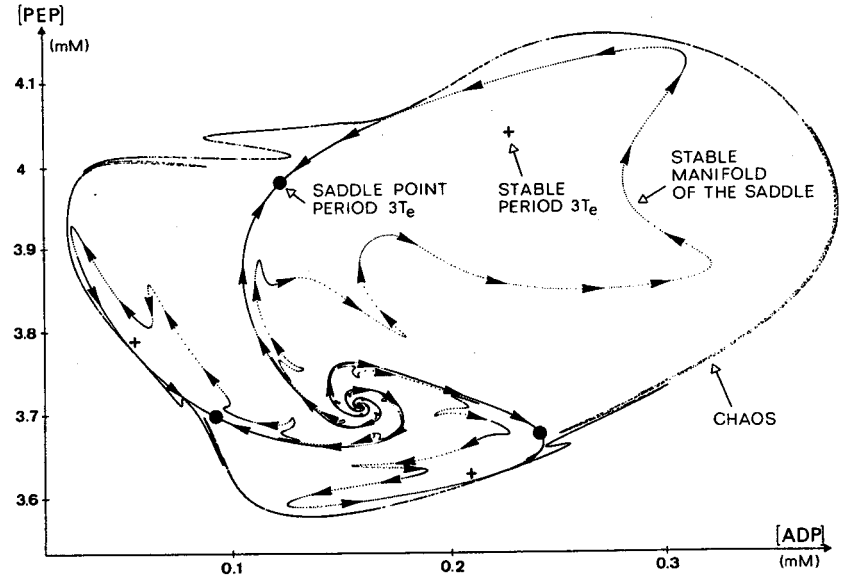


Figure 4

lision. As time proceeds, the system describes an orbit lying on the attractor shown in Fig. 5a. The crisis occurs when the orbit touches the surface shown in Fig. 5b. This surface consists of three tori having a common closed curve. For a better understanding of Figs. 4 and 5, note that the chaotic region and the stable manifold on Fig. 4 are Poincaré sections of Figs 5a and 5b, respectively.

These complex dynamic processes make possible the sudden disappearance of a disordered time pattern in favour of a periodic oscillation.

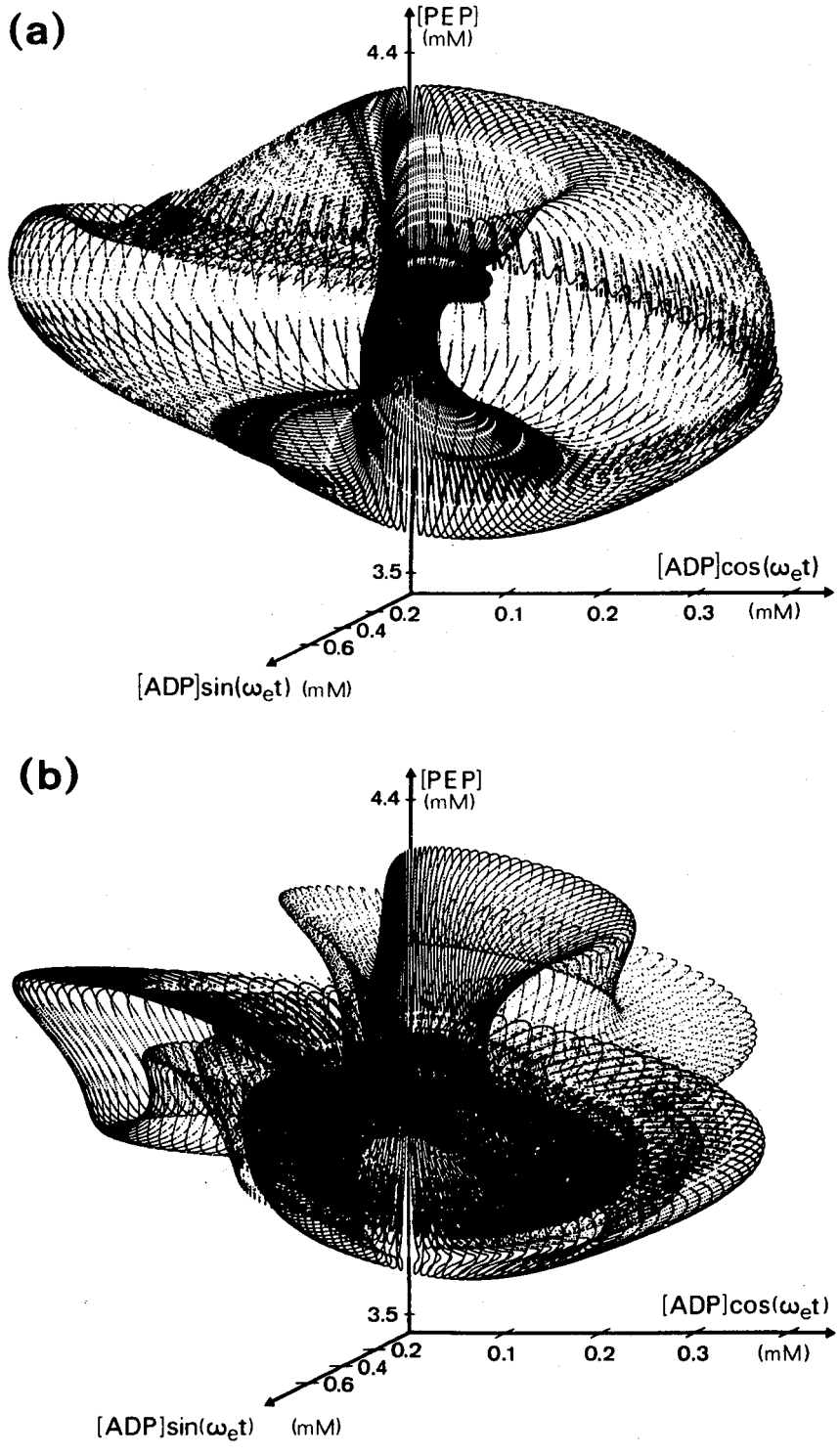

(a)
[PEP]
(mM)
4.4
3.5
[ADP]cos(ωet)
0.2
0.1
0.2
0.3
(mM)
0.4
0.6
[ADP]sin(ωet) (mM)
(b)
[PEP]
(mM)
4.4
3.5
[ADP]cos(ωet)
0.2
0.1
0.2
0.3
(mM)
0.4
0.6
[ADP]sin(ωet) (mM)

Figure 5

DISCUSSION

Using glycolysis under periodic excitation as a model system for coupled oscillators, we have discussed here the possibilities of transitions between biochemical time patterns. The variety of these transitions may be of fundamental importance in the organization and self-organization of living systems.

As shown, time patterns may not only be changed by external influences, such as control parameters or phase variable pulses, but also autonomously, if the conditions for a crisis are fulfilled. Experimentally, however, it may be difficult to distinguish between a change in the control parameters and a crisis, since a small uncontrollable drift of a control parameter close to a bifurcation point may look like a crisis, or vice versa. In fact, we have experimentally observed a sudden breakdown of chaotic oscillations in experiments with glycolyzing yeast extracts and it is undecided whether this breakdown is caused by the drift of a control parameter, or by a crisis.

In view of the present results, we postulate metabolic regulatory processes controlling the avoidance or the occurence of chaos, as well as regulatory processes controlling the transitions between periodic modes. These processes may rely, as in classical biochemistry, on a change of a control parameter. Furthermore, regulation through metabolic pulses is possible. In addition, a transition through crisis may occur in the absence of external influences. In the presence of fluctuations, oscillations may also die away without a specific regulatory cause if the attractor is close to its basin boundary and a fluctuation drives the system across this boundary. Such a phenomenon will have pathological consequences whenever an oscillation is related to a vital process, e.g. heart beating.

The role of chaos in biology is not yet explored. Chaos is assumed to be relevant in the occurence of neural disturbances (23). On the other hand, it has been shown that searching processes in complex computing systems can be greatly improved by introducing randomness (24). In biology, a search may be useful in chemotaxis, evolutionary processes and complex neural phenomenae, such as searching memory and decision making. The function of a « generator of internal diversity » in learning by selection has been discussed in (25). In predator-prey systems, randomness decreases the predictability of behaviour and may thus serve as a disguising mechanism (26). Furthermore, chaos has been used to model cell cycle variability in cell populations (27).

Up to here, we have described the effects of coupling of periodic processes in time. Such a coupling is widespread in biological systems, but so is also the coupling between time patterns and spatial phenomenae, including transport processes and cellular differentiation (see 3). Therefore, the results presented here should be regarded in the context of spatial organization. In this connection, it should be mentioned that chaotic time patterns have recently been found to be related to spatial disorders in the mutant FR17 of the slime mould D. discoideum. Experiments show that this organism displays both chaotic cAMP oscillations as well

as aberrant stalks and fruiting bodies (28). A transition to periodic, i.e. to spatially ordered behaviour may be accomplished by the addition of phosphodiesterase (29).

The system described in this contribution, consisting of only two glycolytic enzymes, leads to a large variety of dynamic states and to a number of possibilities for changing between one dynamic state and another. In view of the complexity of biological processes in general, contrasting with the simplicity of the system investigated, the significance of these dynamic phenomenae should not be underestimated. Further study of coupled nonlinear phenomenae, as those exemplified here, may thus have far reaching consequences in the understanding of living nature.

ACKNOWLEDGEMENTS

We thank Miss H. Kessel for programming assistance and Mrs B. Plettenberg for efficient typing of the manuscript.

REFERENCES

(1) HESS, B., BOITEUX, A., Annual Rev. Biochem. *40*, 237-258, 1971.

(2) *Biological and Biochemical Oscillations* (ed. by Chance, B., Pye, E.K., Ghosh, A.K., Hess B., Academic Press, N.Y., London, 1973).

(3) HESS, B., Trends in Biochemical Sciences, Sept., 193-195, 1977.

(4) RAPP, P.E., J. Exp. Biol. *81*, 281-306, 1979.

(5) OLSEN, L.F., DEGN, H., Nature *267*, 177-178 (1977).

(6) OLSEN, L.F., DEGN, H., Quart. Rev. Biophys. *18*, 165-225, 1985.

(7) MARKUS, M., KUSCHMITZ, D., HESS, B., FEBS Lett. *172*, 235-238, 1984.

(8) MARKUS, M., MÜLLER, S.C., HESS B., Ber. Bunsenges. Phys. Chem. *89*, 651-654, 1985.

(9) MARKUS, M., KUSCHMITZ, D., HESS, B., Biophysical Chemistry *22*, 95-105, 1985.

(10) HESS, B., BOITEUX, A., KUSCHMITZ, D., in : *Biological Oxidations* (ed. by Sund, H. and Ullrich, V., Springer-Verlag, Berlin, 1983) pp. 249-266.

(11) HESS, B., KUSCHMITZ, D., MARKUS, M., in : *Dynamics of Biochemical Systems* (ed. by Ricard, J. and Cornish-Bowden, A., Plenum Press, N.Y., 1984) pp. 213-226.

(12) GERISCH, G., HESS, B., Proc. Natl. Acad. Sci. USA *71*, 2118-2122, 1974.

(13) MARKUS, M., HESS, B., Proc. Natl. Acad. Sci. USA *81*, 4394-4398, 1984.

(14) HESS, B., MARKUS, M., Ber. Bunsenges. Phys. Chem. *89*, 642-651, 1985.

(15) GREBOGI, C., OTT, E., YORKE, J.A., Physica 7D, 181-200, 1983.

(16) GREBOGI, C., OTT, E., YORKE, J.A., Phys. Rev. Lett. *50*, 935-938, 1983.

(17) FARMER, J.D., OTT, E., YORKE, J.A., Physica 7D, 153-180, 1983.

(18) MARKUS, M., HESS, B., in : *Temporal Order* (ed. by Rensing, L. and Jaeger, N.I., Springer-Verlag, Berlin, 1985), pp. 191-193.

(19) HESS, B., MARKUS, M., in : *Synergetics — from Microscopic to Macroscopic Order* (ed. by Frehland, F., Springer-Verlag, Berlin, 1984) pp. 6-16.

(20) SHIMADA, I., NAGASHIMA, T., Progr. Theor. Phys. *61*, 1605-1616, 1979.

(21) HAYASHI, C., *Nonlinear Oscillations in Physical Systems* (Mc Graw Hill, N.Y., 1964).

(22) MARKUS, M., HESS, B., in : *Dynamics of Biochemical Systems* (ed. by Keleti, T. and Trón, L., Akadémiai Kiadó, Budapest), in press.

(23) RAPP, P.E., ZIMMERMAN, I.D., ALBANO, A.M., deGUZMAN, G.C., GREENBAUM, N.N., BASHORE, T.R., in : *Nonlinear Oscillations in Chemistry and Biology* (ed. by Othmer, H.G., Springer-Verlag, N.Y.), in press.

(24) KOLATA, G., Science *223*, 917-919, 1984.

(25) CHANGEUX, J.-P., HEIDMAN, T., PLATTE, P., in : *The Biology of Learning* (ed. by Marler, P. and Terrace, H.S., Springer-Verlag, Berlin, 1984) pp. 115-133.

(26) SCHULMEISTER, Th., Studia biophysica *105*, 5-10, 1985.

(27) MACKEY, M.C., in : *Temporal Order* (ed. by Rensing, L. and Jaeger, N.I., Springer-Verlag, Berlin, 1985) pp. 315-320.

(28) COUKELL, M.B., CHAN, F.K., FEBS Lett. *110*, 39-42, 1980.

(29) MARTIEL, J.L., GOLDBETER, A., Nature *313*, 590-592, 1985.

Figure captions

Figure 1.

Hysteretic response of glycolysis upon variation of the input flux amplitude A. The input frequency is $\omega_e=2\pi/T_e=14$ min^{-1}. T is the response period. λ_{max} is the maximum Liapunov exponent per unit time. $[PEP]_{max}$ is the maximum value of [PEP] during the oscillations. The integers at the curves are equal to T/T_e. QP and C indicate quasiperiodicity and chaos, respectively. The arrow p correponds to Fig. 2. The arrow p' corresponds to Figs. 3 to 5.

Figure 2

Switching processes between two time patterns by additions and substractions of PEP. The conditions correspond to those of the arrow p in Fig. 1. The periods of the oscillations are indicated below the time patterns (T_e : input period).

Figure 3

Autonomous transition from a chaotic to a periodic time pattern due to a crisis. The conditions are those of arrow p' in Fig. 1.

Figure 4

Stroboscopic plot at a crisis. The chaotic attractor touches its basin boundary. This boundary is given by the stable manifold of the saddle. The conditions are those of arrow p' in Fig. 1.

Figure 5

(a) Chaotic attractor obtained upon rotation of the [PEP]-[ADP]-plane shown in Fig. 4. The plane is rotated around the [PEP]-axis so that one rotation is completed in one input period. The scale on the [PEP]-axis is linear.

(b) Basin boundary of the chaotic attractor shown in (a).

Self-Sustained Oscillations and Chaotic Behaviour in Kidney Pressure Regulation

Klaus S. Jensen
Physics Laboratory III — The Technical University of Denmark

Erik Mosekilde
Physics Laboratory III — The Technical University of Denmark

Niels-Henrik Holstein-Rathlou
Institute for Experimental Medicine — University of Copenhagen

1. INTRODUCTION

Complementary to the concept of homeostasis which has dominated physiological thinking for such a long time, unstable phenomena are increasingly being recognized as significant in the regulation and function of normal physiological systems[1]. The beating of the heart and the ovarian cycle have long been acknowledged as self-sustained oscillations. But investigations performed during the last decade have revealed the existence of a great variety of additional biological rythms with periods ranging from fractions of a second to several hours or even days. Examples are hormonal regulations[2-4] and neuronal function[5], with secretion of luteinizing hormone, for instance, showing a typical period of 2-3 hours.

From a physical point of view, these phenomena are related to the fact, that physiological systems are open thermodynamic systems which are maintained in a nonequilibrium condition through dissipation of energy[6].

Nonequilibrium thermodynamic conditions can also give rise to more complex temporal behaviours. Chaotic phenomena can thus arise either as a limit cycle becomes unstable and develops through a cascade of

period-doubling bifurcations[7], by interactions between two or more instability generating mechanisms[8] or through other possible roads[9]. Chaotic phenomena have been experimentally observed, for instance, in nerve membranes and in nerve cell bodies[10].

Recent experimental investigations[11-13] have revealed an oscillatory behaviour of the proximal intratubular pressure in rat kidneys. With typical periods of 0.3-0.5 min, these oscillations are far too slow to be caused by breathing or heart beating, and detailed analyses suggest that they arise from an instability in the regulation of glomerular filtration[14]. While for normal rats, the oscillations have the typical appearance of a limit cycle, highly irregular oscillations are observed for spontaneously hypertensive rats (SHR).

Figure 1a shows a typical example of the experimental results obtained for normal rats. The intratubular pressure as measured by a thin micropipette is found to oscillate with an amplitude of about 5 mmHg around a mean value of approximately 12 mmHg. Respiration and heart beating are reflected as a superimposed, small-amplitude ripple with a much higher frequencies (l Hz and 4 Hz, respectively). Figure 1b shows the spectral distribution of the experimental results. The spectrum is dominated by a single sharp peak at 28 mHz. The sharpness of this peak is a measure of the coherence of the oscillations, which, considering the noisy environment, is quite astonishing and typical for a self-sustained oscillation.

Figure 2 shows similar results obtained for genetically hypertensive rats (SHR). The intratubular pressure here oscillates in a highly irregular fashion, and the frequency analysis reveals a broad-banded noisy spectrum with strong components in the low-frequency end. This is characteristic for a chaotic signal.

The purpose of this paper is to describe a dynamical model which can account for the above observations. We show how the nephron pressure- and flow regulation becomes unstable because of a delayed action at the juxtaglomerular apparatus. In this apparatus, a signal from the distal tubule, presumably representing the sodium concentration of preurine leaving the loop of Henle, is processed through a series of reactions to finally influence the nephron blood supply. With realistic parameter values, the model exhibits a limit cycle behaviour. Variation of those parameters which are likely to differ between normal and hypertensive rats causes the model to develop chaotic behaviour through a Feigenbaum cascade of period-doubling bifurcations.

Our analysis extends recent work by Holstein-Rathlou and Leyssac[14] in which existing but hitherto separate physiological models of glomerular filtration, proximal tubular dynamics, and tubuloglomerular feedback were integrated with the purpose of showing that the total feedback system for physiological parameter values can exhibit self-sustained oscillations.

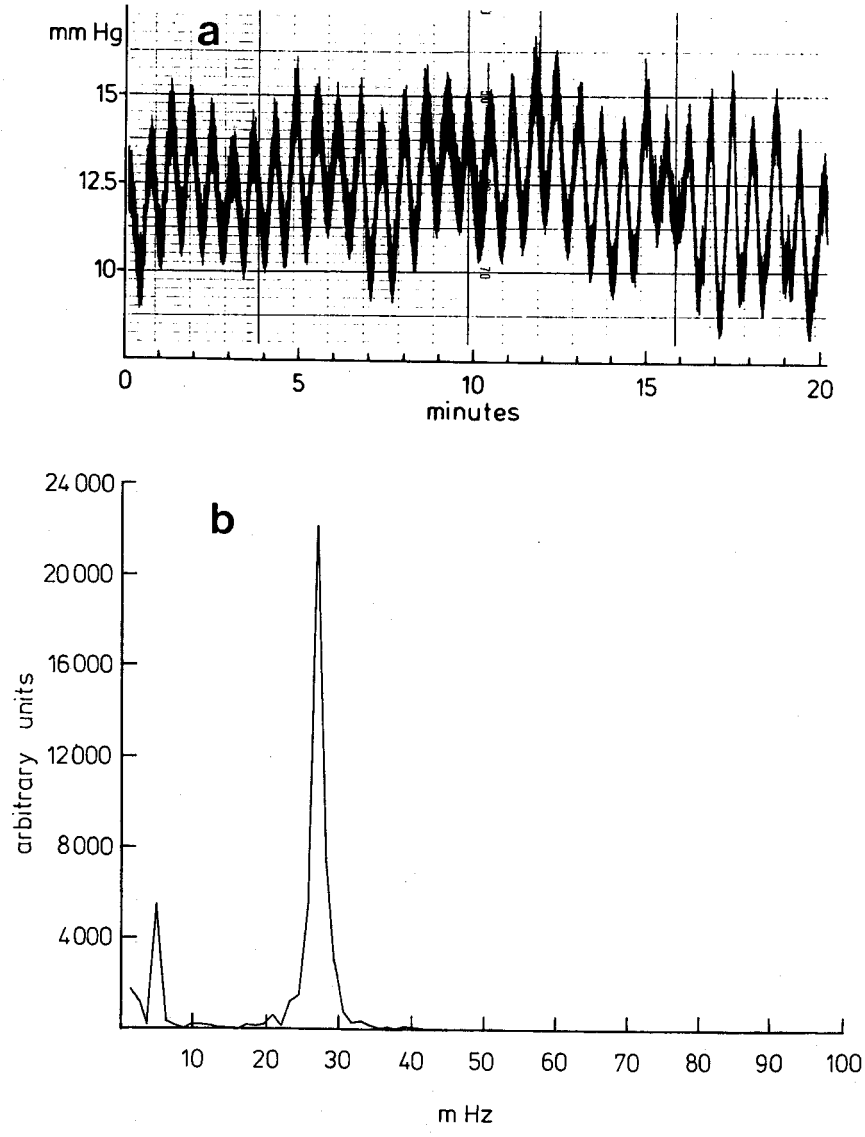


Figure 1
Experimental results for the variation in intratubular pressure for normal rats (a)
Frequency distribution of the experimental results (b).

2. THE MODEL

Urine formation serves the combined purposes of (i) excretion of metabolic biproducts and foreign chemicals and (ii) regulation of the body stores of water and salts. Urine is formed from blood by a process of filtration followed by selective reabsorption and secretion. The basic

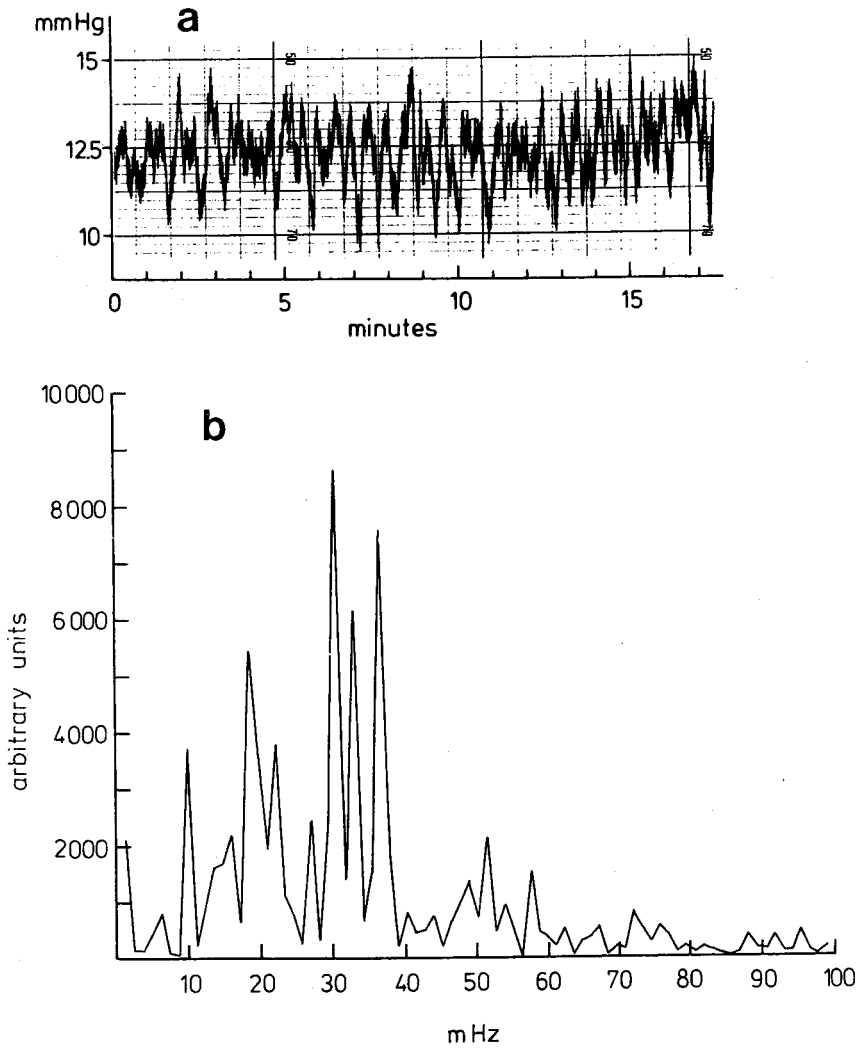


Figure 2
Experimental results for the variation in intratubular pressure for hypertensive rats (a).
Spectral distribution of the experimental results (b).

functional unit is the nephron of which there are 30000 in a rat kidney (1 mill. in a human kidney). For readers not familiar with renal physiology, figure 3 illustrates the main structure of a nephron.

The closed end of the nephron contains 20-40 capillary loops arranged in parallel. This anatomical configuration, the glomerulus, is supplied with blood through the afferent arteriole, a short vessel which is capable of changing its diameter and thereby the flow resistance. The glomerulus acts as a filter through which blood constituents with a molecular weight

of less than 68000 can pass into the proximal tubule. In rats, typically 25-35 % of the plasma is fittered out as the blood passes the glomerulus, and since blood cells and proteins are retained, the blood becomes more viscous, and at the same time, the colloid osmotic pressure rises. The blood leaves the glomerular system through the efferent arteriole to pass through a new capillary system which embraces the tubule.

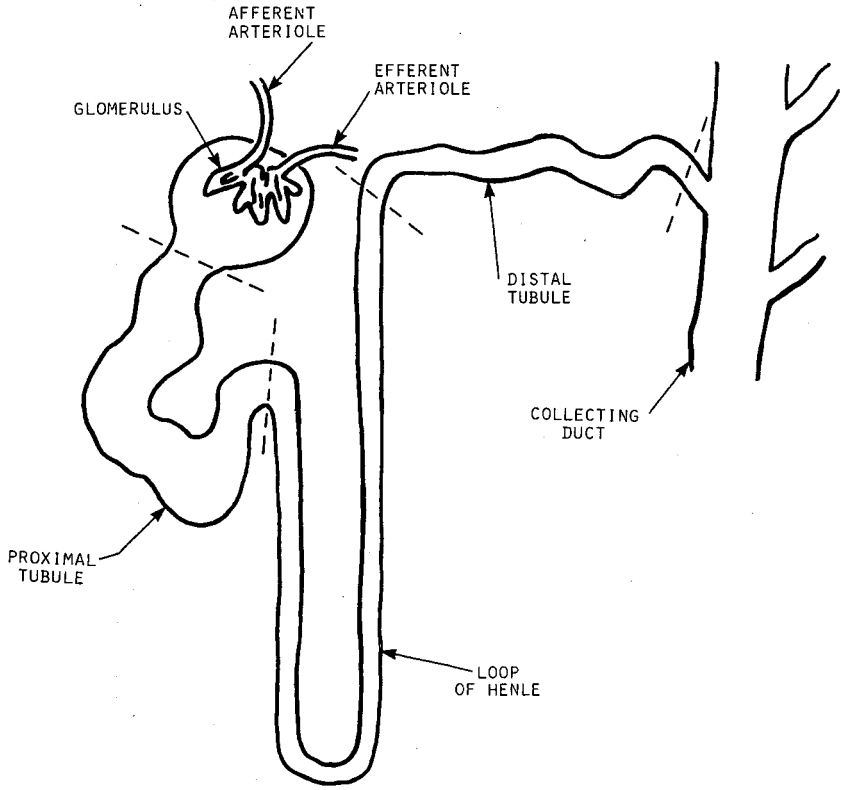


Figure 3.
The main structure of the functional unit of a kidney : a nephron.

By passive and active transport over the tubular wall, the constituents are changed as a consequence of selective reabsorption and secretion as the filtrate passes through the various segments of the tubule : the proximal tubule, the loop of Henle, the distal tubule, and the collecting duct. In practice, more than 95 % of the water and salts are reabsorbed while the metabolic biproducts remain in the tubular fluid and are excreted in the urine.

It has long been recognized that the kidneys are capable of compensating for changes in arterial blood pressure, and that this ability partly rests with feedback mechanisms in the individual nephrons. By adjusting the flow resistance in the afferent arteriole, the pressure in the proximal tubule can thus be kept at a relatively constant level[15,16].

Development of the so-called micropuncture technique in the 60'ies and 70'ies has made it possible to investigate intratubular pressures and flows in more detail. Various experimental designs have provided important information about different mechanisms involved in the total feedback system. In a typical experimental set-up[17,18], a paraffin block is inserted into the middle of the proximal tubule with micropipettes on either side. The steady state magnitude of the single nephron glomerular filtration rate (SNGFR), i.e. the flow of filtrate into the proximal tubule, is then determined as a function of an externally forced, step-wise increased flow into the loop of Henle. Presumably because of the implied assumption that the system is stable, most experiments have concentrated on steady state relations, and only with the recent measurements performed at the Institute for Experimental Medicine in Copenhagen[11,13] has it become clear that the total system can be unstable in its normal operation range.

Figure 4 gives an overview of the dominant feedback structure in nephron pressure- and flow regulation. The glomerular hydrostatic pressure is determined by division of the arterial to venous pressure difference between the afferent and the efferent arterioles. As a simplification, we consider the venous pressure and the efferent resistance as constants, and the glomerular hydrostatic pressure is therefore effectively determined by the arterial pressure and the flow resistance in the afferent arteriole. The difference between the glomerular filtration pressure (glomerular hydrostatic pressure minus colloid osmotic pressure) and the intratubular pressure controls the rate of filtration into the proximal tubule. The filtration rate again influences the two pressures by changing the respective stores. The proximal intratubular pressure determines the flow of filtrate into the loop of Henle, and via the juxtaglomerular apparatus, the Henle flow again determines the afferent arteriolar resistance.

If the Henle flow increases, the afferent arteriolar resistance also increases, and the glomerular filtration pressure and the rate of filtration are reduced. Altogether, the feedback is thus negative. Due to a relatively high gain factor, and due to a delay in the response at the juxtaglomerular apparatus to changes in the Henle flow, the system tends to exhibit self-sustained oscillations in its normal operation range. In the following we discuss the model assumptions in more detail as we present the applied equations one by one.

The blood flow to the glomerular capillaries is given by

$$BF = (P_a - P_g)/R_a \qquad (1)$$

where P_g is the glomerular hydrostatic pressure, and P_a and R_a are the arterial blood pressure and the afferent arteriolar resistance, respectively. The corresponding plasma flow may be found by multiplying the blood flow by a factor $(1-H_a)$, where H_a (the hematocrit) is the volume fraction of red blood cells in the arterial blood. In the glomerulus, some of the plasma is filtered and enters the proximal tubulus.

The glomerular capillaries are short low-compliant vessels with a transit time of only a few tenths of a second[19]. If C_a and C_e denote the plasma

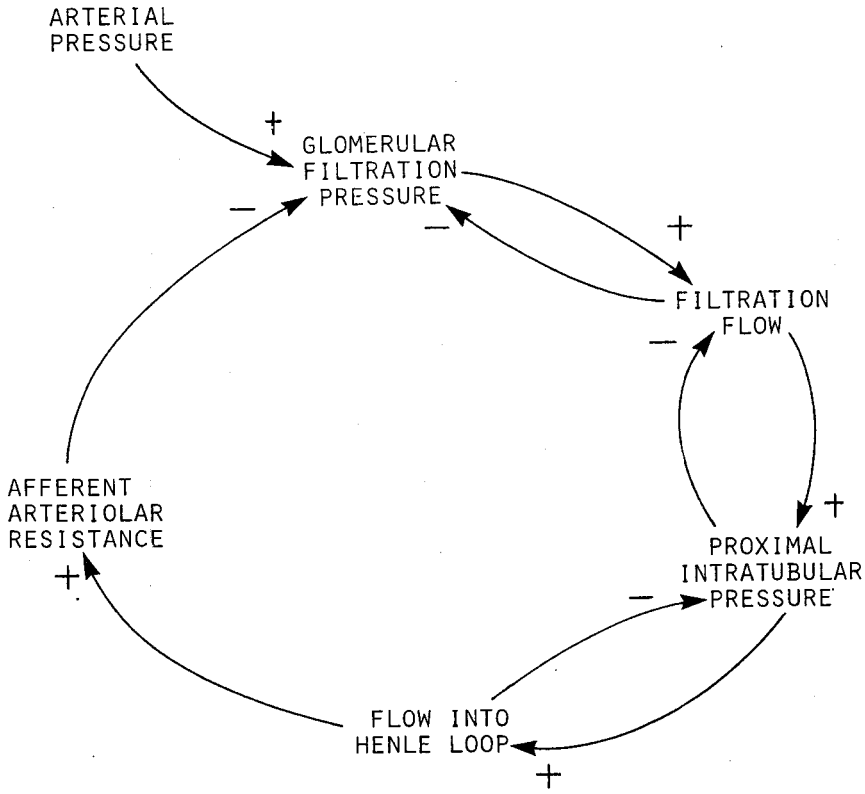


Figure 4
The main negative feedback involved in nephron auto-regulation.

protein concentration in afferent and efferent arteriolar blood, respectively, mass conservation requires :

$$SNGFR = BF\,(1-H_a)\,(1-C_a/C_e) \qquad (2)$$

SNGFR is here the filtration flow.

In order to evaluate C_e, we must consider the filtration process in more detail. Because of the difference in plasma protein concentration across the glomerular capillary wall, there is a colloid osmotic pressure difference in addition to the hydrostatic one. The colloid osmotic pressure as a function of the protein concentration is well represented by[20]

$$\pi(C) = aC + bC^2 \qquad (3)$$

with a and b being experimentally determined constants.

The glomerulus may now be idealized as a single capillary of length L. If F(x) denotes the flow of plasma through a cross section at position x along this capillary, the filtration process is described by[20]

$$-\frac{d\,F(x)}{dx} = (K_f/L)\,[P_g - P_t - \pi\,(C(x))] \tag{4}$$

where k_f, the ultrafiltration coefficient, is a measure of the hydraulic conductance of the capillary wall. P_t is the hydrostatic pressure in the proximal tubule. In the first part of the capillary, the colloid osmotic pressure amounts to about half the hydrostatic pressure difference $P_g - P_t$. As filtration takes place, $\pi(C)$ increases, and the net pressure difference decreases.

If we assume that pressure equilibration has occured before the end of the capillary, we have

$$\pi(C_e) = P_g - P_t, \tag{5}$$

and C_e is simply given by

$$C_e = \sqrt{(a/2b)^2 + (P_g - P_t)/b} - a/2b \tag{6}$$

Experimental studies[19] have shown that because of the short transit time, rat glomeruli can not always maintain complete filtration equilibrium. Eq. (6) must therefore be considered a first, although workable approximation. The shortness of the transit time compared with the oscillation period for the total system does permit the capillary flow to be considered as quasi-stationary such that $F(x)C(x) = F_aC_a$. This condition can be used to eliminate $F(x)$ from (4). By integration, an implicit expression for C_e is then obtained from which C_e can be determined by numerical means[14].

From division of the arterial to venous pressure difference over the afferent and the efferent resistances we have

$$P_g = P_v + R_e\,(BF - SNGFR) \tag{7}$$

where P_v and R_e are the venous pressure and the efferent resistance, respectively. R_e depends exponentially upon the effective hematocrit[21], and therefore on the magnitude of the filtration flow (SNGFR). This dependence has not been considered in our model which means that we have effectively assumed a smaller loop-gain than in the real physiological system. Combining Eqs. (1), (2), (6) and (7), we obtain a cubic expression for P_g) in terms of P_a, P_v, R_a, R_e, C_a, H_a, P_t, a and b. This equation is solved numerically for each simulation step in our model by means of a separate iteration procedure.

The proximal tubule is considered a compliant tube with little or no flow resistance. Hence, the intratubular pressure satisfies the equation

$$\frac{d}{dt}P_t = (SNGFR - F_{reab} - F_{hen})/C_{tub} \tag{8}$$

with F_{reab} and F_{hen} denoting the reabsorption in the proximal tubule and the flow into the loop of Henle, respectively. C_{tub} is the compliance of the tubule. Reabsorption in the proximal tubule is assumed to be constant, while the flow of filtrate into the loop of Henle is approximated by

$$F_{hen} = (P_t - P_d)/R_{hen} \tag{9}$$

with P_d and R_{hen} being the pressure in the distal tubule and the resistance of the loop of Henle, respectively. Both P_d and R_{hen} are taken to be constant. It should be noted, though, that experimental studies indicate a non-linear relation[22] between F_{hen} and P_t due to changes in P_d and due to the compliant nature of the Henle loop.

To close the feedback, we must express the afferent resistance R_a in terms of the Henle flow F_{hen}. The relation between these two variables can not be measured directly. By means of the previously mentioned open-loop experiments[17,18] it has been possible, however, to determine the steady state relation between F_{hen} and SNGFR. Using Eqs. (2), (6) and (7) it is then possible to express R_a in terms of F_{hen}. The results of such a calculation are shown in figure 5 where the transfer function measures R_a in terms of its equilibrium value.

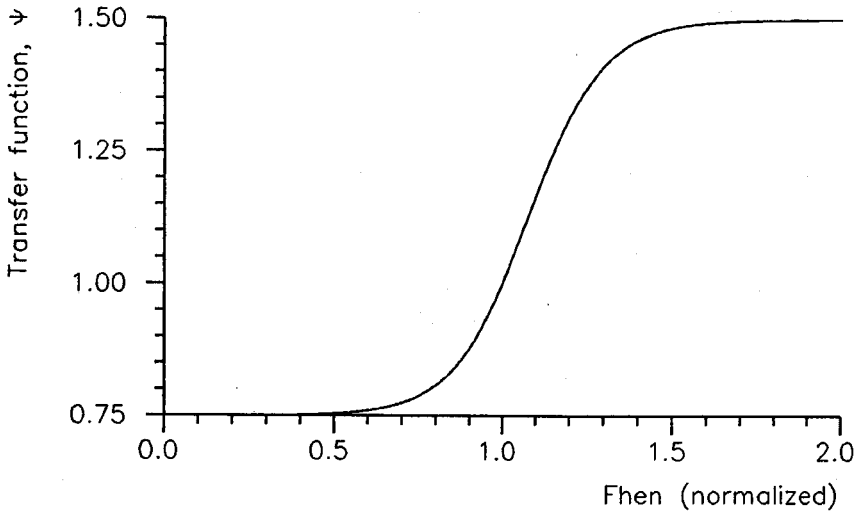


Figure 5
Steady state afferent resistance (normalized) as a function of the normalized Henle flow. The function is approximated by Eqs. (10) and (11) with $\psi_{max} = 1.50$, $\psi_{min} = 0.75$ and $\alpha = 9$.

The S-shaped character of the transfer function is expressive of the physiological restrictions on both the maximum and the minimum arteriolar diameter. In the model we have used the following sigmoidal approximation to the steady state relation between the normalized afferent resistance and the normalized Henle flow[15].

$$\psi(y) = \psi_{max} \frac{\psi_{max} - \psi_{min}}{1 + \exp(\alpha(y-s))} \tag{10}$$

with

$$S = 1 - \alpha^{-1} \ln\{(1 - \psi_{min})/(\psi_{max} - 1)\} \tag{11}$$

y is here the Henle flow normalized relative to its equilibrium value. ψ_{max}, ψ_{min} and α are parameters defining the upper and lower saturation levels and the steepness of the response, respectively.

The physiological mechanisms responsible for adjusting the afferent arteriolar resistance R_a in response to changes in the Henle flow are not known in detail. A widely accepted theory[16,23] is that the arteriolar wall is stimulated by a signal arising from specialized cells situated where the terminal part of the loop of Henle passes the afferent arteriole. Some of these cells, the macula densa cells, are assumed to be sensitive to Na^+– and Cl^-–ions, and the activity is assumed to depend upon the Na^+–concentration of pre-urine leaving the loop of Henle[23]. Due to active absorption in the loop of Henle, the Na^+- and Cl^-– concentrations in the exiting fluid will depend on the flow rate.

The sodium concentration at the macula densa cells does not change immediately upon a change in the Henle flow. A further delay is introduced in transmission of the signal from the macula densa cells to the afferent arteriole, presumably because this transmission involves a series of processes. In the model, these delays are described as a single third order lag with a delay time T between F_{hen} and ψ.

Only limited information is available about the dynamical aspects of arteriolar response to external excitation. However, it has been observed[24] that step-wise changes in vasoconstrictor nerve activity can give rise to large-amplitude, damped oscillations in arteriolar resistance with typical frequencies around 0.03 Hz. The simplest equation capable of describing this feature is a second order ordinary differential equation. For physiological reasons it is reasonable to assume, that both the force constant and the damping increase when the diameter of an arteriole approaches both its maximum and its minimum values.

In the model we have assumed that the inherent oscillatory behaviour of the afferent resistance can be described by

$$\frac{d^2}{dt^2} R_a + 2d\,\omega\,\sqrt{K(R_a)}\,\frac{d}{dt} R_a + \omega^2\,K(R_a)\,(R_a - \psi R_{ao}) = 0 \qquad (12)$$

where $K(R_a)$ is the force konstant normalized relative to its value in equilibrium. d is the damping coefficient, and ω the resonance angular frequency near equilibrium. R_{ao} is the equilibrium value of the afferent resistance.

Figure 6 shows the assumed variation of the relative force constant $K(R_a)$ as a function of R_a/R_{ao}. Figure 7 illustrates the reaction of the normalized arteriolar resistance to a 10 % increase in the transfer function ω. Isolated from the rest of the system, our model resistance is seen to respond to external stimuli by performing damped oscillations with a period of about 30 sec.

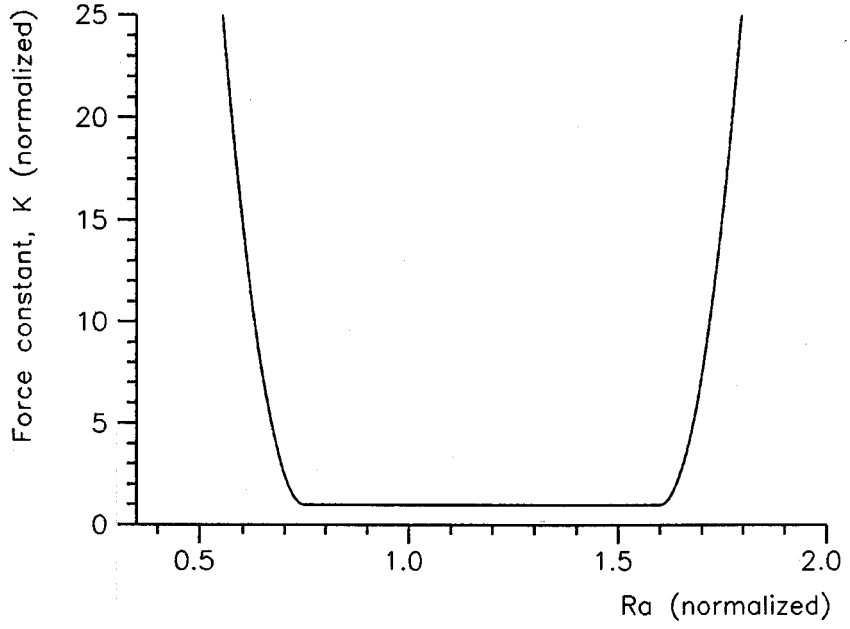


Figure 6
Variation of the relative force constant $K(R_a)$ as a function of the normalized arteriolar resistance.

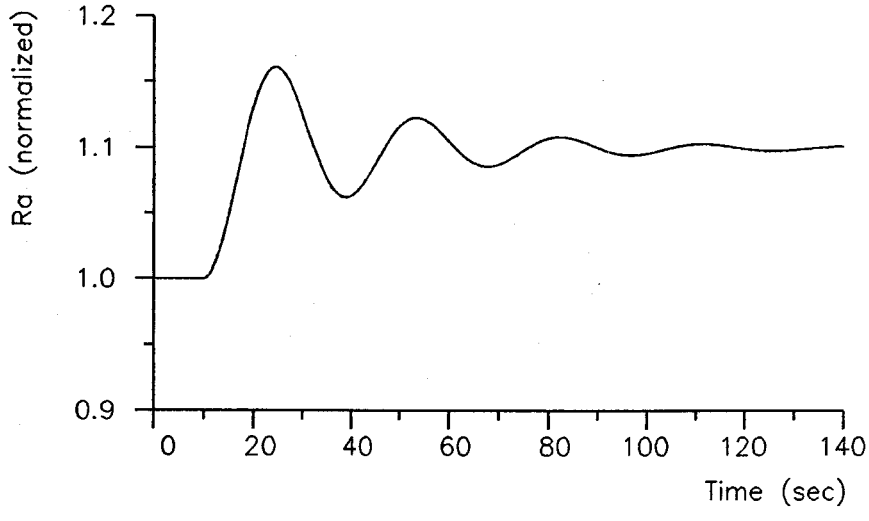


Figure 7
Step response of the afferent resistance.
The transfer function ψ is increased from 1.0 to 1.1 at t=10 sec.

3. SIMULATION RESULTS

Our model has been simulated with the following standard values of the various parameters :

P_a	= 100 mmHg	R_{ao}	= 18.3 mmHg·sec/nl
P_v	= 10 mmHg	R_e	= 14 mmHg·sec/nl
P_d	= 4.5 mmHg	R_{hen}	= 40 mmHg·sec/nl
F_{reab}	= 0.3 nl/sec	F_{heno}	= 0.2 nl/sec
C_a	= 54 g/l	H_a	= 0.5
a	= 0.163 mmHg·l/g	b	= $2.9 \cdot 10^{-3}$ mmHg·l^2/g^2
C_{tub}	= 0.4 nl/mmHg	T	= 4.5 sec
ω	= 0.22 sec^{-1}	d	= 0.16
ψ_{max}	= 1.5	ψ_{min}	= 0.75
α	= 9		

Values available in the literature[14,17,21] have been used where possible. Our base case simulation is therefore believed to correspond to normal physiological conditions. The damping coefficient d may be a little smaller than found experimentally[24].

In comparison with the experimental results in figure 1, figure 8 presents the results obtained with our model for the above parameter values. 8a gives the temporal variation of the intratubular pressure, and 8b shows the corresponding power spectrum. Note, that we are here using a logarithmic scale. The integration was performed with a fourth-order Runge-Kutta method using a fixed step size of 0.05 sec. To generate the power spectrum, the calculated intratubular pressure was sampled each 0.1 sec over a period of 819.2 sec and hereafter Fourier transformed.

Figure 8c shows a projection of the phase portrait into the plane defined by the afferent resistance R_a and the intratubular pressure P_t. It is clear, that because of the delayed, negative feedback in the model, the inherent tendency of the afferent resistance to oscillate has been destabilized, and the total system performs self-sustained oscillations with a period of about 25 sec.

Indirect evidence suggest that genetically hypertensive rats have a reduced transport rate of Na^+–ions across the macula densa wall[13]. In our model, this corresponds to an increased delay time T. Moreover, hypersensitive rats are supposed to have hypersensitive arteriolar walls. This corresponds to a decreased damping coefficient d and an increased steepness α for the transfer function Ψ. In our model we have used a damping coefficient which is somewhat smaller than reported experimental values for normal arterioles[24]. We have hereafter investigated changes in model behaviour as the delay time T is gradually increased from its normal value of 4.5 sec to 9.0 sec.

Figures 9–11 display the obtained simulation results for a delay time of T=6.5 sec, 7.46 sec, and 9.0 sec, respectively. In each figure we have shown (a) the proximal tubular pressure P_t as a function of time, (b) the logarithmic power spectrum of $P_t(t)$, and (c) a projection of the phase portrait into the R_a–P_t–plane. It is observed, how the model develops through a series of period-doubling bifurcations to reach a chaotic state

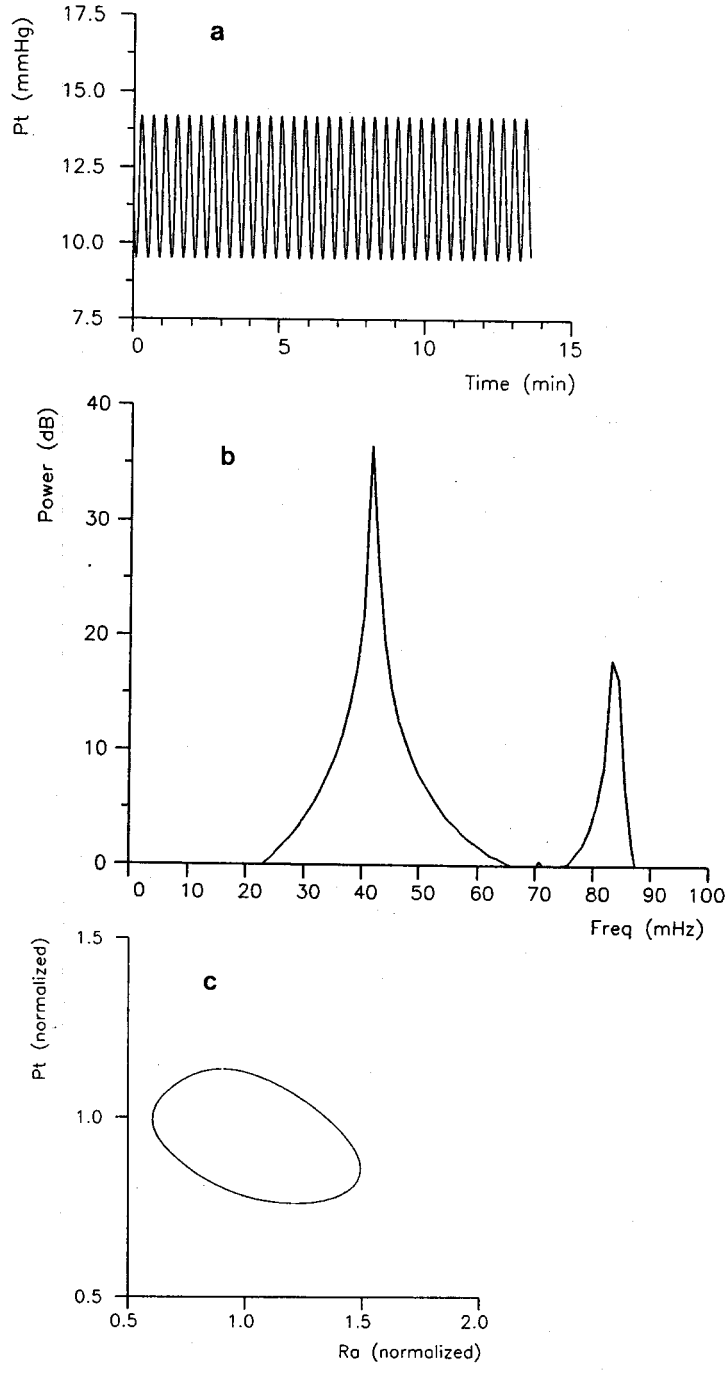


Figure 8
Simulation results obtained with standard values of the various parameters :
a) the proximal intratubular pressure as a function of time
b) logarithmic power spectrum of signal in a)
c) phase plot obtained from the time series of P_t and R_a. The model has an unstable equilibrium point in (1,1).

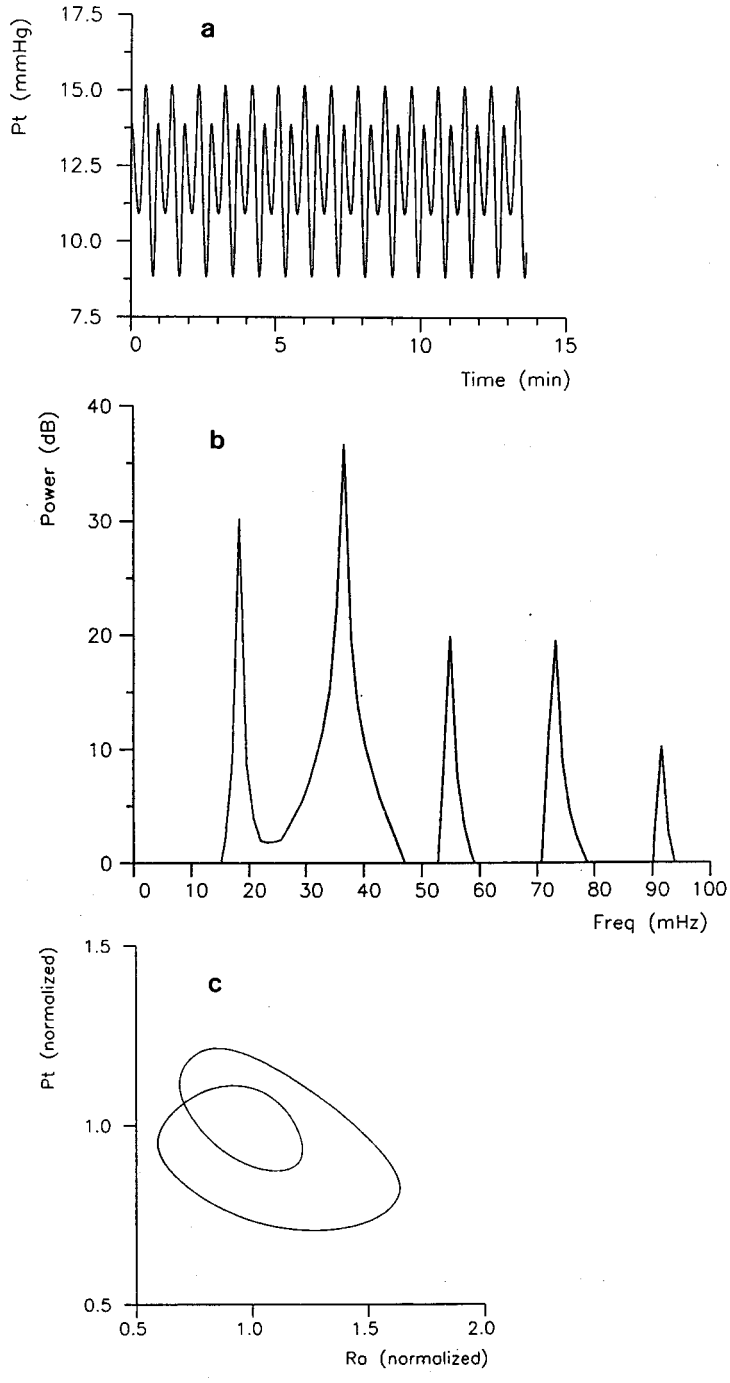


Figure 9
Simulation results similar to those of figure 8, only with a delay time of T= 6.5 sec. In figure 8 the delay time is T= 4.5 sec. The first period-doubling has taken place.

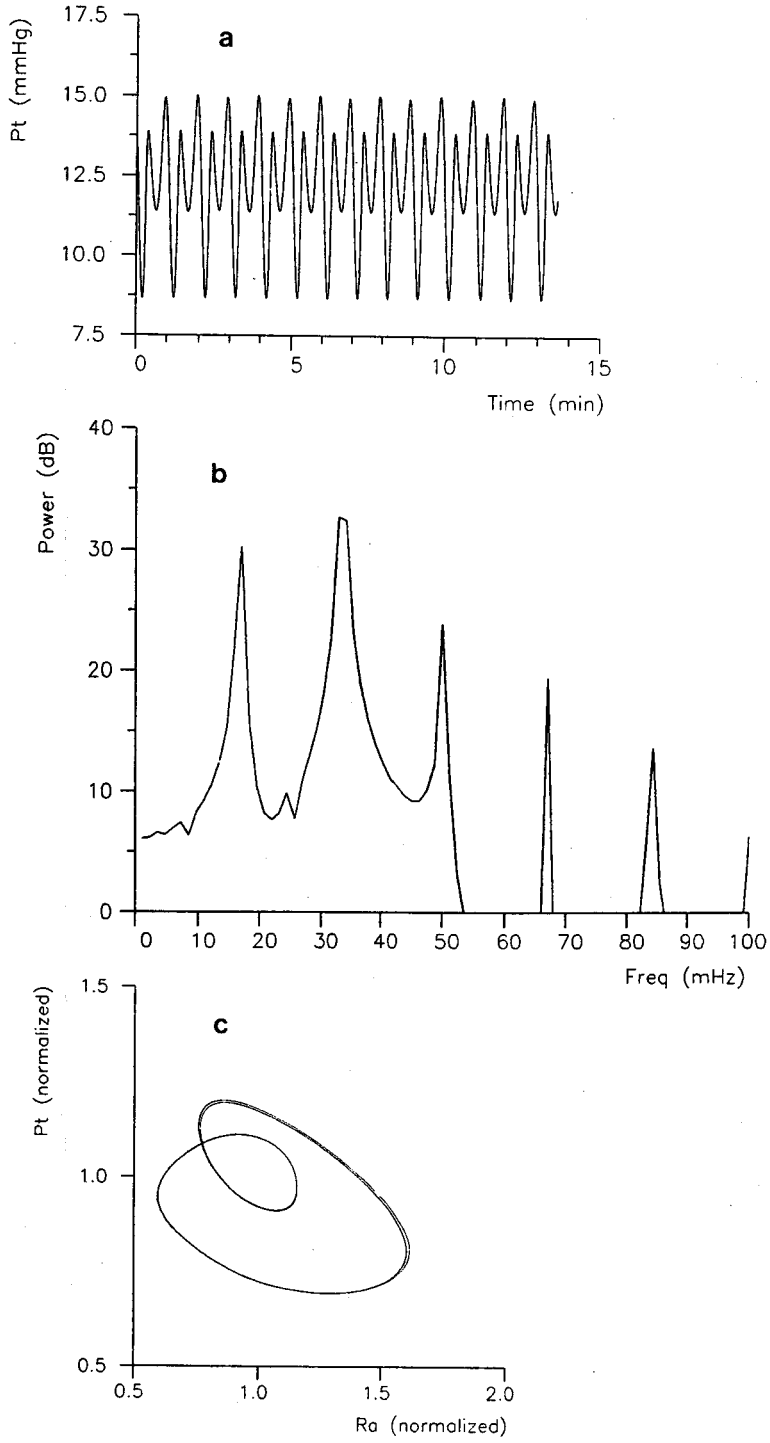


Figure 10
With a delay time of T=7.46 sec, a second period-doubling has occured.

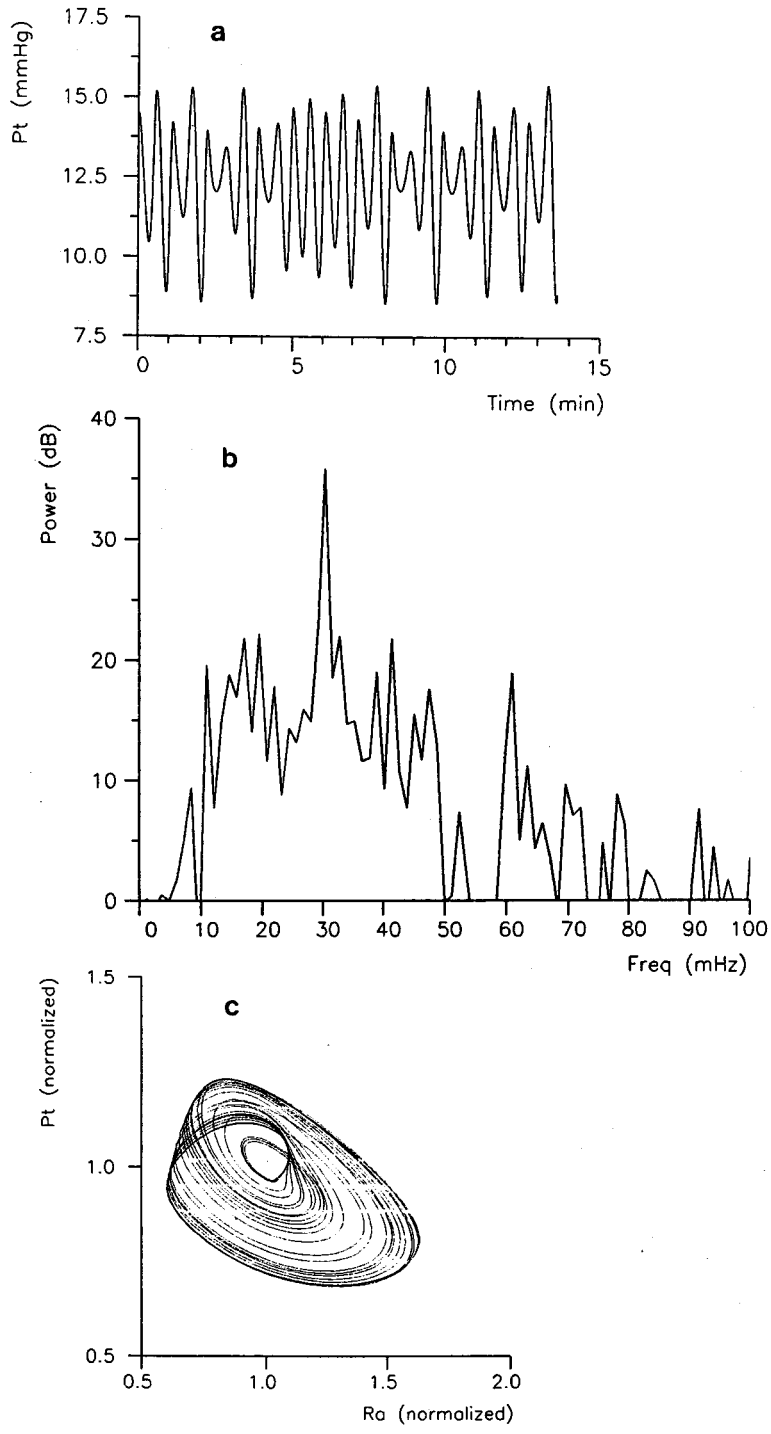


Figure 11
With a delay time of T=9.0 sec, the model has become chaotic.

before T=9.0 sec. We also observe that there is a reasonable, qualitative agreement between the spectral distribution of $P_t(t)$ obtained with the model, and the experimental distribution in figure 2.

The chaotic phase plot of figure 11 has some of the typical appearance of a Rössler band[25]. This is to be expected from the structure of our model. It is still a question, of cause, whether the measured irregular pressure variations represent chaotic behaviour in the biological feedback system. We are presently studying this question in more detail by constructing phase portraits, Poincaré sections and return maps for the measured time series.

4. DISCUSSION

As discussed in the introduction, instabilities are increasingly being recognized as playing a significant role in the regulation and function of normal physiological systems. From a thermodynamic point of view, these instabilities are caused by the fact that physiological systems are open systems which are maintained in a far-from-equilibrium condition through dissipation of energy[6].

In a nephron, for instance, the cells in the tubular wall depend on a steady supply of energy to perform their reabsorption of water and salts from the filtrate. In addition, the heart continuously pumps blood through the afferent-efferent arteriolar system and hereby maintains a hydrostatic pressure difference between capillary blood and intratubular fluid which permits filtration to take place.

Since most physiological rhythms presumably have evolved under selective pressure, one can speculate about the advantages to the physiological system of an oscillatory behaviour. In certain cases, the same structure can perform different functions during various phases of a cycle. This may also be associated with a need for an accurate syncronization of physiological processes. In other cases, the overall efficiency can be improved (the avarage energy requirements reduced) through periodic shifts between an active an a dormant phase. It is also possible that biological rhythms can serve to gard a system against long-term drift.

We have not so far been able to point to particular advantages of the observed oscillatory behaviour in nephron pressure- and flow regulation. During regulatory changes in R_a, however, the integrated net error in renal filtration volume may be minimized for a given (relatively long) delay time in the feedback loop by having a large number of incoherent oscillations instead of a stable, synchronous regulation. It is also possible that the filtration capacity is increased by driving the nephron into an oscillatory state.

REFERENCES

1. A.T. WINFREE, *The Geometry of Biological Time*, Springer, New York, 1980.

2. F.P. ALFORD, H.W.G. BAKER, Y.C. PATEL, *Temporal Patterns of Circulating Hormones as Assessed by Continuous Blood Sampling*, J. Clin. Endocrinol. Metab. *36*, 108-116, 1973.

3. W.R. SMITH, *Hypothalamic Regulation of Pituitary Secretion of Luteinizing Hormone-II, Feedback Control of Gonadotropin Secretion*, Bull. Math. Biology *42*, 57-78, 1980.

4. J.F. SASSIN, D.C. PARKER, J.W. MACE, *Human Growth Hormone Release : Relation to Slow-Wave Sleep and Sleep Waking Cycles*, Science *165*, 513-515, 1969.

5. W.C. ORR, H.J. HOFFMANN, F.W. HEGGE, *Ultradian Rhythms in Extended Performance*, Aerospace Med *45*, 995-1000, 1974.

6. G. NICOLIS, I. PRIGOGINE, *Self-Organization in Nonequilibrium Systems*, Wiley, New York, 1977.

7. M.J. FEIGENBAUM, *Universal Behaviour in Nonlinear Systems*, Los Alamos Science 1, 4027, 1980.

8. O. DECROLY, A. GOLDBETER, *Birhythmicity, Chaos, and Other Patterns of Temporal Self-Organization in a Multiply Regulated Biochemical System*, Proc. Natl. Acad. Sci. USA *79*, 6917-6921, 1982.

9. J.P. ECKMANN, *Roads to Turbulence in Dissipative Dynamical Systems*, Rev. Modern. Phys. *53*, 643-654, 1981.

10. H. HAYASHI, S. ISHIZUKA, M. OHTA, K. HIRAKAWA, *Chaotic Behaviour in Ochidium Giant Neuron under Sinusoidal Stimulation*, Phys. Lett. *88A*, 265-266, 1982.

11. P.P. LEYSSAC, L. BAUMBACH, *An Oscillating Intratubular Response to Alterations in Henle Loop Flow in the Rat Kidney*, Acta Physiol. Scand, *117*, 415-419, 1983.

12. N.-H. HOLSTEIN-RATHLOU, P.P. LEYSSAC, *TGF-Mediated Oscillations in the Proximal Intratubular Pressure : Differences between Spontaneously Hypertensive Rats and Wistar-Kyoto Rats*, Acta Physiol. Scand., 126, 333-339, 1986.

13. P.P. LEYSSAC, N.H. HOLSTEIN-RATHLOU, *Effects of Various Transport Inhibitors on Oscillating TGF Pressure Responses in the Rat*, Pfluegers Arch. (in press).

14. N.-H. HOLSTEIN-RATHLOU, P.P. LEYSSAC, *Oscillations in the Proximal Intratubular Pressure : A Mathematical Model*, submitted to Am. J. Physiol.

15. P.K. JENSEN, K. STEVEN, *A Model Study of Tubuloglomerular Feedback Mechanism : Effector Site and Influence on Renal Autoregulation*, Acta Physiol. Scand. *115*, 295-300, 1982.

16. J.P. BRIGGS, F.S. WRIGHT, *Feedback Control of Glomerular Filtration Rate : Site of the Effector Mechanism*, Am. J. Physiol. *236*, F40-F47, 1979.

17. J. SCHNERMANN, F.S. WRIGHT, J.M. DAVIS, W. von STACKELBERG, G. GRILL, *Regulation of Superficial Nephron Filtration Rate by Tubuloglomerular Feedback*, Pfluegers Arch. *318*, 147-175, 1970.

18. J.P. BRIGGS, *A Simple Steady State Model for Feedback Control of Glomerular Filtration Rate*, Kidney Intern. *22*, Suppl 12, 143-150, 1982.

19. M. STEINHAUSEN, B. ZIMMERHACKL, H. THEDERAN, R. DUSSEL, N. PARIKH, H.-V. ESSLINGER, G. von HAGENS, D. KOMITOWSKI, F.D. DALLENBACH, *Intraglomerular Microcirculation : Measurements of Single Glomerular Loop Flow in Rats*, Kidney Intern. *20*, 230-239, 1981.

20. W.M. DEEN, C.R. ROBERTSON, B.M. BRENNER, *A Model of Glomerular Ultrafiltration in the Rat*, Am. J. Physiol. *223*, 1178-1183, 1972.

21. P.K. JENSEN, O. CHRISTENSEN, K. STEVEN, *A Mathematical Model of Fluid Transport in the Kidney*, Acta Physiol. Scand. *112*, 373-385, 1981.

22. D.A. HABERLE, J.M. DAVIS, *Interrelationship Between Proximal Tubular Hydrodynamics and Tubuloglomerular Feedback in the Rat Kidney*, Kidney Intern. *22*, Suppl. 12, 193-197, 1982.

23. K. THURAU, *Influence of Sodium Concentration at Macula Densa Cells on Tubular Sodium Load*, N.Y. Acad. Sci. *139*, 338-399, 1966.

24. M. ROSENBAUM, D. RACE, *Frequency-Response Characteristics of Vascular Resistance Vessels*, Am. J. Physiol. *215*, N° 6, 1397-1402, 1968.

25. O.E. RÖSSLER, *An Equation for Continuous Chaos*, Phys. Lett. 57A, 397-398, 1976.

Reversibility in the Light of Evolution

Michael Conrad

Departments of Computer Science and Biological Sciences — Wayne State University — Detroit, Michigan, USA

1. DARWIN VERSUS NEWTON

Darwin and Wallace formulated their theory of evolution by variation and natural selection in terms of concepts such as organism, inheritance, environment, and mutability. Attempts to express the theory in physiochemical terms have in recent years attracted a great deal of attention[1,2]. This has been an important step in the direction of the unification of physics and biology. At the same time it has served to sharpen the dissonance between the conceptual schemes that dominate in these two areas of science.

The chief point of conflict is between the probabilistic nature of evolutionary theory and the determinism of microscopic equations of motion[3]. The explanatory power of evolutionary theory rests on the fact that it depends on no extrinsic source of intelligence. It is thus essential that the variations be random. Nonrandom variations slow down the rate of evolution since they prevent a system from exploring all regions of its phase space. If the variations could direct the evolution in particular ways the theory would lose its explanatory power. Explanations would have to be sought in the deeper microscopic dynamics, and reasons would have to be given as to why these dynamics were set up so intelligently. Alternatively, if evolutionary theory is to be maintained in its current form in the face of a microscopic dynamics that is truly deterministic and reversible, matters would have to be organized in just such a way that all this underlying detail is irrelevant to the main features of the outcome.

We can choose to view the fluctuations as representing our ignorance about the detailed dynamics of a system of particles. But this puts us in the peculiar posture of basing the fundamental theory of biology on a postulate of serendipitous ignorance. This postulate imposes an interpretation on organic evolution, human cognition, and human history that is inherently arguable from the philosophical point of view. Yet it cannot be derived in a rigorous way from reversible mechanics.

Could the postulate of ignorance be replaced by an epistemologically more satisfactory or at least philosophically more neutral postulate ? It is evident that the structure of physics depends quite delicately on assumptions about reversibility and the origin of uncertainty. Many physicists are understandably reluctant to introduce features that might disturb the consistentcy of this structure, insofar as it is in fact consistent. On the other hand, it is also necessary to consider the conceptual coherence of science as a whole. Recent interest in the problem of introducing irreversibility into microscopic dynamics[4,5] has perhaps been motivated by this broader consideration. I will follow along this line in the present paper, first looking at the alternatives from the standpoint of the requirements of evolutionary theory and then considering how the reversible dynamics displayed by simple physical systems might be a special case of an irreversible dynamics in which measurement is an intrinsic feature.

2. DARWIN AND DISSIPATIVE STRUCTURES

Organisms are capable of reproducing with variation. The environment selects which organisms actually reproduce by virtue of the fact that the resources necessary for reproduction are limited. According to the Darwin-Wallace mechanism the historical development of biological forms is a consequence of the variation and selection process (subject to the additional proviso that acquired characteristics are not inherited in any significant way).

To express this in somewhat more physical sounding language let us suppose that the organism is an open system defined by nonlinearities, constraints, and tailored boundary conditions. Even in the absence of any unusual boundary conditions it is possible for ordered dynamical states to appear in the presence of suitable nonlinearities. Darwinian evolution occurs when the regime is quasi-autocatalytic. The deviations from perfect autocatalysis are connected with fluctuations that alter the defining nonlinearities and boundary conditions. The finiteness of the environment is represented as a global constraint on the available resources. This distorts the rate constants of the autocatalytic regimes, leading to a historical development that corresponds to the classificatory effect of natural selection in the Darwin-Wallace terminology. The accumulation may be pictured as a sequence of dissipation increasing fluctuations that drive the system to more organized regimes, followed by a dynamics that is dominated by minimum entropy production[2].

Fluctuations occur because transitions from typical microscopic configurations to atypical ones become more and more likely as the spatial scale becomes smaller. Such local violations of the second law of thermodyna-

mics are the basis of all chemical reactions. They control the manner in which the system approaches equilibrium, but have no effect on the final equilibrium if the system is closed. If the system is open the addition of a catalyst can influence the final state in a significant way. Evolution occurs when the dynamics are sensitive to the fluctuations. Sensitivity here means that the nonlinearities and constraints that define the dynamics amplify the effects of fluctuation to such an extent that they are themselves altered by them. This picture makes it clear that fluctuation amplified by constraint is the essence of evolutionary search.

We can choose to view these fluctuations as representing our ignorance about the microscopic details. But as already observed, this puts us in the position of basing the fundamental theory of biology on a postulate of ignorance. It is important to recognize that the issue here has to do with whether probabilistic events have a deterministic description, not with the entropy changes that occur in the course of evolution. Entropy will increase whenever a system is initially in an atypical microscopic configuration, one that corresponds to a state of low macroscopic symmetry. The chance of making a transition from the atypical to the typical set of microscopic configurations is vastly greater than the chance of making the reverse transition. Nothing special need be assumed about the initial atypical state for a global entropy increase to drive a local entropy decrease, other than that it correspond to a state of high potential energy. Why the initial state should be atypical is as much a problem for the student of cosmic evolution as for the student of biological evolution.

3. ALTERNATIVE INTERPRETATIONS

The ergodic hypothesis provided the classical underpinning of the postulate of ignorance. The supposition is that in the absence of constraints the state of a system will map into itself in such a way as to become completely mixed. If so, the system would spend an equal amount of time at each accessible point in its phase space. Strictly speaking this will not happen in three dimensions, though it might be a good enough approximation to yield macroscopic thermodynamics (see for example[6]).A system with irreversible dynamics would also undergo mixing through an ergodic mechanism. In this case the increase in entropy would have a basis both in the microscopic dynamics and in the passage to a macroscopic description. If the ergodic argument were rigorously true the probabilities that form the basis of evolutionary theory would in principle be derivable from an underlying deterministic dynamics without introducing any de novo physical assumption. This is not the case, however.

There are at least three other ideas pertinent to the origin of irreversibility. The first comes from computer science. Computers are themselves physical systems. How much can be computed is subject to physical limitations on space, time, speed of operation, and reliability[7,9]. The dynamics of a real physical system will always outstrip the ability of a programmable machine to compute its behavior[10,11]. This confers some degree of objectivity on the observer's loss of information about microscopic details.

The discovery of deterministic chaos suggests that dynamical instabilities could be a source of mixing. The microdynamics would have to be nonlinear in a way that roughly speaking is associated with stretching, folding, and rotation[12]. Alternatively, it is possible to suppose that the microscopic dynamics are reversible, but subject to constraints that confer initial condition sensitivity. If the system has a chaotic attractor it is truly irreversible. Dynamical instabilities would not confer a fundamental status on the probabilities that occur in evolutionary theory, but they would make these probabilities effectively irreducible to any finer level of description. Ascertaining whether probabilities in fact have their origin in dynamical instabilities or in irreducible accidents is probably undecidable from the operational point of view. It also implies that the question of whether microscopic reversibility extends to macroscopic systems is open since in practice it would be impossible to perform a calculation that would either confirm or disconfirm this extension.

Measurement — the process of obtaining information about the microsystem — provides the third major mechanism of irreversibility. In standard quantum theories the development of the wave function in time is deterministic and reversible. However, in order to connect the wave function to observation it is necessary to apply a projection operator that is entropy increasing and therefore in principle underivable from the equations of motion without additional assumption. The probabilities that occur in evolutionary theory would have a fundamental, irreducible basis if they were generated as a consequence of measurement (assuming hidden variable interpretations are excluded). This is especially attractive since organisms appear to make measurements[13]. There are serious conceptual problems, however. If an observed system, such as an organism, obeys the equation of motion, how can it make measurements ? If one chooses to view the randomization accompanying the measurement as an ergodic phenomenon it is necessary to fall back on a postulate of ignorance, in which case the primary experience of measurement assumes an illusory status. Or one has to admit some breakdown of microscopic reversibility and of the quantum mechanical superposition principle as well.

4. FORCE VERSUS MEASUREMENT

Let us briefly review the quantum mechanical equations for motion and measurement. The time dependent Schrödinger equation may be written

$$H_{op}\,\psi\,(x,t) = i\hbar\frac{\partial\psi(t)}{\partial t} \qquad (1)$$

where $\psi(x,t)$ is the wave function and H_{op} is the Hamiltonian operator corresponding to the sum of the kinetic and potential energies. This is a unitary transformation in the sense that $\psi(x,t)$ can be regarded as undergoing only rotations and mirror reflections in a given coordinate system. The wave function itself can be expressed as a superposition of a complete set of eigenfunctions

$$\psi\,(x,t) = \Sigma_n C_n u_n(x) \qquad (2)$$

where the $u_n(x)$ are eigenfunctions and the coefficients C_n might equal $e^{(C_n - iH_n t/\hbar)}$. If a system can be described by the wave function (2) it is a pure case. If less information is available the system might be described as a mixture of wave functions

$$\psi^k(x,t) = \sum_k w_k \psi_k(x,t) \tag{3}$$

where there are k wave functions each with probability w_k. The past and future of a pure case and of a mixture are both uniquely determined.

The projection postulate that relates the wave function to ordinary experience may be expressed as

$$\bar{a} = \int_{-\infty}^{\infty} \psi^*(x,t) A_{op} \psi(x,t) dx \tag{4}$$

where $\bar{a}$ is the expectation value of some variable, and A_{op} is the operator corresponding to this variable. Eq. (4) and the superposition expressed by Eq. (2) are consistent if $C_i^* C_i$ is interpreted as the probability that the eigenvalue associated with the eigenfunction $u_i(x)$ will be produced by a measurement. This is why the process of measurement may be thought of as producing a randomization of phases such that all the eigenfunctions other than $u_i(x)$ disappear. The loss of information about relative phases associated with this reduction of the wave function is associated with an increase in entropy. In general measurement transforms a pure case into a mixture, or alters a mixture so that its ensemble of wave functions is enlarged. This irreversible transformation is inconsistent with the equation of motion (1) since a mixture can never be produced by a superposition of pure cases[14]. Of course the entropy will also increase through an ergodic mechanism if the initial pure state is not a member of the typical set that corresponds to the situation of greatest macroscopic symmetry.

Two types of influences are clearly at work. The first, expressed in Eq. (1), derives from Newton's idea that the change in the change produced by an influence is constant (influence $\propto \frac{d^2x}{dt^2}$). This was the departure from Aristotelian physics that led to the problem of recapturing the irreversibility that predominates in the macroscopic world. The second influence derives from the unpicturability of the world in terms of classical variables. The conversion from the unpicturable superposition of states to picturable variables reintroduces irreversibility and also probability into physics in a fundamental way through the square law measurement operation. The standard procedures of quantum mechanics keep the two types of influence separate. The reversible « time evolution » of the system, generated by H_{op}, is governed through the potential function by fundamental forces of nature (electromagnetic, gravitational, strong, weak). The influences connected with measurement enter only when the wave function is collapsed by the experimenter, and this whole process is aggregated into the projection operation.

5. A MIXED MODEL

Now I would like to consider how it might be possible to construct a model that combines the reversibility of time evolution in a classical force field with the irreversibility of measuremennt. The ground rule is

that the irreversibility should not be introduced in an artificial way, for example by merely introducing a dissipative term into the potential. Thus the irreversibility should enter through a collapse process that confers a fundamental status on probabilities (because of our interest in biological phenomena such as evolution). I am not going to approach the problem by attempting to develop a general theory. Instead I will propose a mechanistic picture in which force depends on vacuum structure in a way that would lead to wave function collapse in the nonequilibrium state, but that would allow for reversible time evolution in a stable situation. The model is in its nature naive, first because it is picturable, and second because it leads to a variety of open questions that could only be dealt with adequately through a fully developed theoretical analysis.

The basis of our picture is a Dirac type vacuum, comprising vacuum fermions in a negative energy state. Forces between positive energy particles (to be called absorbers) are mediated by exchange of virtual bosons, as in conventional field theories. However, bosons will be identified with virtual pair production processes that propagate in this vacuum (or plenum) in a chain-like manner rather than with undecomposable oscillations. Such fluctuations can originate or terminate in the neighborhood of positive energy particles, but not in free space due to the requirement for conservation of momentum. Roughly speaking, propagation occurs as a consequence of the fact that macroscopic violations of conservation of energy and momentum should be forbidden. Force in this picture is dependent on the density of vacuum fermions. This density is influenced by positive energy mass and by charge. In a static situation the density is fixed and the dependence of the potential on vacuum density is hidden. In an out of equilibrium situation (resulting from significant changes in the distribution of mass and charge) the dependence on the vacuum density appears, and this means a dependence on the wave function of the vacuum. This introduces a complex (absorptive) potential, or alternatively, a measurement type square law into the potential.

Consider two absorbing particles separated by a distance r and denote the probability that a virtual excitation initiated in the neighborhood of one of them is absorbed in the neighborhood of the other by P(i,f). The force between these two particles should be proportional to this probability. Let us tentatively assume that the virtual energy exchange that must occur over any given interval of time in order to achieve any given coupling strength increases linearly with distance. The simplest linear dependence is given by $\Delta E_{req} = Kr$, where K is a proportionality constant with dimensions of force. This is compatible with the fact that the coupling strength should depend only on the momentum exchanged and with the fact that ΔE_{req} should decrease with distance (since the coupling strength decreases with distance). Since ΔE_{req} is supplied by a fluctuation energy we can also write $\Delta E_{req} \sim \hbar/\tau$, where ΔE_{req} corresponds to σE, the standard deviations of the energies that would be obtained by measurements made at some point in the time interval τ. Setting $Kr = Kc\tau$ and equating this to $\hbar/\tau$ gives $\tau^2 \sim \hbar/Kc$. The linearity assumption thus implies that the fluctuation time is fixed by K, independent of

the virtual fluctuation in energy ΔE_{req}. This agrees with our naive mechanistic picture of the exchange process as a series of N steps each extending over a length equal to the distance between the vacuum fermions. Denoting this length by r_K, we can write $r = Nr_K = Nc\tau_K$, where the subscripts denote the fact that the fluctuation intervals are fixed by K.

Let p(a,b) denote the probability that a vacuum particle initially in the highest negative energy state is promoted to a positive energy state as a result of its undergoing an uncertainty fluctuation in the neighbourhood of a positive energy particle or, in free space, as a result of a transfer of energy and momentum from a neighbouring transient pair production process. According to first order, time dependent perturbation theory the probability that a system undergoes a transition from state a to state b in a time interval $t = \tau_K$ in the presence of a perturbation constant over that time interval is given by the scattering formula

$$p(a,b) = 4|\langle b|V|a\rangle|^2 \sin^2 \{1/2\ (\Delta E/\hbar)\ \tau_K\} / \Delta E^2 \qquad (5)$$

where $\Delta E = E_a - E_b$. The probability of undergoing a transition increases with τ_K, in accordance with intuition. However, as τ_K increases, the probability of undergoing a transition to an equal energy state increases as τ_K^2, consistent with the fact that conservation of energy should hold rigorously over macroscopic times (see[15,16]). If τ_K is really the time interval of the fluctuation we can take ΔE as $\sigma E \sim \hbar / \tau_K$. Eq. (5) then reduces to

$$P(a,b) \sim \frac{Ar_K^2}{C^2\hbar^2} \qquad (6)$$

where $A = .92|\langle b|V|a\rangle|^2$.

The matrix element, $\langle b|V|a\rangle = \int \psi_b^* V_{op} \psi_a d^n q.$, can here be interpreted in the fashion of a Feynman diagram, with ψ_a representing the probability amplitude that the vacuum fermion has coordinate $q_1,\ldots,q_n$ in state a, V_{op} is the amplitude that the perturbation potential acts on the particle, and ψ_b^* is the amplitude that the particle enters into excited state b. However, we interpret the perturbation potential as the transient excitation energy of a neighbouring vacuum fermion rather than as a potential function associated with a classical force field.

To express P(i,f) in terms of p(a,b) we can picture the absorbing particle as being somewhere on a spherical surface centered around the originating particle and of area $4\pi r^2 = 4\pi N^2 r_K^2$. The chance that a chain of transient excitations initiated at the center is absorbed is proportional to N^{-2}. Thus $P(i,f) \sim p(a,b)/N^2$, giving

$$P(i,f) \sim \frac{ABr_K^2}{c^2\hbar^2 N^2} \qquad (7)$$

where B is a proportionality constant. On substituting $N = r/r_K$ this becomes

$$P(i,f) \sim \frac{ABr_K^4}{c^2\hbar^2 r^2} \qquad (8)$$

Recalling that force should be proportional to this probability (or, alternatively, to the number of exchanges) we obtain the inverse square law

$$F \sim \frac{Cr^4 k}{c^2\hbar^2 r^2} \tag{9}$$

where C is a merged proportionality constant. Since this has the right form it lends some credence to the linearity assumption and to our use of Eq. (5).

Taking the vacuum density as $\varrho_K = 1/r_K^3$ Eq. (9) can be written

$$F \sim \frac{C}{c^2\hbar^2\varrho^{4/3} k\, r^2} \tag{10}$$

This means that the force between two particles decreases as the vacuum density decreases. This is plausible when it is recalled that the probability of undergoing a transition to an equal energy state increases as the square of τ_K, which means that the likelihood that the excitation will actually be annihilated by the absorbing particle increases. This also means that second order (shorter time scale) processes could not contribute significantly to the force.

If we suppose that our inverse square law represents a gravitational interaction between two particles of mass m, we can set F equal to $-Gm^2/r^2$. Solving for m,

$$m \sim \pm \frac{C_{grav}^{1/2}}{G^{1/2} c\hbar \varrho_{grav}^{2/3}} \tag{11}$$

If F represents a Coulomb interaction between two charges we obtain a similar expression for q, with C_{grav} replaced by C_{em} and ϱ_{grav} replaced by ϱ_{em}. Increase in m or q should thus depress ϱ_{grav} and ϱ_{em} in the surrounding space. As m and q increase ϱ_{grav} and ϱ_{em} must decrease. This is the basis for our earlier statement that mass and charge depress the vacuum density in the surrounding space.

Since ϱ_K is global property a conceptually proper description should actually begin with a wave function for the whole vacuum structure. For this purpose let us write each ψ_a as ψ_a (q_j,t), where q_j represents the coordinates of vacuum particle j at time t. We suppose that the wave function of the vacuum as a whole, to be denoted by $\psi_v(q_1,\ldots,q_n,t)$ can be constructed as a Slater determinant of these vacuum particle wave functions. We can then use the interpretive postulate of quantum mechanics to write $\varrho_K = |\psi_v(q_1,\ldots,q_n,t)|^2$, which means that our inverse square law can be expressed as

$$F \sim \frac{B}{c^2\hbar^2\{|\psi_v(q_1,,\ldots,q_n,t)|^2\}^{4/3} r^2} \tag{12}$$

The new and important feature is the introduction of the irreversible projection process into a force law. The time symmetry of any system governed by a potential function that incorporates such a collapse process must be broken. When the vacuum structure reaches equilibrium the dependence of ψ_v on t disappears ; further phase randomization is not necessary for the force law to be definite and thus the system enters into a regime that is adequately described by reversible time evolution equations.

6. TENABILITY OF THE MODEL

It is evident that our approach deviates from the usual procedure for setting up a physical problem. Ordinarily the investigator begins with a description of the particles and forces in terms of classical variables. Then this is converted to a nonclassical description in terms of operators and wave functions. In our model force is dependent on the structure of the vacuum and therefore its classical description must embed a projection process. Whether the particular model presented is sufficiently compatible with well founded principles of physics to be taken seriously is another question. The important issue, from the standpoint of the conceptual coherence of physics and biology, is whether some member of the class of measurement-embedding models could meet all the requisite criteria. Without going into technical details, I would like to point out a number of features which suggest that this class is worth considering.

The first point to note is that our force law could have been obtained from the linearity assumption without any presupposition about the physical mechanism. The quantity ϱ_K would, however, still represent the density of potential excitations in the vacuum and would therefore presumably be associated with vacuum particle density. The appearance of fluctuation times and lengths in the force law would still require the propagation of these excitations to have a chain-like character.

This chain is initiated when an observable particle emits a virtual particle (or, more accurately, when a spontaneous vacuum fluctuation is absorbed by an observable particle). This is accompanied by a transient violation of energy conservation equal to $\hbar/\tau_K$, but momentum is conserved since the momentum change of the absorbing particle balances the momentum $\hbar/x_K$ associated with the fluctuation, independent of the choice of coordinate system. The energy fluctuation must disappear within a time interval τ_K. This is accompanied by the disappearance of the momentum fluctuation. Since momentum has been conserved up to this point the annihilation of the momentum fluctuation leads to a momentum violation of $\hbar/x_K$, restricted to the space interval x_K. Outside this interval the annihilated momentum must reappear, leading to a reappearance of the energy fluctuation. The process thus repeats until the momentum violation is annihilated by a second absorbing particle. The propagation process is accompanied by a wave of contraction and expansion of the vacuum density due to the fact that x_K and τ_K are increased in the vicinity of mass. Real bosons can be interpreted as propagation processes in which the accompanying vacuum density wave carries real energy.

The interpretation of this propagation process in terms of particle-hole excitations is not a logical necessity. However, this interpretation can be shown to satisfy the requirement for zero rest mass. As is well known, it is possible to derive the more general force laws of electromagnetism from the inverse square law by applying relativistic transformations and some additional assumptions, such as charge conservation[17,19]. Force laws with a noninverse square character would also arise from inhomogeneities in the vacuum density. Infinite self energies do not occur since $r=Nr_K$ cannot go to zero unless $N=0$, in which case no

interaction occurs. Since mass is defined through forces mediated by vacuum particles, the latter can develop mass only if they act as transient absorbers. This leads to rather delicate relations between vacuum particle mass and vacuum particle density.

Since F_{grav} is weaker than F_{Coul}, ϱ_{grav} must be greater than ϱ_{Coul} (about 10^{32} times greater for two interacting electrons and about 10^{28} times greater for two interacting protons). This suggests a two fluid model of the vacuum in which gravitational force is due to the altering effect of mass on the density of all surrounding vacuum particles and the electromagnetic force is dependent on the density of only those vacuum particles that are charged.

7. DETERMINISM VERSUS INDETERMINISM

If force laws embed measurement-like collapse processes, it is evident that this is not manifested in laboratory experiments involving a small number of particles. Presumably the generation of irreducible probability and irreversibility would only be prominent in large scale physical processes, such as those that occurred early in the history of the universe or that involve significant enough events to disturb the equilibrium between observable matter and vacuum density. The processes that occur in biological systems are not on such a scale. However, we need only recall that fluctuation amplified by nonlinearity is the essence of evolutionary search. The large scale effects of mutations that occur in DNA molecules of a present day organism provide a dramatic example. In measurement embedding models of the type proposed this elaborate amplification machinery unmasks projection operations that are only latent under ordinary physical conditions.

Our use of the term measurement needs some clarification. Amplification machinery is always necessary when microphysical phenomena are measured. Becoming macroscopic is part of becoming classical. The issue in measurement theory has been where the collapse process occurs — in the amplification machinery, in the organism, or only in the mind of the observer. When measurements are performed some features of the system become more definite, with accompanying phase randomization. How much more definite they become depends on how much randomization occurs. But according to the uncertainty principle the conjugate features become less definite. The initial conditions necessary for calculating the system's future behavior must as a whole become more uncertain, with the extra uncertainty associated with the entropy increase. The process of becoming indefinite, which we can call wave function expansion, is just as much a part of the measurement cycle as wave function collapse. In measurement embedding models the process of wave function collapse and wave function expansion occur in all systems in nature, with concomitant probability generation. These processes are not prominent enough under ordinary physical conditions to effect the dynamics of simple systems ; but this situation changes radically when these systems are coupled to the sensitive dynamics of measurement machinery.

On this picture the probabilistic aspect of evolutionary theory *can* have an ultimately irreducible status that would not be possible if the laws of nature were reversible and if measurement entered only in an extrinsic way. The objection may be raised that a probabilistic interpretation is contingent on our having excluded hidden variables — for example, subquantum fields with deterministic chaos and action at a distance[20,21]. This, however, would not vitiate our point. Whether nature is deterministic or indeterministic would be undecidable, in the sense of Poincare's conventionalism. The situation would be similar to that of gravitation, where it may well be undecidable whether a simple force law operates in a complicated space or a complicated force law operates in a simple space. From the standpoint of the architecture of science as a whole the advantage of force laws that embed measurement is not that they decide the inherently arguable issue of determinism versus indeterminism in favor of the latter, but that they do not decide it in favor of the former.

REFERENCES

(1) Eigen, M. (1971) *Naturwissenschaften 58*, 465-523.

(2) Prigogine, I., Nicolis, G. and Babloyantz, A. (1972) *Physics Today* November and December, 23-28 and 38-44.

(3) Conrad, M. (1983) *Adaptability* (Plenum, New York).

(4) Prigogine, I. and George, Cl. (1983) *Proc. Natl. Acad. Sci USA 80*, 4590-4594.

(5) George, Cl., Mayne, F. and Prigogine, I. (1984) To appear in *Advances in Chemical Physics.*

(6) Khinchin, A.I. (1949) *Mathematical Foundations of Statistical Mechanics* (Dover, New York).

(7) Bremerman, H.J. (1962) In : *Self-Organization Systems*, eds. Yovits, M.C., Jacobi, G.T. and Goldstein (Spartan Books, Washington, D.C.).

(8) Landauer, R. (1982) *Int. J. Theor. Phys. 21*, 3-4, 282-297.

(9) Conrad, M. and Liberman, E.A. (1982) *J. Theoret. Biol. 98*, 239-252.

(10) Conrad, M. (1984) *BioSystems 16*, 3-4, 345-363.

(11) Conrad, M. (1985) *Comm. ACM 28*, 5, 464-480.

(12) Rössler, O. (1983) *Z. Naturforsch. 38a*, 788-801.

(13) Pattee, H.H. (1973) In : *The Physical Principles of Neuronal and Organismic Behavior*, eds., Conrad, M. and Magar, M. (Gordon and Breach, New York).

(14) von Neumann, J. (1955) *Mathematical Foundations of Quantum Mechanics* (Princeton Univ. Press, Princeton, N.J.)

(15) Dirac, P.A.M. (1958) *The Principles of Quantum Mechanics*, Fourth Edition (Oxford Univ. Press, Oxford, England).

(16) Schiff, L.I. (1955) *Quantum Mechanics*, Second Edition (McGraw-Hill, New York).

(17) Rosser, W.G.V. (1960) *Contemp. Physics 1*, 453.

(18) Frisch, D.H. and Wilets, L. (1969) *Am. J. Phys. 24*, 574.

(19) Tessman, J.R. (1966) *Am. J. Phys. 34*, 1948.

(20) Bohm, D. (1981) *Wholeness and the Implicate Order* (Routledge and Kegan Paul, London).

(21) Conrad, M., Home, D. and Josephson, B.D. (1985) To appear in *Proceedings of the Urbino Conference on Microphysical Reality and Quantum Formalism.*

II

Dynamical Systems, Information and Cognitive Sciences.

Randomness and Gödel's Theorem

Gregory J. Chaitin
IBM Research Division

Our view of the physical world has certainly changed radically during the past hundred years, as unpredictability, randomness and complexity have replaced the comfortable world of classical physics. Amazingly enough, the same thing has occurred in the world of pure mathematics, in fact, in number theory, a branch of mathematics that is concerned with the properties of the positive integers. How can an uncertainty principle apply to number theory, which has been called the queen of mathematics, and is a discipline that goes back to the ancient Greeks and is concerned with such things as the primes and their properties ?

Following Davis (1982), consider an equation of the form

$$P(x, n, y_1, \ldots, y_m) = 0,$$

where P is a polynomial with integer coefficients, and x, n, m, y_1, ... , y_m are positive integers. Here n is to be regarded as a parameter, and for each value of n we are interested in the set D_n of those values of x for which there exist y_1 to y_m such that $P = 0$. Thus a particular polynomial P with integer coefficients in $m + 2$ variables serves to define a set D_n of values of x as a function of the choice of the parameter n.

The study of equations of this sort goes back to the ancient Greeks, and the particular type of equation we have described is called a polynomial Diophantine equation.

One of the most remarkable mathematical results of this century has been the discovery that there is a « universal » polynomial P such that by varying the parameter n, the corresponding set D_n of solutions that is obtained can be any set of positive integers that can be generated by a

computer program. In particular, there is a value of n such that the set of prime numbers is obtained. This immediately yields a prime-generating polynomial

$$x[1 - (P(x, n, y_1, \ldots, y_m))^2],$$

whose set of positive values, as the values of x and y_1 to y_m vary over all the positive integers, is precisely equal to the primes. This is a remarkable result that surely would have amazed Fermat and Euler, and it is obtained as a trivial corollary to a much more general theorem !

The proof that there is such a universal P may be regarded as the culmination of Gödel's original proof of his famous incompleteness theorem. In thinking about P, it is helpful to regard the parameter n as the Gödel number of a computer program, and to regard the set of solutions x as the output of this computer program, and to think of the auxiliary variables y_1 to y_m as a kind of multidimensional time variable. In other words,

$$P(x, n, y_1, \ldots, y_m) = 0$$

if and only if the nth computer program outputs the positive integer x at time $(y_1, \ldots, y_m)$.

Let us prove Gödel's incompleteness theorem by making use of this universal polynomial P and Cantor's famous diagonal method, which Cantor originally used to prove that the real numbers are more numerous than the integers. Recall that D_n denotes the set of positive integers x for which there exist positive integers y_1 to y_m such that $P = 0$. I.e.,

$$D_n = \left\{ x \mid (\exists\, y_1, \ldots, y_m)[P(x, n, y_1, \ldots, y_m) = 0] \right\}.$$

Consider the « diagonal » set

$$V = \left\{ n \mid n \notin D_n \right\}$$

of all those positive integers n that are not contained in the corresponding set D_n. It is easy to see that V cannot be generated by a computer program, because V differs from the set generated by the nth computer program regarding the membership of n. It follows that there can be no algorithm for deciding, given n, whether or not the equation

$$P(n, n, y_1, \ldots, y_m) = 0$$

has a solution. And if there cannot be an algorithm for deciding if this equation has a solution, no fixed system of axioms and rules of inference can permit one to prove whether or not it has a solution. For if there were a formal axiomatic theory for proving whether or not there is a solution, given any particular value of n one could in principle use this formal theory to decide if there is a solution, by searching through all possible proofs within the formal theory in size order, until a proof is found one way or another. It follows that no single set of axioms and rules of inference suffice to enable one to prove whether or not a polynomial Diophantine equation has a solution. This is a version of Gödel's incompleteness theorem.

What does this have to do with randomness, uncertainty and unpredictability ? The point is that the solvability or unsolvability of the equation

$$P(n, n, y_1, \ldots, y_m) = 0$$

in positive integers is in a sense mathematically uncertain and jumps around unpredictably as the parameter n varies. In fact, it is possible to construct another polynomial P' with integer coefficients for which the situation is much more dramatic.

Instead of asking whether $P' = 0$ can be solved, consider the question of whether or not there are infinitely many solutions. Let D'_n be the set of positive integers x such that

$$P'(x, n, y_1, \ldots, y_m) = 0$$

has a solution. P' has the remarkable property that the truth or falsity of the assertion that the set D'_n is infinite, is completely random. Indeed, this infinite sequence of true/false values is indistinguishable from the result of successive independent tosses of an unbiased coin. In other words, the truth or falsity of each of these assertions is an independent mathematical fact with probability one-half ! These independent facts cannot be compressed into a smaller amount of information, i.e., they are irreducible mathematical information. In order to be able to prove whether or not D'_n is infinite for the first k values of the parameter n, one needs at least k bits of axioms and rules of inference, i.e., the formal theory must be based on at least k independent choices between equally likely alternative assumptions. In other words, a system of axioms and rules of inference, considered as a computer program for generating theorems, must be at least k bits in size if it enables one to prove whether or not D'_n is infinite for $n = 1, 2, 3, \ldots, k$.

This is a dramatic extension of Gödel's theorem. Number theory, the queen of mathematics, is infected with uncertainty and randomness ! Simple properties of Diophantine equations escape the power of any particular formal axiomatic theory ! To mathematicians, accustomed as they often are to believe that mathematics offers absolute certainty, this may appear to be a serious blow. Mathematicians often deride the non-rigorous reasoning used by physicists, but perhaps they have something to learn from them. Physicists know that new experiments, new domains of experience, often require fundamentally new physical principles. They have a more pragmatic attitude to truth than mathematicians do. Perhaps mathematicians should acquire some of this flexibility from their colleagues in the physical sciences !

APPENDIX

Let me say a few words about where P' comes from. P' is closely related to the fascinating random real number which I like to call Ω. Ω is defined to be the halting probability of a universal Turing machine when its program is chosen by coin tossing, more precisely, when a program n bits in size has probability 2^{-n} [see Gardner (1979)]. One could in principle try running larger and larger programs for longer and longer amounts of time on the universal Turing machine. Thus if a program ever halts,

one would eventually discover this ; if the program is n bits in size, this would contribute 2^{-n} more to the total halting probability Ω. Hence Ω can be obtained as the limit from below of a computable sequence $r_1 \leqslant r_2 \leqslant r_3 \leqslant \ldots$ of rational numbers :

$$\Omega = \lim_{k \to \infty} r_k ;$$

this sequence converges very slowly, in fact, in a certain sense, as slowly as possible. The polynomial P' is constructed from the sequence r_k by using the theorem that « a set of tuples of positive integers is Diophantine if and only if it is recursively enumerable » [see Davis (1982)] : the equation

$$P'(k, n, y_1, \ldots, y_m) = 0$$

has a solution if and only if the nth bit of the base-two expansion of r_k is a « 1 ». Thus D'_n, the set of x such that

$$P'(x, n, y_1, \ldots, y_m) = 0$$

has a solution, is infinite if and only if the nth bit of the base-two expansion of Ω is a « 1 ». Knowing whether or not D'_n is infinite for $n = 1, 2, 3, \ldots, k$ is therefore equivalent to knowing the first k bits of Ω.

BIBLIOGRAPHY

(1) G.J. CHAITIN (1975), « *Randomness and mathematical proof* », *Scientific American* 232 (5), pp. 47-52.

(2) M. DAVIS (1978), « *What is a computation ?* », *Mathematics Today : Twelve Informal Essays*, L. A. Steen, Springer-Verlag, New York, pp. 241-267.

(3) D.R. HOFSTADTER (1979), *Gödel, Escher, Bach : an Eternal Golden Braid*, Basic Books, New York.

(4) M. GARDNER (1979), « *The random number Ω bids fair to hold the mysteries of the universe* », Mathematical Games Dept., *Scientific American* 241 (5), pp. 20-34.

(5) G.J. CHAITIN (1982), « *Gödel's theorem and information* », *International Journal of Theoretical Physics* 22, pp. 941-954.

(6) M. DAVIS (1982), « *Hilbert's tenth problem is unsolvable* », *Computability & Unsolvability*, Dover, New York, pp. 199-235.

Transmission and Amplification of Information in a Changing Environment : the Case of Insect Societies.

Jacques M. Pasteels
Laboratoire de Biologie Animale et Cellulaire — Université Libre de Bruxelles

Jean-Louis Deneubourg
Laboratoire de Chimie-Physique II — Université Libre de Bruxelles

Simon Goss
Laboratoire de Chimie-Physique II — Université Libre de Bruxelles

1. INTRODUCTION

The central question in understanding how an insect society functions is the following :

How individual behaviour which often looks so erratic, leads to a behaviour far more deterministic at the level of the society taken as a whole ?

The French verb « fourmiller » describes well the impression of inefficiency and confusion exhibited by the ants in their social behaviour. It literally means « to behave like ants » and is translated in Harrap's New Standard French and English Dictionnary by « to move hither and thither ». If the ants indeed « fourmillent » when building their nest or collecting their food, how is it that ant nests are well organized and structured, and their foraging patterns are well defined ?

This apparent paradox between low efficiency at the individual level and structure at the society level has always puzzled biologists, and at best has only been approached by a few authors with remarkable intuition, e.g. Grassé's stimergy theory to explain building behaviour in termites (Grassé, 1959).

The attitude of most social insect experts, consciously or not, has been to try to demonstrate that social insect behaviour is less erratic than it appears. Of course, ants do not behave randomly in a completely uncoordinated way. A division of labour is usually observed between workers depending on their morphology and age, and they do communicate with each other by mechanical and chemical signals (pheromones). Communication and division of labour, however are not deterministic but probabilistic in nature, and ants are not simple deterministic automata. In spite of this, descriptions of social behaviour are usually reported in deterministic terms especially in studies of comunication behaviour.

Consider for example figs 1a and 1b. Fig 1a gives the sequence of acts performed by an ant communicating the discovery of a food source as described in the literature. The ant drinks, and returns to the nest drawing a chemical trail. In the nest, it invites and excites other workers by mechanical or chemical signals and unloads the food it collected. It then returns to the food source, the recruited ants following the trail. This is a highly idealistic description of mass recruitment (see below) and fig. 1b gives a more realistic description of the ants' behaviour. This flow diagram was constructed by pooling the behaviour of 30 individuals which discovered a food source and recruited nest mates (Champagne et al., 1984). The arrangement is the same as in fig. 1a. It describes their behaviour around the source, on the way back to the nest, in the nest, and finally on their way back to the food source. Each circle represents a different behaviour, some completely unrelated to the sequence, like exploration or grooming. The width of the arrows is proportional to the frequency of transition between the acts. No two ants behaved the same. Communication, like recruitment is obviously not at all deterministic, but includes a large stochastic component.

The stochastic component present in communication is often ignored or considered to be an unavoidable side effect of animal behaviour. Oster and Wilson (1978), have suggested that ants can afford such variance in their behaviour because of their great number. If one ant fails, another will succeed possibly by more indirect means. Some inefficiency due to variance in behaviour is compensated for by a greater reliability of the system, the goal being reached in different ways. This is true, but stochastic events could have more creative consequences in the development of behavioral patterns at the societal level.

The duality between stochastic behaviour at a microscopic level and determinism at a macroscopic level is not peculiar to insect societies, but occurs in other biological and non-biological systems involving numerous sub-units. It is now well-established that both non-linear processes (e.g. communication) and stochastic events are relevent to the understanding of the system's behaviour at the macroscopic level (Nicolis and Prigogine, 1977). We thus investigated the possibility that stochastic events could have a function in insect societies when linked to an amplifying mechanism, i.e. that from their interactions, patterns and structures could emerge which are adaptive for the society.

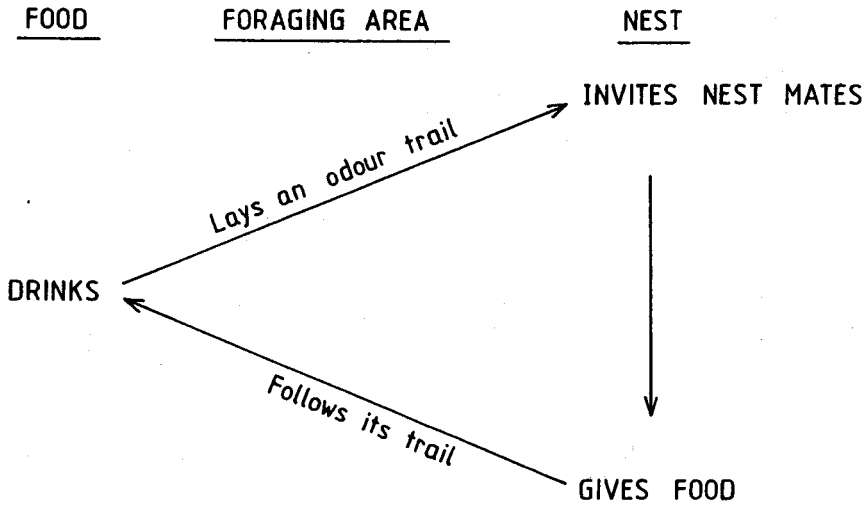


Figure 1a
Flow diagram of the behaviour of first recruiters in *Tapinoma erraticum*
Idealized sequence

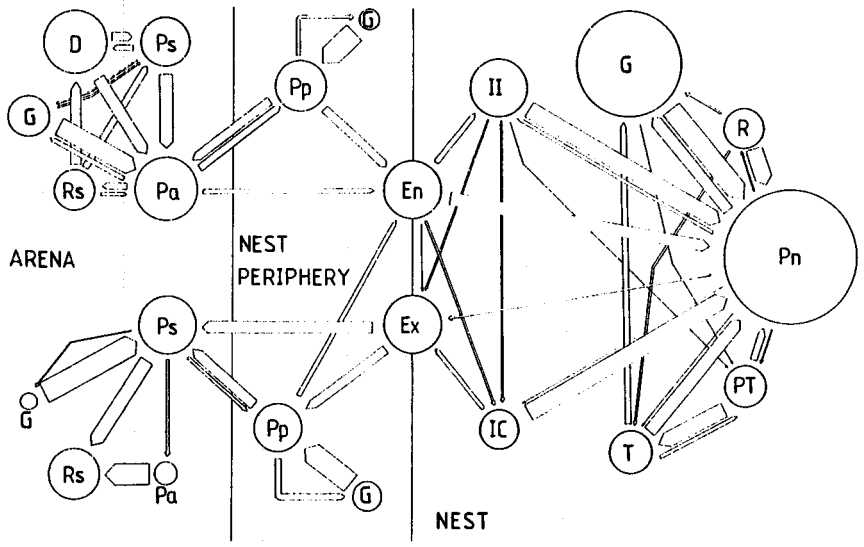


Figure 1b
Flow diagram of the behaviour of first recruiters in *Tapinoma erraticum*
Diagram constructed by pooling the behaviour of 30 different ants. Each circle represents one act performed in the arena, the nest periphery, while returning to the nest (above), while returning to the food source (below), while still in the nest. The arrow's width is proportional to the transition frequency between the acts.

D : drinks ; En : enters ; Ex : exits ; G : grooming ; II : individual invitation ; IC : collective invitation ; Pa, Pn, Pp, Ps : patrolling in the arena, nest, nest's periphery and around the source ; PT : trophallaxy posture ; R : recognition ; Rs : returns to the source ; T : trophallaxy.

The interplay between stochastic behaviour, number of participants and non-linear processes will be illustrated first by describing how ants can cope with unpredictable resources whilst foraging for food. Other self-organisation mechanisms, i.e. specialization resulting from the development of long-term memory, will then be discussed in more predictable situations.

2. THE MECHANISM OF MASS RECRUITMENT

Mass recruitment as described in fig. 1 is a very simple process, albeit subtler than it first looks. Recruitment is autocatalytic as successful recruits, and only those, lay down trail pheromone on their way back to the nest. Successful recruits are those reaching the food source and able to feed. A very simple negative feed-back stops the recruitment. If no more food is available, recruits are no longer successful and stop reinforcing the trail. The trail pheromone evaporates and recruitment comes to an end (Wilson, 1962).

For a finite population the simplest description of an autocatalytic process of this kind is given by Verhulst's equation. Indeed this equation can to some extent describe the population dynamics of ants recruited in the foraging area (fig. 2 ; Verhaeghe and Deneubourg, 1983).

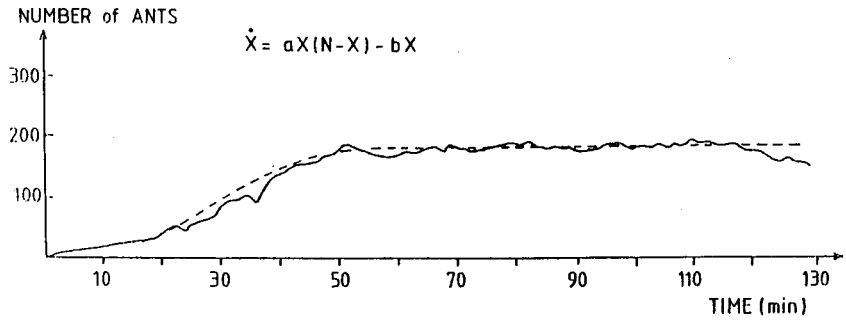


Figure 2
Fitting of a food recruitment experiment in *Tetramorium impurum* with the logistic equation.

N : number of ants ; X : recruited ants ; a : recruitment rate ; b : departure time. Value of parameters : N=270 ; $a=8\ 10^{-4}$; $b=72\ 10^{-3}$. Experiment : solid line ; fitting : dashed line.

Recruitment is not necessarily accurate and the error level varies considerably from species to species in the same experimental conditions, as illustrated in table 1. In one species, 74 % of the ants recruited by the first scout reached the food source, whereas only 18 % did so in another species in exactly the same experimental set-up. Also, trail-following behaviour is far more accurate in the first ant (Verhaeghe et al., 1980).

What is the biological meaning of such variance in accuracy ? At first sight, it is tempting to conclude that the most accurate system of communication is the best and that one ant is simply more efficient than the other. This is certainly true when only one source is present in the fora-

Table 1

Recruitment accuracy of a first recruiter
(from Verhaeghe et al. 1980)

	Tetramorium impurum	**Tapinoma erraticum**
Successful recruits (%)	18	74
Length of trail followed (%)	17	68

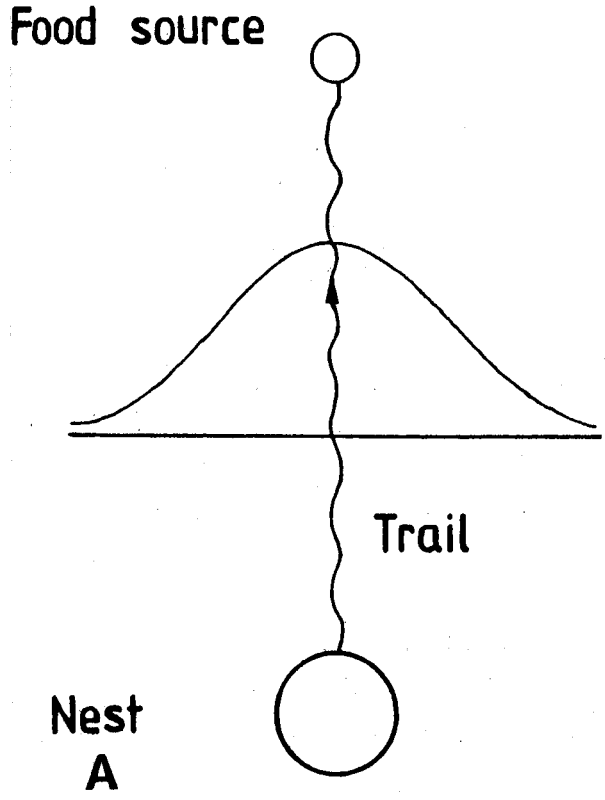


Figure 3a
Trail recruitment to a single source
The distribution of the recruits about the trail is represented by a Gaussian

ging area. In fig. 3a, the distribution of lost ant is described by a Gaussian. Experiments confirmed that it is a fair approximation of the reality. Of course, the smaller the variance, the better the recruitment. Natural environments, however, are more complex and several sources could be simultaneously available to the ants, as schematized in fig. 3b. Lost ants

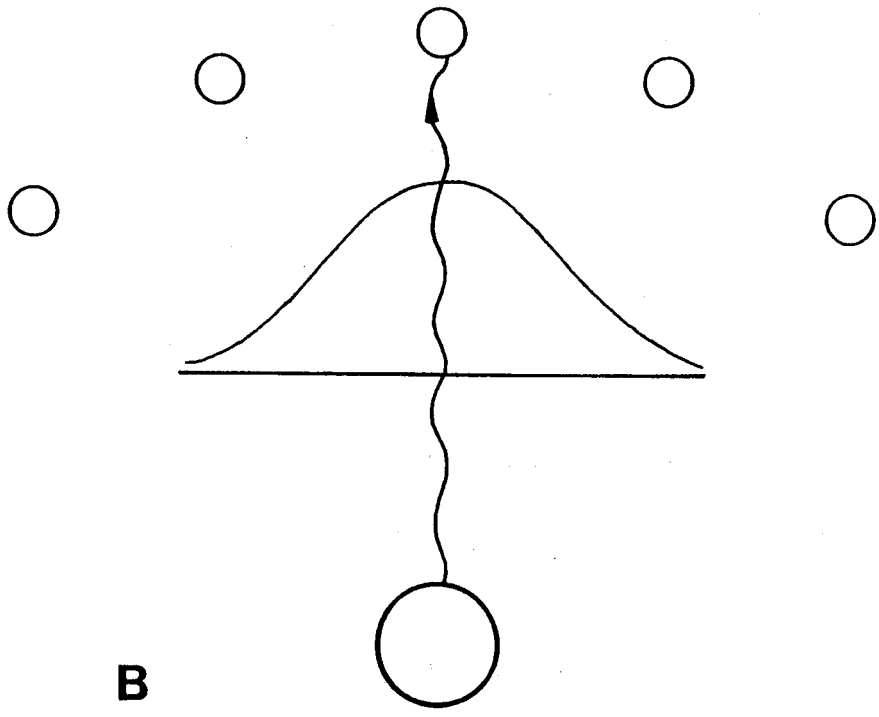


Figure 3b
When several sources are aggregated,
« lost » ants discover new sources and induce additional recruitments.

discover these additional sources and induce new recruitments. The question then is : what is the best strategy between high accuracy allowing immediate exploitation and low accuracy allowing new discoveries ?

2.1. A strategy of inaccurate communication

The best compromise between accuracy and discovery was investigated with the help of a mathematical model depicting the idealistic situation described in fig. 3b : food is divided into identical sources, evenly spaced and equidistant from the nest. The model (fig. 4) is based on the logistic equation, but allows a certain level of error. The ants reaching a given source are not the only successful recruits to that source. A fraction of the lost ants that were recruited to other sources find by chance the first source. Ants recruit only to the source from which they just drank. Foraging efficiency is measured by the time needed to collect the food, and accuracy by the inverse of the Gaussian standard deviation.

Simulations of the model (fig. 5) demonstrate that when one source is present, the more accurate the communication, the more efficient it is, but as soon as food is divided, there is an optimal error level which minimizes the time needed for collecting the food. The more divided the food, the noisier the optimal communication system. Stochastic

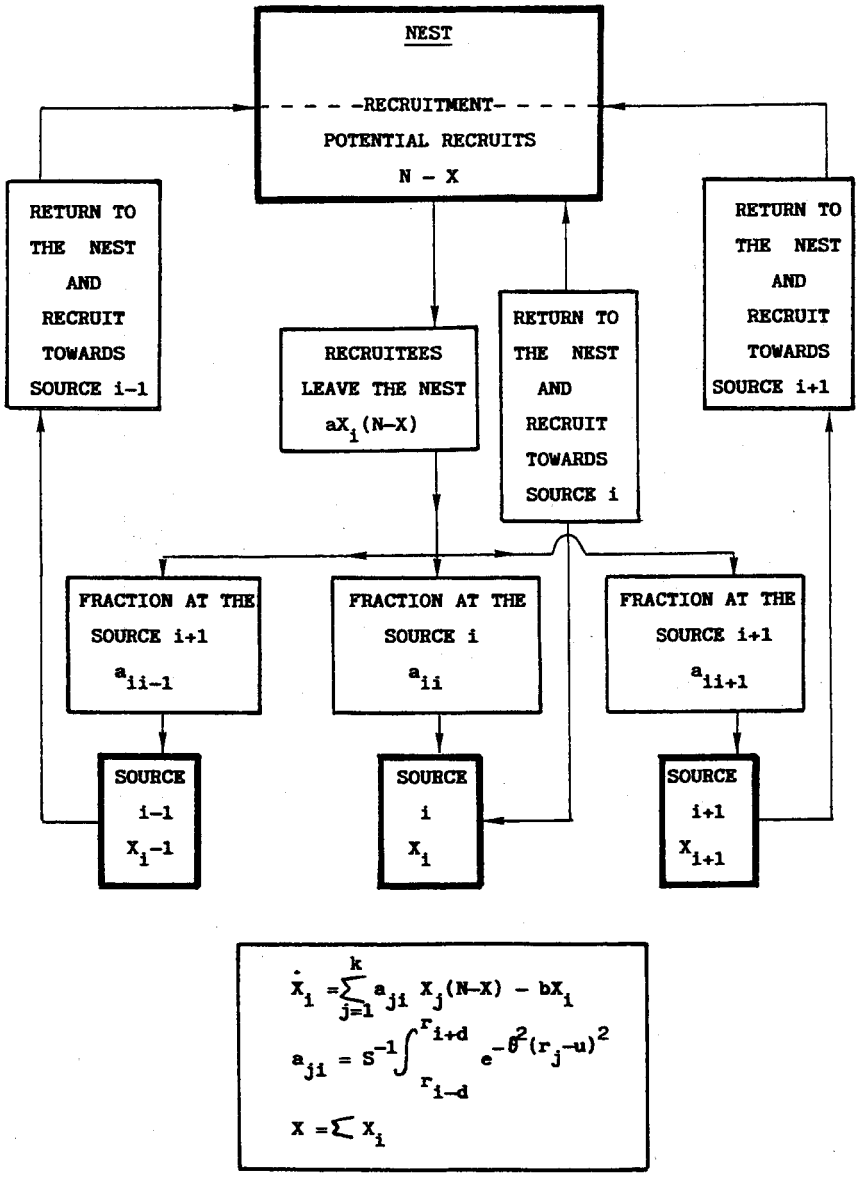


Figure 4

X_i number of ants around the food source i, a_{ii} is the recruitment rate for the ant recruited to i and effectively arriving at i and a_{ji} is the recruitment rate for the ants recruited to j and erroneously going to i.

$X = \sum_{i=1}^{k} X_i$ and K is the number of food sources.
a_{ji} and a_{ii} are calculated according to the hypothesis that the ants going to the sources distribute themselves with a Gaussian distribution centred on the food sources. S is a normalization factor, 2d the diameter of the food source and $r_i - r_j$ the distance between the food source i and j.

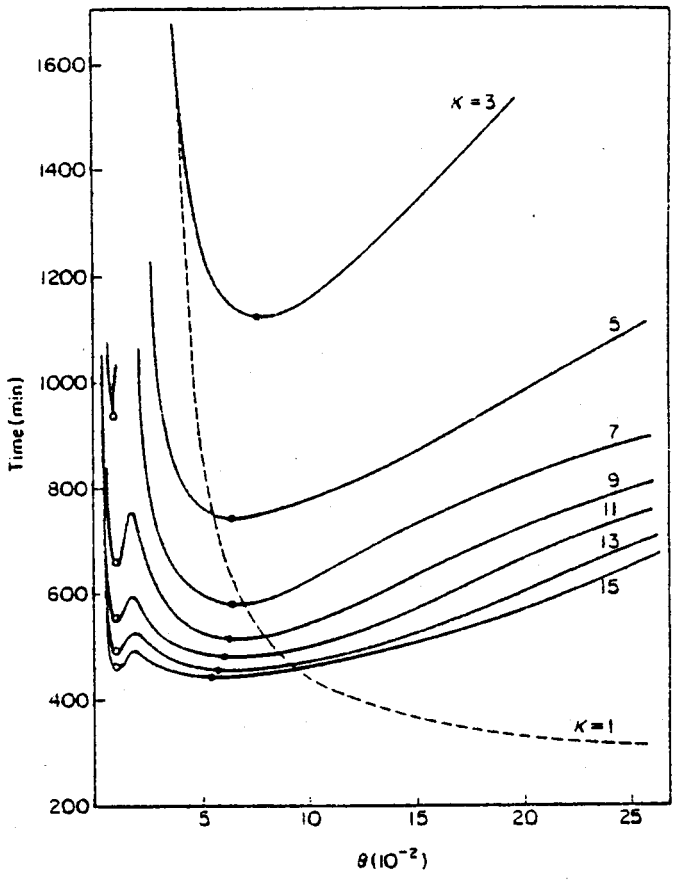


Figure 5

Simulations of the model described in fig. 4
Collection time of food with different numbers of food sources (K). All food offered is collected. Value of parameters ; N=650 ; Q (food quantity)=1ml ; $a=10^{-3}$; $b=16\ 10^{-3}$; 2d=1.4 cm.

events are not just an unavoidable by-product of animal behaviour, but they can impove the system's functioning (Pasteels et al., 1982 ; Deneubourg et al., 1983). If some inaccuracy is adaptive, we must expect that the intrinsic level of accuracy in communication could be tuned to the average food distribution in the natural environments of the different species. This possibility has been recently suggested to interpret variations in accuracy in the honey bee's dance language (Gould et al., 1985).

The domestic bee has several dances to communicate the food source location, depending on the distance of the sources from the nest. The waggling dance is performed when food sources are far from the nest, i.e. hundreds of meters from it. These dances contain accurate information on direction via the orientation of the waggling runs. At short dis-

tances of a few decameters, the dance is circular and contains no directional information. For intermediate distances, divergent dances are observed in which a range of directions is given by the angle formed by the waggling runs. The wider the angle, the closer the food source. For example, at one hundred meters the angle is about 30°, dropping to less than 5° for food 700 meters or more away. The bee is thus able to adjust the recruitment accuracy according to the food source's distance. This seems well adapted to collecting scattered but aggregate food sources like flowering herbs in temperate regions. This interpretation is supported by the fact that the dances of tropical honey bees living in forests are far more accurate, even over short distances. These bees forage on far more concentrated resources like a single flowering tree (Gould et al., 1985).

2.2. A collective selection of the best resources

Food sources are not necessarily equally rewarding and their distribution in time and space is often unpredictible. How do the ants cope with these fluctuations ? The answer again is the interplay of stochastic components in their behaviour and amplifying processes during communication. This was experimentally explored by offering food sources in large arenas to ants. The food sources consisted of more or less concentrated sugar solutions, given simultaneously or in succession in either finite or unlimited quantities (Parro, 1981 ;Pasteels et al., in prep.).

Recruitment to a 1M sugar solution is not very accurate and lost ants explored the whole foraging arena (fig. 6). When two sources of unequal quality were given at the same time, both sources were discovered, but

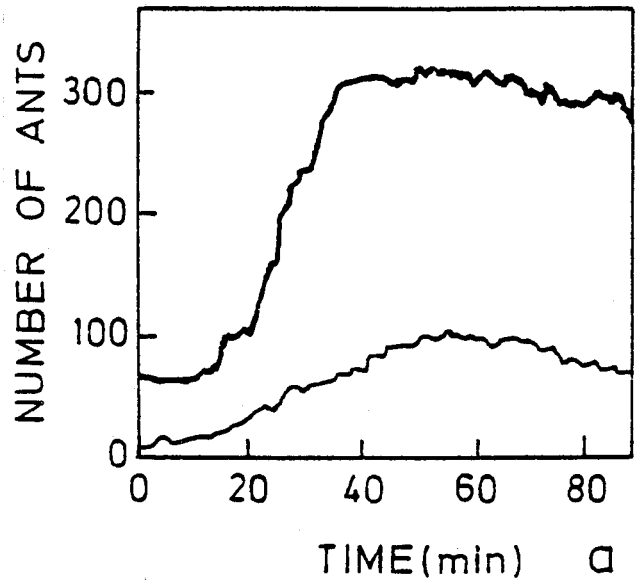


Figure 6a
Recruitment for one 1M sucrose food source in *Tetramorium caespitum*
——— total population on the arena
——— population at the food source

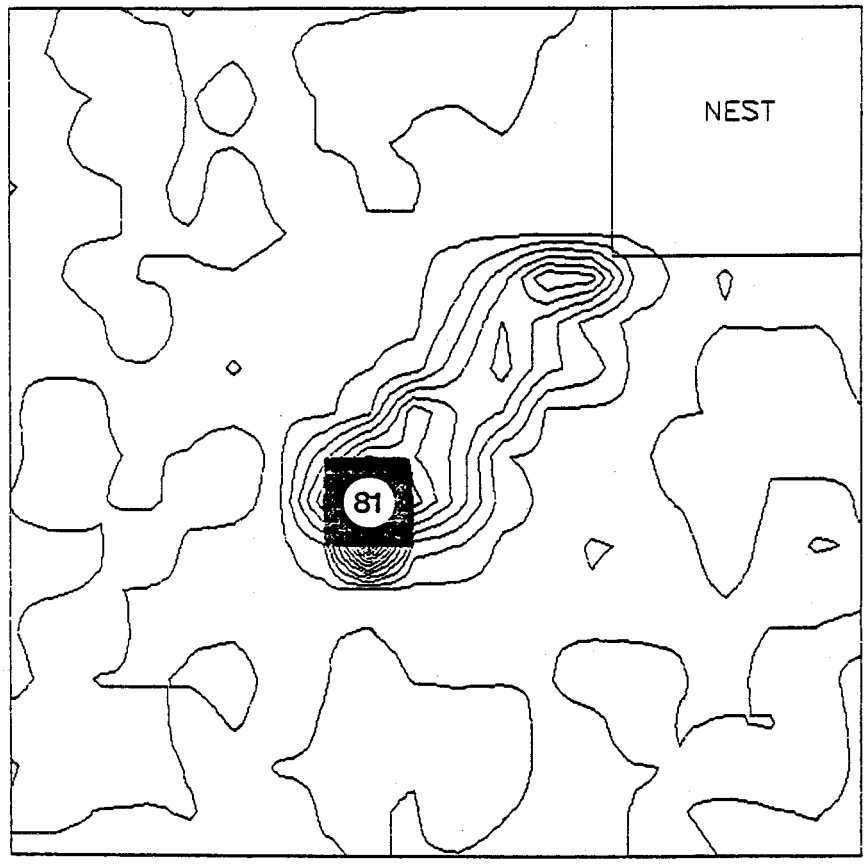


Figure 6b
Recruitment for one 1M sucrose food source in *Tetramorium caespitum*
Distribution of the ants on the arena.
Average values for the last 30 minutes. Isodensity curves at 0.02, 0.08, 0.14... ants/cm^2. The number of ants at the food source is indicated explicitly.

the most rewarding source was always exploited more than the other, and the more so the greater the difference in food quality. The ants are indeed able to adjust their recruitment rates to food quality : recruiting ants lay down more trail pheromone when food sources are of a better quality (Hantgartner, 1969). It must be emphasized that the selection of the richer source is not the result of an individual decision comparing the source's quality. It is a collective decision resulting indirectly from the competition between two autocatalytic processes with different growth rates.

The interplay between error during communication and amplification is best illustrated when the sources were given in succession (fig. 7). Ants were recruited to a first 0.1 M sucrose solution. One hour later, when the recruitment was well underway, a richer 1M solution was offered in the arena. Patrolling ants which missed the first target soon discovered the more rewarding source and a quick shift in the main recruitment

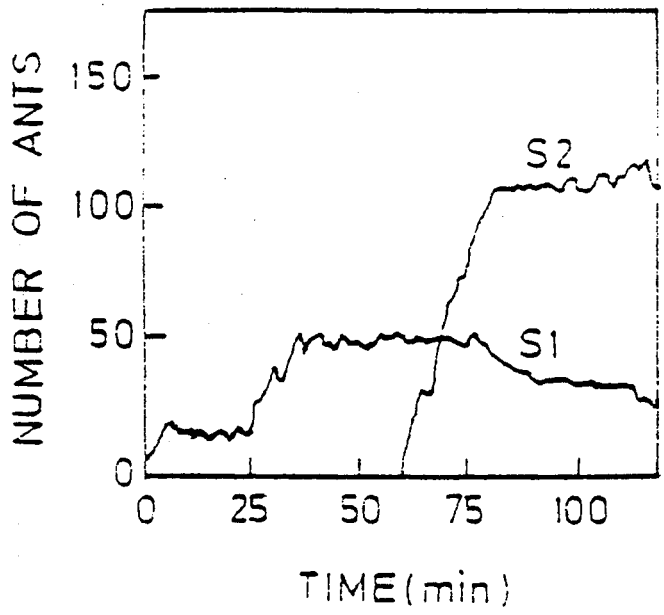


Figure 7
Recruitment in *T. caespitum* for two successively introduced food sources (0.1M sugar solution (S1) followed 1 hour later by 1M (S2)). Population at each food source.

direction occured. Not only were more ants recruited to the new source, but the first source was progressively neglected.

All these results can be described with a system of logistic equations with constant recruitment rate and error level, i.e. the variance around the trail remains constant during the whole recruitment process. Although, the level of accuracy must be greater when ants recruit to a better source.

One experiment demonstrated that the mechanism of mass recruitment is even more sophisticated. When 2 identical sources were offered at the same time, the number of ants around each source grew at an equal rate until an abrupt change in the kinetic occured, one source becoming more exploited than the other (fig. 8). This was not the result of an asymmetry in the experimental set up, since one or the other source may be more exploited during successive repetitions of the experiment. This asymmetry did not appear when small colonies were used or if the experimental set-up was manipulated so as to increase the error level. We thus have here a classical example of a bifurcation phenomenon : below a critical value, exploitation is symmetrical, above this value asymmetry occurs (Nicolis and Prigogine, 1977).

Such bifurcation could not be predicted by a system of logistic equations with constant recruitment rates and constant error levels. During mass recruitment, the recruitment rate and error level cannot stay constant, since the trail is reinforced by new recruits. The variance about the trail

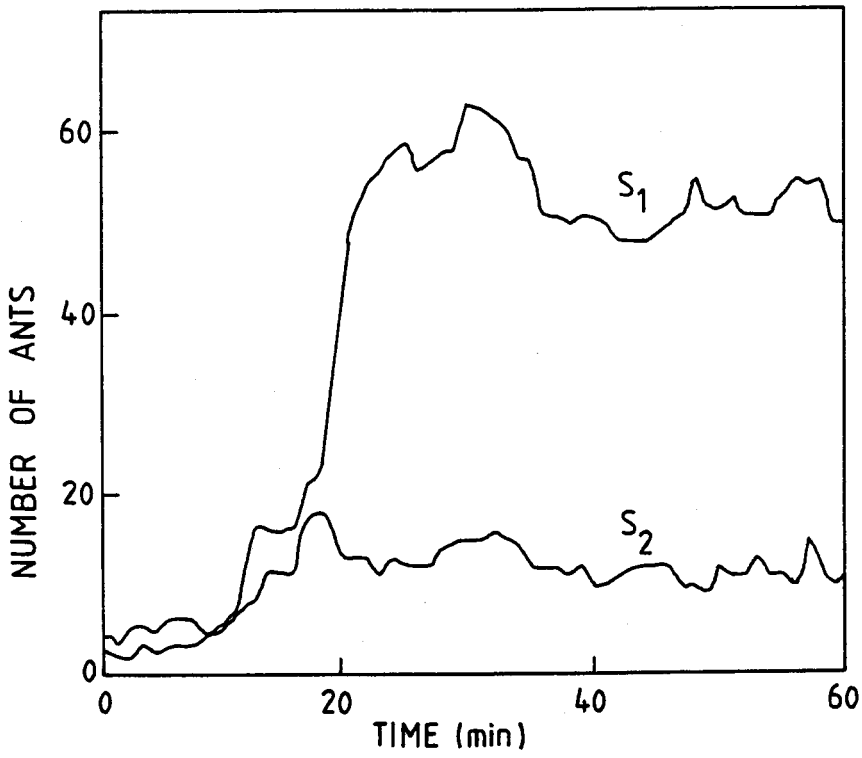


Figure 8
Recruitment in *T. caespitum* for two simultaneously introduced 1M food sources

decreases as the recruitment proceeds and communication increases in accuracy (fig. 9).

An asymmetrical exploitation of two identical sources offers several biological advantages. Both sources are exploited in sequence. A trail with maximal accuracy is built towards one source, the other is kept in reserve, being progressively more exploited as the first vanishes without any interruption in collection. Moreover, the model developed below (section 2.4) demonstrates that with the same number of participants, recruitment is more efficient with one heavily used trail and one less used trail than it would be with two equal intermediate trails.

The ants are thus remarkably able to adjust their recruitment efforts to unpredictable situations. Discoveries are always possible thanks to the stochastic component in communication. Exploitation of the best resource is possible because recruitment is a non-linear process during which accuracy increases if the food source is more rewarding.

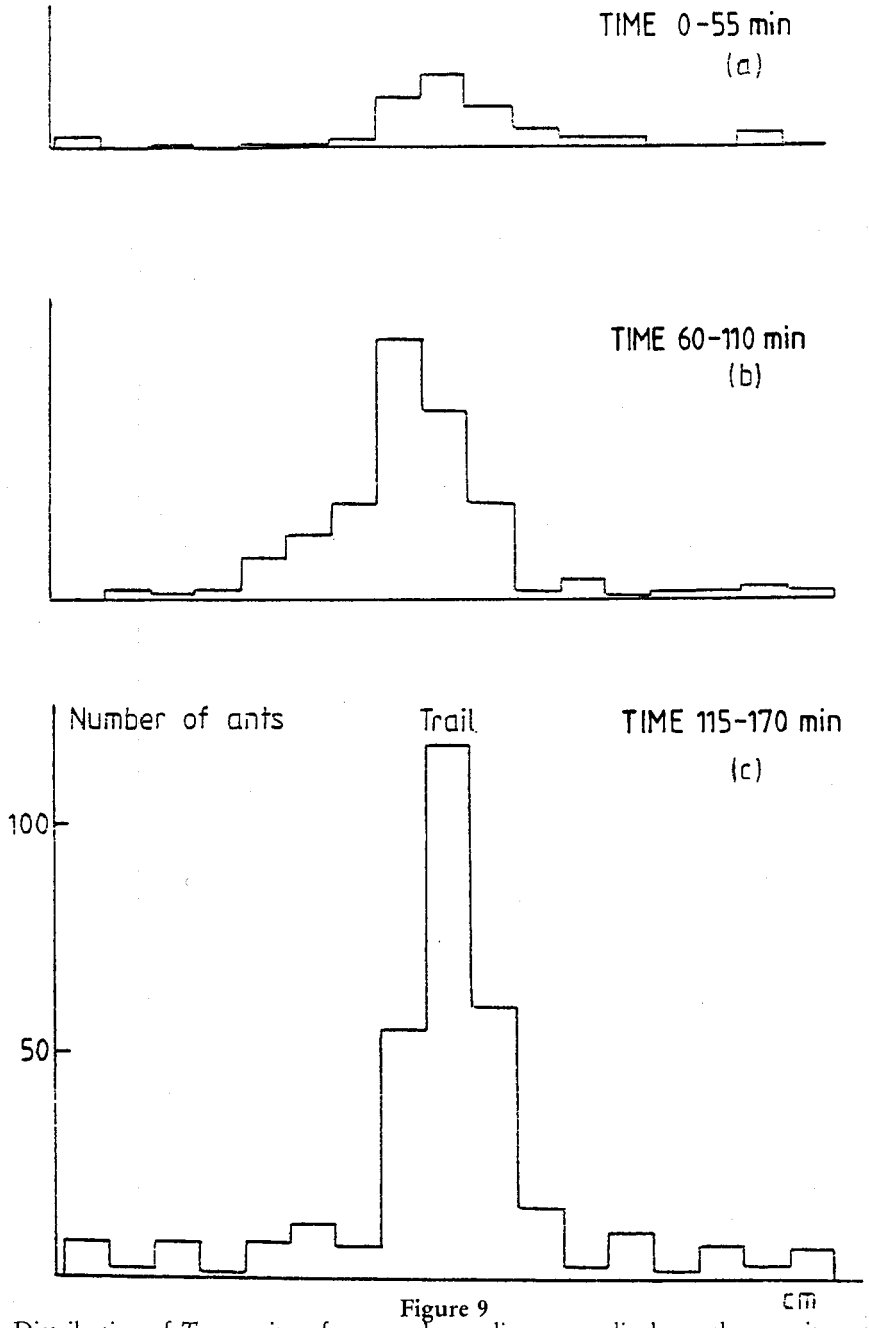


Figure 9
Distribution of *T. caespitum* foragers along a line perpendicular to the recruitment trail, half-way along its length

If snowball mechanisms during mass recruitment are the motor which allows the society to exploit the best discoveries, the stochastic component in communication is the society's imagination making these discoveries possible.

2.3. *How do errors occur during trail following behaviour ?*

Hangartner (1967) showed that orientation along a trail is mainly due to osmotropotaxis, meaning that the ant perceives different concentrations in trail pheromone with its left and right antennae. If the difference reaches a threshold value, the ant moves so as to balance the perceptions of both antennae. Due to the inertia of the movement, it overcompensates and must then correct in the opposite direction, and so on. The result is a sinuous walk, characteristic of an ant following a trail. Osmotropotaxis is not the sole mechanism involved, but if it is indeed one of the main orientation mechanisms, several predictions can be made. Success in orientation must depend on the angle at which the ants reach the trail, and this dependence should be stronger when the concentration in pheromone is low. The loss of the trail could arise when too large of an angular deviation from the main direction occurs, and must be more frequent at low pheromone concentrations.

These predictions were experimentally explored by offering ants circular trails drawn with pure pheromone (Pasteels et al., 1986). The results confirmed the predictions. Acute leading angles increase the probability of successful trail orientation, and more so, the lower the concentration in trail pheromone. For a given concentration in trail pheromone, the number of ants still present on the trail is a negative exponential function of the length of trail followed (fig. 10). Thus the probability of an ant losing a trail remains constant with time, independent of learning or sensorial adaptation. This could be due to the tropotactic mechanism itself inducing a sinuous walk during which the concentrations perceived by the two antennae are periodically reserved. No saturation of the sense organs can occur, and loss of the trail is probably due only to casual movements of the ants.



Figure 10
Logarithm of number of ants still on the trail versus length of trail followed trail concentration = 410^{-12} g/cm.

The mean length of trail followed increases non-linearly with trail pheromone concentration (fig. 11). This increase is best approximated by a power function. However, above a certain level further increase in pheromone concentration impairs the orientation (fig. 12). When the odour tunnel becomes too wide, saturation of the sense organs could take place.

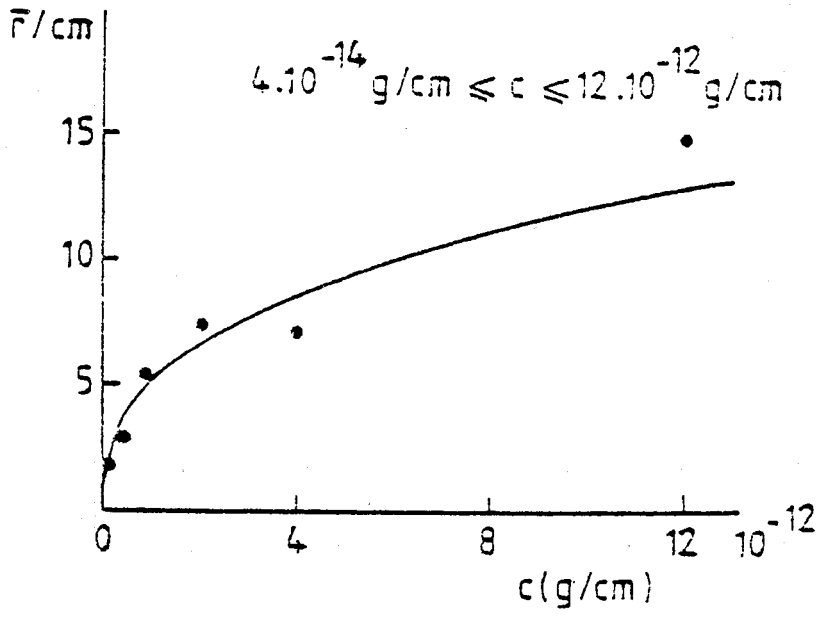


Figure 11
Mean length (r) of trail followed by *M. rubra* as a function of the amount of pheromone (c) on trail.

2.4. A model of mass recruitment

A dynamic model of mass recruitment was developed in which the recruitment rate and error level are non-linear functions of the number of recruited ants (Deneubourg et al. in preparation, fig. 13). This model is a far better description of mass recruitment than the simple logistic equation. For example, it predicts the well-known phenomena of hysteresis and bifurcation (fig. 14). If the number of participants is too small, no recruitment occurs. When the society reaches an intermediate size, or when the error-level is high, two identical sources may be exploited evenly. Asymmetry occurs as the number of participants increases still more. Most of these predictions are already confirmed by experiments.

3. MASS COMMUNICATION

The ant societies described so far have been characterised by a large number of identical foragers which exhibit a simple and random beha-

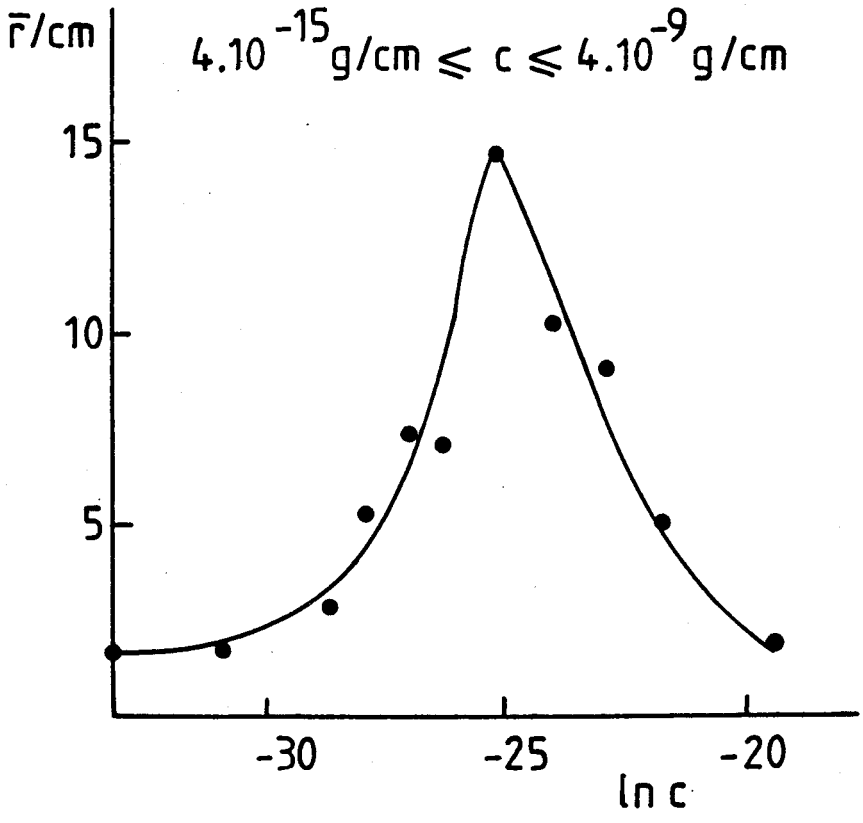
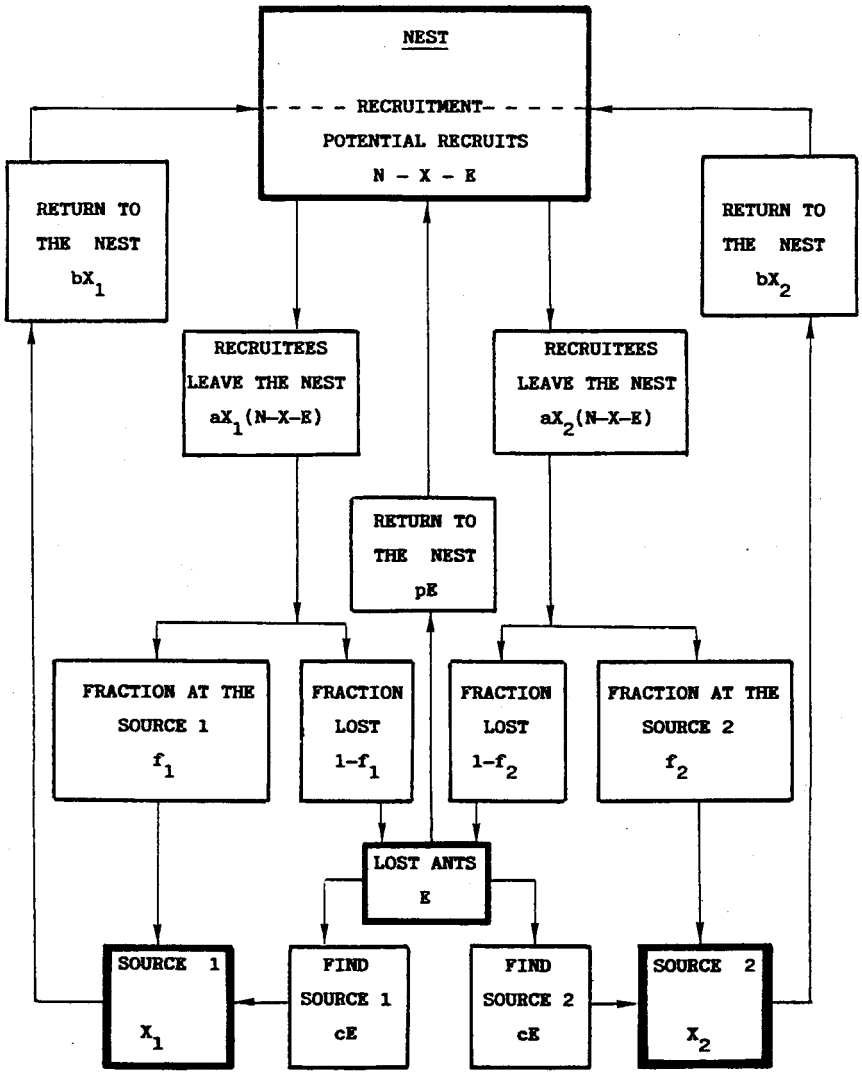


Figure 12
Mean length (r) of trail followed by the ants (*M. rubra*) as a function of the natural logarithm of the amount of pheromone (c) on trail.

viour, with strong communication. The communication and interaction between individuals generates collective, coherent and efficient structures which are characterised by larger scales of time and magnitude than the corresponding ones at the individual scale.

Exploration, discovery, amplification, exploitation and selection of the best resources result from a collective decision, related to the concept of mass communication introduced by Wilson (1971).

Mass communication not only regulates the ants' foraging behaviour, it also regulates other social activities too, e.g. building behaviour. At least in termites, the first stage in building activity appears to be disorder behaviour among the workers. But as soon as nest material accumulates by chance to a level above a threshold value, it attracts workers and a feed-back loop appears leading to coherent structures. This is another example of self-organization process in social insects. If the density of termites is too low, building behaviour is observed, but no coherent structures develop (Grassé, 1959 ; Deneubourg, 1977 ; Bruisma and Leuthold, 1977).



$$\dot{X}_1 = aX_1f_1(N-X-E) - bX_1 + cE$$
$$\dot{X}_2 = aX_2f_2(N-X-E) - bX_2 + cE$$
$$\dot{E} = a(X_1(1-f_1)+X_2(1-f_2))(N-X-E) - pE - 2cE$$
$$f_i = X_i/(g + X_i)$$

Figure 13
E is the number of lost ants. X_1 and X_2 represent the number of ants at source 1 and 2, with $X=X_1+X_2$. N is the number of potential and actual foragers. a is the number of ants recruited per ant at the source, per potential forager, per minute. f_i describes the trail following efficiency. The successful ants, X, remain on average b^{-1} minutes at the source. The lost ants, E, have a probability c of finding the source per minute, or else return to the nest after an average p^{-1} minutes.

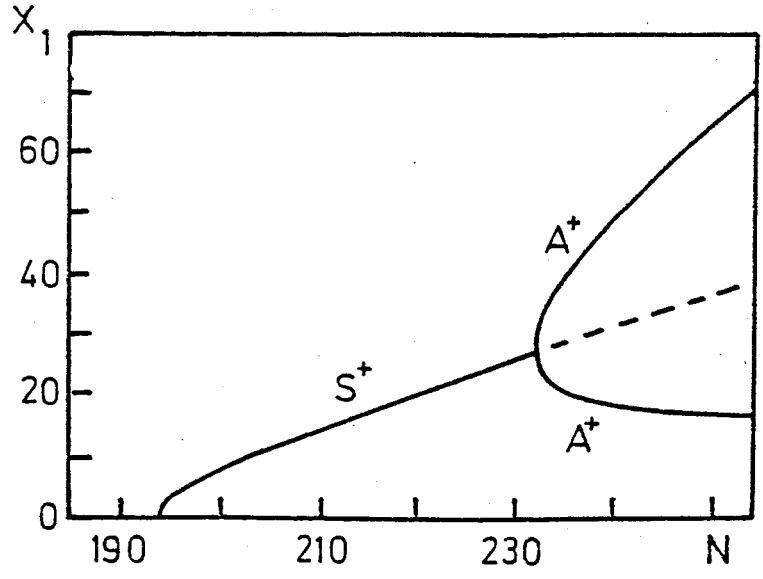


Figure 14a
Two sources, variable error. The number of ants at one of the sources is presented as a function of N, for symmetrical (S^-, S^+) and asymmetrical (A^-, A^+) steady state solutions. Stable solutions are presented by the solid curves and unstable ones by the dashed curves.

As N increases (the other parameters remaining fixed), the system passes, in fig. 14a, through :

(1°) No recruitment
(2°) Recruitment, always successful, with symmetrical exploitation of the two food sources
(3°) Recruitment, always successful, with asymmetrical exploitation of the two food sources.

4. RANDOM OR DETERMINISTIC BEHAVIOUR ?

Is mass communication the only possible way in which the society may function ? Obviously not. There are societies in which the behaviour of the individuals appears far more deterministic either in their communication or in their task specialisation.

A system allowing a constant adaptation of mass foraging to new discoveries is only possible in large societies with numerous participants. For small societies, to bet on errors would be inefficient or even catastrophic. Determinism and accuracy seem a better strategy for small societies. Also when important food resources are highly predictible over long periods of time, continuous recruitment of workers which forage largely at random would be awfully inefficient. A far better solution is to have specialized workers which know what to do.

Deterministic communication and task specialization will be discussed in the next two sections.

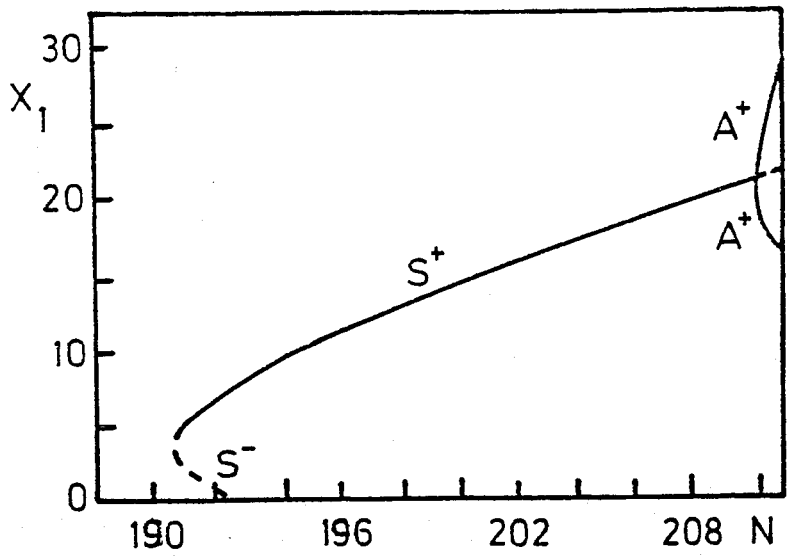


Figure 14b

In fig. 14 b we observe the following sequence :

(1°) No recruitment
(2°) Recruitment, only realized when a critical number (S^-) of ants at the source is reached, with symmetrical exploitation
(3°) Recruitment, always successful with symmetrical exploitation of the two food sources
(4°) Recruitment, always successful, with asymmetrical exploitation of the two food sources.

5. TANDEM AND GROUP RECRUITMENT

Other recruitment techniques are known in ants, in which the recruiting ant actually leads one (tandem) or a small group of recruits by mechanical or short range chemical signals. Depending on the species, an orientation trail may or not be drawn (review in Hölldobler, 1978).

In these techniques, the recruitment rate is lower than in mass recruitment, but accuracy is high since the leader already « knows » the way, and the recruits are in physical contact with the leader. In some species, feed-back mechanisms increase still further the accuracy of the recruitment (Möglich et al., 1974) : should the tandem be broken, and the recruitment ant lose touch with the recruiter, the leader stops and emits an attractant pheromone from its extruded sting. The tandem moves again when the recruited ant is once more in physical contact with the leader by its antennae (fig. 15).

Available data indicate that tandem recruitment is used by small societies. For such societies, it is tempting to speculate that the workers are too few to be lost. It seems more important that the recruits reach the food source than to develop mass foraging. Most societies are of course small during the early stages of their development, and mass recruitment

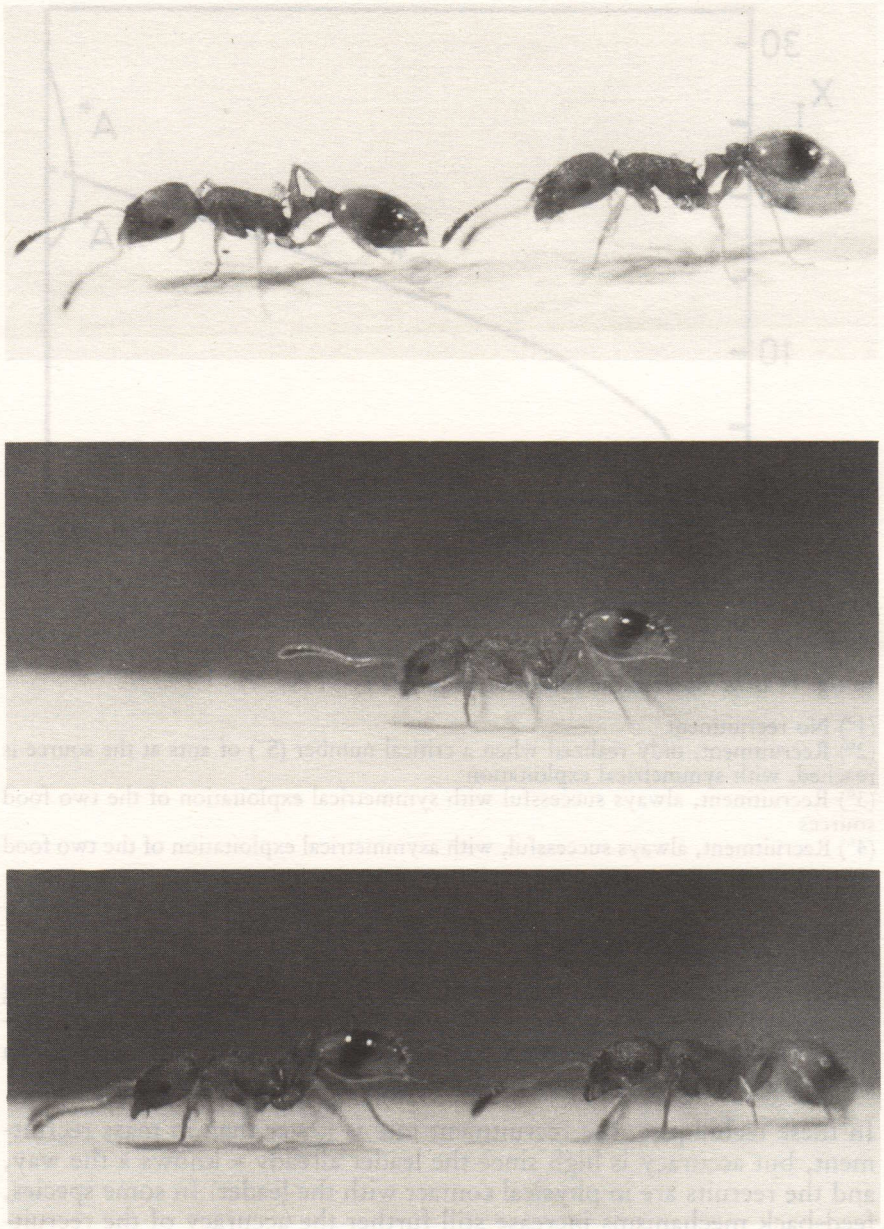


Figure 15
Tandem recruitment in *Leptothorax unifasciatus*. If the tandem is broken, the recruiter stops and attracts the recruit with a pheromonbe from its extruded sting (Photographs taken by S. Aron).

would be only efficient in large mature societies. There are, however, several ways for a society using mass recruitment to avoid being small even in early stages, for example to reproduce by budding (e.g. *Monomorium pharaonis*) and not by an isolated foundress or by being a temporary parasite of other societies (e.g. *Lasius fuliginosus*). In this latter

case, the foundress kills the foreign queen and benefits from the foreign worker force until they are progressively replaced by her own descendants.

Some species use a mixed group and mass mechanism of recruitment. The recruiter both draws a trail and leads a group of recruits (e.g. *Tetramorium*, Verhaeghe, 1982). When the colony is small, group recruitment could prevail, but when the number of participants increases trail recruitment could replace group recruitment. This possibility has not yet received experimental confirmation. However, over the course of one recruitment, group recruitment prevails during the early stages when there are few participants, and mass recruitment dominates during the later stages of the recruitment when the participants are in greater number (fig. 16, Pasteels et al., in preparation ; Verhaeghe, 1982).

6. *SPECIALIZATION BY THE ACQUISITION OF LONG-TERM MEMORY*

The repetition of the same task may make an ant a specialist in that task. Even if we do not know exactly how this kind of long-term memory develops, it seems likely that its development is autocatalytic : the more an ant performs one specific task, the greater the probability that it will continue to perform this task in the future. More or less deterministic patterns will then emerge from the interplay between three factors :

— first, the number of participants influencing the probability that an ant will perform a particular task, due to the competition between individuals for that task ;

— second, more or less stochastic events in the environment inducing a different spatio-temporal distribution of tasks to perform, or different reinforcement patterns ;

— and third, the ants' learning capacity.

In the same way that the level of accuracy in communication could be tuned to environmental and colony constraints, the learning rate and memory capacity for specific activities could also be tuned to both these constraints. This is, however, a nearly completely unexplored field of research.

Short-term memory is frequently observed in ants foraging individually : the forager will return to the neighbourhood where it has just found some food, and it will continue to do so unless it deviates from that location during an unsuccessful search and finds some food elsewhere (Hölldobler, 1980 ; Roisin, 1982). Long-term territorial fidelity, however, is observed in some species in which the foraging ground is partitioned between individual foragers (Wehner et al., 1983 ; Fresneau, 1985) (fig. 17). This long-term memory could evolve by successive reinforcements when the animal makes discoveries, but this remains to be proved.

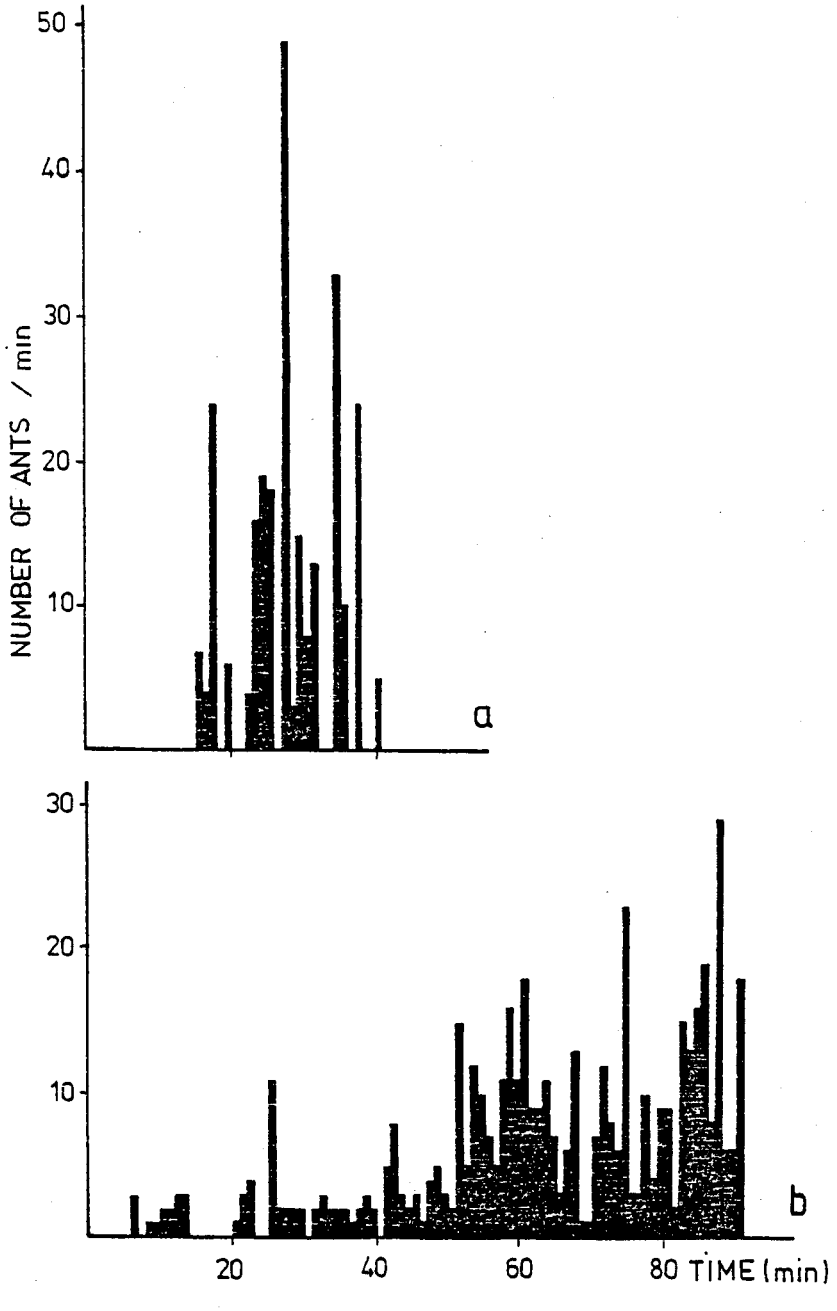


Figure 16
Same recruitment as in fig. 6. Number of ants moving in groups (a) and alone (b).

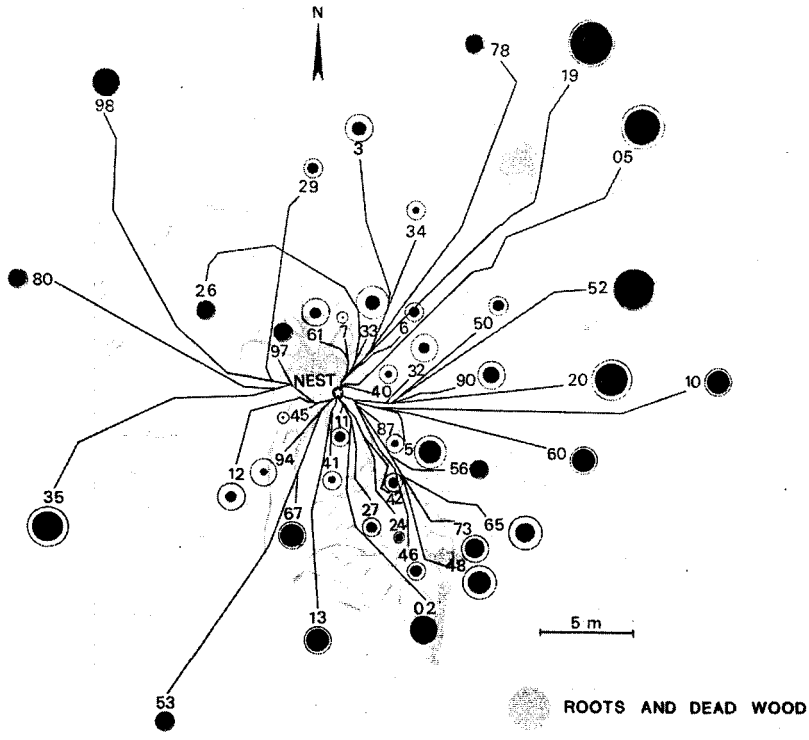


Figure 17
Map of foraging routes taken by 43 foragers of *Neo-ponera apicalis* over one month period. The numbers at the end of routes are those with which the individuals were identified in the field. Open circles : total number of individual excursions. Solid circles : percent of returns with prey. (From D. Fresneau).

One of the most spectacular examples of specialization due to long-term memory is given by some European species which collect honey dew from aphid colonies on trees, which is a highly predictible and stable resource (Rosengren, 1971 ; Hennault-Riche et al., 1979). These species build permanent roads (fig. 18) and the foragers exhibit not only road fidelity, but also fidelity to one particular branch in a tree. Each spring the road system is rebuilt with great accuracy by old foragers which have kept the memory of its direction during the winter. Their memory is so good that it was the ants which helped the biologists find again the sticks used the previous year as signposts for the road. The same ants marked with a spot of paint while on one road the preceding year were observed again on that same road the next spring. Young, naive foragers are not specialized but become so in the course of the season.

Division of labour due to learning could occur in other activities like brood care or emigration. This is a subject barely touched, but the fol-

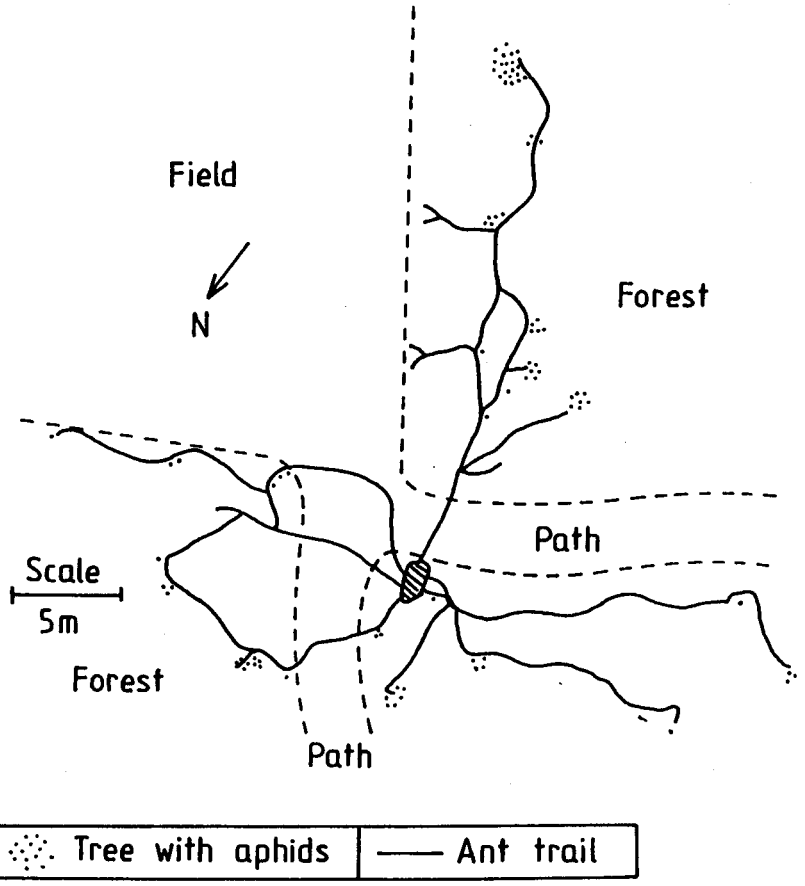


Figure 18
Permanent road of *Lasius fuliginosius* leading to aphid colonies

lowing experiment illustrates this possibility. Division of labour was assessed in ant societies during nest moving where the eggs, larvae and queen must be carried to the new nest-site (Abraham et al., 1983). This is an ideal circumstance to study division of labour, because all the members of the society must participate in the move in one way or the other. Worker activity was measured by the number of transports of either larvae or queen. All the workers from one society were individually labelled and followed for more than eight months during which moves were periodically induced. In fig. 19, ants were ranked according to their level of activity. There was a strong difference in the activity level of workers, although they were chosen from a mother colony as identical as possible in both age and size. This distribution remains stable from one move to another. The society was then split into two halves, one half containing all the active workers, the other all the inactive ones.

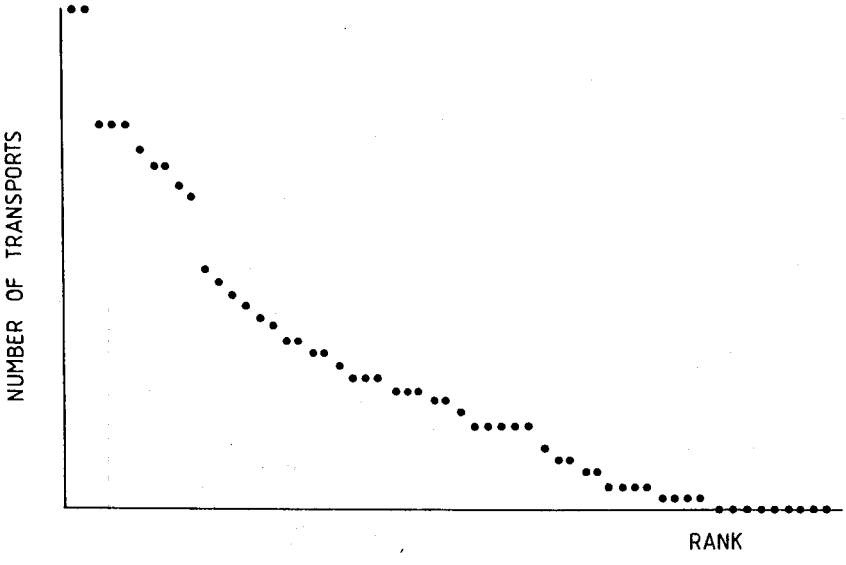


Figure 19
Rank order of ants (*M. rubra*) based on their transport activity during nest moving.

The number of eggs and larvae to transport was made proportional to the number of transporting workers. Each half colony was then induced to move, after which the ants were again mixed together and tested.

The results are summarized in fig. 20 in which each dot represents the relative activity of the two groups measured during series of five moves. Before the second trial, the society was split. The difference in activity between the two groups decreased when each group moved separately. This was due to the fact that in the active half colony, some ants were forced to rest due to the competition for work, and the inactive ants in the other half colony were of course obliged to work. When the ants were mixed again before the third trial, a temporary memory of their last experience was observed. The ants did not immediately return to their original level of activity of before the split. Further evolution of the system will not be discussed here, since it involves other parameters like change in the society's demography. The important point is that past activity influences future activity in a conservative way, suggesting that division of labour may be due in part to a learning process in the broad sense.

7. CONCLUDING REMARKS

Two different self organization processes in social insects were discussed both resulting from the interplay between stochastic events and snowball mechanisms. They suggest at first sight antinomic foraging strategies and social systems.

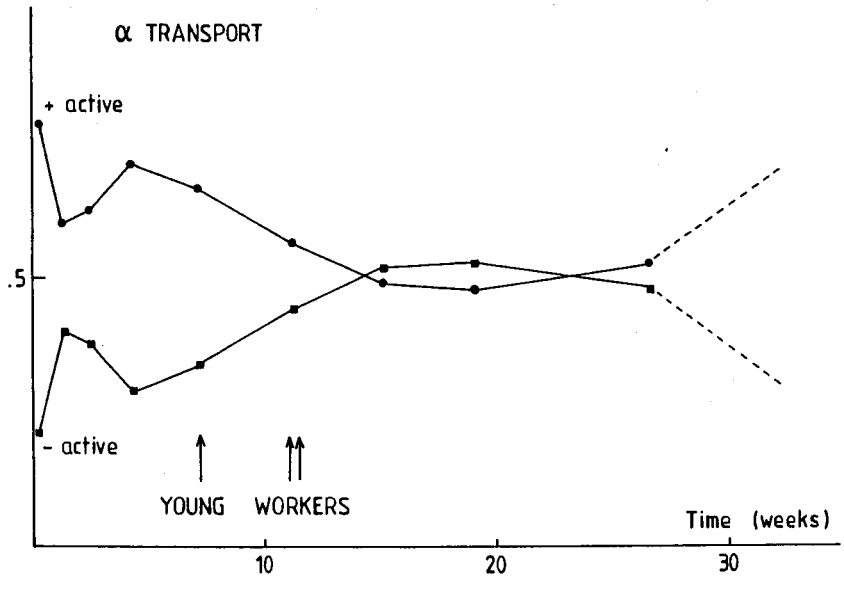


Figure 20
Relative activity of two groups of workers (*M. rubra*) during successive trials. Each trial is a series of five moves. Before the second trial the colony is devided in two groups : the most and the least active ants of the previous trial. The two groups are fused again after the second trial.

The first process, designated as mass communication and best exemplified by mass recruitment, suggests large societies of unspecialized workers, exhibiting simple and random behaviour. They forage on short-lived, unpredictible resources. The second, less documented case, is the development of long-term memory for particular tasks. It suggests small societies of highly specialized workers, exhibiting a more complex and deterministic behaviour on long lasting, or large societies foraging predictible resources.

It must be emphasized that there is no such dichotomy in nature, but a continuum from one kind of social organization to the other, or a complex mixture of both. Only some foragers may show a territorial fidelity, whereas others remain unspecialized, possibly available for unpredictable events. In *Lasius fuliginosus*, specialized foragers follow permanent roads to the stable aphid colonies, but mass recruitment to less permanent and unpredictable resources, like earth worms, are possible. It is indeed the challenge for sociobiologists to understand how both mechanisms, together with other possible ones, operate in the functioning of an insect society. The concepts necessary for such analyses are presently available.

There are between 10.000 and 20.000 living ant species, exhibiting a large diversity of life styles, living in all kinds of terrestrial environments, with colony sizes ranging from 10 to several millions of individuals. This extraordinary diversity provides all possible natural examples allowing one to test theoretical predictions. Ant societies are not only amazing

natural curiosities, but also real life models which constitute an ideal opportunity to investigate how patterns and structures emerge in complex biological systems made up of more or less numerous and autonomous sub-units.

ACKNOWLEDGMENTS

Many thanks are due to our colleagues M. Abraham, S. Aron, Ph. Champagne, Cl. Detrain, D. Fresneau, M. Parro, Y. Quinet and J.-C. Verhaeghe for many stimulating discussions and for having allowed to reproduce some of their figures. We wish also to thank Professors I. Prigogine and G. Nicolis for their constant interest in this study. The technical help of MMrs C. Fernandez, E. Mete, N. Van Mol, and Mr. P. Kinet is deeply acknowledged.

REFERENCES

M. ABRAHAM, J.L. DENEUBOURG, J.M. PASTEELS, 1984, *Act. Coll. Ins. Soc.*, 1 : 19 – 25.

O.H. BRUISMA, R.H. LEUTHOLD, 1978, *Proc. VIII Cong. IUSSI, Wageningen*, pp. 257-258.

Ph. CHAMPAGNE, J.L. DENEUBOURG, J.C. VERHAEGHE, J.M. PASTEELS, 1984, *Act. Coll. Ins. Soc.*, 1 : 31 – 37.

J.L. DENEUBOURG, 1977, *Ins. Soc.*, 24 : 117 – 130.

J.L. DENEUBOURG, J.M. PASTEELS, J.C. VERHAEGHE, 1983, *J. Theor. Biol.*, 105 : 259 – 271.

D. FRESNEAU, 1985, *Ins. Soc.*, 32 : 109 – 116.

J.L. GOULD, F.C. DYER, W.F. TOWNE, 1985, *Forstchr. Zool.*, 31 : 141 – 161.

P.-P. GRASSÉ, 1959, *Ins. Soc.*, 6 : 41 – 83.

W. HANGARTNER, 1967, *Z. Vergl. Physiol.*, 57 : 103 – 136.

W. HANGARTNER, 1969, *Z. Vergl. Physiol.*, 62 : 111 – 120.

B. HENNAULT-RICHE, G. JOSENS, J.M. PASTEELS, 1979, *C.R. UIEIS sect. franç., Lausanne*, 71–78.

B. HÖLLDOBLER, 1978, *Adv. St. Behav.*, 8 : 75 – 115.

B. HÖLLDOBLER, 1980, *Science*, 210 : 86 – 88.

M. MÖGLICH, U. MASCHWITZ, B. HÖLLDOBLER, 1974, *Science*, 186 : 1046 – 1047.

G. NICOLIS, I. PRIGOGINE, 1977, *Self Organization in Non Equilibrium Systems*, New-York, Wiley.

G.F. OSTER, E.O. WILSON, 1978, *Caste and Ecology in the Social Insects.* Princeton University Press, Princeton, New Jersey.

M. PARRO, 1981, *Mémoire de licence*, Université Libre de Bruxelles.

J.M. PASTEELS, J.L. DENEUBOURG, J.C. VERHAEGHE, J.L. BOEVÉ, Y. QUINET, 1986, in : *Mechanisms of Perception and Orientation to Insect Olfactory Signals*, M.C. Birch and T.L. Payne eds. Oxford University Press, Oxford, 131 – 137.

J.M. PASTEELS, J.C. VERHAEGHE, J.L. DENEUBOURG, 1982, in : *The Biology of Social Insects*, M.D. Breed, C.D. Michener, H.E. Evans eds. Westvieuw Press, Boulder 297 – 301.

Y. ROISIN, 1982, *Mémoire de licence*, Université Libre de Bruxelles.

R. ROSENGREN, 1971, *Act. Zool. Fennica*, 133 : 1 – 106.

J.C. VERHAEGHE, Ph. CHAMPAGNE, J.M. PASTEELS, 1980, *Biol. Ecol. Mediter.*, 167 – 168.

J.C. VERHAEGHE, J.L. DENEUBOURG, 1983, *Ins. Soc.*, 30 : 347 – 360.

R. WEHNER, R. HARKNESS, P. SCHMID-HEMPEL, 1983, *Foraging strategies in individually searching ants Cataglyphis bicolor* (Hym. Formicidae). Gustav Fischer Verlag, Stuttgart.

E.O. WILSON, 1962, *Anim. Behav.*, 10 : 134 – 144.

E.O. WILSON, 1971, *The Insect Societies*, Belknap Press of Harvard University Press, Cambridge.

A Study Program of Chaotic Dynamics Applied to Information Processing

John S. Nicolis
Dpt. of Electrical Engineering — University of Patras — Greece

1. INTRODUCTION

In this brief report, we present some ideas developed in the last ten years on the possible applications of chaotic dynamics, to information processing and the dynamics of pattern formation and pattern recognition.

Let us start by dividing, for the shake of the argument, all cognitive tasks into two broad categories :

a) « Exercises » or « deductive » problems-whose outcomes always exist, are unique and are implicitly contained in the initial premises-provided that these premises form a complete and self consistent set, and

b) « Complex » problems, of the « Inductive » type whose outcomes may or may not exist, they may be not unique and are not necessarily nested in the initial conditions but are in part generated by the evolution of an iteractive and recursive procedure described by some underlying dynamical system. Under such circumstances we say that if there is asymtotic stability in the dynamical flow in state space, then there is a *discriminable pattern* or a « solution ».

In dealing with problems of the a) category one goes through regressive steps on a sequence which converges, eventually « halts » and gives the solution. No *new* information is produced during this process although admittedly, the agent who deals with the problem may be ignorant of how to turn the « implicit » into « explicit » or how to choose amongst different strategies of approach. We bypass this issue here.

In problems of the b) category, convergence cannot be guaranteed. Such problems are « dynamical » and their outcome can only be understood

in a historical that is, an evolutionary context. When we say that the information required for the solution of an « inductive » problem is not « hidden » exclusively in the initial conditions — so there is *no* a priori truth (« α–ληθεια' ») to be revealed — we mean that at certain moments the evolving scenario that is the trajectory of the system in state space *branches-off* or bifurcates into multiple asymptotically stable regimes (stationary, periodic, quasi-periodic or aperiodic) in ways which depend on critical values of a number of control parameters.

These control parameters stand for the coupling among the variables, and their numerical values are shifted by ever-present (internal and/or external) fluctuations. When these parameters acquire their « critical » values the coupling between the variables concerned changes in ways which precipitate instabilities-meaning that infinitesimally small changes in the parameters lead to finite changes in the variables. Beyond these instabilities the (new) structure(s) may not retain any trace of memory of the previous regimes-although it (they) evolved from them via quite deterministic processes. Similar « loss of memory » takes place in systems without bifurcations but with a sensitive dependence of their flows to initial conditions.
It is exactly this mixture of determinism and (long or short-term) unpredictability which makes the problems of « inductive » type so relevant to information sciences.

Most systems relevant to information sciences (e.g. linguistics) contain multiple, hierarchical (see below) feedback-loops-which inevitably make them « Self »-referential. If such systems do not possess « closure » or logical consistency they are « paradoxical ». The « truth(s) » — if any — revealed by such systems is associated with « Self-consistency » and this characteristic as we are going to see implies the existence of stable attractors in the iterative and recursive system we call « self-referential sentence ». The search for self-consistency of a self-referential sentence then, leads us to the formulation of a boundary value problem. If so, the suggestion follows immediately that « logic » or « rationality », so far a static concept, can be dynamically implimented and replaced with « evolutionary stability ».

2. STATEMENT OF THE PROBLEM

In a series of 1-6) papers we developed the thesis that chaos may provide the connection between information theory and physics, the main issue being that information is a physical entity and as such it should not be treated in absentia of its dynamical (hardware) substratum.
How can we explain on the basis of known dynamical laws and processes of physics the communication transaction taking place between two systems — i.e. not only the energetic exchanges but also the information ones ? How can we dynamically implement the generation, stable storage, reliable retrieval and dissipation of information ? In other words : what are the physical (dynamical) agents mediating symbolic interactions ?

In the first place we have to clarify what we mean when we assert that two systems « communicate ». For a physicist information is admittedly

an almost « metaphysical » word. Indeed in the current parlance of the mathematical theory of communication (invented by Claude Shannon 37 years ago and not improved considerably ever since) information processing is to be considered as a sequence of (one-to-one) *mappings* from the members of a set belonging to a multidimensional abstract signal-space onto the members of another set belonging to an abstract space of a *different* dimensionality ; the « solution » of the problem of information transfer from set I to set II or more concretely from place A at time τ to place B at time $\tau + \Delta\tau$ boils down essentially to a *variational principle* which etries to find the optimum mapping (or code) in order to compromise two conflicting requirements : speed of transfer and reliability or degree of errorless reproduction of an abstract pattern of digits sent from I and received by II : in such a formulation all physics is left out.

Such a mentality of treating communication processes serves admirably the whole group of applications associated with typical communication engineering design. People grow restless, however, when the above formalism is applied to fields such as biology, cognitive psychology and in general to interactions between « self »-organized systems. The reason is that in such cases complete decoupling between physical and symbolic interactions (or between hardware and software) is no more possible or meaningful and people start asking questions about the ongoings at the *interface* between structure and function.

In such cases information processing is associated :

a) With cascades of mapping — transcriptions and/or translations — (*not* necessarily one-to-one but also multi-to-few, i.e. hierarchical) from one *physical* entity onto another. For example in the generation of speech a pattern residing in a sequence of nerve pulses incident upon the vocal chords is *mapped* onto an air sound wave. This sound wave is intercepted by a microphone which *maps* the pattern onto an electromagnetic, wave etc.

b) With low-energy triggers giving rise to high-energy effects (when informational channels intersect effector system) or, to put it another way with extreme sensitivity to initial conditions.

There is *physical* justification (see e.g. 6) — 8)) for the hierarchical or *non* one-to-one mapping between « object » and « image » that is the fact that physical perception *is* ambiguous. There are three distinct reasons for this :

a) The loss of evanescent waves : half of the encoded information in the spherical wave scattered from the object is irretrievably lost due to the impossibility of designing receptors for both propagating *and* evanescent waves, into which a spherical wave is analysed.

b) The existence of fluctuations around the average number of photons carried by each normal mode or electromagnetic degree of freedom.

c) The finite aperture of the perceiving instrument : This finiteness has two results : it distorts like a spatial filter a portion of propagating components of the impinging spherical wave and second it attributes a finite

number of degrees of freedom on the receiving image. The ambiguity here is due to the fact that the spherical wave which mediates originally the information transfer from the transmitter to the receiver possesses an infinite number of degrees of freedom.

At first glance it seems that there exist irreconciliable differences between Soft- and Hardware[1]. Specifically the difficulty lies in the fact that at the Hardware (physical) levels, processes follow *laws* with universal validity whereas at the Software (symbolic) levels processes follow *rules* whose relativity and flexibility are manifested even at the levels of a writer struggling for « le mot juste ».

What is the difference between laws and rules ? Can we « reduce » the second to the first ? Our conjecture at this point is that the issue involves essentially *different time scales* : if ones allows a deterministic physical law to operate *long enough* the result may be chaotic. However the time scale is quite different say for gravitational turbulence, « cardiac » turbulence (fibrillation), hydrodynamic turbulence or symbolic turbulence (Language ?) to develop.

The time scale beyond which chaos emerges is in fact depending not only on the intrinsic nonlinearities and the stability properties of the system involved but also on extrinsic reasons (initial and boundary conditions) : depending on the case this time scale may be large ($\sim 10^{15}$ years) or small ($\sim$ a few minutes).

The above theorizing aims at bridging the collosal difference in « behavior » between say 1) a physical body which when faced with the task of moving from place A to place B follows simply and « compulsively » Maupertuis' variational principle and 2) an animal which « chooses » the specific pathway anew every time ; in this second case the animal follows a rule or algorithm which may spontaneously modify itself depending on slight changes in the initial conditions and the parameters of the situation. We claim in short that symbolic processes *are* chaotic manifestations of collective properties of physical systems — at a higher hierarchical level ; « Decision making » and « free will »[2] are *deterministic* since they are harnessed by natural laws but *unpredictable* since they are chaotic ; the same thing could be claimed for music and language.

Be as it may we shall argue that the most important aim of the communication transaction is not just *copying* but *simulation* : this essentially refers to a process of *modelling* a physical system by another one with the obvious goal of predicting and controlling it. We say that a physical system I (see Fig. 1) simulates another (II) when it is capable, out of a finite length time series it receives, of constructing minimal length algorithms or *compressed* descriptions of II, which algorithms when introduced as inputs (programs) in a finite-stage machine will allow it to furnish as an output the behaviour of system II for all times future *and* past.

Since the prerequisite for simulation is the compressibility of the received time series from the partner-system (or the environment) we, in this paper, indulge in a new dynamical « paradigm » namely *dissipative chaos*, which through the gadgetry of « strange attractors » promises in a single stroke to do this : to carry out very complex functional

repertories with very simple « hardware » and to behave as a versatile (albeit somewhat uncontrollable) information processor. The new paradigm comes forth if one realizes that information is not only produced via cascading bifurcations giving rise to broken symmetry, but also via successive iterations giving rise to ever-increasing resolution. Beyond a certain resolution interval, microcscopic fluctuations are no longer smeared out ; they get amplified and pass from lower to higher hierarchical levels. It is these fluctuations which essentially account for *new* information generated by an evolving hierarchical system.

In short, the purpose here is to develop an evolutionary approach for the basic problem of generation, storage and dissipation of information in physical and biological systems via dynamical processes. These processes are associated with branching-offs or cascading bifurcations giving rise to broken symmetry as well as chaotic dissipative dynamics giving rise to decreasing and increasing resolution. From an attractor, simulating some aspects of a cognitive system (or a « processor ») there exist *two* basic requirements : big storage capacity and good compressibility. Conventional singularities such as stable steady states and stable limit cycles, possessing dimensionality zero and one, respectively are very poor as information storage units but ideal as information compression devices.

Strange attractors on the other hand via an harmonious combination of expanding and converging trajectories in state space (« positive » λ_+, and « negative » λ_- Lyapounov exponents, respectively) can in principle comply with both requirements : they may possess a considerable information dimensionality — and this idiom makes them suitable for dynamical information storage ; on the other hand, being « attractors », ($|\lambda_-| > \lambda_+$) they can serve at the same time as information compressors. (See following Sections for details)

We are going to see that (dissipative) chaotic dynamics offers an alternative and perhaps more elegant way towards the formation of hierarchy and classification : whereas the time honored statistical-mechanical methods lead to hierarchical structures through the collective properties of a Master equation (e.g. moments and crossmoments of the probability density function describing the dynamics at the « microscopic » level) dissipative chaotic systems accomplish the same thing differently : they do it by reducing the number of degrees of freedom in state space-where the dynamical system starting from a set of initial conditions, of dimensionality N sooner or later locks — in at a compact subset of state space (the « attractor ») — of *much lower dimensionality.*

For System I to be able, in the above sense, to simulate or model System II (and vice versa), a number of sine qua non prerequisites are demanded.

In the first place, the system must be hierarchical, i.e. possess at least a hardware (energetic-structural) level and a software (cognitive-symbolic, -functional) level. Information, in the form of discrete time series emanating from the S-level of System II, impinges at level H of System I, where a set of convolutions (cross-correlations) takes place among the time series received and the intrinsic dynamics at level H is chaotic. The

result of such convolutions is the emergence of collective properties possessing by far fewer degrees of freedom than S_2 or H_1. These collective properties are now transported at the level S_1.

The communication or mutual simulation between the two systems may go on, therefore, as the evolution in time of a set of pairs of coupled nonlinear integral-difference equations (each associated with a given value fo ξ and ϱ).

$$\left.\begin{aligned} S_1\,(n+1) &= \lambda S_1\,(n) + < H_1^{\xi}\,(n)\, S_2^{\varrho}\,(n)> \\ S_2\,(n+1) &= \lambda S_2\,(n) + < H_2^{\varrho}\,(n)\, S_1\,(n)> \end{aligned}\right\} \qquad \lambda > 1$$

$H_j\,(n+1) = f_j\,(H_j(n))$ or $H_j = f_j\,(H)$ being the intrinsic dynamics at the hardware level(s).

Each pair stands in eqs (1) for one statistical moment. For each pair of the above equations the process clearly amounts to a cascading iteration on S_1 and S_2. S_1, S_2 stand for the descrete (in general multidimensional) variable at the Software level(s) H_1, H_2 stand for the (discrete or continuous) variables at the hardware levels and n is discrete time ; ξ and ϱ are positive integers. Finally <.> stand for time averaging.

The harmonious co-evolution of (dynamical) bare variables and their collective properties in time is a demand for any linguistic scheme which includes semantic connotations i.e. context and meaning : this ongoing iterative interactions (Fig. 1) between a collective property (a « set ») with each of its own members and vice-versa creates the « context » and at the same time is responsible for the phenomenon of self-reference. Due to a difference in the time scales at the H and S levels we can say that the S levels may exercise history depending, feedforward control on the H levels. Indeed the integral component brings forth « memory » effects.

We have to emphasize that the above model is quite different in philosophy and physics from the usual « Artificial intelligence » (A.I.) paradigm. A.I. are local[3] theories dealing essentially with logical circuitry ; in them deductive processes take place where the outcomes are implicitly included in the initial premises. Chaotic processes on the other hand are more akin to inductive reasoning ; most of the information is *not* contained in the initial conditions but is produced « en route » via branching offs of the dynamical flow itself.

Hierarchy — like self-organization (or the ability of a system to compress environmental time series and construct minimal — length algorithms for the simulation of the environment involved) — is an *emergent* property of nonlinear dynamical systems evolving under a discontinuous interplay among the basic (four) types of physical interactions. It is through a study of such an interplay that one can historically understand the ability to communicate as an evolutionary, emergent property of complex large-scale systems (something which runs *against* the traditional worldview dictated by a static universe running down to his « thermal death »). That is to say, in an expanding universe the interplay of the physical interactions gives rise in a concomitant way to entropy

production, progressive differentiation, increasing complexity, and increasing organization. The cosmological, thermodynamic, electromagnetic and signal-processing « arrows of time » all point in the same direction : the direction of increasing differentiation with time, and the emergence of cognitive systems which not only can simulate the external environment but, in a certain incomplete way, even parts of themselves.

Let us now deliberate on the concepts of hierarchy, complexity and organization.

A. Hierarchy

By a hierarchical system we mean, in general, an ensemble of interacting parts that is composed of (and is analysable or decomposable into) successively nested sets of interacting subunits.

Each set of interacting components-forming a distinct hierarchical level-affords a specific state-space description with variables and parameters pertaining to that particular level. The interacting variables (and/or parameters) at a higher hierarchical level are « collective properties » (statistical moments or convolutions) of the dynamics going on at the level just underneath. Therefore as we move to a higher level, we (usually but not always — see last section) witness a tremendous reduction in the number of degrees of freedom. The higher level receives selective information from « below » and in turn it exercises on the dynamics of the lower level (efferent) feedforward control commands — which harness the further dynamical deliberations on that lower level. (Hierarchical mapping)

Although two systems do not have to be hierarchically predisposed in order to exchange energy, in order to enter in a communication transaction they certainly do. (We need one hierarchical level to perform a cross correlation and another higher one in order to « cognize » it).

B. Complexity

A system may be complex either on a structural or a functional level. Structural complexity increases with the number of interacting sub-units ; the percentage of mutual connectedness among them, in dyadic or plural fashion ; and the variance of the probability density function of the strengths of interactions between the individual sub-units.

On a functional level complexity increases with the minimum length of the (most compressed) algorithm from which one can retrieve the full behavior of the system. In this sence, as we are going to see in the paper, the functional complexity of a system increases with the length of its « phylogenetics tail », namely its evolutionary history marked by a discrete sequence of bifurcation episodes. In that sense, then, a virus or an orange is a more complex system, functionaly speaking, than a big star, since in order to understand the function to each one of them you already need as a prerequisite the existence of a supernova. But let us begin from the begining :

Consider a macroscopic physical system at its lowest hierarchical level, namely the level of equilibrium thermal (molecular) *mixing*. At that par-

ticular state the system possesses the maximum number of degrees of freedom or the fullest possible symmetry — exactly equal to the symmetry of the microscopic laws of motion of its constituent interacting parts. The symmetry is then maximum, the number of degrees of freedom likewise and the complexity is minimum since the number of bits required to specify a system at thermodynamic equilibrium is the shortest possible : to reproduce such a system one may start from almost *any* initial condition.

Imagine now that the system comes in contact with the environment ; it may again settle to equilibrium when new conditions for minimization of its free energy are met ; or it may happen that as a consequence of internal and/or external microperturbations (which are inevitable due to the enormous number of the constituent components of the system and its environment) the one and only stable steady state of equilibrium may become deleted. Beyond a certain critical distance from equilibrium (which the system franchises triggered by fluctuations) at least two new, alternative stationary configurations (containing their own symmetrical branches) emerge. Suppose the system settles in one of them by a more or less random « choice ». At this metastable state the system now forms, say, a spatial static pattern (i.e. a crystal) which, although it may appear quite symmetrical, has a number of degrees of freedom far below the one corresponding to the symmetry of the laws of motion of its constituent molecules. Nevertheless, we do not yet speak of « broken symmetry » but rather of « repealed » symmetry[ref.14] since the probability of « tunneling » to the symmetrical metastable state remains finite (although it rapidly decreases with the scale of the system).
If however by coupling the system with an external asymmetric field we perturb the bifurcation diagram so that the alternative mirror symmetrical configurations get repressed[ref.19] we do witness a genuine broken symmetry and the asymmetrical configuration at the corresponding state may be stable for a wide range of the control parameters. We say that the complexity of the system has increased meaning that the length of instructions or the minimum amount of bits an external observer requires for the replication of the system has *increased.* We can imagine a succession (cascade) of such symmetry breaking, bifurcation « episodes » stimulated by the shifting of the numerical values of the control parameters beyond critical thresholds ; these broken symmetries increase the complexity of the system by sequentially reducing the number of degrees of freedom of the new pattern(s), (via triggering long-range spatial-temporal correlations). The greater the reduction of symmetry i.e. the greater the number of *constraints* a pattern incorporates, the greater the length of the minimum program necessary for the « complete » description of the system — which would lead to its replication.

So we may consider that each pattern (at a hardware level) is associated with a number of digits (at a software level) — we could classify them as beloging to two categories « pre »-digits and « post »-digits. Predigits compose the set of the successive critical (threshold) numerical values of the control parameters, which triggered the corresponding sequence of bifurcations whose final product of broken symmetry is the pattern under consideration. The more abstract the pattern is the longer

the sequence of pre-digits needed to mark (and preserve its « dotted » or « punctuated » evolutionary history. We may say that the pre-digit set is the evolutionary « zip-code » of the pattern. If the set of pre-digits constitutes the « address » of the pattern the set of post-digits constitutes its « ID card » or rather its « fingerprints » : the set of post-digits contains the set of the amplitudes of the « order parameters »[ref.23] or the dominant eigenfunctions which emerge after an instability. These eigenfunctions form the collective or macroscopic properties of the pattern concerned.

If a pattern dynamically produced as above is now described or coded using a string of { 0,1 } symbols the length of the shortest such string can be deduced as follows : let p be the a priori probability of 0's and 1-p that of the 1's in a fair coin tossing realization of length N digits (bits). There are $\binom{N}{pN}$ such sequences, or

$$\frac{N!}{[(pN)!]\quad[N(1-p)!]}$$

Hence $\log_2\binom{N}{pN}$ bits suffice to uniquely specify such a sequence, or

$$\frac{\log_2\binom{N}{pN}}{N}\ \frac{\text{bits}}{\text{per digit}}\ ;\ \text{using Stirling's formula}$$

$\xi! \simeq 2\xi\ \xi^{\xi}.e^{-\xi}$, $\xi >> 0$ and applying it above, we get the expression $S(p) = -p\log_2 p-(1-p)\log_2(1-p)\ \frac{\text{bits}}{\text{per digit}}$

for uniquely specifying an N sequence of 0 and 1 with p N symbols 0 and (1−p)N symbols 1. (as $N \to \infty$)
However this is not all : to reconstruct a given *finite*length N sequence of O's and l's where p is the a priori probability of O's, one needs to specify the number of zeros which for N finite is not exactly known that is one needs an extra amount information $\log_2 (N+1)$ bits long[4]. So the total *complexity* of the pattern is

$N' \sim \log_2(N+1) + \log_2\binom{N}{pN} \sim \log_2 N + NS(p)$ for $N >> 1$ or just

$$S(p)\ \frac{\text{bits}}{\text{per digit}}\ \text{as}\ N \to \infty$$

Chaitin[ref.21] showed clearly that the overwhelming majority of the above sequence are incompressible i.e. random.

Indeed, consider the number of sequences of given length N which are compressible up to K bits i.e. $N' \leq N-K$. This number equals

$$\Sigma = 2^1 + 2^2 + \ldots + 2^{N-K-1} = 2^{N-K} - 2$$

The percentage of compressible sequences is thus

$$\eta = \frac{\Sigma}{2^N} \sim 2^{-K}(N >> 1)$$

So, for K = 10 for example, only one in a thousand sequences can be specified by an algorithm 10 bits shorter that the sequence itself.

This fact prompts us to suggest a *dynamical* gadgetry for the memory of our receiver. Instead of storing in the memory bank of this receiver the full repertoire of the transmitter messages (something quite impractical for large N), one rather implements at the hardware level of the decoding system an *algorithm* i.e. a non-linear recurrent relationship (map) which upon request i.e. after being triggered from the transmitter of the partner-system puts forth « here and now » a whole set of aperiodic strings of digits — whose « grammar » is tempered by the shape of the map. Most of them will be pseudorandom sequences and hence the probability of matching any incoming « packet » of given length N will go to one, asymptotically with time.

C. Organization

A cognitive device capable of compressing and simulating a physical phenomenon possesses a more or less advanced degree of « self »-organization.

In general, self-organization is associated with the spontaneous emergence of a long-range spatial and/or temporal coherence amont the variables of the (organized) system. Nevertheless, the system may be quite « innocent » of this self-organization, namely it may not use it further or may not know how to use it or may not know that it knows how to use it... The point is that a self-organized system becomes a *cognitive* system only beyond a minimum number (namely two) of hierarchical levels ; (this means that one-to-one mapping does not produce cognition).
The sine qua non prerequisite for the emergence of a self-organizing *cognitive* system is the ability of this self-organized system to *replicate* another system (including[5] itself or parts of itself). This self-replication ability is in turn *the* indication that the system knows how to *compress* and *simulate*. Examples : some chemical reactions constitute dynamical systems which are self-organizing but are *not* cognitive. The genome of any biological organism on the other hand *is* a cognitive system (the simplest one) since via transcription and translation (a process involving two hierarchical levels, the level of nucleic acids and the level of polypeptides) it simulates and eventually reproduces itself. Clearly then, Cognition is instrumental in compressing the complexity of *another* system (including parts of itself) via the formation of collective properties (cross-correlations or statistical crossmoments) — thereby dramatically reducing the initial number of the degrees of freedom of the system under observation. Alternatively, as we are going to see, an attractor acting on any point of its Basin (set of initial conditions) does *de facto* the same.

Concluding then this discussion we suggest that *communication* is involved in triggering the emergence of new spatial-temporal patterns via a sequence of instabilities producing bifurcations which lead to broken symmetry, and subsequently in creating and maintaining multiple attractors in dynamical systems, which attractors categorize and compress such patterns.

3. WHAT IS A COGNITIVE CHANNEL ?

The human mind possesses the unique capability of « mapping » the external (as well as part of the organism's internal) world — that is, it

« compresses » long and complex strings of impinging environmental stimuli (« observations ») — and then uses these « minimal length algorithms » in order to simulate physical phenomena — thereby revealing the « laws of nature ».

We may theorize that this process of « self »-organization and category formation is implemented via a set of coexisting (strange) attractors in the cognizant apparatus — each one of which attracts (and therefore compresses) whole subsets of « initial conditions » the sum-total of which constitute the set of external stimuli (Fig. 2).

This set of initial conditions forms the « Basin » of the attractors and the processes of partition and category formation in the mind involves the topology of the separatrices among the individual subsets of the Basin.
Let us then try to sketch the *dynamical equivalent* of an information channel mediating the mapping bnetween a set of messages of the « transmitter » and a set of « regimes » of the « receiving » or rather the cognitive apparatus — playing the role of memory bank. (Let us remind ourselves again that an information channel according to Shannon's « static » information theory provides a one-to-one mapping between the transmitter and the receiver and is in no way associated with the act of « cognition ». The sole role of a classical information channel is the mere copying of the transmitted patterns, *as they are*, at the receiving end with the greatest possible reliability).

We propose that the set of messages should be *identified* as a set of initial conditions or the Basin of the attractors in an N-dimensional state space where N stands for the number of variables mediating the dynamical phenomenon under consideration.

The receiver, or, more generally the cognitive apparatus is identified as the set of attractors, of much lower dimensionality embedded as invariant compact subsets in the same N-dimensional state space.

Falling into the attractor via (short or long) transients in state space corresponds to the act of progressive signal abstraction (or compression). The issue is : what about the dynamical isomorph of the « channel » ? or, how is the set of initial conditions, that is the messages, *partitioned among different categories*, that is among more than one coexisting attractors (steady states, limit cycles, tori or strange attractors) ? One has to realize here that the mapping between messages and attractors (or « templates ») is *not* one-to-one : whole subsets (of initial conditions) of the Basin converge towards one and only one attractor. Alternatively some points of the Basin may undergo a perpetual (or too long) transient trajectory without converging anywhere, or they may even be repelled to infinity. How can we tell which subset of the Basin will land into which attractor ?

The problem in this general form is, of course, of formidable difficulty. In a very elementary case (two simple attractors, namely two steady states) the issue has been recently addressed. One considers a two-dimensional non invertible map (the « channel ») and finds a fractal separatrix between the two subsets of the Basin[ref.16,17].

The essential outcome of the above work is that the precision with which a given initial condition is known may play an extraordinary sensitive role on the predictability of the final attractor upon which this initial state will eventually land. This means that an extraordinary high accuracy is sometimes necessary in our measurements for the reliable prediction of the pertinent attractor (« compressor ») into which the observed phenomenon will be categorised : clearly then, the set of attractors play in our model the role of dynamically stored patterns in the brain against which whole sets of externally impinning stimuli have to be « convoluted ». [This process may not be accomplished in one single hierarchical step ; in a second step the attractors (if numerous at the lower level) will play the role of the members of a hyper-Basin towards a new hierarchy of fewer hyper-attractors and so on.]

Knowing the state space dimensionality of the impinging messages *and* the (fractal) dimensionalities of the corresponding attractors one can evaluate in each case the degree of (average) *compressibility* which a given N-dimensional stimulus is subjected to.

Furthermore, investigating the topology of the separatrices of the Basin one can calculate the uncertainty (or the information gained thereof) associated with the category to which the given stimulus belongs. Together, *the entropy of partitioning* and *the degree of compressibility* give the *two* essential macroparameters characterizing the *cognitive* process at a given hierarchical level[6].

Once *the mapping* (that is the « linguistic scheme » adopted by the individual) is known, both the above parameters can in principle be calculated.

We have appreciated already in the previous section the fact that the percentage of compressible sequences in a string say of 0's and 1's of length N is extremely small. Considering the 2^N a priori equiprobable sequences of length N (the members of the Basin) the number of compressible sequences to say K bits is just

$$\Sigma = 2+2^2 + \ldots + 2^{N-K-1} = 2^{N-K} - 2$$

and the percentage of compressible sequences is

$$\eta = \frac{\Sigma}{2^N} \sim 2^{-K}.$$

So, although in principle there is a numerous set of attractors at each cognitive level, the percentage of those which really deserve the name, that is possess an information dimensionality c = N − K bits (K ~ 0(N)), is just $\eta = 2^{-K}$, that is insignificantly small.

The moral is that the human mind can simulate a very small subset of natural phenomena. Their overwhelming majority literally « passes through » our heads undetected.

Even so how come that our templates-attractors *really* have as Basin the set of *externally* impinging stimuli ? To accept this one has to a priori digest the hypothesis that *there exists* an « homology » between some

patterns of the external world and the categories of the software of the human brain.

But if the human brain is just an unsuspected evolutionary byproduct of the « laws of Nature » how come that its software can simulate the very laws which, acting uppon matter, triggered its formation ?

An example of a fractal Basin boundary (separatrix)

Grebogi et al.[ref.16,17] give the following example of two-dimensional non-invertible map :

$$\theta_{n+1} = 2\theta_n \pmod{2\pi}$$

$$z_{n+1} = \lambda z_n + \cos\theta_n,\ 2 > \lambda > 1,\ 0 \leqslant \theta \leqslant 2\pi.$$

The Jacobian matrix of this map has eigenvalues 2 and λ. Since both exceed 1 there are no attractors with finite z. In fact all initial conditions will produce orbits which will asymptotically go either to $z = +\infty$ or $z = -\infty$ as $n \to \infty$. So we consider $z = +\infty$ and $z = -\infty$ as two coexisting attractors and ask what is their Basin boundary. This means : what is $f(\theta)$ such that initial conditions in $z > f(\theta)$ lead asymptotically the orbits to $z = +\infty$ while those in $z < f(\theta)$ lead the orbits to $z = -\infty$.

Grebogi et al., starting from the above map, find the expression

$$z = f(\theta) = -\sum_{k=1} \lambda^{-(K+1)} \cos(2^K\theta)$$

Since $\lambda > 1$ the above sum converges.

Now investigate $df/d\theta$. From the above expression we get

$$\frac{df}{d\theta} = \frac{1}{2}\sum_{k=1}^{\infty} \left(\frac{2}{\lambda}\right)^{K+1} \sin(2^k\theta).$$

Since $\lambda < 2$ this sum diverges and so $f(\theta)$ is *not* differentiable. In fact Grebogi et al. find that the separatrix $z = f(\theta)$ has an infinite length and a fractal dimensionality

$$d = 2 - \frac{\ln\lambda}{\ln 2} = 2 - \log_2 \lambda.$$

A final important aspect of the information dimensionality of multiple existing attractors-« classifiers » is the problem of their invariance under smooth coordinate transformations. In other words we are interested in knowing if the degree of abstraction or compressibility of the Basin of initial conditions possesses universality or not. We do not possess a general answer to this question. Ott et al.[ref.20] have recently investigated the possible invariance of a number of « dimensions » of strange attractors and have found some specific coordinate transformations ensuring the invariance of only the information dimensionality. Clearly much more work has to be conducted along this direction.

4. CHAOTIC DYNAMICS OF PATTERN FORMATION

The question « where exactly information is stored ? » seems to us to receive a rather original answer from a series of well known experiments on the perception of random dot interference patterns performed several years ago by Glass and Perez[ref.15].

The essential conclusions we draw from those experiments is that the informations lies neither in the externally impinging pattern (which may be just a random dot aggregation on the plane) nor on the dynamically stored pattern — replica of the receiver (which again may be random) ; it rather emerges from the flow in the neighborhood of the *singularities* — as deduced from the pertinent eigenvalues of the linear or non linear, discrete or continuous transformation (mapping or flow) — which transformations the processor *imposes between* the incoming and the internal pattern.

So one is inclined to consider, at least in a number of cases where such experimental documentation exists, that what we « really » see or cognize in the world is perhaps just the *attractors* (or the repellors) of the maps or the flows which the processor generates by way of *relating* external stimuli with internal activity. Biological processors belonging to the same species, possess identical nervous systems and they accordingly tend to generate more or less identical maps or flows — thereby creating as it were by « social consensus » universal pattern classification schemes.

Not only in Perception but also in Cognition and Linguistics emerging stable patterns may be identified as eigenvalues of non-linear operators which operators are involved in the recursive character of these systems. For example recent work in game theory as applied to biology and animal contests[ref.13] as well as to the process of mutual simulation between competing hierarchical systems-each trying to « break » the code of the opponent first[ref.11,12] shows that rationality in such tournaments can be completely replaced by evolutionary stability. This points to the emergence of evolutionary stable strategies (E.S.S.) which are point attractors for the probabilities of moves adopted by the contestants : such asymtotically stable strategies lead to a stable population polymorphism or stable behavioral repertoires adopted by a single individual.

In Linguistics the situation is different since, here, unlike animal or finite state automata games, self-reference is involved. We conclude with two examples which show that self-consistency hence logic, is provided also by the existence of stable attractors.

Take as a first example the symbolic self-referential loop :

« This sentence contains X_0 0's, X_1 1's, X_2 2's, X_3 3's, X_4 4's, X_5 5's, X_6 6's, X_7 7's, X_8 8's, X_9 9's ».

Let the 10-dimensional vector X (X_0, X_1, X_2, ...X_9) start from a randomly selected initial condition and then provide the next iteration by checking the truth of the previous one.
For example start from X_a (0,1,2,3,4,5,6,7,8,9).
The next iteration gives X_b (2,2,2,2,2,2,2,2,2,2)
Next comes : X_c (1,1,11,1,1,1,1,1,1,1), followed
by X_d (1,12,1,1,1,1,1,1,1,1) and finally
X^e(1,11,2,1,1,1,1,1,1,1) which obviously is *the* attractor corresponding to the selected initial condition and which makes the self-preferential sentence above, self-consistent that is, true.

(Through exaustive search one can find another point attractor as well as a limit cycle of period 2.
Here the basin of attraction is partitioned into three subsets).
As a second example take the sentence :
« This sentence has *X* letters »
Start iterating from X_a = « one »
Then $F(X_a) = 25$ where F is the number of letters after putting in the sentence the given initial condition.
The second « seed » then is X_b = « twenty five » which leads to $F(x_b) = 32$. Likewise X_c = « thirty two » leading to $F(X_c) = 31$ and finally X_d = « thirty one » lands the system to $F(X_d)$ = « 31 » — which is the attractor corresponding to the initially chosen seed « one ». (There is another point attractor corresponding to a different subset of the basin).
Note finally that while the truth assessed by the second sentence is « English bound » (translated e.g. into French it becomes false) the first sentence conveys as « translinguistic » truth.

5. COLLISIONS VERSUS ITERATIONS : A BASIC ISOMORPHISM BETWEEN PHYSICAL AND SYMBOLIC DYNAMICS.

The parallelism between a rarefied *gas* of finite number hard spheres anbd a recurrent *algorithm* as far as the ergodic and mixing properties are concerned is an excellent example of an isomorphism between the Kolmogorov-Sinai entropy of the two systems. Consider first the physical system composed of N hard spheres. This system of moving scatterers is sensitive to initial conditions in the following sense :

Take the collision process between a pair of spheres. An angular deviation θ on the trajectory of a sphere changes after a collision to an angle θ' where $\theta' \sim \frac{2\delta}{R}\theta$ where δ is the mean free path and R is diameter of each sphere. Let τ be the intercollisional time (which is of the order of 10^{-10} sec). After a time t the deviation of the trajectory of one sphere in the 6N dimensional state space will grow as $(\frac{2}{R}\theta)^{t/\tau}$, so the maximum (positive) Lyapounov exponent of the system is $\lambda \sim) \frac{1}{\tau} \log \frac{2\delta}{R}$.
Since the Kolmogorov-Sinai entropy is taken equal to the sum of positive Lyapounov exponents whose number is of the order of the phase space dimension if the particles are mutually independent, the K-S entropy will be $S \sim N.\lambda$ where N is of the order of the Avogadro's number. Taking into account a typical rarified « Lorentz gas » we get :
$S \sim 10^{34} \frac{bits}{sec}$ — which represents essentially the *new* information (or variety) *generated* by the evolving system per unit time — or the information needed to follow the second-to-second trajectory of the system.

Consider next within the unit interval the very simple (piece-wise linear) recursive relation $X_{n+1} = 2X_n \pmod 1$. Start iterating from any arbitrary initial value X_0 which written in dyadic form is say, X_0 = .1110101011011.. Each iteration moves the dot one position to the right. After ξ iterations the value of the number will be

$$X_\xi = 2^\xi X_0.$$

The initial value X_0 can never be known exactly — even in theory. Why ? Because this would amount in asking an infinite number of questions (and getting an infinite number of answers in the yes-no form) that is it would amount of performing an infinite number of measurements. The entropy of a continuous distribution is infinite. Given that each measurement requires a finite time (commensurate with the switching time or the relaxation time of the particle-composed « decision making » element of the receiver) getting knowledge of any number X_0 with infinite accuracy would require an infinite amount of time.

Now $(\Delta X_\xi) = 2^\xi(\Delta X_0)$ which means that the initial uncertainty ΔX_0 (however small, yet finite) is amplified *exponentially* with the number of the iterations. If X_0 is known with an accuracy $\sim 10^{-\tau}$ after $\xi_c = \tau \log_2 10$ iterations the initial condition X_0 is completely forgotten and from then on the recurrent algorithm functions as a « pseudo-random generator » producing an essentially incompressible sequence of digits — although the relationship- responsible for such a sequence is quite deterministic.

The K–S entropy here equals $\log_2 2 = 1$ bit/iteration. This « software » system is also ergodic and mixing.

NOTES

(1) See also section 5

(2) What we really mean by « free will » is the simulation of a set of alternative pathways of action and a choice making amongst them.

(3) A.I. is essentially characterized by fine-graining : it processes first fine details of an object and synthesizes those details to obtain the object. In chaotic processing the input pattern is first coarse-grained and then through successive iterations one is led towards the elucidation of finer details.

(4) N + 1 digits are needed instead of N, just in order to make the difference between N *zeros* and N *ones*.

(5) It is sometimes asserted that no self-organizing system can fully simulate itself ; this, they say, could give rise to the « paradox of self-reference » that is to a perpetual non-converging tautology. For higher cognitive functions this assertions remains undecidable. For genetic self-replication, tautology is transcended via the spontaneous folding of linear polypeptide chains and the resulting formation of stereospecific proteins.
This process of symmetry breaking creates *new* information in the system and dispels self-reference.

(6) According to our model, the EEG is the pacemaking agent which permutates the attractors on a time division multiplexing basis.

REFERENCES

1. J.S. NICOLIS, *Sketch for a dynamical theory of language*, Kybernetes, 1982, Vol. *11*, pp. 123-132.

2. J.S. NICOLIS, *Shold a reliable information processor be Chaotic ?*, Kybernetes, 1982, Vol. *11*, pp. 269-274.

3. D. GIKAS and J.S. NICOLIS, *Stochasticity from Deterministic Dynamics An Excit Example of Generation of Markovian strings*, Z. Phys. B-Condensed Matter 47, 279-284, 1982.

4. J.S. NICOLIS, G. MAYER-KRESS, G. HAUBS, *Non-uniform chaotic dynamics with implications to information processing*, Z. Naturforch *38a*, pp. 1157-1169.

5. J.S. NICOLIS, *The role of chaos in reliable information processing*, Journal of the Franklin Institute 317, 1984, pp. 289-307.

6. J.S. NICOLIS, *Chaotic dynamics of information processing with relevance to cognitive brain function*, Kybernetes, 1985, in press.

7. J.S. NICOLIS, I. TSUDA, *Chaotic dynamics of information processing. The Magic Number Seven Plus, Minus Two revisited*, Bull of Math. Biology *47*/3 in press, 1985.

8. J.S. NICOLIS, *Dynamics of hierarchical Systems*, Vol. *25*, Synergetics Springer-Verlag, 1986.

9. J.S. NICOLIS, *Chaotic dynamics applied to information processing*, Rep. on Progress in Physics, to appear. (1985)

10. J.S. NICOLIS, J. MILIAS-ARGITIS, C. CARABALIS, *Dynamic simulation of the game of chick*, Kybernetes, 1983, Vol*12*, pp. 9-20.

11. J.S. NICOLIS, E.N. PROTONOTARIOS, M. THEOLOGOU, *Communication between two-self-organizing systems modelled by controlled Markov chanins*, Int. J. Man-Machine-Studies *10*, 1978, pp. 343-366.

12. J.S. NICOLIS, E.N. PROTONOTARIOS, *Bifurcations in non-negotiable games : a paradigm of self-organization in cognitive systems*, Int. J. Bio-Medical Comuting *10*, 1979, pp.417-447.

13. John MAYNARD-SMITH, *Evolution and the theory of Games*, Cambridge Univ. Press, 1982.

14. P.W. ANDERSON, *More is different*, 1972, Science *177*, p. 393.

15. L. GLASS, R. PEREZ, *Perception of Random Dot Interference Patterns*, *Nature* 246, December 7, pp. 360-369.

16. C. GREBOGI, E. OTT, J. YORKE, 1983a, *Phys. Rev. Lett.* 50, pp. 935-938.

17. C. GREBOGI, S. McDONALD, E. OTT, J. YORKE, 1983b., *Phys. Lett.* 99A.

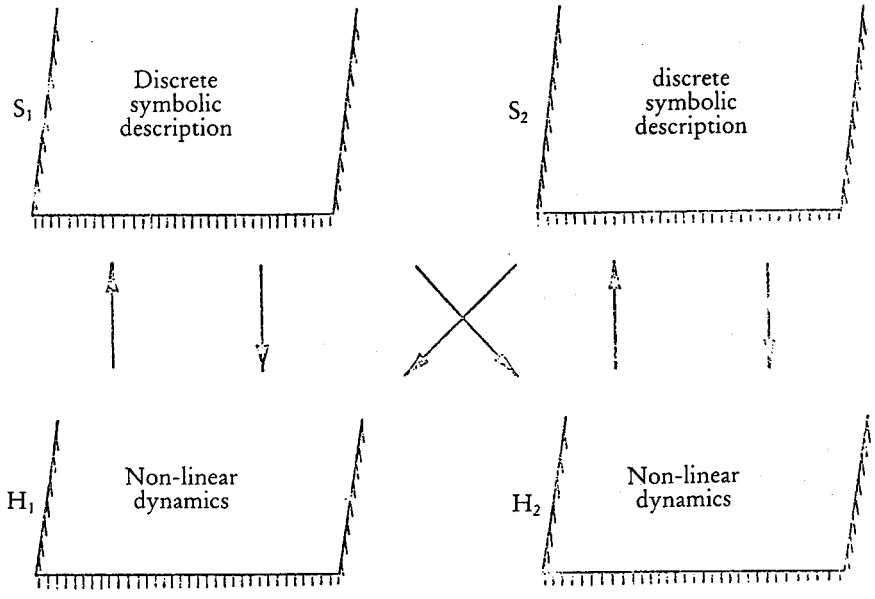
18. G. NICOLIS, I. PRIGOGINE, 1977, *Self-Organization in Non-Equilibrium Systems*, Wiley, New York.

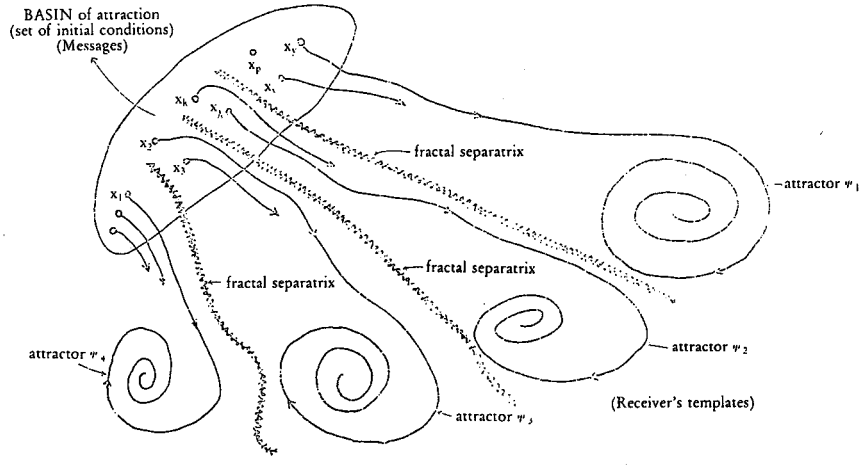
19. G. NICOLIS, I. PRIGOGINE, 1981, *Symmetry Breaking and Pattern Selection in Far-From-Equilibrium Systems*, Proc. Nat. Ac. Sci. USA 78, N° 2, pp. 659-663.

20. E. OTT, W.D. WITTERS, J.A. YORKE, 1984, *Is the Dimension of Chaotic Attractors Invariant Under Coordinate Changes ?, Journ. of Statistist.*, Physics 36, pp. 687-697.

21. G. CHAITIN, 1975, Sci. Am., May, p. 47.

22. G. NICOLIS, *Dissipative Systems*, Rep. on Prog. in Phys. in press.


S1
Discrete
symbolic
description
S2
discrete
symbolic
description
H1
Non-linear
dynamics
H2
Non-linear
dynamics


BASIN of attraction
(set of initial conditions)
(Messages)
fractal separatrix
fractal separatrix
fractal separatrix
attractor ψ1
attractor ψ2
attractor ψ3
attractor ψ4
(Receiver's templates)

Order, Complexity, and Disorder

Terry Hogg and Bernardo A. Huberman
Xerox Palo Alto Research Center, Palo Alto

The spectacular success of the reductionist program in the natural sciences highlights the remaining challenge of explaining the behavior of complex structures in terms of their parts. Started a century ago, statistical mechanics provided a powerful method to deal with infinite degrees of freedom by using the laws of large numbers in computing observables. Such an approach, although successful in treating many-body systems, fails when applied to systems with many, but not infinite degrees of freedom displaying self-organized behavior. On the other hand, recent advances in nonlinear dynamical systems theory have allowed for increased knowledge of the time evolution of non-linear systems which to external probes act as composed of one or two degrees of freedom.

Thus, it is fair to state that the present understanding of collective behavior in complex systems remains an open problem in science and cannot be satisfactorily explained by either one of the above paradigms. Examples are given by living systems, formal organizations, and finite state machines. The typical problems that one encounters include the description of their dynamic properties, their adaptation and self-organization, and fault tolerance. Although considerable progress has been made in understanding the characteristics of such systems, a quantifiable, physical measure of their complexity is still lacking. Very much in the same spirit that allows for the macroscopic state of matter to be parametrized by a few quantities such as temperature and pressure, one desires a complexity measure which, besides being quantifiable, encodes only the relevant properties of the system. This is in contrast to measures which are based on the detailed specification of the system, like computational complexity, where the length of the minimal algorithm needed to recons-

truct an exact copy of the structure quantifies its degree of complexity[1-4]. Although useful in contexts such as cryptography[5], formal languages and automata theory[6], the analysis of algorithms[7], and biology[8-10] such information theoretic measures do not address intuitive notions of complexity in physical or biological systems, nor can they always be explicitly computed.

When a system has only a small number of parts the characterization of its behavior can be performed directly. This is also true for ordered many-body systems, where the existence of symmetries and constraints simplify the problem of description considerably. Conversely, statistical techniques are often applicable when there are very many parts which give rise to an effectively random behavior. In this case one deals with a small set of averaged quantities rather than the detailed behavior of each part, and obtains results which become exact in the limit of an infinite number of parts. In either case, it is relatively simple to describe the behavior of the entire system based on knowledge about the functioning of the parts. There remains however, the intermediate and difficult case in which many partial constraints act on a large but finite set of parts. In this case, the resulting complex structure is neither completely ordered nor disordered. Such systems are characterized by the presence of many relevant length and time scales, a fact which makes their analysis very difficult.

These ideas are simply illustrated with examples drawn from many branches of science. Just to name a few, in physics one speaks of crystalline materials and gases as being relatively simple when compared with liquids or amorphous structures. Equivalently, ordered magnets or disordered ones (i.e. paramagnets) are construed to be simpler than a magnetic system near its critical point. Similarly, in linguistics the grammars of formal languages, as well as those of random strings generated by typing monkeys, possess a degree of simplicity which is in marked contrast with that of natural languages[11].

A particularly relevant case is that of biological systems, which have been traditionally characterized in terms of their diversity[12-13]. In this context, for example, the very symmetric configurations in shell mollusks, as well as the patternless ones, are construed to be simpler than those with many chiralilty and length scales[14-15].

An additional example is given by digital computing machines in the form of an interacting collection of simple parallel processes. Such structures can give rise to fault-tolerant self-organizing behavior. In this context an important issue is determining how to build useful systems for tasks such as pattern or speech recognition from computationally simple parts[16], as well as investigating their applicability to optimization[17] and symbolic processing problems[18]. These tasks require flexible systems that can be easily modified to satisfy novel constraints reliably and that are stable against noise or component failures. Once again, different components of the system are partially constrained by the action of others, giving rise to structures which are neither completely ordered nor disordered.

A powerful concept for understanding these systems is that of a hierarchy. This can correspond to the structural layout of a system or, more generally, to clustering the parts by strength of interaction. In particular, if the most strongly interacting components are grouped together and this procedure is repeated with the resulting clusters, one produces a tree reflecting the hierarchy of the system. In physics one sees this type of organization as quarks form hadrons, which in turn form nuclei, then atoms, molecules, and so on. In biology, this has been the traditional way in which taxonomies are constructed[13]. Similarly, in computation one assembles subroutines into programs which form higher level structures such as operating systems and networks[19].

This generic form of organization possesses two important properties. At the dynamical level, the lower eschelons can, for all practical purposes, be integrated over while the larger scale structures are essentially frozen and act as static constraints. Furthermore, an ultrametric topology provides the proper description of the clusters in terms of subgroup interactions[20-22].

In this paper, we introduce a physical measure of complexity which embodies the notions we have discussed above. It applies to discrete systems and in particular, provides a precise, readily computable measure of complexity. Furthermore, it has the important property of assigning a low value to both ordered and disordered[23] systems while becoming maximal for intermediate cases, as illustrated in Fig. 1. Finally we discuss the implications of this work for the dynamics of these systems and in particular to the evolution of ecological systems[24].

We now describe our measure of complexity and, after illustrating it with a number of examples, we show that it agrees with the arguments presented above. This measure is based on examining the diversity of a given structure as determined by the number of different kinds of interactions among the parts that exist in the corresponding hierarchy at all levels. Diversity in relationships is the essential feature that makes for complex behavior in systems made up of simple elementary parts[19]. In other words, the behavior of systems which have a highly differentiated set of relationships will be harder to unravel from the properties of their parts than that of simple systems. For the sake of definiteness, we will focus on structures composed at the lowest level of a collection of identical units, in which case the corresponding hierarchy is described by an unlabeled tree with constant depth[25]. In this case only the structures of the various clusters are relevant, rather than the nature of their lowest level components.

Consider a tree T of the general form illustrated in Fig. 2. It can correspond to any of the hierarchies discussed above. This tree consists of b subtrees T_1, ..., T_b, with b the number of branches at the root of the tree. Since we are interested in quantifying the number of different interactions or relationships, repeated copies of the same (i.e. isomorphic) subtree will not contribute to the count. Thus it is necessary to find the set of nonisomorphic subtrees T_{i1}, ..., T_{ik} where $k \leqslant b$ is the number of such subtrees. The diversity measure, D(T), counts both the interactions between subtrees and those within. Specifically, to compute it we first

multiply the values of D for all nonisomorphic subtrees. This counts the number of distinct interactions within the clusters represented by each of the subtrees. This is then multiplied by N_k, the number of ways the k different clusters themselves can interact. This last step captures the large scale interactions that take place at the top of the hierarchy. For full generality our count includes all n-ary interactions, although similar results are obtained if the count is restricted to pairwise interactions. The number of nary interactions among k distinct objects is just the number of ways of choosing n out of the k objects, i.e. the binomial coefficient ${}_kC_n$. Thus we have

$$N_k = K + {}_kC_2 + {}_kC_3 + \ldots + {}_kC_k = 2^k - 1. \quad (1)$$

Combining these steps we obtain

$$D(T) = (2^k - 1) \prod_{j=1}^{k} D(T_{ij}) \quad (2)$$

where j ranges from 1 to k, the number of nonisomorphic subtrees. When a tree has no subtrees D(T)=1, representing a single self-interaction.

This measure of diversity can now be used to quantify the complexity of a given system. Since the simplest structures have only one kind of interaction, which we arbitrarily set to have zero complexity, we define the complexity of a tree, C(T), as

$$C(T) = D(T) - 1. \quad (3)$$

We should note that a forest of trees can be considered as a single tree with one more row in which the trees of the forest are the subtrees of the new tree. Thus, our measure can also applied to a forest of trees with the value obtained from the corresponding combined tree.

The definition of Eq. (2) can be illustrated with the trees of Fig. 3. For the binary tree with three levels of clustering, T_3, shown in Fig. 3a), each of the subtrees is the same, i.e. another binary tree. Thus $D(T_3)=1$. A similar argument shows that in general D will be 1 for any tree with a constant branching ratio at each node.

By contrast, the tree T of Fig. 3b), has two distinct subtrees, for each of which D = 1. Thus $D(T) = (2^2 - 1) = 3$ representing the self-interaction of each of the two kinds of clusters, and the single (pairwise) interaction between them.

This measure gives a value for the absolute complexity of a hierarchical structure. One can also define a measure of relative complexity which adjusts for the size of the structure. This is necessary for comparing the complexity of structures with different sizes. In particular, for a tree T in a class S of structures, the relative complexity of T is given by

$$r(T) = C(T) / (\text{max value of C in the class S}) \quad (4)$$

Values of r near 1 indicate complex structures. We should mention that this distinction between absolute and relative complexity also arises in discussions of the algorithmic complexity of a string of bits of length n. Specifically, the relative algorithmic complexity of a string is obtained by comparing its complexity (measured in terms of the length of a

minimal description) to the maximum possible for strings of that length. This maximum is roughly n when the length becomes large. A string is considered complex (or random) by this measure when its relative algorithmic complexity is near one.

In the case of the hierarchies considered here, the natural class of structures consists of all those that have a fixed number *n* of basic parts and *m* levels of clustering. We denote the number of elements in this class of trees by I^n_m and the class itself by $\mathcal{V}^n_m$. The I numbers can be computed from the following recursion relation

$$I^n_m = \sum_{\{\lambda\}} \prod_{i=1}^{n} (I^i_{m-1} + \lambda_i - 1)C\lambda_i \qquad (5)$$

where the sum is over all integer partitions of n and $I^n_1 = 1$. The partitions are specified by the non-negative values of the λ's, which satisfy

$$n = \sum_{i=1}^{n} i\lambda_i \qquad (6)$$

This relation can be seen by noting that each tree with n inputs and m levels of clustering defines a partition of n. That is, the subtrees can be grouped according to how many leaves each has an λ_i is just the number of subtrees with i leaves. Eq. (6) simply states that the entire tree has n leaves. Now each subtree with i leaves, containing m — 1 levels of clustering, can be filled in with any of I^i_{m-1} possible trees. This however, overcounts since not all assignments produce a new tree. This can be corrected by ordering all members of $\mathcal{V}^i_{m-1}$ in an arbitrary sequence. The λ_i subtrees to be filled from this class are then selected so that each choice does not precede the previous choice in this sequence.

For example, if $\lambda_i = 2$ and there are three possible subtrees, say {A B C}, then the distinct trees are produced from the choices (A A), (A B), (A C), (B B), (B C) and (C C). The remaining choices, namely (B A), (C A) and (C B), do not give new trees and are excluded by the sequencing requirement.

The general case of making λ choices from I possibilities is the number of ways of placing balls into I boxes allowing many balls in each box. The number of balls in box i specifies how many times the ith tree is chosen. Thus there will be $_{(I+\lambda-1)}C\lambda$ ways to make the choice. Combining factors for all parts of the original partition, and then summing over all possible partitions gives the recursion relation of Eq.(5).

We should also mention that by applying techniques used to randomly construct trees[26] an alternative recursion relation from the I numbers can, be obtained, namely

$$n\, I^n_m = \sum_{k=0}^{n-1} I^k_m \sum_{d/(n-k)} d\, I^d_{m-1} \qquad (7)$$

with $I_{0m} = 1$. The sum over d includes only those values from I to n-k which exactly divide n-k. Some of these numbers are listed in Table 1.

We can now see that our measure does indeed correspond to the physical complexity discussed above by comparing its maximum to the value one obtains for ordered and disordered trees. First we notice that ordered trees will have a low value of D due to their high symmetry. This implies that there will be few nonisomorphic subtrees at each level. Moreover,

since for a completely ordered tree all subtrees will be the same, it follows that D = 1 and C = O.

Since disordered trees are not so readily characterized, we have examined their diversity by computing D(T) for a large set of randomly selected trees and comparing its value to the maximum for the class. Fig. 4 shows a specific example of such a computation for the class of trees with 10 inputs and 3 levels of clustering. The maximum values of D(T) for other classes of trees are shown in Table 2. With this and other numerical computations we verified that the complexity of disordered trees is indeed much lower than the maximum. Thus our measure of complexity has the desired physical behavior outlined in Fig. 1.

There remains to establish the connection that exists between the complexity of a given system and its dynamics or evolution. Since one relevant property is adaptability, such a connection throws some light on important issues of ecology[27-28] and biological science[29]. In the particular case of ecosystems, it has been argued that diversity, as measured by the number of existing species in an ecosystem, initially grows as empty niches are occupied, only to be reduced at later times through competition and predation[30]. A typical example is given by the process of succession in plant communities which leads to a climax forest[24]. In this case, many competing species are eventually replaced by a few dominant ones. This in turn suggests that the hierarchical relationships between species[31] will also undergo a process of initial increase and eventual decrease in its diversity as defined by Eq. (2).

More generally, one can consider the relationship between, the organization of a system and its adaptability. Adaptability is characterized by the ability to satisfy variations in constraints with minimal changes in the structure. Within this context one can state that systems that are most *adaptable* will tend to be very complex by our measure. This follows from the fact that extreme diversity is most able to satisfy changes in constraints, a fact which has been advocated in biological evolution by a number of authors[12-28]. Secondly, systems which are most *adapted* to given constraints tend to have a lower complexity than *adaptable* ones, as the development of specific connections between their parts will likely lower the diversity of their structures. Therefore one might conclude that as a system adapts to more static constraints, it lowers its complexity.

In summary, we have introduced a quantitative and easily calculable measure of complexity of hierarchical structures which embodies the intuitive notion that complexity is intermediate between perfectly ordered and completely disordered systems. It thus describes the diversity of structures whose organization not only involve many length and time scales but also a high degree of differentiated relationships. This measure could be easily applied to phylogenies based on the principe of maximum parsimony[13-32]. Moreover, this measure may provide a quantitative characterization of the process of adaptive evolution. This will apply to biological systems if species diversity implies a diversified hierarchy. Furthermore, it may provide insights into the dynamics of adaptive computing structures.

ACKNOWLEDGMENTS

We are indebted to Dr. J. McGlade for pointing out the connection between our measure of complexity and the behavior of ecological systems, and for numerous and fruitful discussions. This research was partially supported by ONR contract N00014-82-0699.

REFERENCES

(1) R. SOLOMONOFF, Inform. and Control 7, 1, 1964.

(2) J.HARTMANIS and R.E. STEARNS, Trans. Am. Math. Soc. 117, 285-306, 1965.

(3) A.N. KOLMOGOROV, Probl. Inform. Transm. 1, 1-5, 1965.

(4) J.F. TRAUB, G.W. WASILKOWSKI and H. WOZNIAKOWSKI, *Information, Uncertainty, Complexity*, Addison-Wesley, 1983.

(5) H.J. BAKER and F.C. PIPER, IEEE Proc. A129, 357, 1982.

(6) *Automata Theory and Formal Languages*, Edited by J. Goos and J. Hartmanis, Lecture notes in computer science 33, Springer 1975.

(7) R. SEDGEWICK, *Algorithms*, Addison-Wesley, 1983.

(8)L. CATLIN, *Information Theory and the Living System*, Columbia University Press, N.Y., 1972.

(9) H. ATLAN, J. Theor. Biol. 21, 45-70, 1968.

(10) C. BENNETT, in *Emergence Syntheses in Science*, D. Pines Ed., Santa Fe Institute Press, 297-313, 1985.

(11) T. WINOGRAD, *Language as a Cognitive Process-Syntax*, Vol. 1, Addison-Wesley, 1983.

(12) E. MAYR, Fed. Proc. Am. Soc. Exp. Biol. 23, 1231-1235, 1964.

(13) D. PENNY, L.R. FOULDS and M.D. HENDY, Nature 297, 197-200, 1982.

(14) A. WRIGLEY, Proc. Matacol. Soc. London 27, 206-211, 1948.

(15) S. WOLFRAM, Nature 311, 419-424, 1984.

(16) B.A. HUBERMAN and T. HOGG, Phys. Rev. Lett. 52, 1048-1051, 1984, and T. HOOG and B.A. HUBERMAN, J. Stat. Phys. 41, 115-119, 1985.

(17) J.J. HOPFIELD and D. TANK (preprint, 1985).

(18) D. TOURETZKY and G. HINTON, in Proceedings of IJCAI, LA, 238-243, August 1985.

(19) H.A. SIMON, Proc. Am. Phil. Soc. 106, 467, 1962.

(20) C.J. JARDINE, N. JARDINE and R. SIBSON, Math. Biosci. 1, 173-179, 1967.

(21) B.A. HUBERMAN and M. KERSZBERG, J. Phys. A18, L331-L336, 1985.

(22) T. HOGG and B.A. HUBERMAN, Phys. Rev. A.32, 2338-2344, 1985.

(23) We should emphasize that the notion of disorder used here refers to the physical notion of a disordered system as measured by its statistical properties. In this context both pseudo-random and random (in the sense of algorithmic complexity) structures are disordered.

(24) C.J. KREBS, *Ecology*, Harper and Row, New York, 1977.

(25) If the elementary parts are all distinct, then the natural class will be trees with labeled leaves. The corresponding numbers are the H numbers introduced in ref. 22. In this case all subtrees are distinct, unless specific parts or labels can be reused.

(26) A. NIJENHUIS and H.S. WILF, *Combinatorial Algorithms*, Academic Press, 1978.

(27) P.M. ALLEN, in *Ecosystems Theory for Biological Oceanography*, Edited by R.E. Ulanowicz and T. Platt, Canadian Bulletin of Fisheries and Aquatic Sciences 213, 3-26, 1985.

(28) J.M. McGLADE and P.M. ALLEN, in R. Mahon, ed. *Towards the Inclusion of Fishery Interactions in Management Advice*, Can. Tech. Fish. Aquat. Sci. N° 1347.

(29) F. JACOB, *The possible and the Actual*, Pantheon, 1982.

(30) R.H. MacARTHUR and E.O. WILSON, *The Theory of Island Biography*, Princeton University Press, Princeton, 1967.

(31) R. MAY, *Stability and Complexity in Model Ecosystems*, Princeton University Press, Princeton, 1974.

(32)Other systems may require a different counting factor at each node of the tree instead of the value defined in Eq.(1) which counts all n-ary relationships.

Table 1

Number of unlabeled trees with *n* leaves and *m* levels of clustering, I^N_m, for various values of *n* and *m*.

n = m	1	2	3	4	5	6
1	1	1	1	1	1	1
2	1	2	3	5	7	11
3	1	3	6	14	27	58
4	1	4	10	30	75	206
5	1	5	15	55	170	571
6	1	6	21	91	336	1.337

Table 2

Maximum diversity for unlabeled trees with *n* leaves and *m* levels of clustering for various values of *n* and *m*

n = m	1	2	3	4	5	6
1	1	1	1	1	1	1
2	1	1	3	3	3	7
3	1	1	3	9	9	21
4	1	1	3	9	27	27
5	1	1	3	9	27	81
6	1	1	3	9	27	81

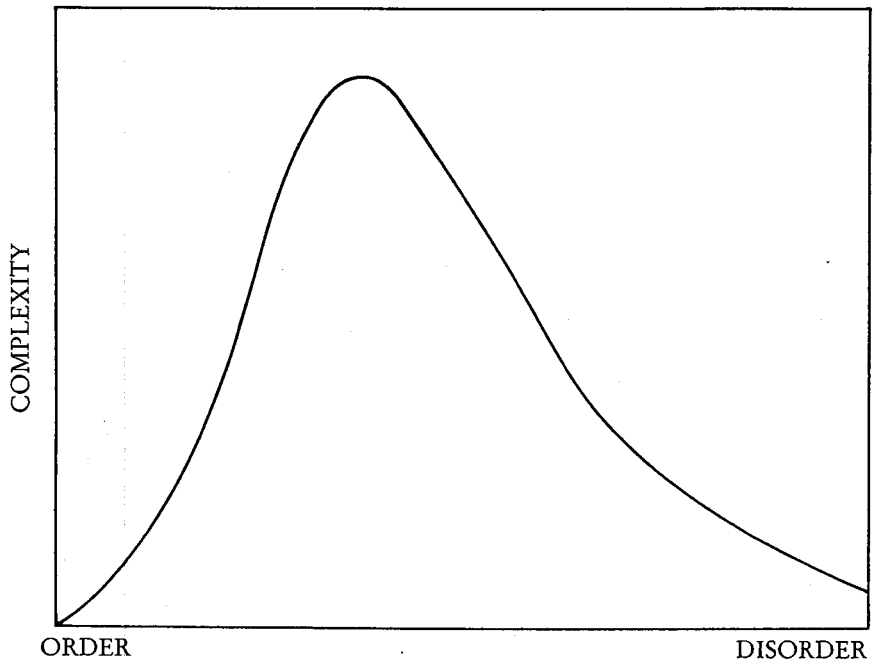


Figure 1
Schematic variation of physical complexity with the degree of disorder of a system.

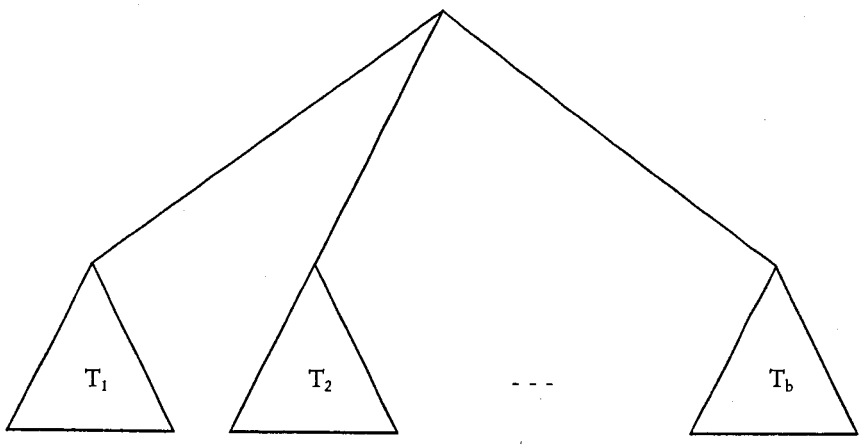


Figure 2
Generic tree containing b subtrees

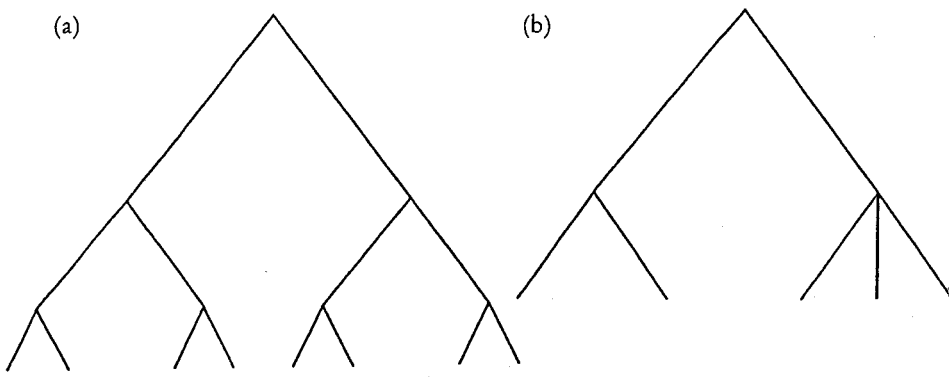


Figure 3
a) Binary tree with three levels of clustering

b) Tree containing distinct subtrees

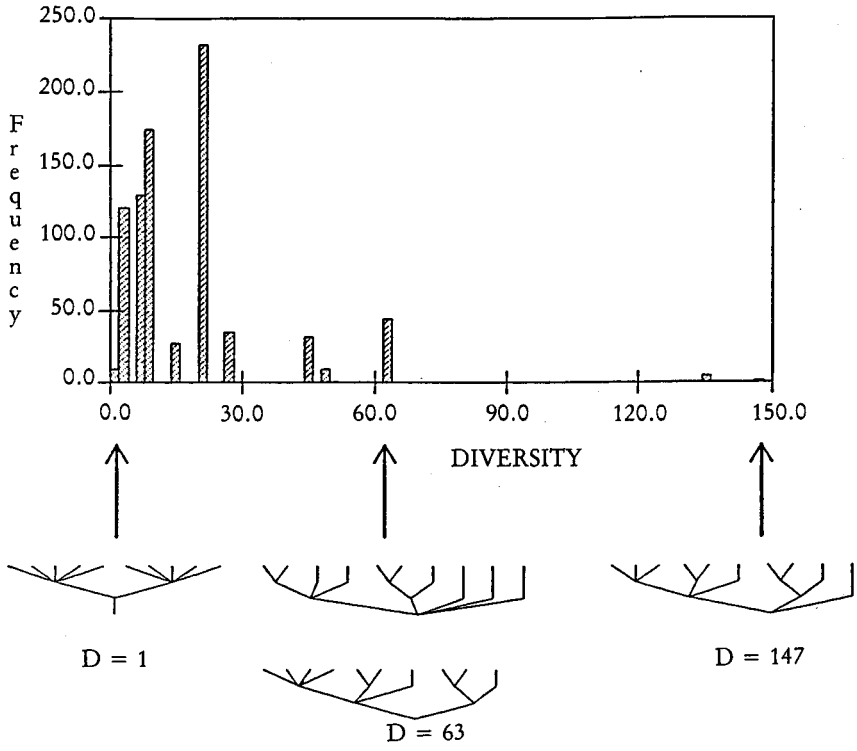


Figure 4
Diversity of the class of trees with 10 inputs and 3 levels of clustering.

This class contains 817 trees, with a maximum diversity of 147. The histogram shows the number of trees corresponding to each value of the diversity as given by Eq. (3). Some examples are shown above. Specifically, one of the nine trees with D = 1, two of the 44 trees with D = 63 and the unique maximal tree with D = 147 are shown.

Intelligence Quotient of Artificial Intelligences

Toshiyuki Furukawa
Faculty of Medicine, University of Tokyo

IS THE COMPUTER A GREAT IDIOT ?

For a long time the computer has been considered only a high-speed calculating machine with no intelligence. Someone has exquisitely personified it as a great idiot. Of course, a computer has an extremely high speed calculating capability and also a large memory capacity. A computer is able to access the arbitrary data in its memory at a moment's notice, and work for a whole day tirelessly.

Despite these advantages, especially its high-speed performance, the computer is generally thought to be honest but slow on the uptake. It should be noted, however, that the reason the computer has been so unfairly underestimated stems mainly from its lack of conversation capability or, in other words, the immaturity of man-machine interface.

According to experience in the development of an expert system, the « FLUIDEX-consultation system for fluid therapy », man-machine conversation is much too retarded and time wasting. Even in pioneer artificial intelligence systems the Achilles heel is the waste of time. In addition, these system often store and adhere to too much worthless information. It is clear, however, that the responsibility for this falls squarely on the men who create the artificial intelligence knowledge structures. One can recognize the effectiveness of every effort to shorten and simplify conversation between man and machine if one participates in the study of practical artificial intelligence use.

CAN THE INTELLIGENCE QUOTIENTS OF ARTIFICIAL INTELLIGENCES BE MEASURED ?

Let us next consider the intelligence quotients of individuals measured. We know a priori that intelligence is an abstract concept which cannot be measured directly. The popular IQ tests consist of various questionnaires (verbal tests) and performance tests, but none of them represents the intelligence itself. One can, however, estimate the intelligence of individuals from their responses to separate problems. Accordingly, the same IQ tests for human beings might be applied to describe the abilities of computers and artificial intelligence systems.

There are several kinds of verbal tests. Information tests ask common sense questions. Comprehension tests inquire into the rather deep understanding of such things as customs, cultures, morals ethics and laws. Arithmetic tests consist of primary problem of mathematics. Similarity tests ask for the identification of the common characteristics between two words. Digit span tests involve the memorization of several series of numeric digits and their repetition in regular order, followed by repetition in reverse order. Vocabulary tests ask the ordinary accepted meanings of words and phrases.

Performance tests come in a similar variety. Digit symbol tests require a comparison of symbols to numerical digits following an example. Picture completion test call for the discovery of the missing part of each scene in a cartoon. Block design tests involve the assembly of blocks of the same size and with the same colored faces into required patterns. Arrangement tests ask for the composition of logical story using animated scenes. Assembly tests require the completion of a picture from dispersed parts as in a puzzle. Maze tests are mazes, as the name suggests.

To the extent that the computer is a high-speed calculator, either artificial intelligence systems or computers can obtain the highest scores on arithmetic and digit span test. Similarly, since computers have a huge memory, they can be expected to achieve high scores on vocabulary and similarity tests, and even on information tests. If the machine is equipped with a pattern recognition capability, digit symbol, block design, assembly and maze tests can also be easily solved.

Comprehension, picture completion and arrangement tests are essentially difficult for the machines. In the worst cases, artificial intelligence systems cannot score at all on these tests.

On the whole, today's artificial intelligences can score highest on all the tests except the above three. Therefore, even from the most pessimistic point of view, the intelligence quotient of an artificial intelligence can be estimated at as high as 140 or more when compared to an adult human. It can be said that artificial intelligence systems already have the capability of being the best partners for human beings.

WHAT CAN ARTIFICIAL INTELLIGENCE SYSTEMS LEARN FROM THE HUMAN BRAIN ?

This question refers to the requirements for the future development of artificial intelligences. Studies on the human brain have revailed the localization of some specific functions. The cortical areas are divided into primary sensory areas (somatic, visual and auditory), interpretative areas (somatic, visual, auditory and general), motor areas, the Broca's speech area and the personal area. In general, the more complex the function, the wider the specific cortex area to which it is assigned. For example, the voluntary motor areas and somatic sensory areas correponding to the fingers and lips occupy the broad area of the central fissure.

Usually, in a fully developed artificial intelligence, a considerable part of the program is allotted to logical mechanisms which manage human-like interpretations. The brain also allots fairly wide cortex areas to interpretative functions. Unfortunately for today's computers, not only sensory but also motor organs are almost entirely lacking. Even, a large computer with more than 1.000 key-board terminals is like a monster which merely has 1.000 fingers. Even when the computer is equipped with 100 robot eye cameras, it is only an imaginary beast Argos with 100 stemma. We might call the artificial intelligence of today an unfortunate disabled with an unrecognized high intelligence equivalent to an intelligence quotient of 140 or more. This suggests that human beings may be going to succeed in creating an equal intellectual partner for the future.

III

Dynamical Systems, Economic Theory and Games.

Testing for Deterministic Chaos in Economic Data

William A. Brock
Social Systems Research Institute — University of Wisconsin-Madison

1. INTRODUCTION

Recently there has been a lot of interest in nonlinear deterministic economic models that generate highly irregular trajectories (for example, Benhabib and Day (1983) ; Dana and Malgrange (1986) ; Grandmont (1983) ; Day (1982) ; Stutzer (1980) ; Day and Shafer (1983) ; Deneckere and Pelikan (1984) ; Boldrin and Montrucchio (1985). Intense exploration of low dimensional deterministic dynamical systems models has been going on in physics and chemistry, Swinney (1983), ecology and biology May, 1976, population dynamics, Brillinger et al. (1980), climatology, Nicolis and Nicolis (1984), and so on. Prigogine (1980), gives a strong argument in general for nonlinear modelling in science.

The literature cited above relies heavily on mathematical literature on « chaos » and nonlinear dynamics. See Collet and Eckman (1980), and Guckenheimer and Holmes (1983), for overviews of the relevant mathematical literature and economic literature respectively.

The main reason for this recent explosion of interest in nonlinear dynamics by the applied sciences is that the trajectories generated by some nonlinear difference equations look completely random to the naked eye. A particularly dramatic example was given by Sakai and Tokumaru (1980). They show that most trajectories of the difference equation $x_{t+1} = F(x_t)$, x_o given where

$$F(x) \equiv x/a,\ x\varepsilon[0,a],\ F(x) \equiv (1-x)/(1-a),\quad x\varepsilon[a,1],\ 0 < a < 1 \qquad (1)$$

generate the same autocorrelation coefficients as the first-order AR process

$$v_{t+1} = (2a-1)\, v_t + u_{t+1},\ \{u_t\} \text{ i.i.d.} \qquad (2)$$

Bunow and Weiss (1979), present chaotic first order difference equations that generate spectra indistinguishable from that generated by a sequence of i.i.d. random variables.

Brock and Chamberlain (1984) show that given any spectral measure G there is a deterministic overlapping generations economy whose equilibrium trajectory generates an empirical spectrum that approximates G. Hence linear time series methods (spectral analysis and autocovariance functions) may not be able to observationally distinguish between deterministic and random systems. We come to the subject of this article.

The subject of this article is to discuss tests that are potentially capable of distinguishing between certainty and uncertainty. We will argue that such tests can be usefully applied to economic data.

This paper is organized as follows. Section 1 contains the introduction. The second section gives tests for deterministic chaos that can be applied to a time series of data. Section 3 reports applications of the tests to postwar U.S. business cycle data and to the Wölfer sunspot series. Section 4 is a short summary. Finally an expanded version of this paper is available in Brock (1985).

2. *TESTS FOR CHAOS*

In order to demonstrate that an apparently random time series $\{a_t\}$ is actually deterministic chaos the researcher must show (1) the dimension is low and (2) there is a positive lyapunov exponent. To explain we need

Definition 2.1 (Takens, 1983). The time series of real numbers $\{a_t\}_{t=1}^{\infty}$ has a *smoothly deterministic explanation* if there exists a *system* (h,F,x_o) such that $h : R^n \to R$, $F : R^n \to R^n$ are smooth and

$$a_t = h(x_t),\ x_t = F(x_{t-1}),\ t=1,2,\ldots ;\ x_o \text{ given.} \tag{2.1}$$

There is an analogue of this definition for continuous time systems but, for space reasons, we concentrate on discrete time systems in this paper. The following definition of Guckenheimer and Holmes (1983, p. 256) will be needed.

Definition 2.2 : An *attractor* Λ of the deterministic dynamical system F is an indecomposable closed invariant set Λ with the property that, given $\varepsilon > 0$, there is a set U of positive Lebesque measure in the ε-neighbourhood of Λ such that $x \in U$ implies that the w-limit set of x is contained in Λ, and the forward orbit of x is contained in U. Here the w-limit set of x is the set of accumulation points of the forward orbit of x.

In addition we shall always assume that F and its attractor Λ possess an ergodic invariant measure ϱ (Guckenheimer and Holmes, 1983, pp. 280-288) with compact support. When needed we will assume that ϱ is normalized so that $\varrho(\Lambda) = 1$ and ϱ has a smooth density.

We reserve the term *chaotic attractor* for attractors whose largest Lyapunov exponent (to be defined below) is positive.

Defitiniton 2.3. (Guckenheimer and Holmes, 1983, pp. 283-284). Let $F : R^n \to R^n$ define a discrete dynamical system. Fix $x \varepsilon R^n$. Suppose that there are subspaces $V_i^{(1)} V_i^{(2)} \ldots V_i^{(n)}$ in the tangent space at $F^i(x)$ and numbers $\mu_1 \geqslant \mu_2 \geqslant \ldots \geqslant \mu_n$ that depend on (x, v) with the properties that

a) $DF(V_i^{(j)}) = V_{i+1}^{(j)}$

b) $\dim V_i^{(j)} = n+1-j$

c) $\lim_{N\to\infty} (1/N) \ln ||D_x F^N(v)|| = \mu_j$ for all $v \varepsilon V_o^{(j)} - V_o^{(j+1)}$.

Then the μ_j are called the *Lyapunov exponents* of F. Here i = 0, 1, 2, ... For later use we need the Oseledec multiplicative ergodic theorem.

Theorem 2.1 : (Oseledec, 1968 ; Guckenheimer and Holmes, 1983, p. 284 ; Benettin et al., 1976). Let F be C^1 and let DF be Hölder continuous for some exponent θ. Let F and its attractor Λ possess an ergodic invariant (Guckenheimer and Holmes, 1983, p. 280) measure ϱ. Then there is a ϱ measurable set $\Lambda_1 \subset \Lambda$ such that $\varrho(\Lambda_1) = \varrho(\Lambda)$ such that for all $x \varepsilon \Lambda_1$, Lyapunov exponents exist.

Remark : Since $\dim V^{(1)} = n+1-1 = n$ therefore, as pointed out by Bennettin et. al (1976, p. 2339), if one chooses the vector v in (c) « at random » one may expect to find $\mu_j = \mu_1$. Furthermore ergodicity of ϱ implies that μ_j is independent of the initial condition, j=1,2,...,n. In particular the largest Lyapunov exponent does not depend upon the particular trajectory under scrutiny.

Example 2.1 : The tent maps (1.1) have as invariant measure $\mu(dx) = j(x)dx$, $j(x) = 1$, $x \varepsilon [0,1]$. The logistic map $x_{t+1} = 4x_t(1-x_t)$ has invariant measure $\mu(dx) = j(x)dx$, $j(x) = 1/(\pi(x(1-x))^{1/2})$. The largest Lyapunov exponent for these maps is given by

$$\lambda = \int_0^1 \ln|F'(x)| \mu(dx) \qquad (2.2)$$

so that for the tent map $F(x) = 2x$, $x_t \varepsilon [0,1/2]$, $F(x) = 2(1-x), x \varepsilon [\frac{1}{2}, 1]$, we get $\lambda = \ln 2$.

With this preparation we may explain the two tasks that the researcher must do to test for deterministic chaos in time series data : (1) show that the dimension of the time series is low, (2) show that the largest Lyapunov exponent is positive.

Task I : Calculating Dimension of a Time Series

In order to motivate notions of dimensions suppose that $\{a_t\}_{t=1}^{\infty}$ has a deterministic explanation by the system (h,F,x_o) and look at the m-history starting at t :

$$a_t^m \equiv (a_t, a_{t+1}, \ldots, a_{t+m-1}) \equiv (h(x_t)), \ldots, h(F^{m-1}(x_t)) \equiv J_m(x_t) \qquad (2.3).$$

Hence $J_m : \Lambda \subset R^n \to R^m$. In order to grasp what is coming look at examples (1.1). Here $\Lambda = [0,1]$ and $j(x) = 1$ for the tent maps. Thus $J_m : [0,1] \to R^m$, x_t is distributed according to μ which is uniform on [0,1] so that the dimension of $J_m(\Lambda) \equiv \{J_m(x), x \varepsilon \Lambda\}$ is *one* for any sensible notion of dimension. This is so because $J_m(\Lambda)$ is a one dimensional arc embedded in R^m provided that J_m is a smooth map. Furthermore Takens (1980) has shown that generically J_m is 1-1 from Λ to $J_m(\Lambda)$ if $m \geqslant 2n+1$.

The same reasoning applies whatever the dimensions of Λ. Calculate the dimension D_m of $J_m(\Lambda)$ for all « embedding dimensions » m and find $\lim_{m \to \infty} D_m$.

We now face the practical problem of calculating the dimension of $J^m(\Lambda)$ for each m from a finite data set $\{a_t\}_{t=1}^N$. After much experimentation discussed in Brock (1985) the natural science community seems to have settled on the Grassberger-Procaccia (1983) correlation dimension α_m as one of the most useful dimension measures. It is defined for $\varepsilon > 0$ by

$$\alpha_m \equiv \lim_{\varepsilon \to 0} \ln C_m(\varepsilon)/\ln \varepsilon \tag{2.4}$$

$$C_m(\varepsilon) \equiv \#\{(i,j)/||a_i^m - a_j^m|| < \varepsilon,\ 1 \leq i \leq N_m,\ 1 \leq j \leq N_m\}/N_m^2 \tag{2.5}$$

$$N_m \equiv N-(m-1) \tag{2.6}.$$

Here #A denotes the cardinality of set A and ln is the natural logarithm.

Natural scientists (e.g. Swinney, 1985, and references) construct Grassberger-Procaccia dimension plots of $\ln C_m(\varepsilon)$ against $\ln \varepsilon$ and attempt to measure the slopes of $\hat{\alpha}_m$ of these G–P plots for each embedding dimension m. After constructing these plots they look to see if $\hat{\alpha}_m$ levels off to some $\hat{\alpha}$ as $m \to \infty$.

This procedure requires skill and judgment for two reasons. First since $\ln C_m(\varepsilon) = 0$ for large ε you must « estimate » the slope of the G-P plot over a range of ε of moderate to small size. Second Brock and Dechert (1985) prove that if there is noise in the data of positive variance even if the variance is small then for each embedding dimension m, $\alpha_m = m$ almost surely. Therefore the slopes $\hat{\alpha}_m$ of the dimension plots must be estimated over ε's larger than the scale ε of any noise that is present in the data set.

In practice a range of ε's over which the slope $\hat{\alpha}_m$ of the G-P plot appears to be « stable » is chosen by « eyeballing ». The noise level ε is « estimated » by hunting for an ε small enough so that $\hat{\alpha}_m \sim m$ when $\hat{\alpha}_m$ is estimated over ε in $[0,\varepsilon]$ (Ben–Mizrachi et al., 1984). This procedure is applied to data in Section 3 of this article.

Task II : Calculating the Largest Lyapunov Exponent

Calculation of the largest Lyapunov exponent $\lambda \equiv \mu_1$ is based upon the formulae (a–c) of Definition 2.3 and the Oseledec Theorem 2.1. We briefly explain the Wolf et al. (1984) algorithm which we use in our empirical work.

For each embedding dimension we use the time series $\{a_t\}_{t=1}^N$ to form a time series $\{a_t^m\}_{t=1}^{N_m}$ of m–histories. Start the algorithm by locating the nearest neighbour $a_{t_1}^m$ a_1^m to the initial m–history a_1^m.
Let $d_1^{(1)} = ||a_{t_1}^m - a_1^m||$. Note that $d_1^{(1)}$ is smallest *positive* distance $||a_t^m - a_1^m||$. Select an evolution time q and set $d_2^{(1)} = ||a_{t_1+q}^m - a_{1+q}^m||$ and store $g_1(q) = d_2^{(1)}/d_1^{(1)}$. This ends the first iteration. We are now ready to enter the main program loop.

Ideally, in order to start the second iteration we would like to find a new m–history $a_{t_2}^m$ near a_{1+q}^m whose angle $\theta(a_{t_2}^m - a_{1+q}^m,\ a_{t_1+q}^m - a_{1+q}^m)$

close to zero. In this way we mimic the definition 2.3(c) of Lyapunov exponent as closely as possible with $a^m_t - a^m_l$ playing the role of v. Definition 2.3(b) shows that except for hairline cases $\lim(1/N)\ln||D_xF^N(v)||=\mu_1$, the *largest* Lyapunov exponent, because the set $V_o^{(1)}-V_o^{(2)}$ has full Lebesque measure.

Motivated by this strategy we choose t_2 to minimize the penalty function

$$p(a^m_t-a^m_{l+q},\ a^m_{t_1+q}-a^m_{l+q}) \equiv ||a^m_t-a^m_{l+q}|| + w|\theta(a^m_t-a^m_{l+q},\ a^m_{t_1+q}-a^m_{l+q})| \quad (21.7)$$

subject to the nondegeneracy requirement $a^m_t \neq a^m_{l+q}$. Here w is a penalty weight on the deviation $|\theta|$ from zero. Store

$$g_2(q)\ d_2^{(2)}/d_1^{(2)},\ d_1^{(2)} \equiv ||a^m_{t_2}-a^m_{l+q}||,\ d_2^{(2)} \equiv ||a^m_{t_2+q}-a^m_{l+2q}||. \quad (2.8)$$

This ends iteration two. Continue in this manner.

For iteration K store

$$g_k(q) \equiv d_2^{(k)}/d_1^{(k)},\ d_1^{(k)} \equiv ||a^m_{t_k}-a^m_{l+(k-1)q}||,\ d_2^{(k)} \equiv ||a^m_{t_k+q}-a^m_{l+kq}||, \quad (2.9)$$

where t_k minimizes

$$p(a^m_t-a^m_{l+(k-1)q},\ a^m_{t_{k-1}+q}-a^m_{l+(k-1)q})$$

subject to $a^m_t \neq a^m_{l+(k-1)q}$. Continue until $k=K$ where K solves $\max\{k|1+kq \leq N_m\}$. Set

$$\lambda_q \equiv \frac{1}{K}\sum_{k=1}^{K}[\ln(d_2^{(k)}/d_1^{(k)})/q] \quad (2.10)$$

Brock and Dechert (1985) locate sufficient conditions on the system (h,F,x_o) that deterministically explains the data under the null hypothesis of a deterministic explanation that enables them to prove that an idealized form of the above algorithm converges to the largest Lyapunov exponent λ for almost all starting vectors. See the Wolf et al. paper for a discussion of numerical experience with this algorithm.

We wish to make four points concerning this algorithm. First the ratio $g_k(q)\ d_2^{(k)}/d_1^{(k)}$ will tend to increase with q in the case of a chaotic attractor. This is so because it estimates local spreading which increases as q increases. For example for the tent map $F(x) = 2x$, $x\varepsilon[0,1/2]$, $F(x) = 2-2x$, $x\varepsilon[1/2,1]$ you will get $d_2/d_1 = 2^q$ provided that d_1 is small. Second, even if small noise is present the mean of $g_k(q)$ will tend to rise with q.

Third, if $\{a_t\}$ is a sequence of independently and identically distributed random numbers there will be no tendency for the mean over k of $g_k(q)$ to increase with q and $\hat{\lambda}_q$ will tend to decrease. These two properties show up in Table 3 for random numbers. For small data sets we recommend that the investigator display the results as we did in Table 3 below. Else the analyst might hastily conclude that, for example, the sunspot series analyzed in Tables 2-3 is generated by a chaotic attractor.

3. EMPIRICAL APPLICATION OF THESE IDEAS

Empirical calculation of the Grassberger-Procaccia (1983) correlation dimension α and the largest Lyapunov exponent λ is the procedure typically used in natural science to test for the presence of chaos in time

series data (Swinney, 1985 ; Wolf et al., 1984). Economists must deal with time series much shorter than the 10.000-30.000 observations typically used in natural science work and, furthermore, economic time series are probably noisier. The problem is especially acute in business cycle analysis.

Brock and Sayers (1985) test for non nolinearities in quarterly data on U.S. real gnp and U.S. real gross private domestic investment by (a) calculating the Grassberger-Procaccia (1983) correlation dimension and estimating the largest Lyapunov exponent for various embedding dimensions ; (b) calculating measures of asymmetry such as, Blatt (1978), and measures of skewness and kurtosis. For lack of space, we only describe some of the results for U.S. quarterly data from 1947 : 1 to 1985 :1 (1972 = 100), and for Wölfer's sunspot numbers 1749-1924.

Let x_t = real GNP at quarter t and y_t = real gross private investment at quarter t. The data was detrended by the following OLS regressions :

$$\log x_t = 2.681 - .003671t + ex_t \tag{3.1}$$

$$\log y_t = 1.851 + .003765t + ey_t. \tag{3.2}$$

In view of the well known result that autoregressive models of order two (AR(2)) fit detrended U.S. real GNP well we fit two AR(2) models :

$$ex_t = 1.36\ ex_{t-1} - .42\ ex_{t-2} + \delta x_t,\ R^2 = .933 \tag{3.3a}$$

$$\sigma_{ex} = .0167,\ \sigma_{\delta x} = .0043,\ \sigma_{ex}/\sigma_{\delta x} = 3.88,\ sk_{ex} = -.244,\ sk_{\delta x} = -.086,\ k_{ex} = 2.19,\ k_{\delta x} = 4.05. \tag{3.3b}$$

$$ey_t = 1.12\ ey_{t-1} - .31\ ey_{t-2} + \delta y_t,\ R^2 = .760 \tag{3.4a}$$

$$\sigma_{ey} = .051,\ \delta_{\delta y} = .025,\ \sigma_{ey}/\sigma_{\delta y} = 2.03,\ sk_{ey} = -.46,\ sk_{\delta y} = -.57,\ k_{ey} = 2.78,\ k_{\delta y} = 5.14. \tag{3.4b}$$

Have σ_A, sk_A, k_A denote standard deviation of A, skewness of A, kurtosis of A. Standard errors are not reported because the second coefficient changes a lot when we fit AR(3) and AR(4) to this data.

For $\{ex_t\}$, $\{ey_t\}$, $\{\delta x_t\}$, $\{\delta y_t\}$ we calculated, for embedding dimension d,

$$C(\varepsilon) \equiv \frac{1}{n_d} \#\{(i,j)/||a_i^d - a_j^d|| < \varepsilon\},\quad C^*(\varepsilon) = \frac{1}{n_d^*} \#\{(i,j)|i\ j, ||a_i^d - a_j^d|| < \varepsilon\} \tag{3.5}$$

$$\alpha(\varepsilon) \equiv \ln C(\varepsilon)/\ln\varepsilon,\ \alpha^*(\varepsilon) \equiv \ln C^*(\varepsilon)/\ln\varepsilon, \tag{3.6}$$

$$SC(\varepsilon_k, \varepsilon_{k-1}) \equiv (\ln C(\varepsilon_k) - \ln C(\varepsilon_{k-1}))/(\ln\varepsilon_k - \ln\varepsilon_{k-1}) \tag{3.7}$$

$$SC^*(\varepsilon_k, \varepsilon_{k-1}) \equiv (\ln C^*(\varepsilon_k) - \ln C^*(\varepsilon_{k-1}))/(\ln\varepsilon_k - \ln\varepsilon_{k-1}) \tag{3.8}$$

and estimated the largest Lyapunov exponent λ. Here $n_d^* \equiv n_d^2 - n_d$, $|S|$ denotes the cardinality of set S ; $n_d = n-(d-1)$ is the number of d-histories $a_t^d = (a_t^d, \ldots, a_{t+d-1})$ constructed from the sample of length n, and $1 \leqslant i, j \leqslant n_d$.

Three difficulties emerged. First a standard of comparison was needed so that meaning could be attached to a « large » or « small » dimension or Lyapunov exponent. To take care of this difficulty, we calculated the

same measures of dimension and Lyapunov exponent for normal pseudo random numbers of the same mean and standard deviation for each series.

Second, the utility of the measure $\alpha(\varepsilon)$ (suppressing obvious subscripts for ease in typing) is based on « the power law conjecture ».

$$C(\varepsilon) \cong K(\varepsilon)\varepsilon^{\alpha(\varepsilon)},\ \alpha(\varepsilon) \to \alpha, K(\varepsilon) \to K,\ \varepsilon \to 0. \tag{3.9}$$

So that

$$\ln C(\varepsilon)/\ln\varepsilon = \ln K(\varepsilon)/\ln\varepsilon + \alpha(\varepsilon) \to \alpha,\ \varepsilon \to 0. \tag{3.10}$$

Here the subscript «∞ » on any symbol denotes the limit as n_d or n_d^* goes to infinity. With a time series of length 10.000, $\ln C_{d,n_d}(\varepsilon)/\ln\varepsilon$ may approximate $\ln C_{d,\infty}(\varepsilon)/\ln\varepsilon$ even for small ε but a time series of length 100-200 is a different matter. For example let $\varrho = \min\{||a_i^d - a_j^d|| > 0,\ 1 \leq i,j \leq n_d\} > 0$, be the smallest positive distance. For $\varepsilon < \varrho$, $C_{d,n_d}(\varepsilon) = \frac{1}{n_d}$, $\alpha_{d,n_d}(\varepsilon) = -\ln(n_d)/\ln\varepsilon \to 0$, $\varepsilon \to 0$. Note also that the slope estimator $SC_{d,n_d}(\varepsilon_k,\varepsilon_{k-1}) = 0$ for $\varepsilon_k, \varepsilon_{k-1} < \varrho$. This is true even if $\{a_t\}_{t=1}^{n_d}$ is a sequence of random numbers for which the theoretical value of α and SC is d. Hence α and SC are poor estimators of the underlying dimension especially for small data sets.

Since α^*, SC^* give values much closer to the theoretical value of d for random numbers therefore we calculated the measures α^*, SC^* as well. These measures attempt to remove the distortion in a possible power law caused by the fact $C_{d,n_d}(\varepsilon) = 1/n_d$ on $[0, \varrho)$.

There is the further problem of how to report the results of dimension calculations for the small data sets used in business cycle analysis. That is to say what constitutes a « confidence interval » and « rejection region » under a given null hypothesis for an « estimator » where we know very little about the sampling distribution and where generating an empirical distribution under various null hypothesis is expensive ? We shall handle this problem as we have seen it handled in the natural science literature that we have seen to date. That is to try to report our findings in such a way that the reader may make her own judgment as to the significance or stability of any estimate. Due to lack of space we can only report the highlights of the results of Brock and Sayers (1985) here.

Third, even less is known about the small sample properties of our estimate of the largest Lyapunov exponent. Hence we calculated a measure of *cumulative* spreading between initially close trajectories for (a) our data set, (b) random numbers of the same mean and variance as our data set, and (c) a time series generated by the tent map (1.1) with a = 1/2.

In a nutschell our results indicate the following : there is not enough information in the 1947 :1–1985 :1 data set for the Grassberger-Procaccia type dimension measures to reject the null hypothesis that detrended real GNP and real GPDI are generated by an AR(2) process as estimated in (3.3) and (3.4). To put it another way we need more data to establish that U.S. real GNP and real GPDI are generated (in the main) by a low dimension chaotic deterministic dynamical system.

It may be helpful to future researchers in this area to describe the chain of thought that lead us to this conclusion. First we calculated measures of dimension for $\{ex_t\}$, $\{ey_t\}$ and got evidence of low dimension as revealed in the tables below.

Second, our colleague Don Hester suggested that AR(2) models like (3.3) and (3.4) might generate a low dimension estimate if they fit the data well. Although a theorem of Brock and Dechert (1985) assures us that the dimension of an infinite data set generated by (3.3) or (3.4) must be infinity a data set of length 153 may be a long way from infinity especially if the rate of convergence of dimension estimates is slow.

It is easy to prove that if, say, $\{ex_t\}$ has a smoothly deterministic explanation

$$ex_t = h(z_t),\ z_t = F(z_{t-1}),\ z_0 \text{ given}, \tag{3.11}$$

then

$$\delta x_t = h(F^2(z_{t-2})) - 1.36\ h(F(z_{t-2})) + .42\ h(z_{t-2}) \equiv M(z_{t-2}). \tag{3.12}$$

Hence the dimension of $\{\delta x_t\}$ should be the same as the dimension of $\{ex_t\}$ if $\{ex_t\}$ has a deterministic explanation. In fact Table 1 below shows evidence that the dimension of $\{\delta x_t\}$ is slightly (?) smaller than the « large » dimension of a sequence $\{\delta x_t\}$ of the same length of normal pseudo random numbers with the same mean and variance. This is evidence against the chaotic low dimensional deterministic dynamical system hypothesis.

Third, faced with this mixed evidence we generated sequences $\{\delta x_t\}$, $\{\delta y_t\}$ of normal pseudo random numbers with the same mean and variance as $\{\delta x_t\}$, $\{\delta y_t\}$ and generated two simulated AR(2) time series $\{\tilde{ex}_t\}$, $\{\tilde{ey}_t\}$ of the same length as $\{ex_t\}$, $\{ey_t\}$ using the fitted values of the AR(2) models in (3.3) and (3.4). Grassberger-Procaccia measures of dimension were calculated for $\{\tilde{ex}_t\}$ and $\{\tilde{ey}_t\}$ and compared with those calculated for $\{ex_t\}$ and $\{ey_t\}$. There appears to be no significant difference to the naked eye. See the table for real GNP. A qualitatively similar result holds for real GPDI (Brock and Sayers, 1985).

Fourth, the same battery of tests were applied to the Wölfer sunspot series (Anderson, 1971, and Priestley, 1981) which displays asymmetries in that the average gradient of rise from trough to peak is greater than the average gradient of fall from peak to trough across « cycles » (Priestley, 1981, p. 882). Furthermore, an AR(2) model fits the sunspot series quite well although a better fit is obtained with a bilinear model according to the Akaike criterion (Priestley, 1981, p. 884). The story is similar to that of real GNP and real GPDI. See Tables 2, 3.

Fifth, the same procedure was applied to the Lyapunov exponent calculations. Look at Table 3 which contains calculations for Wolfer's sunspot numbers. We calculated $\bar{\lambda}_q$ from (2.10) for $q=1,2,\ldots,9$ as well as the mean $\bar{g}$ and the standard deviation σ_g of $\{g_k(q)\}_{k=1}^{k}$. Notice that if $\{a_t\}$ is a sequence of i.i.d. random numbers then from (2.9) and (2.10) $\ln(d_2^{(k)}/d_1^{(k)})/q \to 0$, $q \to \infty$. Hence $\bar{\lambda}_q \to 0$, $q \to \infty$ in this case. But, in contrast, if there is deterministic chaos present in the data we should see

a tendency of g (q) to grow with q. This is so because g(q) is a measure of spreading of nearby trajectories.

Look at Table 3 where three typical runs with pseudo random numbers were done in order to get a perspective. Notice the absence of any tendency of g(2) to grow with q and the tendency of $\hat{\lambda}_q$ to fall with q for each of the three runs. Now turn to the same calculations with the sunspot numbers. If there are instabilities i.e., deterministic chaos present in the sunspot data we should see a tendency of $\hat{g}(q)$ to grow with q and $\hat{\lambda}_q$ should not fall with q. Our own naked eye sees little difference from the three runs of random numbers. The same pattern emerged in Lyapunov exponent calculations from the AR_2 residuals from real GNP and real GPDI.

Conclusion : dimension tests and Lyapunov exponent calculations have a hard time rejecting a linear model where there is a lot of variation of the data « *within* the regression plane » relative to the variation of the data « *around* the regression plane », when the data set is small. Brock and Dechert (1985) formalize this empirical finding into a theorem. Our findings reported above should not be construed as negative to the attempt to find evidence of significant nonlinearities in economic data.

This is so for several reasons. First we used aggregate data. Hence many non linearities at the microlevel may have been « washed out ». Second the data set 1947 :1-1985 :1 is a period where severity of recessions has fallen (Zarnowitz and Moore, 1984, pp. 12-15) and growth cycles (Moore and Zarnowitz, 1984, Table 8) look quite symmetric throughout this period. Therefore our results square with DeLong and Summers (1984) who find little evidence of significant skewness of growth rates of GNP and industrial production for the postwar period. The failure to find large kurtosis agrees with Blanchard and Watson (1984). This does not contradict Blatt because he looks at different series over different periods for example pig iron production data from Burns and Mitchell (Blatt, 1981, p. 231).

Third Brock and Sayers (1985) have examined US employment data and have found tentative evidence for nonlinear structure in that data.

We summarize the main message for empirical application of these methods. It is *not* enough when you are working with short data sets to report low dimension and positive Lyapunov exponents to make the case for deterministic chaos in your data. Plausible alternative hypotheses (such as a linear model for aggregative macroeconomic time series data) must be tested. This can be done by fitting a plausible model to the data (such as an AR_2 model used here) and calculating the dimension and Lyapunov exponents of the residuals. We may state the

Residual Test for Deterministic Chaos : If $\{a_t\}_{t=1}^{\infty}$ has a deterministic explanation and you fit a smooth time series model with a finite number of leads and lags to $\{a_t\}_{t=1}^{\infty}$ then, except for hairline cases, the residuals $\{\delta_t\}_{t=1}^{\infty}$ of that model must have the same largest Lyapunov exponent and the same dimension as $\{a_t\}_{t=1}^{\infty}$.

This is true because, exactly as in (3.11) and (3.12) δ_t can be written as some smooth function $M(z_{t-L})$ of z_{t-L} for some finite lag L. Replacing « h » by « M » in Takens' (1980) and Brock and Dechert's (1985) theorems finishes the proof. This is so because, except for hairline cases, the dimension of Λ and the largest Lyapunov exponent of (F, Λ, ϱ) is *independent* of the observer function h.

We have not seen the residual test used before. It appears to be new. In any event we believe, in the absence, of explicit reconstruction of the dynamics $a^m_{t+1} = J_m(F(J_m^{-1}(a^m_t)))$ that the residual test could be used to test claims of chaos in the natural sciences as well as in economics.

4. SUMMARY

In this paper we briefly reviewed methods devised by the natural sciences to test for the presence of low dimensional nonlinear chaos in time series data. The tests consists of two parts. First calculate some notion of dimension and show that it is small. Second calculate and estimate of the largest Lyapunov exponent and show that it is positive.

In this paper we amended and adapted these tests to maximize the information available in the short data sets available in business cycle analysis. Our findings indicate that there is not enough information available in U.S. real GNP, real gross private domestic investment, and Wölfer's sunspot series for the two part test discussed here to reject the null hypothesis that the series under scrutiny were generated by an AR(2) process. But Brock and Sayers' (1985) move extensive nonlinearity tests *do* find evidence of nonlinearity.

Finally a more extensive discussion of these methods is available in Brock (1985) and a more extensive discussion of empirical applications to business cycles and labor statistics is in Brock and Sayers (1985) and Sayers (1985). Scheinkman's (1985) paper applies these methods to the stock market. Scheinkman finds strong evidence of *non* independence in stock returns.

A list of references that I have found useful in my literature search as well as references to other studies that I have found is in the reference section.

We close with a final caveat : The empirical results reported here are tentative. For example, we are not satisfied with our present algorithm to calculate Lyapunov exponents for the small data sets that we are analyzing. Our group is extending and updating our findings reported here.

ACKNOWLEDGMENTS

This paper draws heavily from my paper « Distinguishing Random and Deterministic Systems : An Abridged Version », *Journal of Economic Theory*, October, 1986. Research support of NSF Grant #SES-8420872 and the Wisconsin Graduate School is gratefully acknowledged. I thank the participants at the Brussels and Paris conferences for their stimulating comments on this work. J.M. Grandmont, G. Laroque and M. Kurz made especially helpful comments. I especially thank W. Dechert,

Don Hester, C. Sayers and J. Scheinkman. James Ramsey of NYU informed me that he is testing for chaos in financial data. W. Barnett and P. Chen, « The Aggregation-Theoretic Monetary Aggregates Are Chaotic and Have Strange Attractors », Univ. of Texas, Austin Texas, Department of Economics 1986 adduce evidence of low correlation dimension is monetary data. Furthermore I would like to thank the Center for Policy Studies at Monash University, Australia for providing an excellent environment in which some of the work on this paper was done. The usual exemption applies.

Table 1
Dimension Calculations for Detrended U.S. Real GNP
U.S. Real GNP
1947 :1-1985 :1

	ex_t, n_d=134					$\tilde{e}x_t$, n_d=134				
ε	S_{20}	α_{20}	α^*_{20}	SC_{20}	SC^*_{20}	ε	S_{20}	α_{20}	α^*_{20}	SC_{20}	SC^*_{20}
.9	9582	5.96	6.02	2.11	1.97	.9	7246	8.61	8.72	1.54	1.56
$.9^2$	7674	4.03	3.99	2.45	2.65	$.9^2$	6166	5.07	5.14	1.59	1.63
$.9^3$	5936	3.50	3.55	3.17	3.26	$.9^3$	5216	3.91	3.97	2.19	2.25
$.9^4$	4248	3.42	3.48	3.35	3.46	$.9^4$	4144	3.48	3.54	3.13	3.25
$.9^5$	2990	3.40	3.48	3.41	3.59	$.9^5$	2982	3.41	3.48	2.62	2.75
$.9^6$	2090	3.40	3.50	3.14	3.39	$.9^6$	2266	3.27	3.36	2.32	2.49
$.9^7$	1502	3.36	3.48	2.87	3.19	$.9^7$	1774	3.14	3.23	3.54	3.88
$.9^8$	1112	3.30	3.44	3.13	3.65	$.9^8$	1224	3.19	3.32	2.91	3.32
$.9^9$	800	3.28	3.47	2.38	2.95	$.9^9$	902	3.15	3.32	2.61	3.13
$.9^{10}$	622	3.19	3.41			$.9^{10}$	686	3.10	3.30		
	δx_t, n_d=132					$\delta \tilde{x}_t$, n_d=132					
ε	S_{20}	α_{20}	α^*_{20}	SC_{20}	SC^*_{20}	ε	S_{20}	α_{20}	α^*_{20}	SC_{20}	SC^*_{20}
.9	5636	11.00	10.87	6.03	6.21	.9	4966	11.91	12.10	6.93	7.22
$.9^2$	2992	8.50	8.54	6.02	7.30	$.9^2$	2392	9.42	9.66	8.16	8.95
$.9^3$	1590	7.67	8.12	7.16	7.24	$.9^3$	1012	9.00	9.42	9.05	11.54
$.9^4$	750	7.54	7.90	6.68	8.98	$.9^4$	390	9.02	9.98	7.13	15.20
$.9^5$	372	7.36	8.12	4.25	7.76	$.9^5$	184	8.64	11.02	2.46	12.19
$.9^6$	238	6.84	8.06	2.55	7.13	$.9^6$	142	7.61	11.79	.69	∞
$.9^7$	182	6.23	7.93	2.49	17.39	$.9^7$	132	-	∞	0	∞ ∞ ∞ ∞
$.9^8$	140	5.76	9.11	.56	∞	$.9^8$	132	-	∞ ∞	0	
$.9^9$	132	5.14	∞ ∞	0	∞ ∞ ∞	$.9^9$	132	-		0	
$.9^{10}$	132	4.62		0		$.9^{10}$	132	-		0	

Embedding dimension d = 20, n_d=132, 134, $S_{20} = \#\{(i,j) | \|a_i^d - a_j^d\| < \varepsilon\}$

Table 2
Dimension Calculations for Wolfer Sunspot Numbers
Wölfer Sunspot Numbers : 1749-1924
From Anderson (1971)

x_t, n_d=157						δx_t, n_d=155					
ε	S_{20}	α_{20}	α^*_{20}	SC_{20}	SC^*_{20}	ε	S_{20}	α_{20}	α^*_{20}	SC_{20}	SC^*_{20}
.9	2627	21.25	21.77	2.98	3.21	.9	1543	26.06	27.00	5.00	5.74
$.9^2$	1919	12.12	12.49	2.93	4.52	$.9^2$	551	11.94	16.37	4.79	6.16
$.9^3$	1409	9.05	9.83	2.98	2.16	$.9^3$	313	10.30	12.97	5.37	8.72
$.9^4$	1029	7.54	7.91	4.61	5.75	$.9^4$	197	9.12	11.91	4.39	12.58
$.9^5$	633	6.95	7.48	3.18	4.51	$.9^5$	179	7.75	12.04	.91	5.31
$.9^6$	453	6.32	6.99	2.81	4.72	$.9^6$	159	6.80	10.92	1.13	17.01
$.9^7$	337	5.82	6.66	2.00	4.17	$.9^7$	155	5.98	11.79	.24	∞
$.9^8$	273	5.34	6.35	1.18	3.07	$.9^8$					
$.9^9$	241	4.88	5.99	.77	2.89						
$.9^{10}$	223	4.47	5.62								

$x_t \equiv s_t - \bar{s}$, $x_t = 1.336\, x_{t-1} - .65\, x_{t-2}$, $R^2 = .802$, $n = 174$, $n_d = 155, 157$

Table 3
Lyapunov Exponents and Other Instability Measures for Wolfer's Sunspot Numbers

Sunspot numbers

q	K	$\hat{\lambda}$	g	σ_g
1	100	.64	3.22	5.67
2	50	.57	8.22	22.81
3	45	.38	5.96	7.67
4	35	.35	7.78	8.68
5	25	.28	8.52	10.71
6	24	.21	6.45	8.20
7	21	.25	7.93	5.45
8	18	.15	4.31	2.55
9	16	.16	5.78	4.54

3 runs on random numbers to compare with sunspot numbers

q	K	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	g_1	g_2	g_3	σ_{g1}	σ_{g2}	σ_{g3}
1	100	1.01	.89	1.01	4.06	4.51	3.95	4.26	10.57	3.77
2	50	.64	.68	.57	5.88	6.32	4.67	6.80	6.64	5.49
3	45	.47	.47	.49	5.84	6.07	8.07	6.01	6.22	10.31
4	35	.36	.33	.35	6.13	6.49	6.55	4.92	10.90	6.01
5	25	.28	.24	.30	5.79	5.21	6.09	3.89	5.02	4.35
6	24	.29	.19	.25	10.66	4.19	7.53	12.52	2.95	9.05
7	21	.19	.18	.23	5.95	6.79	8.16	8.62	9.49	8.24
8	18	.17	.15	.21	5.50	17.87	7.42	4.37	60.42	6.12
9	16	.17	.15	.16	7.07	4.88	8.94	6.16	2.88	13.12

REFERENCES

D. AEYELS, 1981, « Generic Observability of Differentiable Systems », *SIAM J. Control and Optimization*, Vol. 19, N° 5, September.

T.W. ANDERSON, *The Statistical Analysis of Time Series*, Wiley, New York, 1971.

R. ASHLEY and M. HINICH, 1983, « A Diagnostic Test for Nonlinear Serial Dependence in Time Series Fitting Errors », Department of Economics, Virginia Polytechnic Institute and The University of Texas at Austin, Texas.

G. BENETTIN et al., « Kolmogorov Entropy and Numerical Experiments », *Physical Review A*, Vol. 14, N° 6, December 1976, pp. 2338-2345.

G. BENETTIN et al., « Lyapunov Characteristic Exponents for Smooth Dynamical Systems : A Method for Computing All of Them », *Meccanica*, 15, March 1980, p. 9.

J. BENHABIB and R.H. DAY, 1982, « A Characterization of Erratic Dynamics in the Overlapping Generation Model », *Journal of Economic Dynamics and Control*, 4, pp. 37-55.

A. BEN-MIZRACHI et al., 1984, « Characterization of Experimental (Noisy) Strange Attractors », *Physical Review A*, 29, February, p. 975.

R. BHATTACHARYA and M. MAJUMDAR, 1980, « On Global Stability of Some Stochastic Economic Processes : A Synthesis », in Klein, L. et al. (eds.), *Quantitative Economics and Development*, Academic Press, New York, N.Y. 1980, pp. 19-42.

O. BLANCHARD and M. WATSON, « Are All Business Cycles Alike ? » Department of Economics, M.I.T. and Harvard, Jan. 1984.

J.M. BLATT, 1981, *Dynamic Economic Systems*, M.E. Sharpe, Armonk, New York.

J.M. BLATT, 1980, « On the Frisch Model of Business Cycles », *Oxford Economic Papers*, 32(3), 467-479, 11D, E1.

J.M. BLATT, 1978, « On the Econometric Approach to Business-Cycle Analysis », *Oxford Economic Papers*, 30, 292-300, 11C.

M. BOLDRIN and L. MONTRUCCHIO, 1985, « The Emergence of Dynamic Complexities in Models of Optimal Growth : The Role of Impatience », Conference on Nonlinear Economic Dynamics, Paris, June.

D. BRILLINGER et al., 1980, « Empirical Modeling of Population Time Series Data », *Lectures on Mathematics in the Life Sciences*, Vol. 13, *American Math. Society*, 65-90.

D. BRILLINGER and M. HATANAKA, 1969, « Harmonic Analysis of Non-stationary Multivariate Economic Processes », *Econometrica*, Vol. 37, N° 1, January, 131-141.

W. BROCK and C. SAYERS, « Is the Business Cycle Characterized by Deterministic Chaos ? » Department of Economics, University of Wisconsin-Madison, Madison, WI, September 1985.

W. BROCK and L. MIRMAN, 1972, « Optimal Economic Growth and Uncertainty : The Discounted Case », *Journal of Economic Theory*, 4, 1972, 479-513.

W. BROCK, 1982, « Asset Prices in Production Economy », *Economics of Information and Uncertainty*, J.J. McCall (eds.), University of Chicago Press, 1982, 1-43.

W. BROCK and M. MAGILL, 1979, « Dynamics Under Uncertainty », *Econometrica*, Vol. 47, N° 4, July, 843-868.

W. BROCK and G. CHAMBERLAIN, 1984, « Spectral Analysis Cannot Tell a Macro-Econometrician, whether his Time Series came from a Stochastic Economy or A Deterministic Economy », SSRI W.P. #8419, Department of Economics, University of WI, Madison, WI.

W.A. BROCK and W.D. DECHERT,1985, « Theorems on Distinguishing Deterministic and Random Systems », Department of Economics, University of Wisconsin, Madison, Wisconsin and Department of Economics, University of Houston, Houston, Texas.

W. BROCK, 1985, « Distinguishing Random and Deterministic Systems : An Expanded Version », Department of Economics, University of Wisconsin-Madison, August, *Journal of Economic Theory*, October, 1986.

P. CHEN and W. BARNETT, « A Possible Case of Economic Attractor », Dept. of Physics, U.T. Austin 78712, Dec. 1984.

P. COLLET and J. ECKMANN, 1980, *Iterated Maps on the Interval as Dynamical Systems*, Birkhauser, Basel.

J. CRUTCHFIELD and N. PACKARD, 1983, « Symbolic Dynamics of Noisy Chaos », *Physica 7D*, 1983, 201-223.

R.A. DANA and P. MALGRANGE, « The Dynamics of A Discrete Version of A Growth Cycle Model », in J.P. Ancot (ed.), *Analyzing the Structure of Econometric Models*, Amsterdam : M. Nijhoff, 1984.

R. DAY, 1984, « Intrinsic Business Fluctuations and Monetary Policy : Can The Fed Cause Chaos ? », University of Southern California Working Paper.

R.H. DAY, 1982, « Irregular Growth Cycles », *American Economic Review*, 72, June, 406-414.

R.H. DAY and W.J. SHAFER, 1983, « Keynesian Chaos », Department of Economics, University of Southern California.

J. DELONG and L. SUMMERS, 1984, « Are Business Cycles Symmetric ? » Harvard Institute of Economic Research Discussion Paper 1076, August.

R. DENECKERE and S. PELIKAN, 1984, « Competitive Chaos », Institute for Math and its Applications, University of Minnesota, Minneapolis, Minnesota 55455.

J. DOOB, 1953, *Stochastic Processes*, Wiley, New York.

C. FUTIA, 1982, « Invariant Distributions and the Limiting Behaviour of Markovian Economic Models », *Econometrica*, Vol. 50, N° 2, March 1982, 377-408.

H. FROEHLING et al., 1981, « On determining the Dimension of Chaotic Flows », *Physica 3D*, 1981, 605-617.

A. GOLDBERGER, 1984, « Reverse Regression and Salary Discrimination », *The Journal of Human Resources*, XIX, 3, 293-318.

J.M. GRANDMONT, 1983, « On Endogenous Competitive Business Cycles », *Econometrica*, 53, pp. 995-1045.

C. GRANGER and P. NEWBOLD, 1977, *Forecasting Economic Time Series*, Academic Press, New York.

P. GRASSBERGER and I. PROCACCIA, 1983, « Measuring The Strangeness of Strange Attractors », *Physica 9D*, pp. 189-208.

C. GREBOGI et al., 1984, « Strange Attractors That Are Not Chaotic », *Physica 13D*, 261-268.

J. GUCKENHEIMER and P. HOLMES, 1983, *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields*, Springor-Verlng, New York.

J. GUCKENHEIMER, 1982, « Noise in Chaotic Systems », *Nature*, Vol. 298, 22 July 1982, 358-361.

J. GUCKENHEIMER, 1978, « Comments on Catastrophe and Chaos », *Lectures on Mathematics on the Life Sciences*, Vol. 10, *American Mathematical Society*, 1-47.

M.HINICH and D. PATTERSON, 1985, « Evidence of Nonlinearity in Stock Returns », *Journal of Business and Economic Statistics*, 3, January, p. 69.

M. HINICH, 1982, « Testing for Gaussianity and Linearity of a Stationary Time Series », *J. Time Series Analysis* 3(3), 1969-176.

J. JOHNSTON, 1984, *Econometric Methods* (3rd Edition), McGraw-Hill, New York.

R. KALABA and L. TESTFATSION, 1980, « A Least-Squares Model Specification Test for a Class of Dynamic Economic Models with Systematically Varying Parameters », *Journal of Optimization Theory and Applications*, Vol. 32, N° 4, December, 537-567.

R. LITTERMAN and L. WEISS, « Money, Real Interest Rates and Output : A Reinterpretation of Postwar U.S. Data », *Econometrica*, 53, 1, January, 129-156.

J. LONG and C. PLOSSER, 1983, « Real Business Cycles », *Journal of Political Economy*, 91, 39-69.

R. LUCAS, 1975, « An Equilibrium Model of the Business Cycle », *Journal of Political Economy*, 83, 1113-1144.

C. MANSKI, 1984, « Recent Work on Estimation of Econometric Models Under Weak Assumptions », SSRI Working Paper #8403, The University of Wisconsin at Madison, January.

R. MAY, 1976, « Simple Mathematical Models with Very Complicated Dynamics », *Nature*, Vol. 261, June 10, 459-467.

R. MICHENER, 1984, « Permanent Income in General Equilibrium », *Journal of Monetary Economics*, 13, 1979, 297-305.

G. MOORE and V. ZARNOWITZ, « The Development and Role of the National Bureau's Business Cycle Chronologies », NBER Working Paper #1394, Cambridge, MA, July 1984.

S. NEFTCI, 1984, « Are Economic Time Series Asymmetric Over the Business Cycle ? », *Journal of Political Economy*, Vol. 92, N° 2, April, 307-328.

C. NICOLIS and G. NICOLIS, 1984, « Is There a Climatic Attractor ? », *Nature*, Vol. 311, October, 529-532.

V. OSELEDEC, 1968, « A Multiplicative Ergodic Theorem : Lyapunov Characteristic Numbers for Dynamical Systems », *Transactions Moscow Mathematical Society*, 19, 197-231.

N. PACKARD et al., 1980, « Geometry from a Time Series », *Physical Review Letters*, Vol. 45, N° 9, Sept. 1, 1980, 712-715.

Y. PESIN, 1977, « Characteristic Lyapunov Exponents and Smooth Ergodic Theory », *Russian Math. Surveys*, 32 :4, 55-114.

K. PETTIS et al., 1979, « An Intrinsic Dimensionality Estimator from Near-Neighbor Information », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-1, N° 4, January.

E. PRESCOTT and R. MEHRA, 1980, « Recursive Competitive Equilibrium : The Case of Homogeneous Households, *Econometrica* 48 (1980), 1365-1379.

M. PRIESTLY, 1981, *Spectral Analysis of Time Series*, Vol. I, II, Academic Press, New York, 1981.

I. PRIGOGINE, 1980, *From Being to Becoming*, W. Freeman, New York.

J. RAMSEY and A. ALEXANDER, 1982, « The Econometric Approach to Business-Cycle Analysis Reconsidered », Dept. of Economics, New York University, March 1982.

J. ROUX, R. SIMOYI and H. SWINNEY, 1983, « Observation of a Strange Attractor », *Physica 8D*, 1983, 257.

D. RUELLE and F. TAKENS, 1971, « On the Nature of Turbulence », *Comm. Math. Physics*, 20 : 167-192.

H. SAKAI and H. TOKUMARU, 1980, « Autocorrelations of a Certain Chaos », *IEEE Transactions on Acoustics, Speech and Signal Processing*, V. I. ASSP-28, N° 5, Oct., 588-590.

T. SARGENT and C. SIMS, 1977, « Business Cycle Modelling Without Pretending to Have Too Much A Priori Economic Theory » in *New Methods in Business Cycle Research*, Federal Reserve Bank of Minneapolis, 45-110.

T. SARGENT, 1979, *Macroeconomic Theory*, Academic Press, New York.

T. SARGENT, 1981, « Interpreting Economic Time Series », *Journal of Political Economy*, 89 (2), April 213-248.

C. SAYERS, 1985, « Work Stoppages : Measuring the Attractor Dimension », University of Wisconsin, Madison, Department of Economics, April.

J. SCHEINKMAN, 1985, « Can Economic Fluctuations be Explained by Deterministic Motion ? », Dept. of Economics, The University of Chicago.

C. SIMS, 1984, « Martingale-Like Behavior of Prices and Interest Rates », University of Minnesota, Department of Economics.

C. SIMS, 1984, « Discussion of Grandmont », in *Abstracts for the Workshop on Price Adjustment, Quantity Adjustment and Business Cycles*, Institute for Mathematics and its Applications, University of Minnesota, February.

M. STUTZER, 1980, « Chaotic Dynamics and Bifurcations in a Macro Model », *Journal of Economic Dynamics and Control*, 2, 353-376.

H. SWINNEY, 1985, « Observations of Complex Dynamics and Chaos » in Cohen E.G.D. Ed., *Fundamental Problems in Statistical Mechanics VI*, Elsevier, North-Holland, Amsterdam.

H. SWINNEY, 1983, « Observations of Order and Chaos in Nonlinear Systems », *Physica 7D*, N° 5, 1-3, 3-15.

F. TAKENS, 1983, « Distinguishing Deterministic and Random Systems », in Borenblatt G., Iooss G. and Joseph D. (eds.), *Nonlinear Dynamics and Turbulence*, Pitman Advanced Publishing Program, Boston, 315-333.

F. TAKENS, 1980, « Detecting Strange Attractors in Turbulence », in Rand D. and Young L. (eds.), *Dynamical Systems and Turbulence*, Warwick 1980, Lectures Notes in Mathematics N° 898 : Berlin Springer Verlag, 1980, 366-382.

F. TAKENS, 1984, « On The Numerical Determination of the Dimension of an Attractor », unpublished.

A. WOLF, J. SWIFT, H. SWINNEY and J. VASTANO, 1984, « Determining Lyapunov Exponents from a Time Series », Dept. of Physics, U.T. Austin, 78712, Oct.

V. ZARNOWITZ, 1984, « Recent Work on Business Cycles in Historical Perspective : Review of Theories and Evidence », University of Chicago and NBER, Oct.

V. ZARNOWITZ and G. MOORE, « Major Changes in Cyclical Behavior », NBER Working Paper #1395, Cambridge, MA, July 1984.

Economic Theory as a Generator of Measurable Attractors

William Barnett
Stuart Professor — Department of Economics — University of Texas — Austin

Ping Chen
Graduate Research Assistant — Department of Physics — University of Texas — Austin

1. INTRODUCTION

A deep new scientific methodology is beginning to take form as a result of dramatic recent advances in nonlinear dynamics[1]. During the past ten years, the mathematics and physics literatures on strange attractors, bifurcation theory, and deterministic chaos have acquired growing capabilities in many fields[2]. The reason for the rapidly expanding interest in this literature is evident. The recent advances have generated the capability to produce deterministic solution paths that have the characteristics of stochastic processes and hence have the properties usually seen in real data. Those results, in addition to being directly empirically useful, can be derived directly from the relevant structural theory without the need to introduce *ad hoc* additive or exogenous stochastics. Furthermore, the resulting solution paths can exhibit the appearance of unpredictability and of structural change, when in fact there has been no shift in the model's structure at all. In other words, empirically useful results, applicable even to very long run data, become available without the need to introduce *any* « free parameters ». In fact a measure theoretic approach exists for interpreting chaotic solution paths over the set of initial conditions as realizations of a stochastic process, so that mathematical chaos can be viewed as a means of producing *a deterministic origin for stochastic processes*. Hence the literature on deterministic chaos can pro-

duce a theoretically deep explanation for « stochastic » phenomena. Structural theory and stochastics become linked.

Once it is found that observed data can be explained from a chaotic attractor model, the possibility then exists of recovering information about the unknown model which produced the observed data path. In this paper we report the results of an ongoing research project that has been successful in identifying economic chaotic attractors from long time series of unusually high quality data. In particular, those attractors can explain the dynamical behavior of the broad Divisia monetary aggregates and hence can reveal information about the nature of the dynamical system that produces the observed monetary services path over time. The revealed information produces a measure of the degree of controllability and of the degree of aggregation error for each of the various monetary aggregates considered. We shall use these results as a means of comparing the usefulness of the official simple sum monetary aggregates, the Divisia demand monetary aggregates, and the new Divisia supply monetary aggregates in terms of their controllability and degree of embedded aggregation error. Although there previously have been a number of successful applications of deterministic chaotic dynamics in economic theory, we believe that our results represent the first clearly successful empirical application.

2. THE NEW RESULTS FROM NONLINEAR DYNAMICS

2.1. Deterministic Chaos.

The relevant mathematical theory deals with very deep modeling, in the sense that all sources of the system's dynamics are endogenized, so that the system itself then produces its own dynamics. In particular, let us define s_t to be the vector of state variables at period t. The state vector is of dimension S. The state vector is defined to contain all information that is relevant to the behavior of the economy during period t + 1. It therefore follows, almost tautologically, that there must exist a function f (called the « *deterministic dynamical system* » or « law of motion ») such that $s_{t+1} = f(s_t)$ for all t. We need *not* assume that all or any of the state variables are actually observed. We also certainly need not assume that we know f. However, we do observe a variable m_t during period t. Clearly m_t must depend upon the state vector. Hence there must exist an « *observer function* » g such that $m_t = g(s_t)$ for every t. The space that contains all possible paths of the state vector is called the *state space* and has dimension S.

Clearly in reality the dimension of that space is huge, and there is likely to be a tremendous number of choices of state vectors, s_t, and dynamical systems, f, that can explain the existing data on m_t deterministically. However, in the spirit of all economic theory and of all science, the literature on chaotic dynamics seeks to produce the *simplest* useful explanation for the observed data path. Hence we shall be seeking the lowest dimensional state space that can be used to produce a self-generating deterministic explanation of the past, present, and future behavior of the data series m_t. We should expect that the resulting minimum

dimension may be very low, since it has been shown that very simple small neoclassical economic models can produce richly complex chaotic dynamics.

If the function f is monotonically increasing or monotonically decreasing, the state vector will forever trend in one direction. Such simple trends do not produce interesting (i.e. stochastic appearing) results in this literature. As a result, it is assumed that f has at least one turning point.

In applications, we might, for example, assume that our data had been detrended, so that f is not monotonic. Under the very general and highly abstract assumptions just made, it has been shown in the mathematics and physics literatures that m_t can evolve over time in a remarkable number of ways, from simple convergence to a constant, to cycling, to deterministic chaos having the properties of almost any kind of informative stochastic process. Much is now known about all of the possible regimes, including the conditions producing transition between regimes. The source of this theoretical richness is the nonlinearity of f and the feedback produced by the definition of the function f. Feedback induced dynamics is produced by the resulting nonlinear iterations on the function f. Beginning at some initial condition s_o, the sequences or paths produced for s_t and m_t by iteration on f are called the *orbits* or *trajectories*.

For our purposes, chaotic dynamical systems can be viewed as deterministic systems that produce solution paths that appear to be stochastic[3]. The literature becomes noninformative in the case in which the data is *pure* white noise. In that case, any exact deterministic explanation would have to be infinitely complex, so that no possible finite amount of information about the state vector could produce any information whatsoever relative to m_{t+1}.

It is difficult to believe that the dynamic evolution of economic variables, observed to a finite degree of disaggregation, is beyond all possible explanation with a finite number of variables, and hence we believe that the recent developments in nonlinear dynamics should be empirically applicable to economic systems. We therefore believe that it should be possible empirically to recover some information about unknown dynamical economic systems from observed paths of economic variables. Potential exceptions to these conclusions could result from the existence of much (high noise to signal ratio) pure white noise induced by substantial measurement error. In such cases, currently available algorithms cannot always extract the signal from the noise. As a result, data quality is of great importance in empirical applications of mathematical chaos to economics.

Differences in opinions between the various macroeconomic schools of thought amount to differences in opinions about the degree of depth of modeling needed to capture the source of the system's dynamics. Keynesians often argue that the system's dynamics can be produced from within the structure of the private sector. The monetarist model requires endogenizing the central bank's reaction function. The literature on

rational expectations modeling and on the related « new classical economics » has produced its own dynamical models, which typically requires endogenizing private sector expectations (especially about relative prices) consistently with the economy's structure and also endogenizing all sources of governmental shocks to the private sector ; the resulting econometric specifications reveal the economy's « deep parameters ». The linearity assumptions common in models in all three traditions have rarely been viewed as inherent to the economic theory, but rather as simplifying local approximations. In fact formidable literatures on endogenous business cycles exist in each of the three traditions. The empirical inference techniques that we use from chaotic dynamics require no prior assumptions about the depth of modeling needed to endogenize the system's dynamics and hence are *not* dependent upon one's views about the source of the system's dynamics. In fact the techniques are totally nonparametric and specification-free !

2.2. Strange Attractor Theory

Attractor sets, and especially « strange » attractors, play an important role in chaotic dynamics. We now define the concept of a strange attractor. First collect together the cluster (limit) points of the sequence (s_t) into a set, S, called the *attractor*. The sequence (s_t) is produced, through its definition, by successive iterations on the function f. In the limit the orbits will be trapped or « locked » in « by that attractor set, so long as the initial condition, s_0, is sufficiently close to the attractor set[4]. Once trapped by the attractor set (i.e. for sufficiently large t), the orbit will wander within or near the attractor set forever, and the behavior of the trapped paths within the attractor set will be influenced by the geometry of the attractor set.

If the sequence converges to a fixed point, then the attractor is a singleton set containing that one limit point and therefore has dimension zero. We can imagine the sequence as drawn by some attraction to the limit point, so that the sequence necessarily eventually must converge to the limit point. If, however, the attractor set contains two points, the sequence will simultaneously be « attracted » by two separated points, and therefore will oscillate between points in two separate sequences, each of which converges to a different limit point. This can occur when the choice of the model's parameter values causes the height of the turning point of the function f to pass a critical point at which the original single limit point « *bifurcates* » or « fissions » into two cluster points.

The transition to chaos often proceeds as follows. As the parameters of the model are further shifted, the two cluster points, which coincided at their birth from the bifurcation, separate farther and farther apart. Eventually succeeding bifurcation points will be passed at predictable values of the parameters. The number of cluster points doubles each time bifurcation occurs, and every point in the attractor bifurcates simultaneously. The potential richness of the sort of observed behavior that can be produced increases as the number of points in the attractor set increases[5].

In the limit, after the infinite sequence of bifurcations has converged, the attractor set can have an infinite number of points. In such cases, which can produce transition to « chaos », the attractor set usually will be a *fractal*, which is a Cantor set of fractional Hausdorf dimension[6]. We then say that the attractor set is a chaotic or « *strange* » *attractor.* The knob setting at convergence of the sequence of bifurcations to the fractal set is finite, so that convergence of the knob settings to the chaotic regime actually can be attained. Chaos is *not* an exceptional case that can be approached only in an unattainable limit.

It has been shown that the existence of a strange attractor is neither necessary nor sufficient for chaos. Nevertheless, in virtually all known useful applications of chaos, a strange attractor exists. Hence we shall equate the existence of chaotic dynamics with the existence of a strange attractor set drawing the time path towards it. The potential explanatory power of strange attractors is tremendous, since they can « attract » a sequence into deterministic behavior that is sufficiently complex and turbulent to appear to be stochastic.

Since orbits produced from nearby initial conditions may initially diverge rapidly (exponentially), a « statistical » view of the generated phenomena becomes natural. In addition, there is a beautiful form of mathematical order that can be found in the resulting « chaos ». The resulting order is a deeper kind than the more transparent kind encountered prior to the transition to chaos[7].

2.3. Empirical Inference Capabilities of Chaotic Dynamics and Strange Attractors

There are a number of techniques widely used in physics to detect chaos. However, not all of the approaches developed by physicists are applicable in economics, because most of those methods require a large amount of data to assure sufficient precision. In this paper we use only those techniques that are potentially useful for detecting chaos with less than 1000 data points. In this section we discuss each of those techniques, although we first describe the fundamental relationship between phase space and state space.

2.3.1. Phase Space Embedding

It can be shown that strange attractors are never topological manifolds, which have integer dimensions, but rather fractals, which are Cantor sets having noninteger dimensions and zero volumes (relative to Lebesgue measure). A measure theoretic method, based upon Hausdorf measure, exists for measuring the dimension of a fractal. The resulting noninteger-valued dimension of a strange attractor is called its fractal dimension, which measures the number of independent degrees of freedom relevant to the system's dynamical behavior[8]. *The attractor is defined to be « strange » if and only if the attractor's dimension is noninteger valued, i.e., if the attractor is a fractal.*

The attractor set is a subset of the state space, although we typically have no measured values of the state vector, *s*. We only have measured values of m. Hence it is important to have a means of relating fractal dimension to the potential values of m. Takens (1980, p. 369) has done so through an embedding theorem, which can be understood as follows. Select a state vector from the basin set ; iterate on (g,f) repeatedly (n-1)τ times, where time delay, τ, and « embedding dimension », n, are preselected ; and then stack the resulting n values of m_t into an n-dimensional vector, with successive elements of the vector being separated by τ periods. If the first observation in the n-history is m_t, then we can designate the resulting n-history at t by $m(n,\tau)_t$, where we define

$$m(n,\tau)_t = (m_t, m_{t+\tau}, m_{t+2\tau}, \ldots, m_{t+(n-1)\tau}) \qquad (2.1)$$

at eacht t. For notational simplicity below, the t will usually be dropped in $m(n,\tau)_t$. The existence of the argument should be understood.

The space of those n-histories, $m(n,\tau)_t$, for a fixed choice of n is called *phase space*, and the selected value of n is called the *embedding dimension*. As defined in that manner, phase space has delay coordinates. The set of phase space trajectories for all possible initial conditions for the state vector within the basin set is called the *phase portrait* of the system. Sometimes the term phase portrait alternatively is used to refer to the trajectories of the n-histories produced in phase space by the actual observed data.

2.3.2. Dimensionality Measures

Suppose the construction of $\underline{m}(n)_t$ for fixed embedding dimension is repeated for each possible value of $\underline{s}_t$ within the basin set[9]. Designate the resulting set of n-history vectors by J(n). Then the Hausdorf dimension of J(n) is computed. This computation is repeated for successively larger values of n. Takens (1980, p. 369) proved that the limiting dimension of J(n) as $n \to \infty$ equals the fractal dimension of the strange attractor in state space. We shall call that theorem the *Takens Embedding Theorem*. Hence we can define the dimension of the strange attractor in state space in terms solely of n-histories, $m(n)_t$, of observable values of m_t[10]. Hence *a strange attractor will exist in phase space if and only if a strange attractor exists in state space, and the dimensions of the attractors will be the same in both spaces.*

The Hausdorf dimension of a strange attractor, whether computed in phase space or directly in state space, is extremely difficult to compute. As a result, more easily computed approximations have been proposed. The most notable examples are the Grassberger-Procaccia correlation dimension and the information dimension. Physical scientists now consider the Grassberger and Procaccia (1983a,b) method to be the best method[11].

We first define the Grassberger-Procaccia method of measuring the dimension of J(n). The dimension measure resulting from that method is called the correlation dimension of J(n), since the procedure uses the correlation function, $C_n(\varepsilon)$, defined by

$$C^*_n(\varepsilon)=\#\{(i,j) : ||\underline{m}(n)_j-\underline{m}(n)_j||<\varepsilon,\ 1\leq i\leq N_n,\ 1\leq j\leq N_n, i\neq j\}/(N_n^2-N_n), \quad (2.2)$$

where $N_n=N-(n-1)\tau$ is the number of n-histories that can be produced from a sample of size N with time lag τ, and # A denotes the cardinality of (number of distinct points in) the set A. In particular, the Grassberger-Procaccia correlation dimention of J(n) is defined to be

$$D(n) = \lim_{\varepsilon\to O} (\log_2 C^*_n(\varepsilon) / \log_2\varepsilon). \quad (2.3)$$

If we then take the limit as the embedding dimension goes to infinity, as suggested by the Takens Embedding Theorem mentioned above, we get the correlation dimension of the strange attractor in state space,

$$D = \lim_{n\to\infty} D(n). \quad (2.4)$$

In practice, with a well behaved chaotic model the value of D(n) will « saturate » (i.e., attain its limit) at some finite level of n, called the saturation embedding dimension, n_S. The correlation dimension of the strange attractor then is $D(n_S)$. At the opposite extreme, saturation is never attained if the data is white noise. In that case all dimensions of phase space are used by the data at every embedding dimension, and D(n)=n at all n. As a result, D equals infinity for white noise.

The reason for definition 2.3 is easily seen from the fact that at sufficiently small ε, Grassberger and Procaccia showed that the correlation function can be written approximately in the form

$$\log_2 C^*_n(\varepsilon)=\log_2 k + D(n) \log_2 \varepsilon, \quad (2.5)$$

where k is a constant. Hence (2.3) follows by dividing (2.5) by $\log_2 \varepsilon$, and letting $\varepsilon\to O$. From (2.5) we see that we alternatively can-define D(n) by

$$D(n) = \lim_{\varepsilon\to O} (\delta\log_2 C_n(\varepsilon) / \delta \log_2\varepsilon), \quad (2.6)$$

which is, in fact, the form used in practice. The procedure is to select a small value for ε, but not to go all the way to zero, since data noise tends to dominate at extremely small values of ε. As a result, (2.6) must be used , rather than (2.3), since for small nonzero values of ε the intercept of the plot of $\log_2 C_n(\varepsilon) / \log_2 \varepsilon$ against $\log_2 \varepsilon$ remains nonzero.

In particular, $\log_2 C_n(\varepsilon)$ is plotted against $\log_2 \varepsilon$ to permit measurement of the slope, $\delta\log_2 C_n(\varepsilon) / \delta \log_2 \varepsilon$, versus $\log_2 \varepsilon$. At sufficiently low value for $\log_2 \varepsilon$, the slope will reach a plateau before noise begins to dominate and to produce further variation of the slope at even lower values of $\log_2 \varepsilon$. The value of the slope along the plateau is used as D(n)[12]. By (2.4), the limiting value of D(n), approximated by the value of D(n) at the saturation embedding dimension, is the fractal dimension D.

We would expect from the Takens Embedding Theorem, that D will measure the dimension of the strange attractor in state space, despite the fact that D(n) at all finite n is produced from vectors in phase space. In fact Grassberger and Procaccia have shown that D can indeed be written in terms of the state vectors, s_t ($1\leq t\leq N$), as :

$$D = \lim_{\varepsilon \to 0} (\log_2 C^*(\varepsilon) / \log_2 \varepsilon), \qquad (2.7)$$

where

$$C^*(\varepsilon) = \lim_{N \to \infty} \#\{(i,j) : ||\underline{s}_j - \underline{s}_i|| < \varepsilon,\ 1 \leq i \leq N,\ 1 \leq j \leq N,\ i \neq j\} / (N^2 - N). \qquad (2.8)$$

For sufficiently large n and N, the limiting correlation dimension can be found with reasonable accuracy. Grassberger and Procaccia showed that the limiting correlation dimension, D, is less than, but nevertheless closely approximates, the Hausdorf fractal dimension of the strange attractor. We use the Grassberger-Procaccia procedure for computing fractal dimensions for strange attractors.

The next integer larger than that fractal dimension of the strange attractor clearly provides a lower bound to the dimension of the space that can contain the attractor. Furthermore, since the attractor set is a subset of the state space, it follows that the next integer larger than the fractal dimension also provides a lower bound to the dimension of the state space that could contain state vector paths (s_t) capable of producing the observed orbit of m_t. Since the Grassberger-Procaccia correlation dimension, D, is less than the Hausdorf fractal dimension, we see that the correlation dimension, D, of the strange attractor is also a lower bound to the dimension of the state space, S.

All of the above conclusions, including the existence of the postulated limits, will hold, if the state vector orbits are generated in accordance with a deterministic dynamical model consistent with the existence of (f,g). This is a powerful result, since it provides a *lower bound to the number of variables that can be used to produce an entirely theoretical explanation for endogenous evolution of* m_t. That result follows from the fact that $m_t = f(s_t)$, and the dimension of s_t is S. It is important, however, to recognize that the simplest chaotic deterministic model that can produce the observed path of m_t need not be the model which actually did produce the observed data. In fact in experimental applications the fractal dimension is typically very much smaller than the number of variables that actually produced the observed data. Only when the attractor is not chaotic (« strange ») is that lower bound tight.

2.3.3. The Lyapunov Exponent as a Test for Chaos.

The *Lyapunov exponent* provides a different sort of information. The Lyapunov exponent characterizes the nature of the wandering behavior of chaotic orbits. Trajectories initially close to each other are considered, and the long run average rate at which they diverge or converge over time is computed in various locally defined orthogonal directions. It can be shown that the behavior of all of the trajectories drawn by the attractor can be characterized by a finite number of Lyapunov exponents. *The positivity of the largest Lyapunov exponent can be used as the definition of mathematical chaos.* As a result, the sign of the maximum Lyapunov exponent is more important than its numerical value.

The vector of Lyapunov exponents, $(\lambda_1(t),\lambda 2(t),\ldots,\lambda_S(t))$, is called the *Lyapunov spectrum, where* $\lambda_i(t)$ *is independent of t for all* s_t within the

attractor set. Hence the Lyapunov exponents are considered only within the attractor set. The maximum Lyapunov exponent is

$$\bar{\lambda}(t) = \max \{\lambda_j(t) : j = 1,...,S\}. \qquad (2.9)$$

We compute the largest Lyapunov exponents, $\lambda(t)$, using the algorithm in Wolf, Swift, Swinney, and Vastano (1984)[13].

2.3.4. Inspection of the Phase Portraits

An informal exploratory data analysis often is useful in uncovering chaotic dynamics. The approach is to plot the n-histories in a 2-dimensional phase space with preselected time delay. In other words, for n=2 and a preselected value for τ we plot the value of the two dimensional vector $\underline{m}(n,\tau)_t$ in phase space at each t. The geometric pattern that appears is compared with the patterns that previously have been shown to be produced by systems that have known properties[14]. As a result, an initial inspection of phase portraits is useful, although the more formal methods described above should follow.

2.3.5. Inspection of the Autocorrelation Function

Another form of useful preliminary exploratory data analysis is inspection of the autocorrelation function. In particular, let var ($\underline{m}$) be the variance of the sample, $\underline{m} = (m_1,...,m_N)$, and let cov (t,t) be the covariance of m_t and $m_{\bar{t}}$. Then the autocorrelation function is defined to be

$$G(t,\bar{t}) = \text{cov}\ (t,\bar{t})/\text{var}\ (m). \qquad (2.10)$$

It is common to plot $G(t,\bar{t})$ against $\bar{t}$ at fixed t, since the properties of the resulting plot are known for many kinds of models[15].

2.3.6. Time Series Statistical Inference Methods

Computations of correlation dimensions and maximum Lyapunov exponents can produce convincing evidence of deterministic chaos. However, those tests use point estimates having incompletely known statistical properties. The reason is that they are based directly upon the definition of chaos and hence upon deterministic mathematics. Nevertheless, we have observed above that a stochastic interpretation can be given to the orbits produced from chaotic dynamics, and that the distribution of the process can be derived directly from the deterministic dynamics. However few studies currently exploit the stochastic interpretation[16].

3. DATA REQUIREMENTS

3.1. The Issue

As discussed above, data quality is very important in applying the currently existing techniques for uncovering the dynamics of nonlinear systems. The white noise produced from measurement error complicates the use of existing methods for extracting information from strange attractors, since such data error produces fuzzy appearance to strange attractor sets. In addition, it seems reasonable to presume that the variable used as the measured variable, m_t, should be an « important »

one, since the empirical techniques currently available for extracting empirical information from chaotic orbits use the observed orbit of only one variable to reconstruct the dynamics of the complete (simplest compatible) system in state space.

It is important to recognize that there is no inherent reason to believe that economic data is more noisy than data produced in the physical sciences. Many techniques from classical statistics require most variables to be exogenous. But fully nonstochastic exogenous variables usually can be acquired only as control variables in controlled experiments. As a result, it is often believed that inference procedures applied to economic data, which typically is not produced from controlled experiments, are greatly complicated by « noisy » economic data. While simultaneity bias created by endogeneity of many variables is certainly a problem for classical statistical inference procedures, endogeneity is *not* a problem for the use of strange attractor theory. The distinction between endogenous and exogenous variables is irrelevant to chaos, and in fact the power of chaotic dynamics lies in its ability to produce stochastic appearing solution paths. The usefulness of fully deterministic mathematical chaos requires only that the data be of high quality, so that large amounts of purely white noise produced by measurement error not be present.

There is no reason to believe that *completely disaggregated* economic data need inherently be any more noisy, in the sense of possessing high measurement error, than data produced from controlled experiments in the physical sciences. But of course almost all available economic data is aggregated both over goods and over economic agents, and that aggregation can indeed induce white noise into the aggregated data. Hence the problem for the empirical use of deterministic chaos and strange attractor theory is not that economic data is not produced from controlled experiments, but rather that the data is aggregated, and aggregation bias can induce noise into the aggregated data.

Most governmental data is produced from Laspeyres or Paasche indexes, which are only first order approximations to the true aggregation theoretic aggregate. The resulting second order remainder terms could possibly produce a troublesome amount of noise for the nonlinear inference procedures currently available from the physics literature. While the money supply would be an « important » variable, the official Federal Reserve System simple sum monetary aggregates are not even first order approximations, and hence have first order remainder terms, which certainly are too large for our purposes. In addition, the techniques that we seek to use in this research benefit substantially from the largest possible sample sizes.

We believe that a data series potentially well suited to the objectives of this research is the weekly data on the Divisia monetary aggregates. Barnett's Divisia monetary aggregates are second order approximations to the exact aggregation theoretic monetary service flows produced by the corresponding monetary asset components. As a result, the remainder terms are only third order and have been shown to be less than the round-

off error in the component data. Furthermore, the monetary service flows in the economy are important variables. We use both the Divisia demand monetary aggregates and Divisia supply monetary aggregates as data in our tests for chaos. For purposes of comparison, we also use the official simple sum monetary aggregates and IBM common stock daily returns.

3.2. The Divisia Monetary Aggregates

Barnett (1980 ; 1981, ch.7) introduced the use of neoclassical aggregation and index number theory into monetary economics[17]. His Divisia aggregates, based upon Diewert's (1976) superlative class of index numbers, measure the economy's flow of monetary services, as perceived by the users of those monetary services. As a result, the relevant aggregation theory is that produced by neoclassical consumer and factor demand theory. Much empirical research now exists on that subject ; and, as would be expected from the theory, the empirical research is mostly related to demand-side empirical tests. See, e.g., Barnett (1982,1983,1984) and Barnett, Offenbacher, and Spindt (1981,1984). More recently Barnett (1986) has introduced the Divisia supply monetary aggregates, which are based upon supply-side aggregation theory with a conventional neoclassical model of financial intermediary monetary asset supply. As shown both theoretically and empirically in Barnett, Hinich, and Weber (1986), the aggregation-theoretic exact money demand aggregate need not always equal the aggregation-theoretic exact money supply aggregate[18].

6. RESULTS

We applied the tests of chaos described above to the four Divisia demand monetary aggregates, the four Divisia supply monetary aggregates, and the four simple sum monetary aggregates. For comparison, we also produced phase portraits and autocorrelation function lots for 2000 successive observations on IBM daily common stock returns, beginning on July 2, 1962. The monetary aggregate data was weekly data beginning in January 1969[19].

We detrended using log linear detrending. Our data therefore consists of the residuals of an ordinary least squares linear regression of the logarithm of the aggregate on time. Since log linear detrending removes the neutral constant growth rate trend, we believe that log linear detrending is the appropriate form of detrending for these tests with economic data. We use the following abbreviations for the detrended monetary aggregates data series. The Divisia monetary demand aggregates that aggregate over the components of M1, M2, M3 and L are denoted by DDM1, DDM2, DDM3, and DDL, respectively. The corresponding Divisia monetary supply aggregates are denoted by DSM1, DSM2, DSM3, and DSL, respectively. The corresponding official Federal Reserve simple sum monetary aggregates are SSM1, SSM2, SSM3, and SSL, respectively. As a check on robustness to the detrending method, we reran all of our tests using log cubic smoothed data. In that case, our data consists of the residuals of an ordinary least squares regression of the logarithm of

the data on a third order polynomial in time. We designate log cubic smoothing by appending the superscript C to the aggregates symbol. For example, $DDM2^C$ designates log cubic smoothed Divisia demand M2. Although we see no particular theoretical merit to log cubic detrending, it does produce a convenient test for robustness to the theoretically attractive log linear detrending.

The successful cases were DDM2, DDM3, DDL, DSM2, SSM2, and $DDM2^C$. Those data series passed all of our tests for mathematical chaos and for the existence of a strange attractor. In all other cases, a finite correlation dimension could not be found, since no linear region, (2.5), could be found in the plots of $\log_2 C_n(\varepsilon)$ against $\log_2 \varepsilon$ for any low embedding dimension n. In those unsuccessful cases we conclude that either the data is very noisy, or else that the aggregate cannot be controlled using a small number of instruments, since the minimum dimensionality of the state space is extremely large. We now describe the results in detail for the successful cases. For details of these results along with the supporting tables and figures, see Barnett and Chen (1986a,b).

Since the attractors are fractals (strange attractors), the correlation dimensions are all fractional. In each case, the correlation dimension estimate was found by Barnett and Chen (1986a,b) to be between 1.3 and 1.5. Hence the minimum dimension of the state space is 2 for discrete time models in each of our successful cases[20].

7. CONCLUSIONS

The monetary aggregates DDM2 and DDM3 were found to possess the characteristics of mathematical chaos very clearly. The monetary aggregates DDL, DSM2, and SSM2 possess those characteristics rather well, but with more noise being evident. Those results combined with the very low estimated state space dimensions suggest one of the following two conclusions. The remaining aggregates that do not produce well defined dimensions either (1) are low quality noisy aggregates having substantial aggregation or data error, or (2) are good quality aggregates that cannot be controlled by a small number of control variables, so that the state space dimension must be very large.

These results are consistent with the relevant aggregation theory. At the lowest level of aggregation, M1, little substitution has been internalized by aggregation, and hence many explanatory variables should be expected to be in the demand function and supply function for DDM1, DSM1, or SSM1. It therefore should be expected that those aggregates should be volatile and difficult to control without the use of many policy instruments. In other words the minimum dimension of the state space must be large.

The fact that Divisia demand aggregates worked better than the simple sum aggregates is precisely as should be expected from the aggregation theory. The simple sum index is not even a first order approximation to the exact aggregation theoretic aggregate, and such zero'th order approximations produce tremendous aggregation error. We should expect to find extremely high noise in the simple sum monetary aggregates[21].

The fact that DDL possessed more noise than DDM2 and DDM3 probably results from the fact that the quality of the data needed to go from the M3 level of aggregation to the higher L level of aggregation is not as good as that used in the M3 level of aggregation. Similarly the quality of data used in the supply side Divisia monetary aggregates is not as high as that used in computing the demand side Divisia monetary aggregates, since the user cost formulas needed in the computation of the supply side Divisia monetary aggregates contain, in theory, many transactions costs on which no data exists. In addition, the Divisia demand monetary aggregates directly measure the economy's monetary service flow. But the regulatory wedge (created by the existence of required reserves paying no interest) between the Demand and supply Divisia monetary aggregates produces distance between the Divisia supply monetary aggregates and the economy's monetary service flow.

We conclude that DDM2 and DDM3 are high quality aggregates that are likely to be controllable more easily than any of the other monetary aggregates. If, for example, the minimum state dimension of 2 were the *actual* state dimension, then only two instruments would be necessary to permit perfect control of the aggregate. If only one of the state variables were a controlled policy instrument, then it would need only to be used to offset the effects on the monetary aggregate of the one remaining uncontrolled variable. In reality, of course, we should expect that the actual state dimension is much higher than 2, since the lower bound produced by the correlation dimension is very conservative.

We also conclude that the literatures on deterministic chaos and strange attractor are applicable to nonparametric econometric inference, when very high quality, carefully aggregated data, such as DDM2 and DDM3, are used.

NOTES

(1) Furthermore, a strong case for the use of nonlinear dynamics as the fundamental general paradigm in science has been made by Prigogine (1980).

(2) For surveys of much of that literature, see Barnett and Chen (1986a,b) and Nicholis and Nicholis (1984). Applications to economic theory were pioneered by Benhabib and Day (1980,1981,1982) and Stutzer (1980). For additional important contributions to economic theory, see Day (1982,1983) and Grandmont (1985). For an earlier use in empirical economic inference, see Brock (1986). Although that pioneering paper contained the first attempted empirical application to economics, that paper was not successful at detecting deterministic chaos. Our research was successful at detecting chaos, perhaps because our sample sizes are about four times larger than those used by Brock.

(3) Nevertheless, an entirely rigorous and formal definition for chaos exists in the mathematics literature. In particular, a time sequence is « *chaotic* » or « turbulent », if the sequence has the following three properties : (1) *sensitive dependence* upon the initial conditions (called the « *seed* »),(2) a form of stationarity, and (3) nonperiodicity. Sensitive dependence upon initial conditions is defined to mean that if two initial conditions are very close together, then the two induced time paths of m_t will initially diverge exponetially (i.e., as if produced stochastically).

(4) The « *basin* » or « domain of attraction » induced by an attractor set is the closure of the set of initial conditions (« seeds ») such that the orbit originating at the initial condition will eventually be trapped by the attractor set.

(5) If the behavior is periodic, the attractor is a closed loop (a limit cycle). If the behavior is n-periodic, the attractor is an n-torus. In all of those cases, points that are initially close will remain close for all time.

(6) Fractals are the sets made famous by Mandelbrot (1977).

(7) See Prigogine and Stengers (1984). In particular, strange attractors (as well as fractals in general) are charaterized by an infinite regress of detail, with a never-ending nesting of pattern within pattern as one looks increasingly closely at the appearance of the attractor set.

(8) In general, the dimension of a set is the amount of information needed to specify the position of a point on the set to within a given accuracy. If the attractor is a Euclidean object (i.e., topological manifold), any of the methods described below of measuring attractor dimension produces the Euclidean dimension of the attractor.

(9) In empirical applications, the set of initial conditions is restricted to the set of state vector values that produced the set of observed values of m_t.

(10) In fact Takens proved an even stronger result. He proved that there exists a deterministic dynamical system (coupled recursion relation), F_n, in phase space for any fixed embedding dimension, n, such that a one to one correspondence exists between the dynamical properties (in particular, all conjugate invariants) of $s_{t+1} = f(s_t)$ in state space and the dynamical properties of $m(n)_{t+1} = F_n(m(n)_t)$ in phase space. Those invariant properties include the Lyapunov exponent (to be discussed below) as well as the dimension concepts.

(11) See Barnett and Chen (1986a,b) for further background and discussion.

(12) It is important not to exceed the lowest candidate for a saturation embedding dimension, since the number of observations in phase space rapidly declines as the embedding dimension increases. With too few points in phase space, the ability to identify the linear region, (2.5), quickly disappears.

(13) The Fortran computer coding for the algorithm was published in the paper by Wolf, Swift, Swinney, and Vastano (1984). For a formal mathematical definition of the Lyapunov exponents, see Barnett and Chen (1986a,b).

(14) For example, periodic motion with period equal to $T=k\tau$ is known to produce a phase portrait with k points in a closed orbit. At the other extreme, pure white noise produces a uniform cloud of data points in phase space. Chaos produces complex non-uniform phase portraits, such as those of the Henon attractor, the Rossler model, and the Mackey-Glass model.

(15) For example, it is known that the autocorrelation function is an oscillating cosine function for a periodic process. The autocorrelation function is a delta function (spike at the origin) for white noise, since m_t is then totally uncorrelated with m_τ for any $\tau \neq t$. For deterministic chaos, the autocorrelation function first exhibits fast decay at an exponential rate and then residual oscillations with slow decay. In the limit, as the orbit is completely trapped by the attractor, the asymptotic behavior is independent of the initial conditions. In other words, the correlation function has a long fat tail declining asymptotically to zero at the right.

(16) For further discussion of that literature, see Barnett and Chen (1986a,b).

(17) For a fuller presentation of the relevant economic theory, see Barnett (1986).

(18) This potential complication can arise, even if both aggregates are produced from Divisia aggregation over the same components, and *even if the markets in all monetary components are continually cleared.* This paradox is caused by the existence of required reserves having regulated zero rates of return. That regulatory constraint produces an implicit tax on financial intermediation, in terms of foregone investment interest. Financial intermediaries receive zero return on part of their deposits and a market return on the rest. However, depositors receive the same marginal return on every dollar deposited in a given account. Barnett, Hinich, and Weber (1986) showed that subtle dynamic differences exist between the corresponding demand and supply indexes at high frequencies.

(19) At the M1, M2 and M3 levels of monetary aggregation, our data included the 807 weekly observations beginning in January 1969 and ending in July 1984. At the highest level of aggregation, called L by the Federal Reserve, our data included 798

weekly observations beginning in January 1969. Our data source is Fayyad (1986). The remaining 9 observations at the L level of aggregation were not yet available.

(20) Recall that m_t is not an element of the state vector, s_t. As a result, the minimum dimension of the space that could contain the path of the augmented vector, (m_t, s_t), is 3 in each of our successful cases. If we were to impute our results to a continuous time world, then the convention is to use the augmented vector (m_t, s_t). Hence if the results of our computation were imputed to continuous time, the minimum state space dimension in our successful cases would be 3.

(21) The exception was the simple sum M2 aggregate. We do not know the reason for the favorable results with SSM2 ; however, it has been observed previously that M2 is the only simple sum monetary aggregate that has ever been found to perform well in comparisons with the Divisia monetary aggregates in other tests. See Barnett (1982).

REFERENCES

Barnett, William A. (1980), « Economic Monetary Aggregates : An Application of Index Number and Aggregation Theory », *Journal of Econometrics*, vol. 14, September, pp. 11-48.

Barnett, William A. (1981), *Consumer Demand and Labor Supply : Goods, Monetary Assets, and Time*, Amsterdam, North-Holland.

Barnett, William A. (1982), « The Optimal Level of Monetary Aggregation », *Journal of Money, Credit, and Banking*, vol. 14, n° 4, pt. 2, November, pp. 687-710.

Barnett, William A. (1983), « New Indices of Money Supply and the Flexible Laurent Demand System », *Journal of Business and Economic Statistics*, Vol. 1, n° 1, January, pp. 7-23.

Barnett, William A. (1984), « Recent Monetary Policy and the Divisia Monetary Aggregates », *American Statistician*, vol. 38, August, pp. 165-172.

Barnett, William A. (1986), « The Economic Theory of Monetary Aggregation », in *New Approaches to Monetary Economics*, edited by William Barnett and Kenneth Singleton, Cambridge : Cambridge University Press, forthcoming.

Barnett, William A. and Chen, Ping (1986a), « The Aggregation-Theoretic Monetary Aggregates are Chaotic and have Strange Attractors : An Econometric Application of Mathematical Chaos », Center for Economic Research Discussion Paper N° 86-5, University of Texas at Austin, Austin, Texas.

Barnett, William A. and Chen, Ping (1986b), « Deterministic Chaos and Fractal Attractors as Tools for Nonparametric Dynamic Econometric Inference », University of Texas at Austin, Austin, Texas.

Barnett, William A., Hinich, Melvin J., and Weber, Warren E. (1986), « The Regulatory Wedge between the Demand-Side and Supply-Since Aggregation Theoretic Monetary Aggregates, « *Journal of Econometrics*, forthcoming.

Barnett, William A., Offenbacher, Edward K., and Spindt, Paul A. (1981), « New Concepts of Aggregated Money », *Journal of Finance*, vol. 36, May, pp. 497-505.

Barnett, William A., Offenbacher, Edward K., and Spindt, Paul A. (1984), « The New Divisia Monetary Aggregates », *Journal of Political Economy*, vol. 92, n° 6, pp. 1049-1085.

Benhabib, Jess and Day, Richard H. (1980), « Erratic Accumulation », *Economic Letters*, 1980, n° 2, vol. 6, pp. 113-117.

Benhabib, Jess and Day, Richard H. (1981), « Rational Choice and Erratic Behavior », *Review of Economic Studies*, July, vol. 48, pp. 459-71.

Benhabib, Jess and Day, Richard H. (1982), « A Characterization of Erratic Dynamics in the Overlapping Generation Model », *Journal of Economic Dynamics and Control*, 4, pp. 27-55.

Brock, William (1986), « Distinguishing Random and Deterministic Systems : Abridged Version », *Journal of Economic Theory*, forthcoming.

Brock, William and Dechert, W.D. (1986), « Theorems on Distinguishing Deterministic and Random Systems », forthcoming in William Barnett, Ernst Berndt, and Halbert White (eds.), *Dynamic Econometric Modeling*, Proceedings of the Third Austin Symposium in Economics, Cambridge University Press, Cambridge.

Day, Richard (1982), « Irregular Growth Cycles », *American Economic Review*, 72, June, pp. 406-411.

Day, Richard (1983), « The Emergence of Chaos from Classical Economic Growth », *Quarterly Journal of Economics*, May, pp.201-213.

Day, Richard (1985), « Dynamical Systems Theory and Complicated Economic Behavior », *Environment and Planning B : Planning and Design*, vol. 12, pp. 55-64.

Diewert, Erwin W. (1976), « Exact and Superlative Index Numbers », *Journal of Econometrics*, vol. 4, May, pp. 115-145.

Fayyad, Salam (1986), « Monetary Asset Component Grouping and Aggregation : An Inquiry into the Definition of Money », Doctoral Dissertation, University of Texas at Austin, Austin, Texas.

Grandmont, J.M. (1985), « On Endogenous Competitive Business Cycles », *Econometrica*, September, pp. 995-1045.

Grassberger, P. and Procaccia, I. (1983a), « Measuring the Strangeness of Strange Attractors », *Physica 9D*, pp. 189-208.

Grassberger, P. and Procaccia, I. (1983b), « Characterization of Strange Atrractors », *Physical Review Letters*, vol. 50, n° 5, January.

Mandelbrot, B. (1977), *Fractals, Form, Chance, and Dimension*, Freeman, San Francisco.

Nicolis, C. and Nicolis, G. (1984), « Is There a Climatic Attractor ? », *Nature*, vol. 311, October, pp. 529-532.

Prigogine, Ilya (1980), *From Being to Becoming*, W. Freeman, New York.

Prigogine, Ilya and Stengers, Isabelle (1984), *Order out of Chaos : Man's New Dialogue with Nature*, Bantam, New York.

Takens, F. (1980), « Detecting Strange Attractors in Turbulence », in Rand, D. and Young, L. (Eds.). *Dynamical Systems and Turbulence*, Warwick 1980, Lecture Notes in Mathematics N° 898 : Berlin, Springer-Verlag, 366-382.

Wolf, A., Swift, J., Swinney, H., and Vastano, J. (1984), « Determining Lyapunov Exponents from a Time Series », Dept. of Physics, University of Texas at Austin, Austin, Texas 78712.

Aggregation, Actualities and Attractors

Joseph E. Earley
Department of Chemistry, Georgetown University, Washington

Professor William Barnett's paper in this volume illustrates a point that is (or ought to be) of interest to those that deal with the traditional philosophic areas of epistemology and ontology.

Humans survive through ability to recognise and deal with objects in the world. Many such objects appear so clear, natural and distinct that their recognition (if not their control) is unproblematical. Familiar technology gives rise to basic notions used to deal with complex problems. Ordinary rocks and solid building stones provide the conceptual models that appear as atoms, molecules and elementary particles in elaborate discourses of physics and chemistry. As usage becomes further extended, severe problems become apparent. Nowadays even strict « physicalists », such as W.V. Quine[1] recognise that simple elementary-particle notions no longer form the basis of an ontology that is adequate to deal with problems arising from contemporary science and culture. There are important actualities that cannot be dealt with using a simple atomistic approach, as understood in the recent past.

Michel Foucault[2] described cases in which ontological categories (what does exist, what counts as a proper subject for discourse) can be seen to change and develop in an evolutionary way, dependent on modifications in technology and in culture. Which categories of existence may be used in scientific discourse depends on complex linguistic and historical processes. Jacques Derrida[3], against Foucault, claims that webs of interconnectedness of things are so dense that there is no warrant in the nature of the world for adopting one ontology in preference to another. In his metaphysics of « natural complexes », Justus Buchler[4] developed an

approach that seems close to that of Derrida. In this view, any division of appearances (any ontology), is as good as any other. There is no objective criterion for deciding conflict between rival ontologies. Past scientific and technical progress has generally been made by persons who operated with realist[10] ontologies. It is, at best, an open question as to whether the quasi-nihilism of Derrida and Buchler can also be fruitful.

The results reported by Professor Barnett seem consistent with a suggestion that I made elsewhere[5] concerning an ontological approach that is coherent with contemporary science and also seems capable of supporting further development. In my view, the question as to whether a specific entity (or entities of particular classes, angels or apples, say) can properly be said to exist depends on how or whether the entity in question interacts with some other. (To use the word « observer » would be both exessively anthropocentric and needlessly restrictive.) Hans Gadamer[6] makes the point that a word is not a word (i.e., does not exist, in some sense) outside discourse, and discourse requires response. Lacking response, there is no discourse, absent discourse there is no word. (Kant makes a related point, on a non-linguistic basis, in his comment on the second antimony[7]).

Every interaction is intrinsically associated with characteristic spatial and temporal scales. The time-scale appropriate for interactions between quarks is quite different from that involved in the bending of a triatomic molecule, or the development of a new species of butterfly. Each of these real processes must be studied with techniques that take these diverse time-scales into account. The electron-microscope and the radio-telescope both discern coherent actualities, but neither is capable of focussing on the entities seen by the other. Existence, properly so designated, requires that the entity in question persist for the time-periods that are long with respect to the characteristic time-constants of external entities (observers, in some cases), that are in interaction with the entity in question.

One of the main intellectual achievments of our own time is the recognition that self-contained entities (rocks, building-stones, atoms) are inadequate as basic ontological models. Open dynamic systems must also be used as fundamental models, if any degree of general applicability is sought. If, as is argued above, some degree of persistence is required for every actuality, then the presence of an attractor (of whatever complexity) is also necessary. Due to an attractor, long-term behavior of a system may be coherently related with behavior at prior times. Presence of an attractor provides a warrant (necessary but not sufficent) for discrimination (by an appropriate external, or internal, other) of that open dynamic system as an integrated actuality, a unified whole.

Professor Barnett's paper can interpreted, on this basis, as providing some support for the approach of Foucault, as against that of Derrida and of Buchler. In the economic systems discussed, an attractor is apparent when certain measures of money-supply are employed, but no attractor can be detected when another measure of money-supply is used. The crucial difference is how (by what method, to what extent) aggregation[8] of elementary data is carried out. It would be a mistake to

regard this as an isolated example. Ambiguities of this sort are ubiquitous. Each of the elementary economic data referred to by Professor Barnett is a highly complex aggregate, if looked at from the point of view of the interactions discussed by microphysics. Conversely, each of these pieces of data is simple in the extreme, when compared with the basic data of historical research. This economic example illustrates the process that has occurred repeatedly in the identification of physical and social coherences that we now take granted. Coherence over time is not as apparent for complex dynamic systems as it is for bricks, stones and other straight-forward objects. Nevertheless, by patient search, it is often possible (as in the case discussed) to find a point of view from which the intrinsic coherence of the system (the attractor) can be discerned. The aggregation technique does not create the coherence, it makes that coherence apparent. The history of culture can be considered to be the record of a continual, and sporadically succesful, search for means by which the coherences of the world can be detected and comprehended.

It is important to recognize (as Derrida and Buchler decline to do) that the existence of an entity depends on the internal constitution (i.e., the attractor) of the complex dynamical system under attention, as well as on the means by which attention (or other interaction) is applied. Both need to be appropriate in order to provide a warrant for existence-claims. An existence-claim can fail because of lack of internal coherence of the candidate actuality, and also by lack of an interaction appropriate for the details of the coherence that is present.

Epistemology and ontology are inextricably connected, and are also evolving together. As Foucault points out, technical, cultural and linguistic developments lead to identifications of actualities not previously detected. Each such identification makes possible higher-level integrations of which the previously-noted entity may be a component. Myriads of interconnections, pointed out by Derrida, do not negate the existence of persistent coherences (each based on attractors of some sort) of which the world is composed. Such coherences ground the subtle ontologies by which we regulate our lives, by means of which the human species has flourished.

Whitehead[9] wrote, « The problem of evolution is the development of enduring harmonies of enduring shapes of value, which merge into higher attainments of things beyond themselves ». Our problem is evolutionary development of progressively more-adequate epistemologies (and associated ontologies) needed for continued human development.

REFERENCES

(1) W.V. QUINE, « *Theories and Things* », Harvard University Press, Cambridge, Ma., 1981.

(2) Michel FOUCAULT (Translated by A.M. Sheridan-Smith), « *The Archaeology of Knowledge* », Random House, NY, 1972.

Herman L. DREYFUS and Paul RABINOW, « *Michel Foucault* », University of Chicago Press, Chicago, Il., 1982.
(3) Jacques DERRIDA (translated by Alan Bass), « *Writing and Difference* », The University of Chicago Press, 1978 (Chapter Two « *Cogito and the History of Madness* »).
(4) Justus BUCHLER, « *Metaphysics of Natural Complexes* », Columbia University Press, NY, 1966.
(5) Joseph E. EARLEY, « *Self-Organization and Agency in Chemistry and in Process Philosophy* », Process Studies, 1981, 242-258.
(6) Hans G. GADAMER, « *Word and Time* », (a lecture), Washington, DC, 1985.
(7) Emmanuel KANT, « *Prolegomena to Any Future Metaphysics* » (Mehaffy-Carus translation, revised by L.W. Beck), Liberal Arts, NY, 150, p. 87 ff.
(8) J. VAN DAAL and G. MERKIES, « *Aggregation in Economic Research* », D. Reidel, Dordrecht, 1984.
(9) A.N. WHITEHEAD, « *Science in the Modern World* », Macmillan, NY, 1967, p. 94.
(10) P.M. CHURCHLAND and C.A. HOOKER, editors, « *Images of Science, Essays on Realism and Empiricism* », The University of Chicago Press, Chicago, Il., 1985.

Evolutionary Game Dynamics

Peter Schuster
Institut für Theoretische Chemie und Strahlenchemie der Universität Wien

Karl Sigmund
Institut für Mathematik der Universität Wien
IIASA, Laxenburg

Let us start with a « Gedankenexperiment » due to Richard Dawkins (1976), the so-called « Battle of the Sexes ». Imagine a species of birds, say, where the males are tempted to desert their mates, leaving them to « hold the baby ». The males can thus increase their offspring by founding several families, while the females will have to do all the rearing and caring for the brood. The obvious female countermove is to insist on a long engagement period before mating : it will be too late in the season, then, for males to start another matrimonial venture, and they will do well to stick around and share the parental investment.

We can describe this in terms of game theory. The females have two strategies : they can be coy or fast. The males have also two strategies : they can be philandering or faithful (where philandering means : not prepared to accept the long engagement period).

Suppose now that most males are philanderers. Females, then, will do well to be coy. If most females are coy, faithful males will fare better than philanderers, and therefore multiply. But if most males are faithful, there is no point for females to insist on the long engagement period, which after all costs them too. So more and more females will be fast. But this means that philanderers can thrive and spread in their turn.

The cyclic structure of this evolution hints at oscillations in the frequencies of the strategies, rather than at an equilibrium. Such an outcome

cannot be described by a static analysis : a dynamical viewpoint is called forth. Another lesson suggested by this slightly farcical example is that game theory can be an interesting tool for theoretical biology.

The first one to anticipate game theoretical arguments in biological context was R.A. Fischer (1930), the father of population genetics and mathematical statistics, in his discussion of the sex ratio. In most species, genetic control mechanisms lead to a sex ratio of one half, which is not obvious to explain. Why are there so many males around ? As every animal breeder knows, a considerably smaller amount would lead to a higher overall growth rate of the population.

Fisher's answer was quite ingenious. Even though the number of children is not affected by their sex ratio, the number of grandchildren is. If there is a majority of females around, it will pay, in terms of grandchildren, to produce males, and vice versa. Indeed, since a grandchild has one male and one female parent, it is best to invest in the rarer sex. Which sex ratio maximizes the number of grandchildren of an individual depends on the overall sex ratio in the population, and hence is a function of what the others are doing.

Evolutionary game theory got really under way in the early seventies, through the analysis of the « hawk-dove » game by John Maynard-Smith (1974). The problem was to explain conventional fighting in conflicts within a species. It has often been noted by ethologists that « all out » combats are rare, and that « gloved fist » types of encounters prevail. This is obviously to the good of the species, but needs an explanation in terms of Darwinian selection operating on the level of individuals.

Here it is : let « dove » denote the strategy which consists in sticking to conventional assessment of strenght, and « hawk » the strategy leading to escalated fights with risks of injury. A population of doves can obviously be invaded by hawks, since a hawk amongst doves will win all his fights. But a hawk amongst hawks lives on borrowed time. A population of hawks can be invaded by doves, which dont get all they want, but do survive. The success of the strategy decreases with its frequency, so that the likely outcome is an equilibrium. This equilibrium will be characterized by evolutionary stability rather than optimality of mean fitness.

From the « hawk-dove » game onwards, evolutionary game theory proceeded at an ever increasing pace. For a review of the first ten years, we refer to Maynard-Smith's (1982) book. Classical game theory, originated by J. v. Neumann and extensively developed since the forties, was meant to handle the rational analysis of human conflicts. It seems fair to say that many of the initial hopes for applications in economic or military domains have failed to materialize, however. Biological applications, which were initially not expected, seem a safer bet. They entail a shift in the meaning of two basic notions of game theory, namely « strategy » and « payoff ».

« Payoff » in human games is often difficult to quantify, since it is hard to evaluate money, social prestige, health and other factors on a single

utility scale. In biological context, an increase in Darwinian fitness is the only relevant quantity. «Strategy » in the context of war, or of parlour games, suggests a nicely calculated sequence of moves. This aspect of plotting and scheming was replaced by the notion of an innate type of fighting behaviour, and more generally by that of a genetically programmed trait. In fact, any phenotype whose fitness is frequency dependent may be viewed as a strategy.

This leads to a game theory without rationality (in Rapoports (1984) expression), which, as we shall see at the end of this paper, turns out to be quite fruitful for the analysis of human conduct too.

For expository reasons, let us stick to the very simplest type of games, consisting of pairwise encounters between individuals of a population with finitely many phenotypes (The « hawk-dove » game would be a case in point, but the « sex ratio » game not). Let us denote the phenotypes by E_1 to E_n and their frequencies by x_1 to x_n. The state $\underline{x}=(x_1, \ldots, x_n)$ of the population, then, is a point of the simplex S_n, since $x_i \geqslant 0$ and $x_1 + \ldots + x_n = 1$. The boundary of S_n consists of states for which some x_i is 0 i.e. some E_i is missing.

Let us denote by a_{ij} the payoff for an E_i-strategist matched up against an E_j-strategist. The matrix $A=(a_{ij})$ is called the payoff matrix. The average payoff for an E_i strategist within population $\underline{x}$ is

$$(A\underline{x})_i = x_1 a_{i1} + \ldots + x_n a_{in}$$

(since x_j is the probability that its opponent is an E_j-strategist). The average payoff for a member of a population in state y randomly matched against a member of a population in state x is

$$\underline{y}.A\underline{x} = y_1 (A\underline{x})_1 + \ldots + y_n (A\underline{x})_n$$

and thus the mean payoff within an $\underline{x}$-population is $\underline{x}.A\underline{x}$.
It is often convenient to interpret an $\underline{x}$ in S_n not as a population mixture, but as a mixed strategy consisting in playing E_i with frequency x_i.

The central notion of classical game theory is that of a Nash equilibrium. A point $p \epsilon S_n$ is said to be a Nash equilibrium if

$$\underline{x} \,.\, A\underline{p} \leqslant \underline{p}.A\underline{p} \qquad (1)$$

for all $\underline{x} \epsilon S_n$. This means that p is a best reply against itself. In a $\underline{p}$-population, it does not pay to deviate. A rational player has no reason not to play strategy $\underline{p}$. But the state of an animal population is not determined by rational forethought.
A state $\underline{p} \epsilon S_n$ is said to be evolutionarily stable if it is stabilized by selection, i.e. if it is a Nash equilibrium satisfying the supplementary condition that for $\underline{x} \neq \underline{p}$,

$$\underline{x}.A\underline{p} = \underline{p}.A\underline{p} \text{ implies } \underline{p}.A\underline{x} > \underline{x}.A\underline{x} \qquad (2)$$

(If $\underline{x}$ is an alternative best reply against p, then p does better against x than $\underline{x}$ itself). It is easy to show that p is evolutionarily stable if and only if no $x \neq p$ can invade the p-population, in the sense that x fares less well than $\underline{p}$ against the mixture $\underline{y}=\epsilon\underline{x}+(1-\epsilon)\underline{p}$ of a few $\underline{x}$ and many $\underline{p}$: $\underline{x}.A\underline{y} < \underline{p}.A\underline{y}$.

This notion is the central one of evolutionary game dynamics. For the « hawk-dove » example, if G denotes the increase in fitness obtained by winning the contest and C the cost of injury, the payoff for a dove against a hawk is 0, that of a hawk against a dove is G, while a dove against a dove (resp. a hawk against a hawk) wins with probability 1/2 and has G/2 (resp. G-C/2) as expected payoff. For C>G, one checks easily that the unique evolutionarily stable frequency of hawks is G/C. The higher the cost of injury (i.e. the better the species is armed) the fewer the hawks.

Similarly, Fishers argument shows that 1/2 is the unique sex ratio which is evolutionarily stable.

Evolutionary thinking is based on an underlying dynamics. Thus the frequency x_i of phenotype E_i depends on time t. The rate of increase $\dot{x}_i/x_i$ (where $\dot{x}_i = dx_i/dt$) is a measure for the evolutionary success of strategy E_i. So is the difference $(A\underline{x})_i - \underline{x}.A\underline{x}$ between the payoff for E_i and the mean payoff in the population. Since payoff is in terms of reproduction, this suggests equating the two expressions, and leads to the dynamics of Taylor and Jonker (1978)

$$\dot{x}_i = x_i \left((A\underline{x})_i - \underline{x}.A\underline{x} \right) \qquad (3)$$

This equation describes the evolution of the state $\underline{x}$ (in S_n). It is easy to see that the simplex S_n is invariant, as well as its faces. If x_i is 0 at some time, it is 0 forever, which reflects the fact that the model takes only account of the effects of selection, and ignores mutations which would be able to introduce a missing type. Moreover, the derivation of the dynamics is based on the assumption that the progeny of E_i is again of this type : such a « like begets like » argument does not take sexual inheritance mechanisms into account. This dynamics therefore is a rather drastic oversimplification of evolution ; but it reflects the game theoretic aspects very well.

The fixed points $\underline{x}$ of (3), i.e. those for which all x_i vanish, are just those $\underline{x} \in S_n$ for which there exists a c such that $(A\underline{x})_i = c$ whenever $x_i > 0$: all phenotypes present have the same payoff c (which is just $\underline{x}.A\underline{x}$). The fixed points can be easily found by linear algebra, therefore. We shall presently see that they relate well with the game equilibria.

Indeed a point $\underline{p} \in S_n$ is a Nash equilibrium if and only if it is a fixed point of (3) which is saturated. This condition means that $(A\underline{p})_i \leq \underline{p}.A\underline{p}$ whenever $p_i = 0$. (If $p_i = 0$ but $(A\underline{p})_i > \underline{p}.A\underline{p}$, selection would « call for » the introduction of the missing phenotype E_i which would have an edge over all those present).

A point $\underline{p} \in S_n$ is evolutionarily stable if and only if the function $\Sigma p_i \log x_i$ increases along all orbits near $\underline{p}$. Since the unique maximum of this function is at $\underline{p}$, this means that $\underline{p}$ is a fixed point of (3) which is an attractor in the sense that all nearby states converge to $\underline{p}$. However, not every fixed point is Nash, and not every attractor is evolutionarily stable. This means that the dynamic approach yields more than the static one, even if one restricts attention to equilibrium situations.

Moreover, game dynamics allows for asymptotic regimes far more complicated than equilibria. In particular, limit cycles, saddle connections and strange attractors are all « in the game », as soon as four or more phenotypes compete. Local fixed point analysis is not able to handle such situations : a global analysis is called for.

In the evolutionary context, the basic question is survival. Accordingly, the precise structure of the attractor is not always needed (and incidentally often impossibly hard to find) : the types which make it up and the basin of attraction are what count. A typical question, for example, is whether the system is permanent, i.e. whether there exists a threshold value $k>0$ such that $x_i(t)>k$ for all i and all sufficiently large times t whenever $x_i(0) > 0$ for all i. This means that there exists a compact set K in the interior of the state space in which all interior orbits end up, and this is related to the notion of stability in the sense of Lagrange (rather than of Ljapunov). The fate of the trajectories in K is not important : it does not matter whether they converge to an equilibrium, or oscillate endlessly in a regular or irregular fashion. What counts is that there will always be a sizeable frequency (at least k) of each phenotype around : small fluctuations will not be able to drive the state to the boundary, i.e. to wipe out some types. On the other hand, permanence means that the boundary is repelling, and hence that if a phenotype is initiatlly absent, then a mutation introducing it in a small quantity will spread : each phenotype can « invade ». In this emphasis on the boundary of the state space, the concept resembles the notion of resilience introduced by Holling in ecological context (see, e.g. Holling, 1973).

A sufficient condition for permanence, which has been used in many examples, is that the function $\Sigma\, p_i \log x_i$ (for suitable $p_i>0$) increases along orbits near the boundary, at least in the long run. This is the case if the system of inequalities

$$\sum_{x_i = 0} p_i ((Ax)_i - x.Ax) > 0$$

admits a solution $p>0$ for all fixed points x on the boundary of the state space S_n (Jansen, 1986).

A necessary condition for permanence, on the other hand, is that there exists a unique fixed point $\hat{x}$ in the interior of S_n. This $\hat{x}$ need not be an attractor, but is the limit of all time averages along orbits in the interior. If there is no interior fixed point, then all orbits converge to the boundary, i.e. some type will become extinct.

There is a strong similarity between models in game dynamics and ecology. In fact, equation (3) is equivalent to the Lotka Volterra equation

$$\dot{y}_i = y_i (r_i + \Sigma b_{ij} y_j)$$

on R^{n-1}, where the y_i denote the population densities in an ecosystem. Such a system is permanent if the set of states y for which no species increases is disjoint from the convex hull of all boundary fixed points

(at least under the obvious « finite planet » — condition that all orbits are uniformly bounded).

Another basic equation of biomathematics, the so-called selection equation of population genetics (studied by Fisher, Haldane and Wright) is just (3) with the symmetry requirement $a_{ij}=a_{ji}$. It describes the evolution of allele frequencies x_i in a gene pool, with a_{ij} as fitness of the genotype consisting of the alleles i and j. In this case, the average fitness $x.Ax$ will always increase (this is Fisher's Fundamental Theorem of Natural Selection). It leads to the familiar picture of evolution as an uphill movement in the « fitness landscape ». Kimura's Maximum Principle states that moreover the movement follows the line of steepest ascent, so that fitness increases at a maximal rate. Indeed, the introduction of a non-Euclidean metric in S_n, discovered by Shahshahani, allows to view (3) as a gradient system with potential $x.Ax$, provided $a_{ij}=a_{ji}$ (see Akin, 1979). This symmetry condition means, in the game theoretic context, that both players share the outcome fairly : they are partners, not rivals. The absence of a « clash of interest » makes optimization work. More generally, a maximum principle holds if and only if

$$a_{ij}+a_{jk}+a_{ki}=a_{ik}+a_{kj}+a_{ji}$$

holds for all i, j and k. The sex ratio game, for example, satisfies this condition (in a framework where the continuum of sex ratios m between 0 and 1 describes all possible phenotypes) and hence is a gradient. The potential is m(1−m), which is maximized at 1/2.

In more complex situations, it is unlikely that natural selection maximizes fitness. Genetic mechanisms involving recombination at several gene loci, and strategic constraints based on frequency dependence can lead to limit cycles or chaotic attractors, and hence preclude a description in terms of optimization. Game theory offers valuable tools to model such conflicts between opposing tendencies.

The game dynamic approach can be extended in many directions. It can be used to model conflicts between two different populations (rather than within one population) : for example, contests between owners or intruders, arms races between hosts and parasites, or the tug of war between males and females as described in the introductory « battle of the sexes ». It can also apply to games with infinitely many strategies, like the sex ratio game or the « war of attrition » where there is a continuum of options for escalation. It can also take Mendelian structure into account (see e.g. Eshel, 1982). The simplest genetic model of the « battle of the sexes », for example, leads to a limit cycle. Getting rid of the assumption of asexual inheritance is of course highly desirable, but the resulting combination of genetic and strategic viewpoints can become rather intractable. Under some set of genetic assumptions, the analysis may lead to the same outcome as the purely game theoretic considerations ; under other assumptions, to a different one. Since the present knowledge on the effect of genotype upon phenotype is rather scant, any such imbedding of game dynamics in frequency dependent population genetics is bound to be somewhat arbitrary.

But strategies can spread, not only through inheritance, but also through learning (see Maynard-Smith, 1984). In fact, like other branches of sociobiology, evolutionary games offer some hope for applications to human conduct. This is not to suggest that matrimonial disputes can be reduces to the « battle of the sexes » or that the « war of attrition » offers advice for disarmament talks. But evolutionary thinking can shed light on game theoretical models of human affairs. Let us mention the evolution of cooperation, which occurs in animal and human populations. Part of the problem is encapsulated in the famous « prisoners dilemma ». Two prisoners are accused of a joint crime. They are kept in separate cells, and asked to confess. If both confess, they will get sentenced to seven years confinement. If both deny, the judge of instruction will keep them jailed for one year, pending investigation of the charge. If only one confesses, he will become witness of the prosecution and get free, while the other one gets imprisoned for ten years. Quite obviously, the best strategy for each prisoner is to confess — no matter what the other one is doing. But the other one will of course also confess, which leads to an outcome which is bad for both.

Now if games of a similar structure are repeated an arbitrary number of times, one would expect that the players learn to cooperate. This question, which is a favourite subject of game theorists, got an exciting new stimulus when Axelrod and Hamilton (1982) set it in an evolutionary context. In particular, Axelrod (1984) organized computer tournaments between programs for repeated « prisoners dilemma ». The simplest strategy turned out to be the winner : Tit for Tat, which consists in a cooperative first move followed by whatever the other player did in the previous mode. This seems a satisfactory outcome, but many open questions remain. In particular, Tit for Tat is not evolutionarily stable : it can be invaded, through random fluctuations, by the strategy « always cooperate », which is an alternative best reply against Tit for Tat. But once there is a sufficient amount of unconditional cooperation in the population, « defectors » can take an advantage of this, and spread in their turn. Furthermore, Tit for Tat is highly vulnerable against errors : one mistake leads to an endless series of defections. Recent work by Axelrod (1986) shows, in fact, that simplified genetic mechanisms, when tested in such context, are able to evolve strategies which may be superior to Tit for Tat.

This whole branche of game dynamics is developing very fast. Extensive computer simulations are currently the main tool, but Kurka (1986) has recently investigated the qualitative theory of such « supergames » using dynamics similar to (3).

But while the contribution of game dynamics to the evolution of human affairs is still at an early stage, its impact on biological evolution is undisputable. We have mentioned that (3) not only describes changes in frequencies for behavioural phenotypes, but also leads to differential equations which are classics in population genetics and mathematical ecology. Such equations play a central role in prebiotic evolution as well. In particular, they occur in the theory of Eigen and Schuster (1979) on macromolecular selforganization. The variables here are the relative concentra-

tions of selfreproducing RNA-molecules, and the frequency dependence is caused by catalytic effects. Thus « hypercycles », i.e. closed catalytic feedback loops of selfreplicating molecules can be described in terms of (3), using a permutation matrix for A.

That (3) applies to behavioural traits, molecules, individuals and genes is not so astonishing. The underlying structure in each case is the action of selection upon some « replicator » (See Schuster and Sigmund, 1983 and Sigmund, 1985). But the relatedness of the corresponding models does not just reduce to a common formalism. It helps to think in terms of « games » between organisms, genes or molecules.

The theme of this conference is « Laws of Nature and Human Conduct ». Evolutionary game theory uses the panoply of game theory — which was developed to analyze human conduct — towards a better understanding of the laws of natural selection.

REFERENCES

A. AKIN (1979), *The geometry of population genetics*, Lecture notes in Biomathematics, Springer Verlag.

R. AXELROD (1984), *The evolution of cooperation*, Basic Books.

R. AXELROD (1986), to appear.

R. AXELROD and W.D. HAMILTON (1981), *The evolution of cooperation*, Science *211*, 1390-6.

R. DAWKINS (1976), *The selfish gene*, Oxford University Press.

M. EIGEN and P. SCHUSTER (1979), *The hypercycle : A principle of natural selforganization*, Springer Verlag Berlin Heidelberg New York.

E. ESHEL (1982), *Evolutionarily stable strategies and natural selection in Mendelian populations*, Theor. Pop. Biol. *21* 204-17.

R.A. FISHER (1930), *The genetic theory of natural selection*, Oxford : Clarendon Press.

C.S. HOLLING (1973), *Resilience and stability of ecological systems*, Annual Review of Ecology and Systematics *4*, 1-23.

W. JANSEN (1986), *A permanence theorem for replicator equations*, to appear.

P. KURKA (1986), *Game dynamics and evolutionary transitions*, to appear in Biological Cybernetics.

J. MAYNARD-SMITH (1974), *The theory of games and the evolution of animal conflicts*, J. Theor. Biol. *47*, 209-231.

J. MAYNARD-SMITH (1982), *Evolutionary game theory*, Cambridge UP.

J. MAYNARD-SMITH (1984), *Game theory and the evolution of behaviour*, The Behavioural and Brain Sciences *7*, 95-125.

A. RAPOPORT (1984), *Game theory without rationality*, The Behavioural and Brain Sciences *7*, 114-115.

K. SIGMUND (1985), A survey on replicator dynamics, in : Casti, J. and Karlquist, A. : Complexity, Language and Life : Mathematical Approaches, Springer Biomathematics Vol. 16.

P. SCHUSTER and K. SIGMUND (1983), Replicator dynamics, J. Theor. Biol. *100* 533-538.

P. TAYLOR and L. JONKER (1978), *Evolutionarily stable strategies and game dynamics*, Math. Biosciences *40*, 145-156.

The Information Economy and the Restructuring of Human Organizations

Jean Voge
Ingénieur Général des Télécommunications,
Président de l'IDATE

Information, communication and audiovisual technologies are subjects of almost magical fascination. Journalists, businessmen, politicians and ordinary citizens expect them to work countless wonders. A confidence that is only slightly diminished by doubts about intrusion of the media, advertising and computers in private life, or about future unemployment of clerical workers, sometimes with veiled references to Big Brother. In the United States, the world's most economically advanced nation, there is already an « information society » — one that is a model for the other industrialized countries. And yet, without sacrilege, one is forced to note a crisis in this information society. A crisis due simply to growing pains, or a sign of congenital malformation ?

The question cannot be evaded.

1. THE INFORMATION SOCIETY

One job out of two in the United States is now concerned with the production, handling or dissemination of information in one form or another, compared with just one out of eight in 1900 and one out three in 1950. The same proportions are observed if we consider instead the added value of information and communication activities as a proportion of national product (Porat, 1977 ; Jonscher, 1983-84 ; Voge, 1983-84). In other words, there is a proliferation of information, but to what purpose and in answer to what needs ?

Perhaps the first sort of information that comes to mind is that tied to cultural and leisure activities. However, its share in total household

consumption, through expenditures on books, newspapers and periodicals, musical instruments, radio and television products, records and cassettes, theater and cinema shows, and so on, declined from 2,3 % to 1,7 % of national product between 1947 and 1983. This is sufficient proof that the information society is anything but a society of art and culture.

The final consumption of more utilitarian information, through expenditures on medical or legal assistance, financial services, private education, and so on, doubled over the same period from 3,4 % to 6,6 % of national product.

This said, what is the remaining, predominant use and purpose of information production ? A total production that seems to never stop growing, accounting for 13 % of employment and national product in 1900, then 32 % in 1947, 45 % in 1972, and 47 % in 1983.

The answer is simple and clear. The main purpose of all information is the organization and regulation of the production and distribution of physical or material goods and services. Information is generated, for example, in training or educating employees (and consumers), research, consultancy, transactions, advertising, bureaucracy, and protection of society, the population and the environment, against multiple risks and threats. After taking into account this intermediate consumption, the final demand of material goods and services has remained virtually stable, decreasing from 94 % of national product in 1947 to 92 % in 1983. The information society cannot therefore be looked on either as a post-industrial society.

However, the producers of material goods and services, in the farming, manufacturing, transport, hotel, domestic services, health and other sectors, are consuming growing amounts of information. As a consequence, they must divert a growing fraction of their sales revenues and gross added values (labor and capital incomes, including profits) towards information producers and distributors inside and, above all, outside their companies : business, professional and commercial services, advertising agencies, banks, insurance companies, administrations of all sorts, etc. Information-related overheads for the organization and regulation of the material goods and services sector thus rose from 30 % to 45 % of national product between 1947 and 1983. At the same time, the residual added value available in this sector for remunerating, expanding, and modernizing its output declined from 68 % to 53 % of the same national product.

This explosive growth in information costs is a direct consequence of a very important phenomena. The apparent productivity of information or « white-collar » workers, equal to the national product divided by their number, has remained almost constant, at about 32.000 (1972) dollars, since the beginning of the century, whereas the productivity of « blue-collar » and agricultural workers has increased seven-fold. The poor productivity of white-collar workers is rather surprising at first sight, considering the spectacular progress in information and communication technologies over the past twenty years or so. Indeed, telecom-

munications is the only area of the economy in which annual productivity gains since 1973 have remained roughly equal to those achieved during the 1950s and 1960s (4,5 %). The trouble lies in the practical use of these technologies. It takes four on average and sometimes a dozen business telephone calls before at last reaching the right person. Administrative data often lay around for weeks or months on end between two runs through a computer. And it takes longer and longer to retrieve the really usefull facts buried in a proliferating mass of parasitic information. Thus, only 7 % of the information produced in Japan in 1982 was actually used, compared to 11 % ten years earlier.

The reasons for this unfortunate mismatch between modern techniques and their practical uses are examined further on. For the moment, it can simply be noted that the stagnation of information productivity also means that the overall productivity of the economy (national product divided by total number of workers) varies in proportion to the percentage share of the information sector in the national product and total employment. Both overall productivity and this share were multiplied by 3,7 between 1900 and 1983. It is easy to deduce that the added value of information, which corresponds essentially to organizational costs, has on average and in constant dollar terms per U.S. worker grown twice as quickly as the overall productivity, increasing fourteen-fold in 83 years.

C.N. Parkinson was one of the first to reveal this « square law » phenomenon in large administrations. But each of us has painfully observed its inflationary effects on our income taxes, social security payments, and other unavoidable payments. The consequences are even more serious for the production apparatus of the material sector. The remaining added value after the corresponding « quadratic deduction decreases not only in terms of national product share. If it is likewise expressed in constant dollars per worker, a simple calculation shows that it virtually stops rising once the information share exceeds 40 % (a level reached around 1965), and starts decreasing above 50 %. The productivity gains from modernization of facilities are then insufficient to offset the accompanying increase in training, research, and administrative costs. It was therefore almost inevitable that productivity gains in the United States would decline. More than 3 % per year in the early 1960_s, they were only half as great between 1965 and 1973, dropping virtually to zero between 1973 and 1982 and in 1985, while the information share rose asymptopically towards 50 %.

This very macroeconomie approach to analysis of the crisis can be applied separately to the various branches of the economy. It is then possible to distinguish two major sectors : 1) an agricultural and industrial sector grouping farming, fishing, mining, construction, manufacturing, transportation, communications and water, gas and electrie power supply ; and 2) a service sector, also encompassing trade, banking, insurance, real estate and government. The first of these sectors is predominantly material and employs a majority of blue-collar workers, while the second sector is predominantly informational or transactional and has a majority of white-collar workers. In the average production of a

U.S. worker, the added value of the agricultural and industrial sector remained constant between 1965 and 1980, even though the productivity of this sector increased by 27 %. For manufacturing industry alone, and over the same period, the added value per worker dropped by 12 % whereas productivity rose by 32 %. All the added value gains resulting from the productivity gains were therefore transferred to the services sector, where the added value grew by 23 % despite an increase of only 11 % in productivity. The transfer has been particularly severe for the manufacturing industry, where the ratio of added value to investments in new plants and equipment decreased by 43 % between 1960 and 1980, while the same ratio increased by 15 % for services and commerce.

It is hardly surprising under these conditions that, according to Fortune, the profits of industrial companies declined in constant dollar terms by 20 % between 1973 and 1983, whereas profits grew by 23 % for financial companies, by 29 % for commercial banks, and by 53 % for retail traders (Labich, 1984).

All the statistics and trends presented above would be perfectly convergent were it not for the recent economic « recovery » in the United States. Between 1982 and 1984, the productivity of manufacturing industry jumped by 12 %, leading to an average 5 % productivity gain for the economy as a whole, while its added value increased by nearly 16 %. However, close study shows that these results are unfortunately due only to the indirect transfer to the federal budget of part of the organizational costs tied to the material production sector, particularly to industry. The size of the federal budget deficit and its effects on foreign exchange markets, the national debt situation, and the trade balance soon brought this advance to a halt, in the middle of 1984[1]. But it will at least have proven that industry is ready to move forward, and that technology can be the fertilizer for a new growth in productivity, provided it is possible to achieve a lasting and real decrease in the costs of information and communication. In other words, provided it is possible to at last increase the productivity of information. Otherwise, the information society, far from representing the predominance of mind over matter and the triumph of electronics, will remain, as today, a society of never-ending crisis, zero growth in overall productivity, and stagnation or (for many people) a decline in standards of living.

2. *FROM LINEAR TO NON-LINEAR — FROM PYRAMIDS TO INTERACTIVE NETWORKS*

So, how can we increase the productivity of information ? If the technical progress alone does not supply the answer, a solution may be sought in the « science of complexity », which studies the organizational structures of complex systems : their architecture, their birth, their operational economy. This economy is not much dependent to the nature of interactions between the system's components[2]. So the science of complexity is able to integrate (Voge, 1985) theoretical contributions from thermodynamics (Prigogine, 1979), the information and communication networks theory (Shannon, 1950 ; Marcus, 1977) and the science of organizations (Simon, 1962).

Many physical, chemical, biological and technical (man-made) systems or sub-systems reveal linear modes of operation. The flows of matter or current are then related to the applied forces by means of linear equations. The « production of entropy », i.e. the power dissipated (usually as heat) by the irreversible processes of interaction between the system's constituent elements, varies as a square-law function of the forces or flows. And lastly, the « useful » power supplied by such a system when it is driven by an external source of energy (e.g. a motor that supplies mechanical work) is very often a maximum when the production of entropy is equal to 50 % of the consumed power (Odum, 1955).

There are obvious points of similarity between these results and those which hold in the U.S. information economics. The flow of production (national product) is linearly proportional to the « forces of work », the white-collar employement and, as is well known to economists, the physical capital placed at the disposal of blue-collar workers. The production (added value) of information per U.S. worker varies as the square of the global production per worker (productivity), while the useful production (added value of material production) per worker reaches a maximum when the production of information attains 50 % of the global production. There is a mass of publications on the very close relationships between entropy and information. The production of entropy and the production of information both represent, in terms of power or monetary value, the organizational costs corresponding to interactive exchanges, i.e. to all forms of communication within the system.

In thermodynamics, it can be shown that a system is linear in « near-equilibrium » conditions, when the forces and flows are small. However, linear operation is possible with forces and flows of any amplitude if the structure of a system is rigid, as with numerous mechanical devices such as motors, or if constraints exist which prevent any evolution. This is the case for the pyramidical, tree-like structures with centralized control encountered in the distribution networks of the major media, and particularly in conventional cable television networks. Communication is then essentially one-way. Home receivers are linked to production centers by a chain of distribution which resembles an upside-down tree. Signals are transmitted downwards over an expanding network of branches and stems generated at intermediate nodes such as local programming centers, headends, and distribution hubs. These nodes are simple branching points and leave no choice in signal routing. Information flowing downards is sent to all the tree's leafs (home receivers), while any information in the opposite direction can take only the single, hierarchical path. More generally, pyramids of the sort are found in all today's human organizations, whether political or economic. And we all know from personal experience that the higher levels of these pyramids are communication bottlenecks. Information flows easily down the pyramids, but getting a message to the top can be an exhausting experience !

If, instead, our ideal is interactive dialoguing, then meshed or cellular structures of the sort familiar to telecommunications engineers are the best, providing a multiplicity of alternative paths between any two points. The nodes of a meshed network, for example switching

exchanges in the telephone network, are like complex road junctions. For each incoming communication path, a node of the sort offers acess to many outgoing paths corresponding to different routes through the network. Instead of being confined to the privileged, upper reaches of a system, as in pyramidical structures, the control over information transmittal is decentralized through all levels of organization.

Comparison of these two classes of structure and of their economics lies at the heart of the theory of communications. If a telephone network had a pyramidical structure, the number of switchable contacts (such as relays) needed to establish a call between a subscriber and any other with no risk of blocking would increase in linear proportion to the total number of subscribers. With a cellular or meshed structure, the number of contacts can be less, increasing at best in proportion to the logarithm of the number of subscribers, giving a non-linear relationship. The network's operating costs increase as the square of the number of calls in the first case, but the relationship is only slightly worse than linear in the second case. This explains why telephone networks are highly meshed and why futur fifth-generation computers will have meshed internal structures.

The same is not true for human organizations, in which the pyramids are solid and, as may therefore be expected, economics are of the « linear » variety and costs are governed by the square-law relationship. It is also therefore hardly surprising that telecommunication is the only branch of activity in which productivity growth has not been slashed by the crisis, because use is made of meshed networks. In all other branches, the pyramids of human organization and communication place limits on productivity growth.

This said, how can we effect the change from pyramidical to cellular organizations ? Thermodynamics shows that if a system moves sufficiently far from equilibrium with increasing power, and provided no constraints are placed on its evolution, beyond a certain threshold it can spontaneously restructure itself through a series of changes and points of « bifurcation », which recall somewhat the purpose of telephone exchange relay contacts. Under conditions well away from equilibrium, the thermodynamics become non-linear and its effects are those of self-organization. Nature has made excellent use of this fact and a large majority of physical and biological systems have a cellular structure, the spontaneous creation of which can be outlined as follows (Eigen, 1981). Within a natural system, there is a tendency towards spontaneous association of its constituent elements as stable groups, analogous to the formation of crystals in a cooling liquid. The resulting groups enter into competition, as illustrated by living species in the Darwinian model of natural selection. This competition and selection would eliminate all but the dominant groups — the most stable, best adapted or most prolific — if other groups in order to survive did not acquire different qualities or particularities. Such differentiation transforms the competition into cooperation, which gives rise to new groupings of a higher order — in turn leading to new competition and selection. Nature has thus progres-

sively built up a hierarchy of groups of increasing size and complexity : from elementary particles to molecular structures and galaxies ; from biological macromolecules to cells and Homo Sapiens.

Selection and grouping (competition and cooperation) « are but the two aspects or connected parts of one and the same phenomenon » (Teilhard de Chardin, 1955). And this duality is the foundation for the entire architecture of natural systems. To take an example, thought processes involve a series of analyses (selections) and syntheses (groupings) at increasing levels of conceptualization. This Cartesian philosophy is materialized by the specialization of the two halfs of the brain, with the left side more analytic and the right side more synthetic.

Functional economy is not the only advantage of natural structures. By minimizing the production of entropy, they slow as much as possible the inevitable degradation imposed on the universe and all its creatures by the second law of thermodynamics. The same is not alas true for human organizations, the pyramidical structures of which place both the ecology and the economy under great threat. And for those who see hierarchical pyramids as a protection against the risks of instability, vulnerability and even chaos in systems operating far from equilibrium, and hence with small contraints, we can draw a last remark. In a cellular system, each subsystem or level of organization, from local to global, is self-regulated by autonomous feedback loops. This dynamic stabilization is more efficient and much more flexible than centralized, quasi-static control. As far as vulnerability is concerned, recent history shows abundantly that even the most powerful army or police force is often helpless when faced with a cellular organization of guerilla or terrorist forces.

In what precedes, we have attempted to show, using only the most evident proofs, that the disease threatening the information society is a consequence of its poor organizational structure. Having established this diagnosis, the cure is easy in the theory. Today's centralized pyramids must be replaced by cellular, decentralized networks providing multiple paths of communication. Taylorism with its linear and sequential model for production, distribution and information processing must be abandoned in favor of new, convivial schemes of work organization, dominated by interactive loops and in which serial and parallel processing are closely intermingled. The aim should no longer be simply to translate the ideas of leaders into the action of employees or citizens, but instead to mobilize the intelligence of all. In short, a structural change is necessary to create a new « *communication society* » giving systematic priority to interactive dialoguing over one-way communication.

To achieve this, it may be concluded that a maximum reliance should be made on the social and economic forces at work under the pressure of the economic crisis. The Invisible Hand is in the first line of the battle to dismantle the barriers to progress, mobilizing under the ecumenical flag of self-organization both the believers in a free market and those who dream of self-management or conviviality.

NOTE

(1) The U.S. productivity decreased by 0,2 % in 1985.

(2) The telephone network economy is independent to the nature of the calls.

BIBLIOGRAPHY

M. EIGEN et autres, 1981, Scientific American, 244, 78-96.

C. JONSCHER, 1983, Information Economics and Policy, 1, 13-35.

C. JONSCHER, 1984, Bulletin de l'IDATE, 16, 13-29.

K. LABICH, 11 juin 1984, Fortune, 108-109.

M.J. MARCUS, 1977, Proceedings of the Institute of Electrical and Electronic Engineers, 65, 1263-1271.

H. ODUM et R.C. PINKERTON, 1955, American Scientist, 43, 331-343.

M. PORAT, 1977, The Information Economy, Special Publication OT-12, Department of Commerce, volumes 1 et 8.

I. PRIGOGINE et I. STENGERS, 1979, La Nouvelle Alliance, chapitres V et VI, Gallimard, Paris.

C. SHANNON, Juillet 1950, Bell System Technical Journal, 29, 3, 343-349.

H.A. SIMON, 1962, The Architecture of Complexity. Proceedings of the American Philosophical Society, 106, 467-482.

P. TEILHARD DE CHARDIN, 1955, Le Phénomène Humain, Le Seuil, Paris.

J. VOGE, 1983, Information Economics and Policy, 2, 97-114.

J. VOGE, 1984, Bulletin de l'IDATE, 16, 53-96.

J. VOGE, 1985, Management of Complexity, in The Science and Praxis of Complexity, 298-311, The United Nations University.

The Physical Input to Economics

John L.R. Proops
Department of Economics and Management Science
University of Keele — Staffordshire, UK

PHYSICAL SCIENCE

Physical science can be defined as the study of the properties of matter and energy ; physical science is about the world of « things », how they relate to each other, the influences that can be exerted upon them, how they can be transformed. Physical science is probably the most successful intellectual endeavour ever undertaken by man. Over the past 3.000 years, and particularly the past 300 years, enormous improvements have been made in man's understanding of the physical world. In rapid succession substantial advances have been made in areas such as mechanics, optics, thermodynamics, electromagnetism, atomic theory, relativity, quantum theory, and cosmology.

The problem of why physical science has been so spectacularly successful has been an area of debate for many years without a definite answer emerging. However, we might note that physical science has been most successful where relatively straightforward mathematics can be applied. Indeed, the relationship between physical science and mathematics has been one of fruitful symbiosis for many centuries.

A further point to note about physical science is that it has been increasingly possible to embody scientific knowledge to produce valued artefacts, particularly over the past century. In a normal home the impact of modern physical science is everywhere visible. The radio and television, the hi-fi, personal computer, refrigerator, motor car, etc. The list is endless. We are continually reminded that late twentieth century life is to a great extent a product of our knowledge in the physical sciences.

ECONOMICS

I shall not attempt to offer a definition of economics, as I feel all those usually offered are lacking in one or another way. This in itself is significant when compared with the unambiguity and near universal acceptance of the usual definition of physical science. About economics, though, one might say that it is about « people », their motives, behaviour and institutions. In particular it is often felt to be about production, innovation and exchange of valued commodities. Now as has already been noted, the goods and services at the centre of the economic process are generally products of modern knowledge in physical science, and to an increasing extent the process of product innovation is one of finding new applications in production for such knowledge. Thus economics takes as its area of study processes which are not only about human choice and human institutions, but more and more determined by physical science and its advancement and application.

Economics is generally reckoned to be a much younger discipline than physical science, with modern economics being taken to date from Adam Smith or thereabouts. Economics has also, unfortunately, been much less successful than physical science ; after 200 years of study we still have no single accepted theory of the business cycle ! An obvious possible explanation for this relative lack of success is that the object of study for economists is much more complicated than that of physical scientists. In particular the apparently universal constancies of physical science have no parallel in economics. Change is all, with any constancies being statistical in nature, and even then of short duration. Faced with such enormous difficulties modern economics has seemed to respond by seeking to use the tried and tested conceptual framework of physical science. Relatively simple mathematics has been used with models that are generally apparently mechanical in their inspiration, with simple causes giving rise to simple effects. In the rest of this paper I shall explore how economics has used physical science, and assess how far modern work in thermodynamics suggests a new world view which may have great consequences for the practice of economics.

ECONOMICS AND PHYSICAL ANALOGY

In a recent paper MIROWSKI (1984) has argued that the originators of modern neo-classical economics made explicit use of analogy with nineteenth century mechanics. In particular he suggest that the dominance of the Hamiltonian technique in mechanics gave rise directly to the emphasis upon optimisation in what we would now call microeconomics. Although Mirowski may be overstating the correspondence between the techniques of Hamiltonian mechanics and neoclassical theory, the writings of Edgeworth, Jevons, Pareto and Walras all clearly demonstrate the explicit use of analogy with physical science. For example, Jevons explicitly wrote of the « mechanics of utility » and Edgeworth leaned directly on Hamilton's work. The use of analogy did not stop with the formulation of neoclassical theory. The nineteenth century also saw the rise of thermodynamics and statistical mechanics, and these methods too were seized upon by economists, with perhaps the erliest being FISHER (1882), the most distinguished SAMUELSON (1948),

and the most recent BRYANT (1982). However, thermodynamics seems to have been less influential than mechanics, and the use of analogy here have been mainly to supplement neoclassical theory rather than attempt to supplant it. (For a fuller discussion of the uses of analogy see PROOPS, 1985)

THE PHYSICAL FRAMEWORK OF ECONOMIC REASONING

Such direct analogy between physical science and economics can be readily interpreted as the borrowing of conceptual « tools » from a mature discipline by a less mature discipline (and currently sociology is similarly borrowing such conceptual tools from economics). However, as I hinted in the introduction, the economists' debt to physical science is much wider than this. Any scholarly activity requires not only conceptual tools but also a conceptual framework, and that adopted by economics seems to be largely the framework of Newtonian mechanics.

In elementary mechanics when bodies interact they do so in a way that is strictly determined, repeatable and reversible. The solution to such a simple mechanical system would be a deterministic set of equations of motion, with the time variable playing an exactly similar role to a spatial variable. Reversibility in space and reversibility in time have the same status in such a system. This symmetry of the status of the past and the future is readily recognised in modern general equilibrium analysis, and is equally recognised as inappropriate to economics. The essence of the past is that it is known, or knowable, to some degree, while the future is unknown and unknowable except by waiting till it becomes the past. Time is irreversible for economics.

It is apparent that the conceptual approach that treats the economy as a simple Newtonian mechanical system is severely flawed, e.g. Samuelson has noted « In theoretical economics there is no « irreversibility » concept ». Furthermore, this Newtonian mechanical framework has now been largely superseded in physical science through the recognition that even relatively simple mechanical systems become « unguessable » in their evolution if it is accepted that the system's initial conditions can be specified only with a certain degree of accuracy, and not absolutely precisely (see e.g. KADANOFF, 1983). This explicit irreversibility of time was also recognised by the rise of classical thermodynamics, where the concept of entropy was introduced to allow the past of an irreversible process to be explicitly dsistinguished from its future.

The Second Law of Thermodynamics, the Entropy Law, has been called « the Arrow of Time ». However it is not the unique arrow of time, as most activities we observe about us involve not only « running down » but also « building up ». Evolution to organisation acts as a second « Arrow of Time ». For example, the growth of a tree involves the transformation of low entropy sunlight into high entropy waste heat, so it exhibits Second Law temporal irreversibility. But the factor we mostly note about the growth of a tree is the increase in structuring and organisation as it develops (for a fuller discussion see LAYZER, 1976). Time irreversibility also has an evolutionary aspect to it, and until recently physical science, as opposed to biological science, was silent upon such

phenomena. However over the past two decades enormous strides have been taken in arriving at a physical understanding of such evolutionary phenomena. These advances have mainly been inspired by the work of Ilya Prigogine and the « Brussels School » of thermodynamics (see e.g. NICOLIS and PRIGOGINE, 1977).

Classical thermodynamics (perhaps better known as thermostatics) deals with the properties of systems of matter and energy at thermodynamic equilibrium, while Prigogine and his co-workers have concerned themselves with the properties of such systems when they are far from equilibrium. Now at or near thermodynamic equilibrium the Second Law reigns supreme, the behaviour of systems being describable in terms of entropy or one of the related « free energies ». However when systems are explored far from equilibrium then evolution may become the dominant theme, with order spontaneously arising out of disorder through successive bifurcations. The classic example of this is the appearance of convection cells in liquids when the temperature gradient across the liquid is increased. Convection cells are macroscopic phenomena which exhibit ordering and structure and their appearance contradicts the notion that the Second Law offers a unique « arrow of time » (for an application of these ideas to economic structuring see PROOPS, 1983).

We see that irreversible time in physical science is taking a more central role. How is time treated in economics ? Modern general equilibrium analysis in economics takes three types of approach to the problem of time. The first, and simplest, way of dealing with time is to ignore it completely, restricting analysis to statics and comparative statics. I shall call this the « atemporal » approach.

The second method uses the approach of elementary mechanics, allowing economic variables to evolve over time, but where time appears as an essentially reversible variable, i.e. perfect foresight is assumed. I shall call this the « reversible temporal » approach.

Third, the status of the future is explicitly recognised as different from that of the past, but it is assumed to be « guessable » from knowledge of the past. This latter approach involves the explicit modelling of the formation of expectations, and at first sight appears to treat time as the irreversible factor we know it to be. However, time is only irreversible in a very limited sense in such models, for surely the essence of the future is that not only that it is unknown, but in general, aspects of the future are UNEXPECTED. Economies change in ways that, in retrospect are comprehensible and natural, but in prospect are surprising. As this third approach attempts to treat time as irreversible, but only in a limited way, I shall term it the « quasi-irreversible » temporal approach.

However, even modern mechanics and classical thermodynamics treats time in a quasi-irreversible way. Although the future of, say, an individual molecule in a gas is unknowable except in the broadest terms, these theories do allow a precise specification of the properties of the gas as a whole. Only when the possibility for the evolution of far from thermodynamic equilibrium systems to unexpected forms is included does the

true « unguessable » irreversibility of time become accounted for. I shall term this aspect of time « absolutely irreversible ».

Now in this respect a system at thermodynamic equilibrium is unlike an economy because, over time, the nature of the interactions between molecules is assumed not to alter, while for an economy the interactions within an economy do alter, and in an upredictable way. The source of this alteration of interactions is, largely, technical progress. That is, the irreversible evolution of techniques of physical transformation available to an economy means that activities in the future are not guessable from activities in the past except in the trivial case where technical progress is seen as the bringing into use of available and understood but so far unused techniques. It is clear that technical progress is almost never of this form, so not only does the simple Newtonian mechanical conceptual framework prove inadequate for economics, even introducing the irreversibility of time through simple considerations such as the Second Law of Thermodynamics will not be sufficient. Economies need to be granted absolute temporal irreversibility by using the conceptual framework not of Newtonian mechanics, or even the Second Law, but rather the ideas of modern far-from-equilibrium thermodynamics (see FEHL, 1983).

Before exploring the implications of modern physical science as an alternative conceptual framework for economics we should note that the quasi-irreversible temporal approach offered by the Second Law has already been recognised and explored, most notably by GEORGESCU-ROEGEN (1971) and has recently received a popular exposition by RIFKIN and HOWARD (1985). But while economists may slowly be adjusting their conceptual frameworks from the physical science of the seventeeth century to that of the nineteenth century, the appearance of more accessible works by PRIGOGINE (1980) and PRIGOGINE and STENGERS (1984) may mean the physical science of the twentieth century becomes internalised within economics in the twenty first century. To paraphrase Keynes, « practical economists who believe themselves to be quite exempt from any intellectual influence are usually the slaves of some defunct physical scientist ».

TIME, PRODUCTION AND TECHNICAL PROGRESS

I have tried to suggest that the present conceptual framework of economics is flawed as it treats time inadequately. I have also suggested that the two arrows of time, the Second Law and the tendency to evolve, both need to be allowed for if economic theory is to reflect the processes in real economies. So far some progress has been made to include Second Law ideas, but I feel that until evolutionary aspects of time are explicitly included, the conclusions reached from such an approach will be so generally pessimistic, and so much at odds with our historical experience, that the economics profession would prefer not to embrace such ideas.

The problem posed by the Second Law is that it suggests, whatever the initial conditions inherited by an economy, the long run for the economy must be of increasing scarcity as free energy sources and raw material are both degrated. Here two assumptions are being made. First, the

Earth is an isolated system, and second the economic system operates in a steady way with unvarying techniques. Now the Earth is certainly an open system, and indeed the whole biosphere attests to the importance of sunlight. Also, as the Earth is an open system and as sunlight provides an energy flux that causes the Earth to be a system that is far from thermodynamic equilibrium, there has been the possibility for order to arise from chaos and biological evolution is taking place, of which social and economic evolution may be received as a part.

A consequence of concentrating attention upon the Second Law is that mankind seems to be caught in a trap. DALY (1973) suggests that the only solution is to accept that economies must progress towards a steady state if economic activity is to be sustainable in the long term. GEORGESCU-ROEGEN (1971) goes even further and makes the case for a return to an agrarian way of life at much lower population levels than at present. While accepting the overall concept of limitations to activity through Second Law constraints, BOULDING (1976) is more optimistic, stressing the potentiabilities for evolution in what the calls his Fifth Law of Change.

If economics is to open up its conceptual framework to embrace the irreversibility of time through not only dissipative, Second Law processes, but also through the potential to evolution to unexpected outcomes, what are the areas of economic analysis that might benefit from such a conceptual framework ? HICKS (1976) has suggested that the two areas where the irreversibility of time are most crucial are Capital Theory and the theory of Markets. As the direct implications of the nature of the physical world are most clear for the problem of Capital, I shall restrict my remarks to that topic.

The production and use of capital are directly subject to Second Law constraints. To manufacture capital goods one needs to dissipate energy and use up raw materials. The use of capital goods is similarly destructive of natural resources ; the physical irreversibility of the production process is clear. However, the production process is also irreversible in that production is causally ordered (HICKS, 1979). Raw materials must be extracted or harvested, processed, turned into semi-manufactured goods, made up into the final product, and then marketed. A given technology dictates the order of the manufacturing process, so the building up of artefacts is also irreversible.

A further source of irreversibility in production is that production processes are not unchanging. One might identify two sources of technical change, one causing reversible changes in technology and the other irreversible changes. First, if several techniques are available then for given prices of inputs and outputs, only one technique will be in use. Changing prices may cause the technique to be changed to another already available technique. Obviously, if prices revert to their previous values the technique in use is likely to revert also, though probably with a considerable time lag. Such technical change is therefore reversible in principle. On the other hand, over time new techniques are likely to become available and these might well be brought into use even if prices are unaltered, simply because they offer better quality, greater profitabi-

lity, etc. Indeed, the introduction of such new techniques is more likely to cause prices to alter than to be induced by altering prices. The distinction I have drawn between these two types of change of technique is probably not as clear cut as I have drawn it. For instance, ROSENBERG (1982) has pointed out that changing technique is rarely a matter of shifting to an already known alternative. Any technique that is not in current use is to some extent unknown, and bringing in to use an available but unused technique is to some extent a matter of « finding out about its isoquants ».

The full bringing into use of a new technique involves the gaining of new knowledge. Something that was previously impossible is now possible, and remains possible. New inventions cannot be uninvented, and therefore technical progress is irreversible. Further, whereas the irreversibility of the Second Law and the steady state production process involves what I have termed quasi-irreversible temporal procedures, the irreversibility of technical progress is, like the irreversibility of emergence and evolution, absolutely irreversible. Both explicitly recognise that not only is the future unknown, it is often unexpected and unguessable.

IRREVERSIBLE TIME IN ECONOMIC THEORY

If economic theory is to embody truly irreversible time, what form might such theory take ? First, the Second Law must be accepted as providing a physical limit to any given technique. Something cannot be made out of nothing, and any production process causes an irreversible depletion of available energy.

Second, the irreversibility of processes because of the direction of production needs to be explicitly recognised, and perhaps here the approach of Neo-Austrian Capital theory might be an appropriate starting place (see e.g. HICKS, 1973 and FABER, 1979).

Third, technical progress needs to be treated in a more realistic way than at present. As ROSENBERG (1976) has pointed out, technical progress is not manna from heaven, nor can it be said to be entirely induced by market conditions. However, one might note that the Second Law irreversibility of time and the evolutionary irreversibility come together in a recursion through technical progress. Given techniques will, through the Second Law, eventually generates shortages. These in turn will cause new techniques to be brought into use, causing the economy to evolve towards generating different goods by different techniques at different relative prices. The process is never ending and represents something akin to Darwinian evolution. The processes in use are continually adapting to meet new challenges and to fit the new environment that the processes themselves are causing to alter (these ideas are further discussed in FABER and PROOPS, 1985).

As THOBEN (1982) has discussed, this notion of an economy as akin to a biological process is far from novel, but has so far not found favour in mainstream economics because of the overwehlming dominance of the mechanical conceptual framework. Now that it is clear that mechanical systems are special cases of more general dissipative and evolutio-

nary systems perhaps it is time for economics to update and generalise its conceptual framework.

REFERENCES

K.E. BOULDING, 1976, *The great laws of change*, in A.M. Tang, F.M. Westfield and J.S. Worley (Eds.), *Evolution, Welfare, and Time in Economics*, Lexington Books, Toronto, pp. 3-14.

J. BRYANT, 1982, *A thermodynamic approach to economics, Energy Economics*, 4, pp. 36-50.

H.E. DALY, 1973, *Toward a Steady State Economy*, W.H. Freeman, San Francisco.

M. FABER, 1979, *Introduction to Modern Austrian Capital Theory*, Springer Verlag, Heidelberg.

M. FABER and J.L.R. PROOPS, 1985, *Interdisciplinary research between economists and physical scientists : retrospect and prospect, Kyklos 38*, pp. 599-616.

U. FEHL, 1983, *Die Theorie dissipativer Strukturen als Ansatzpunkt for die Analyse von Innovationsproblemen in Alternativen Wirtschaftdordnungen*, in : A. Schuler, H. Leipold and H. Hamel (Eds.), *Innovationsprobleme in Ost und West, Schriften zum Vergleich von Wirtschaftdordnungen, Vol. 33*, Gustav Fischer Verlag, Stuttgart.

I. FISCHER, 1982, *Mathematical investigations of the theory of values and prices, Transactions of the Connecticut Academy of Arts and Sciences 9*, pp. 1-126.

N. GEORGESCU-ROEGEN, 1971, *The Entropy Law and the Economic Process*, Harvard University Press, Cambridge, Mass.

J.R. HICKS, 1973, *Capital and Time : A Neo-Austrian Theory*, Clarendon Press, Oxford.

J.R. HICKS, 1976, *Some questions of time in economics*, in : A.M. Tang, F.M. Westfield and J.S. Worley (Eds.), *Evolution, Welfare, and Time in Economics*, Lexington Books, Toronto, pp. 135-151.

J.R. HICKS, 1979, *Causality in Economics*, Blackwell, Oxford.

L.P. KADANOFF, 1983, *Roads to chaos, Physics Today* (December) pp. 46-53.

D. LAYZER, 1976, *The arrow of time, The Astrophysical Journal 206*, pp. 559-564.

P. MIROWSKI, 1984, *Physics and the « marginalist revolution », Cambridge Journal of Economics 8*, pp. 361-379.

G. NICOLIS and I. PRIGOGINE, 1977, *Self-organization in non-equilibrium systems*, J. Wiley, New York.

I. PRIGOGINE, 1980, *From Being to Becoming*, W.H. Freeman, San Francisco.

I. PRIGOGINE and I STENGERS, 1984, *Order out of Chaos*, Heineman, London.

J.L.R. PROOPS, 1983, *Organisation and dissipation in economic systems, Journal of Social and Biological Structures 6*, pp. 353-366.

J.L.R. PROOPS, 1985, *Thermodynamics and economics : from analogy to physical functioning*, in : J. Sweeney and W. van Gool (Eds.), *Energy and Time in the Economic and Physical Sciences*, North Holland, Amsterdam, forthcoming.

J. RIFKIN and T. HOWARD, 1985, *Entropy : A New World View*, Palladin, London.

N. ROSENBERG, 1976, *Perspectives on Technology*, Cambridge University Press, Cambridge, U.K.

N. ROSENBERG, 1982, *Inside the Black Box : Technology and Economics*, Cambridge University Press, Cambridge, U.K.

P.R. SAMUELSON, 1948, *Foundations of Economic Analysis*, Harvard University Press, Cambridge, Mass.

H. THOBEN, 1982, *Mechanistic and organistic analogies in economics reconsidered, Kyklos 35*, pp. 292-306.

Technology, Variety and Competition

Michael Gibbons
Department of Science and Technology Policy and PREST
University of Manchester

John S. Metcalfe
Department of Economics and PREST — University of Manchester

In their book *Order out of Chaos*, Prigogine and Stengers have stressed the importance of understanding concretely the manifold processes that occur in both the natural and social world. Many of these processes are intrinsically irreversible ones, the inner workings of which cannot be even adequately approximated by equilibrium models in which time is reversible and prospect of these processes exhibiting any new phenomena eliminated from the outset. A more appropriate heuristic, it is argued, is given by the model of dissipative structures which, in interacting with their environment, can by amplifying certain localised fluctuations deflect processes up avenues which are novel but, at the same time, eliminate the conditions which created them in the first place thereby making a return to the initial state an extremely unlikely physical event.

The basic idea that « order » might arise from the « chaos » of fluctuations in the environment — that a process might move from one stable equilibrium to others — is pregnant with analytical possibilities. In particular, the simple admission of « time's arrow » into analytical thinking opens up significant new perspectives for understanding many economic processes that have hitherto been guided by rather simple mechanical models. Economic science is perhaps unique among the social sciences in its preoccupation with equilibrium models. In this paper a framework is developed which explores, albeit in a preliminaryway, the consequences for competitive theory of a models which takes into

account that economic development is a dis-equilibrium process with some of the characteristics of irreversibility that are present in natural processes. Economic development is considered as a sequence of more or less major transformations which test the resilience of any given economic structure to accomodate them. Whether that structure is able to absorb the fluctuations presented by emerging technologies and which will be amplified is dictated, to a large extent, by the competitive process. And that process, as Schumpeter observed long ago, is not inaccurately described as a process of creative destruction.

The framework outlined in this paper owes its origins both to a series of case studies of innovations in UK firms (Georghiou, *et al*) and a preliminary reading of some literature on evolutionary biology (Matthews, Sober). We shall argue that economic progress in capitalist economies is based upon two interdependent processes relating to, first, the mechanisms which generate a variety of technological and organisational forms, and, second, the mechanisms by which different forms acquire economic weight. For the economic impacts of a given technology depend directly upon their scale of application and we are not generally interested in technologies or organisational forms which do not acquire economic significance.

In developing our analysis, we have found biological metaphors useful. The biological concepts of selection, adaptation and mutation can be twinned with their economic counterparts, competition, imitation and innovation, and employed, heuristically, to generate useful insights into the nature of competitive process. Selection and adaptation may be regarded as different mechanisms for reducing variety and mutation a mechanism for generating variety. We argue below that the rate of progress in an industry is proportional to the degree of economic variety contained within it and so depends on the balance between the variety enhancing and variety diminishing aspects of technological competition. Of course, any metaphor must be translated, with care across disciplines, the biological one no less so than its more familiar mechanical counterparts. The speed of operation of the variety generating process differs fundamentally between the natural and economic worlds, and there is nothing in the natural sphere to correspond to the directed, anticipatory behaviour which characterises individuals and organisations in their economic behaviour. Nonetheless, we find the biological metaphor a useful basis for exploring technological change. In particular, the concept of isomorphism, that technologies, like species, are « adjusted » to fit their environmental circumstances, we believe to be of particular analytical value (Hannan and Freeman). How this adjustment takes place, be it by selection or by adaptative behaviour is at the heart of our discussion.

The paper is structured as follows. In Section I we discuss dimensions of firm behaviour relevant to the analysis of technological change. In Section II we discuss the competitive process while Section III is devoted to some analysis of selection processes in simple environments, where we focus upon the way in which technologies compete for economic weight.

1. FIRM BEHAVIOUR AND INNOVATION

An understanding of the mechanism of technical change in contemporary capitalism requires analysis at two different levels : that of the capacity of firms to generate a particular momentum of technological change ; and that of the interaction between firms within a « market » environment in which their specific technological differences are resolved into changes of economic weight. In the study of firm behaviour we have found it useful to divide the treatment of a technology into two components, namely, revealed technological performance and the knowledge base.

The revealed technological performance of a firm consists of its products and the processes it employs to produce them, and how these change over time as a consequence of innovative and imitative behaviour. Revealed performance is a basic unit of analysis because, in our view, competition selects directly with respect to current products and processes, not with respect to firms. Firms are only affected indirectly according to the superiority or otherwise of their revealed performance. Underpinning a firm's revealed performance lies its knowledge base ; that set of human skills and competences which define what can be achieved in practical terms. Within any technological area or paradigm (Dosi), a number of useful distinctions can be made with respect to the underlying knowledge base. While major components are contained within firms, external knowledge held by other institutions, e.g. university departments and government laboratories, is often important, as is any firm's ability to inwardly transfer knowledge from these external sources (Gibbons and Johnston). A core of knowledge will often exist in a codified, publicly available form but for the competitive process it is the proprietary, often tacit, knowledge of individual firms which is of crucial importance. This knowledge is embodied in the firm's employees and its successful application is contingent upon how the organisation pools their individual capacities. In understanding differences between firms in terms of revealed technological performance, it is quite impossible to separate questions of strategy and organisation from questions of technological knowledge proper. The routines and correlative structures of organisations are critical to the process of technological differentiation. Typically, the knowledge base of a firm builds cumulatively, through experience, in house R&D and the transfer of the knowledge from other organisations. Each firm, then, develops, as it were, its own technological landscape. Promising paths of development emerge and give to technological changes a momentum of their own. They offer an agenda for change but, equally, they create inertia with respect to opportunities different from the chosen paths. Thus progress is predominantly incremental and localised. Because technological development is canalised organisations are not infinitely malleable ; rather they tend to develop routinised responses to the opportunities they perceive. However, given uncertainty and the role of human imagination, it is not surprising that we persistently observe firms exploring a shared core of knowledge in quite different ways. Variety, it appears, is the basic norm of technological development. It is also as we will show the driving force behind technological competition.

Although revealed technological performance is our primary unit of analysis, it is to the behaviour of a firm which we must turn to understand how the knowledge base is linked with revealed performance. Thus, we define the « firm » as an organisation articulating a knowledge base in a profit seeking way. We assume that more profit is preferred to less but we do not imagine that any firm has the capacity to enumerate and evaluate every opportunity open to it ; effective search is limited by the cognitive capacity of the organisation. Because information is scarce and costly to acquire and process, the idea of a uniform availability of knowledge is simply incompatible with the idea of technological competition. We stress that in a majority of practical cases the relevant unit of organisation is not the firm *simpliciter* but rather that sub-unit with the responsibility for articulating a particular product range and its associated production methods. Often, relevant components of the knowledge base will lie in different parts of the umbrella organisation, and the capacity to link these components together has often been observed to be an important element in innovative performance (Carter and Williams, Langrish *et al*, Science Policy Research Unit).

In sum, in understanding levels and rates of change of revealed technological performance, it is to the mechanisms by which firms acquire, store and articulate a knowledge base that investigators should look.

The dimensions of Competitive Performance

To develop further the relation between technology and competition we assign to a « firm » three different dimensions of performance. The first dimension is the *efficiency* of the firm as measured by its *revealed* technological performance ; the quality of its products and the productivity with which it produces them. The second dimension we term *fitness*, and define this as the ratio of the firm's growth rate of productive capacity to its profit margin per unit of output. In short, fitness summarises the ability of the firm establish a profit and the willingness to transform these profits into growth. The third dimension, by contrast, is more opaque. It is the *creativity* of the firm, as « measured » by its ability to enhance and *improve* its product range and its methods of production. The more creative firm is the one with the greater rate of improvment in revealed technological performance, whether this is achieved directly by augmenting its knowledge base or indirectly, by applying its existing knowledge base to greater effect. Creativity subsumes both innovation proper, and imitative behaviour.

Now the point to be emphasised is that firms differ in all three dimensions, and it is the differences which provide the clue to the relation between competition and the environment. Within a given technological area, firms reveal products of differing quality and they produce each product with different levels of unit cost. Indeed, one of the enduring phenomena of industrial behaviour is the range of efficiency that firms exhibit, and the gap between best-practice and average practice that this implies (Salter, Iwai). Equally, firms differ in their fitness as we have defined it. Differences in the proportion of internal funds devoted to accumulation, differences in access to external funds, and different abi-

lities to manage the process of capacity expansion (Penrose) determine the relation between growth rates and unit profit margins. Moreover, fitness may change systematically with the size or age of the organisation. Finally, of course, firms differ in their creativity : because they have different perceptions of the agenda for improving revealed performance ; because they have different financial and other resources to devote to R&D ; because they have differential ability to manage the innovation process ; because of differential degrees of organisational inertia ; and, not least, because their differing abilities to capitalise on the effects of serendipity.

It will be observed that the three dimensions are inter-dependent. It is because firms differ in their creativity that they come to differ with respect to their efficiency. Similarly, differences in efficiency partially determine the resources available to fund creativity. Moreover, the decision to allocate profits between capacity expansion or R&D, for example, is central to any firm's long-term market performance. It should be noted that the dimensions of performance have significance quite independently of any postulate of maximising behaviour. Efficiency, fitness and creativity are attributes of firms as they are : how they might otherwise behave is quite irrelevant to the analysis of competitive performance or its relation to technological change. In brief, the approach proposed by Nelson and Winter of decision making based upon satisfactory routines, forms a valuable starting point for understanding the origins of persistently diverse behaviour (Nelson and Winter).

2. CONCEPTS OF COMPETITION

That a firm's command over its product and process technology is crucial to its market performance needs only brief comment. The characteristics of a firm's products and processes relative to those of competing firms is a prime determinant of its profitability and thus its access to resources to enhance its market position. Profits determine the scope for expanding capacity, for influencing market perceptions of its products, via sales effort and advertising, and for improving its knowledge base. As a rule a firm will enjoy above average profitability only if its revealed technological performance is of above average practice levels. That is, if its costs of production are less than the average for the firms with which it competes, and if its product quality is such as to command a price premium relative to that commanded on average by its competitors. Of course, a firm may enjoy above average profitability even if it is below average in one technological dimension, providing there is a compensating excess above the average in another. The firm's technological performance as revealed in its products and processes is the foundation for its competitive performance. Above average profitability provides the means to expand market share relative to rivals, either through a faster rate of capacity expansion, through a greater sales effort which attracts customers away from rivals, or through a greater rate of improvment in the firm's products and processes.

Thus, competition is driven by the differences between firms in technological performance ; differences across which the market mechanism

selects ; differences which are continually changing over time with the evolution of technology. Competition is no dull equilibrium affair defined relative to firms with identical products and production methods. It is a vibrant one, driven by variety in performance and, in the long run, no dimension of variety is more telling than that created by technological diversity.

The nature of the competitive struggle is placed in sharper focus when we recognise that a technology is rarely static. Rather when a technology first emerges it is imperfect and offers an agenda for future development limited by a state of technological maturity. Competitive success then depends not upon making a single innovation representing a stage in the evolution of the technology but in maintaining a momentum of technological advance — exploiting the agenda at least as effectively as do rival firms.

In the following discussion, we will be concerned with competition as a process driven by variety, with market structure a consequence of competition not a test for its existence. This process not only provides scope for entrepreneurship but also for viewing competition as a race between unequals, in which prosperity and survival depend on a firm's position within a (multivariate) distribution of behaviour in the industry. In other words, variety is a statistical phenomenon in which firms must be regarded as members of an ensemble engaged in a competitive process.

Typological vs. Population Thinking.

In economic theory, the debate over uniformity and diversity in the theory of cost and competition is paralleled by a similar debate in evolutionary biology on the relative status of typological and population thinking. The issue in both cases is what sense can be made of variety (Sober, Mayr).

Typological, or essentialist, thinking is the doctrine of ideal types. Every phenomenon has a natural status, composed of its central characteristics, deviations from which are to be treated as accidental containing nothing additional to the central core of properties. Variety is deviation generated by interfering forces. Now within the traditional theory of competition, the postulate of identical cost conditions is meant to capture the essense of behaviour, deviations are devoid of analytic significance and are caused by information imperfections. Pigou's equilibrium firms is an example of such an ideal type.

Unwittingly, or otherwise, some authors such as Downie and Steindl reject essentialist thinking and replace it whith what evolutionary biologists term population thinking. Each variation of form now has an independent, equivalent status. Variety is the natural state of affairs and ideal types are an illusion. Uniformity is here the aberration, to be explained by the operation of interfering forces. Thus competition eliminates variety just as selection does for the evolutionary biologist. In other words, variety has a causal role in explaining the behaviour of a population aggregate. This population perspective has a number of important

consequences. First, the behaviour of the population is a statistical construct, to be explained in terms of levels of and changes in the mean, variance and higher moments of the population distribution. Secondly, the performance of any population member is only significant in terms of its position in the overall distribution of behaviour. Thirdly, the behaviour of the population aggregate need not be reducible to the behaviour of the corresponding population members. The whole is, of course, composed of the parts but changes in the whole do not necessarily entail changes in the corresponding characteristics of the parts. Some of these consequences will be explored in the following analysis.

3. SELECTION, EFFICIENCY AND TECHNOLOGICAL COMPETITION

We have suggested above that the behaviour of firms be interpreted in three dimensions — efficiency, fitness and creativity. We now turn to analyse how, in particular, the efficiency dimension of performance relates to the process of competitive selection. The core of the argument is that we must define both the degree of variety in behaviour and the selection environment, before the rate and direction of competitive change can be established.

In making sense of the role of variety it is convenient to focus upon the following simple case. A number of process technologies are available for producing a single, homogenous product. Each technology is articulated by one or more firms with a given technology base, and firms employing the same technology have identical unit costs of production. To incorporate differential « X-inefficiency » will be an important consideration but it is not essential to this argument. Each technology is of a constant returns kind and we let h_i denote the level of unit cost for the ith technology, and v_i denote the corresponding capital : output ratio, capacity increments being possible in infinitely variable amounts.

Starting from a given technology set, we now determine changes in the relative economic weight of the different technologies, the survival prospects of each, and whether a particular technology will emerge as the dominant design (Utterback and Abernathy). In short, we are to examine the working of a multi-technology diffusion process. For the purposes of this short paper we assume that the environment is of a particularly simple and stable kind. All firms face the same vector of input prices. This vector is independent of the aggregate scale of output of the commodity, and does not vary exogenously during the selection process. The price of the commodity does vary over time, according to the relation between supply potential and demand for the product, in a manner yet to be determined. Market demand is growing at the constant proportionate rate, g_D.

Following convention, each technology is exploited by profit seeking organisations called firms, although each firm, or operating unit, is typically to be considered as a component part of a larger multi-product organisation. At the common set of input prices, each technology will be associated with a different unit cost of production, and thus a diffe-

rent profit margin, since all firms receive the same price for their output. It is these differences in profit margins which are the basis for the selection process, and the changing economic weight of the different technologies.

We begin with the case of no technological change proper. There are neither changes to the technology set, nor variations within each technology in the set. In addition to its unit costs, each firm is characterised by its fitness. Fitness is determined by the ratio of the propensity to accumulate to the capital : output ratio. Letting each firm invest a fraction of its internal profits Λ_i, and supplement those with external funds in the proportion e_i, we define the ith firm's fitness as $f_i = \Lambda_i (1+e_i)/v_i$. Provided the firm is earning a rate of profits in excess of a normal rate, it invests in its own technology at the rate g_i, given by

$$g_i = f_i(p-h_i) \; ; \; h_i < p \qquad (1)$$

At a normal rate of profits, $p = h_i$, fitness drops to zero. Otherwise fitness is independent of the firm's rate of profits and is constant during the selection process. To allow firms to invest in other technologies besides their own is a straightforward extension of the argument (Turner and Soete).

Following Downie, we first analyse the relation between selection and efficiency under conditions of uniform fitness, $f_i = f$. Let s_i, the market share of the ith process technology, measure its economic weight. The aggregate growth rate of capacity, g, is then related to the average profit margin by

$$g = \sum_i s_i g_i = f(p-h) \qquad (2)$$

where $\bar{h}$ is average practice unit cost, defined as $\sum s_i h_i$, with the summation in (2) applying only to those « dynamic » firms with above normal profitability.

The variation over time in the economic weight of the ith technology depends on the growth rate difference $g_i - g$, with

$$\frac{ds_i}{dt} = s_i(g_i-g),$$

which from (1) and (2) becomes

$$\frac{ds_i}{dt} = s_i f(\bar{h}-h_i) \; ; \; h_i < p \qquad (3)$$

The interpretation of (3) is straightforward. The economic weight of a given technology is expanding or contracting according to the position of that technology in the population distribution of efficiency, this distribution depending on the particular environmental vector of input prices. A technology which has unit costs above average declines in economic weight, and conversely if its unit costs are below average. Provided unit costs are less than or equal to the product price the technology survives.

In the case of two production inputs, the argument can be illustrated in terms of Figure 1 (Metcalfe). The solid convex area, t-t, denotes the technology set. It summarizes the extent of variety in terms of the two input : output coefficients a_1 and a_2. The given environment is represented by the relative price lines with slope (w_1/w_2). The number and distribution of technologies within the technology set need not be continuous and sections of the set may be empty. Note that only two technologies are required for the selection process to work. Given the environment, $\propto$ is the best practice technology with unit costs $h_\propto$, and $\bar{h}$ is average practice unit cost with $\bar{a}_1$ and $\bar{a}_2$ the average practice input : output coefficients. The price of the product p, exceeds $\bar{h}$ by the amount required to finance the aggregate growth of the dynamic firms at rate g. On the basis of this information the technology set can be divided into three regions. Within region A, the technologies are redundant and the corresponding firms out of business, since $h_i > p$. Regions B and C contain dynamic firms and their technologies. Within B, any technology generates above normal profits so its capacity expands in absolute terms but, since $h_i > \bar{h}$, its market share and economic weight are declining. Within A are technologies with $h_i < \bar{h}$. Not only are these profitable and expanding in absolute terms but their economic weight, market share, is increasing. At the boundaries between each region we have marginal (zero profit, technologies) distributed along p-p, and dynamically representative (constant market share) technologies distributed along $\bar{h}$-$\bar{h}$. Of course, with a different environment, the technology space will be divided differently and best practice may also be redefined. In short, given the environment, the competitive position of any one technology depends on its distance from average practice, h_i-$\bar{h}$, and this distance is related not to that technology's economic weight, s_i, but to the rate of change of its economic weight, ds_i/dt.

Figure 1 is, of course, a snapshot of a situation at a point in time, and it contains the seeds of its own destruction. The process of competition for economic weight continually redefines $\bar{h}$ and p and the relative subdivision of the technology set. From (3) and the definition of $\bar{h}$ we find that

$$\frac{d\bar{h}}{dt} = \sum_i \frac{ds_i}{dt} h_i = -fV_t(h_i),\ h_i < p \quad (4)$$

where $V_t(h_t)$ is the variance in unit costs of dynamic firms in the technology set. Average practice unit cost is a population statistic, continually reduced by competitive selection and tending towards best practice unit cost as variety is eliminated. There is technical progress at the population level even though there is no technical progress proper. It is also clear that the technology which emerges as the dominant design is the best practice technology. From (4) we see that the greater is the variety of efficiency, the greater is the rate of competitive selection. Furthermore, the rate of reduction in average practice cost depends on the fitness coefficient which varies directly with the flow of internal and external funds into investment and inversely with the capital : output ratio. Ultimately,

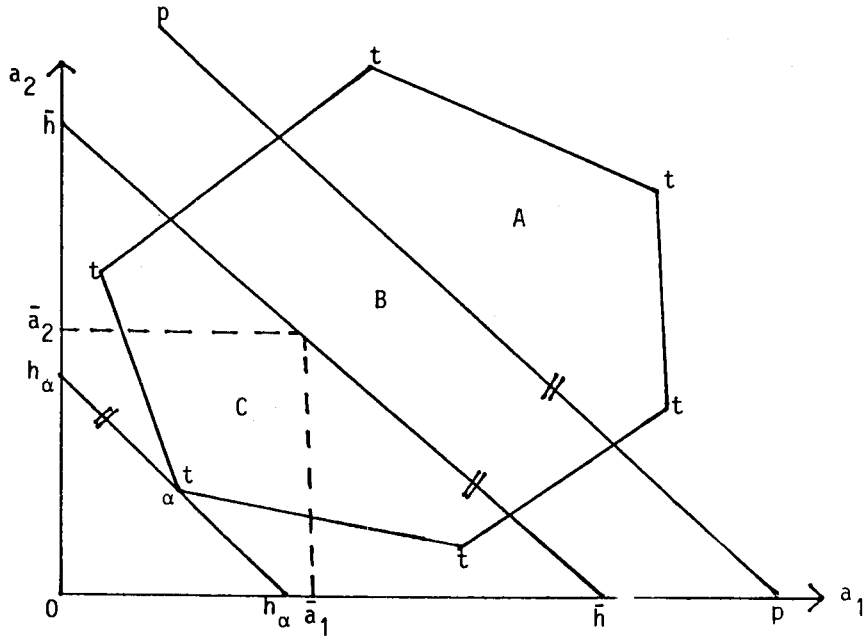


Figure 1

The Technology Set

$V_t(h_i)$ is driven to zero as h converges on h_α. Selection eliminates variety and with it the competition pressures which drive the changes in the population characteristic $\bar{h}$. We should note that (4) has its parallel in evolutionary biology' *viz.*, R.A. Fisher's fundamental theorem of natural selection (Nelson and Winter).

Corresponding to the changes in $\bar{h}$ are changes in the average practice input : output coefficients $\bar{a}_1$ and $\bar{a}_2$. Since $a_j = \sum_i s_i a_{ij}$ it follows that the manner of variation of $\bar{a}_1$, say, is given by

$$\frac{d\bar{a}_1}{dt} = - f[w_1 V_t(a_1) + w_2 C_t(a_1,a_2)] \qquad (5)$$

where, w_1 and w_2 are the input prices, $V_t(a_1)$ is the variance in the use of a_1 and $C_t(a_1,a_2)$ is the co-variance of input use, again across the dynamic firms. A similar formula holds for $d\bar{a}_2/dt$ and combining these results we obtain

$$V_t(h_i) = w_1{}^2 V_t(a_1) + w_2{}^2 V_t(a_2) + 2w_1 w_2 C_t(a_1 a_2) \qquad (6)$$

the familiar decomposition of the variance of a sum. From (5) and (6) we can identify which dimensions of input use contribute most to selection pressure and how pressure is affected by different input price environments. Notice also that since $C_t(a_1 a_2)$ can be negative, it is possible for one (but not both) of the average practive coefficients to regress

during the selection process. It is readily shown that the degree of input-saving bias generated during selection is given by

$$\frac{d}{dt}\log(a_1/a_2) = f\left[\frac{C_t(a_1,h_1)}{\bar{a}_1} - \frac{C_t(a_2,h_i)}{\bar{a}_2}\right] \qquad (7)$$

In each of the expression (4)-(7) we are relating movements in average population behaviour to higher order moments of the population distribution. Without variety in performance, such propositions would be impossible to derive. So far we have established the dynamics of selection and the principle of convergence upon best practice technology in a simple stable environment. We have not as yet determined either the aggregate growth rate of dynamic firms, g, or the price of the product and hence the range over which our population statistics are to be defined. To do this consider Figure 2, in which the price of the commodity is determined by the market demand curve $D_t(p)$ and the corresponding supply function $S_t(p)$ (Salter). Along this supply function, the length of each denotes the amount of capacity invested in that particular technology at date t. At the ruling market equilibrium, the marginal technology is δ with unit costs $h_m = p$. Technology ε is redundant, while the remaining technologies are profitable and earn quasi-rents in proportion to their distance from h_m. It is these quasi-rents which, as with Steindl and Downie, drive the supply curve to the right over time. If we let ϱ_i be the fraction of industry profits attributable to the ith technology, we find that

$$\varrho_i = \left(\frac{h_m - h_i}{h_m - \bar{h}}\right)s_i = \frac{f}{g}(h_m - h_i)s_i \qquad (8)$$

which relates the distribution of profits to market shares, to fitness, to the growth rate, and to distances, from worst practice. In so far as creativity is linked to profits, this profit distribution plays an important part in developing an interaction between selection and technical change.

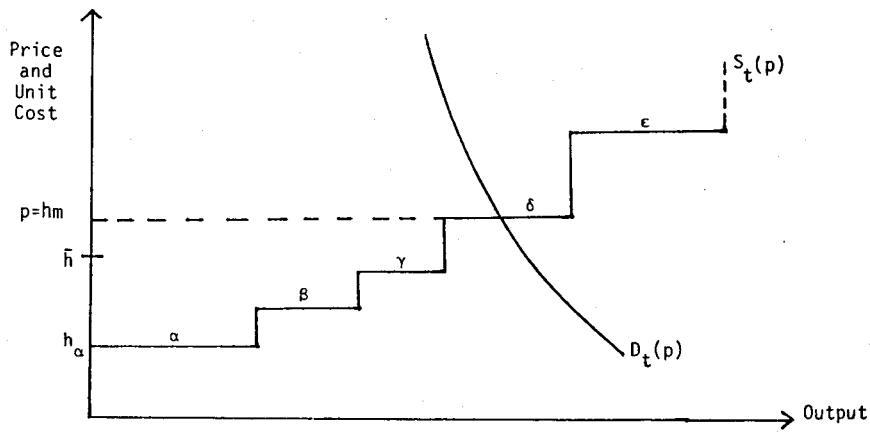


Figure 2

Prices, Costs and Rents

We now appear to have two explanations of the commodity price. A static demand and supply explanation, $p = h_m$, and a dynamic selection version, $p = g/f + \bar{h}$. The connection is provided by the distinction between marginal and dynamic firms. For a marginal firm, capacity accumulation has ceased and changes in its output correspond to changes in the utilization of existing capacity. By contrast, the dynamic firms produce at full capacity and expand capacity according to (1). In Marshallian terminology, the behaviour of the dynamic firms is interpretable by long-period methods of analysis, and that of the marginal firms by short-period methods of analysis. Both methods applied to the same point in the selection process.

We can now distinguish the aggregate growth rate of dynamic firms, g, from that of the marginal firms, g_m, and the average practice unit costs of dynamic firms, $\bar{h}$, from that of all firms with positive economic weight, h_T. If z is the share in total output of the dynamic firms we have $g_D = zg + (1-z)g_m$ and $\bar{h}_T = z\bar{h} + (1-z)h_m$. The rate of change of z is given

$$\frac{dz}{dt} = z(g-g_D) = z[f(h_m-\bar{h}) - g_D] \qquad (9)$$

To interpret (9) we see there are two possible states for the selection process. If, as in Figure 2, the output from marginal technologies is positive, $z < 1$, selection equilibrium is characterized by $p = h_m$, $dp/dt = 0$ and $dg/dt = -f\, d\bar{h}/dt = +f^2V_t(h_i) > 0$. From (9) we have z rising or falling over time as $f(h_m-\bar{h}) \gtrless g_D$ but either way the marginal technology is eventually eliminated. Either it is transferred into the redundant set if $dz/dt > 0$, or its capacity becomes fully utilized if $dz/dt < 0$, and p rises above h_m making the technology profitable. Thus we reach the second possible case in which all technologies generate positive profits and $z = 1$. We then have $g = g_D$ (Downie's case) and $dp/dt = -f\, V_t(h_i)$; price and average practice unit cost fall over time at the same rate.

That we have two possible states for the selection process implies that particular care is needed indefining average practice unit costs in an industry. Depending on z, the values of h_T and $\bar{h}$ may differ substantially and will vary at different rates with changes in z. Two other points are worth noting. First, the transfer of technologies from dynamism to redundancy has no effect on the evolution of $V_t(h_i)$, simply because when technology becomes redundant, its economic weight has already reached zero. Selection in turbulent conditions could clearly affect this conclusion. Secondly, the value of $V_t(h_i)$ is independent of the growth rates g and g_D. This follows immediately, since the range of variety in unit costs, $p - h_\infty$, is defined by D_t and S_t, ie. by the levels of demand and supply potential, not by their rates of change.

We conclude this section with a number of exercises. Consider first the effect of innovations upon the selection process. An innovation is represented by an addition to the technology set simultaneously with the entry of new organizations is quite distinct from entry of capital into the industry, which is already taken care of by the role of external

finance in the definition of fitness. Clearly entry has a shock effect on the existing pattern of market shares but from then on the entrant's fate is determined by its position within the overall distribution of unit costs. Suppose, for example, that at entry, the new technology is superior to average practice but inferior to best practice. Then its diffusion curve, the path of market share over time, will behave as follows. Market share will rise at first, and continue to do so until average practice efficiency has fallen, sufficiently to make the technology dynamically representative. From then on its market share falls progressively towards zero as it drops further behing average practice and moves closer to worst practice. Ultimately, the technology ceases to have economic weight. However, it may still survive in the sense of being intra-marginal and producing an increasing volume of output. In particular, the distinction between a technology having positive economic weight and its surviving is vital to an understanding of selection. With selection, the commodity price tends toward a value p_∞, given by $p_\infty = g_D/f + h_\infty$. In this situation, all technologies satisfying $h < h_i < p_\infty$ will survive but each of them will have zero economic weight. Multi-technology diffusion processes are complex phenomena, and sigmoid diffusion curves are a limited reflection of their dynamics. Indeed, if economic weight is used to measure diffusion, only the dominant design will have a sigmoid diffusion curve with an upper asymptote of unity. All other technologies must ultimately experience declining diffusion levels. As in Gauss's famous principle, in this simple and stable environment all technologies but one are ultimately excluded in terms of their economic weight (Slobodkin). Notice that the survival of non-best-practice technologies is conditional upon g_D being positive, and that the survival range varies inversely with the degree of fitness. Again, we see that the fitness coefficient is a measure of the degree of competitive pressure. Of course, if entrant technologies are superior to existing best-practice, this redefines the selection process rather more drastically. How this works out in terms of our selection principle is not difficult to see.

As our second brief exercise, we note that the selection process implies increasing concentration upon a declining number of technologies. Market structure in this sense is an endogenous consequence of the selection process not an independent test for the existence of competition. Of course, how market structure evolves in terms of firm concentration is more complex and depends ultimately on how firms can acquire and articulate the best practice knowledge base.

4. SUMMARY

Technological variety, fitness and the selection environment define a competitive process in which the best practice technology gradually acquires the status of the dominant design. How quickly this occurs depends on the characteristics of the technology set and the (uniform) fitness of the various firms. Although individual firms do not maximise it can be argued that the environment selects in a maximising, i.e. cost-minimizing way. This does not imply in any sense that the dominant

design firms maximize anything. Selection works with respect to firms as they are, not as they might otherwise by (Alchian). The environment « chooses » the best technology from the available set and the possibility that any technology could be articulated with greater efficiency is simply irrelevant (Winter). In a world of imperfect knowledge how could it be otherwise ?

REFERENCES

A.A. ALCHIAN, 1950, « Uncertainty, Evolution and Economic Theory », *Journal of Political Economy*, Vol. 58.

C.F. CARTER and B.R. WILLIAMS, 1957, *Industry and Technical Progress*, Oxford.

G. DOSI, 1982, « Technological Paradigms and Technological Trajectories », *Research Policy,*

L. GEORGHIOU et al., 1986, *Post Innovation Performance*, Macmillan, London.

M. GIBBONS and R.D. JOHNSTON, « The Role of Science in Technological Innovation », *Research Policy*, *3*, (1974).

M.T. HANNAN and J. FREEMAN, 1977, « The Population Ecology of Organizations », *American Journal of Sociology*, Vol. 82.

K. IWAI, 1984, « Schumpeterian Dynamics », *Journal of Economic Behaviour and Organization*, Vol. 4.

J. LANGRISH *et al.*, 1972, *Wealth from knowledge*, Macmillan, London.

R.C.O. MATHEWS, 1984, « Darwinism and Economic Change », in D.A. Collard *et al.* (eds.), *Economic Theory and Hicksian Themes*, Oxford.

E. MAYR, 1982, *The Growth of Biological Thought*, Harvard.

J.S. METCALFE, 1984, « Technological Innovation and the Competitive Process », *Greek Economic Review*, Vol. 6.

R. NELSON and S. WINTER, 1984, *An Evolutionary Theory of Economic Change*, Harvard.

E. PENROSE, 1959, *The Theory of the Growth of the Firm*, Blackwell, Oxford.

I. PRIGOGINE and I. STENGERS, 1984, *Order out of Chaos*, Heinemann, London.

W.E.G. SALTER, 1966, *Productivity and Technical Change* (2nd Ed.), Cambridge.

Science Policy Research Unit, 1972, *Success and Failure in Industrial Innovation*, Sussex University.

J. SCHUMPETER, 1943, *Capitalism, Socialism and Democracy*, Allen and Unwin, London.

L.B. SLOBODKIN, 1961, *Growth and Regulation in Animal Populations*, Dover, New York.

E. SOBER, 1985, *The Nature of Selection*, MIT Press, Cambridge.

J. STEINDL, 1952, *Maturity and Stagnation in American Capitalism*, Monthly Review Press, London.

R. TURNER and L. SOETE, 1984, « Technology Diffusion and the Rate of Technological Change », *Economic Journal*, Vol. 94.

J.M. UTTERBACK and W.J. ABERNATHY, 1978, « A Dynamic Model of Product and Process Innovation », *Omega*, Vol. 3.

S. WINTER, 1964, « Economic « Natural Selection » and the Theory of the Firm », *Yale Economic Essays*, Vol. 4.

IV

Toward a New « Rationality »

Rationality and Irrationality in the Theory of Knowledge

Yoichiro P. Murakami
University of Tokyo

§ 1. Let me begin this paper with raising a few questions. Suppose I had here an instrument-like black box which is equipped with a screen, a counter-like tube with cable, switches, buttons, measuring meters and so on. And suppose I said that this divice is a mind-reader that I have just invented and completed. When I operate the counter-like tube over a subject's head, the spikes appearing on the screen change their patterns. By reading the change of spike-patterns, I guess what he, the subject, is thinking.

Looking at one spike-pattern, I guess that he is thinking of, say, his lover, at another spike-pattern, I guess that he feels hungry, and still at another pattern, I do that he is mad at me, and so on and so forth. These guesses are sometimes correct, and sometimes not. Now my first question is this. Which people should be categorized as « scientific » ; those who accept without hesitation my description and explanation about my mind-reader, and praise me a great inventor, or those who are very sceptical to my explanation and refuse to accept the data collected by the device as appropriate ?

Needless to say, the answer to this question seems to be quite simple. The scepticists should be called « scientific ». Yes.

Then comes the next question. Almost nobody now knows the name, Cezare Cremonini. But he was very well-known at this time — through the latter half of the 16th century and the 17th century — as a great peripatetic philosopher. He taught at the University of Padova, where G. Galilei also taught mathematics. As is well-known, Galileo constructed his « eyeglasses » (later, telescope) and with it he observed heavenly

phenomena. On the telescopic observations of the heavenly phenomena he made, Galileo published a small book, « Sidereu nuncius » in which he reported that there were hills and seas on the lunar-surface, and the Jupiter had four moons (satelites). Confronted with those reports, the great philosopher, Cremonini, refused to accept the descriptions of the heavenly world that Galileo made, and refused even to look into the newly invented « eyeglasses ». Then, do we categorize him, Cremonini, as « scientific » ? Comparing to the case of my mind-reader, our answer to this question shoud be « yes ».

At Galileo's time, the empirical reliability of the telescope which had been just invented, was rather low, as low as the reliability of my mind-reader. The images were distorted ; sometimes single image was doubled through the telescope. Admittedly the image of the moon looked much larger than as seen by naked-eyes, but the images of stars were not only « not » enlarged but also contracted. Why could we say that the shadowy patterns of the moon's face are hills and seas, while the rim of the moon always looks as a perfect arc even through the telescope ? The theoretical support of telescope was also low at the time, as low as the one to my mind-reader. Geometrical optics was just emerging and Galileo himself made a certain contribution to that. However, in general, it can be safely said that telescopic theory did not exist. In other words, a body of knwoledge shared by most of the intellectuals of early 17th century did not and could not theoretically support the telescopic observations, just as a body of scientific knowledge that we have now does not and connot theoretically support my mind-reader.

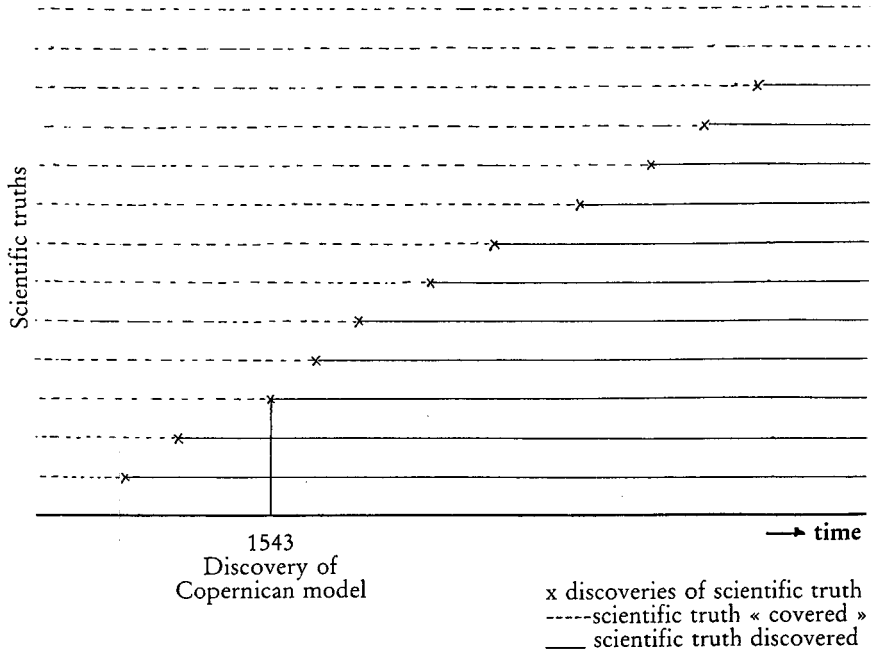
This comparative observation illustrates us a tremendous difficulty we meet, when we define « scientific », when we distinguish « scientific » from « unscientific », and when we demarcate « scientific » from « non-scientific ». Half century ago, the definition of the word « scientific » was rather self-evident. A body of knowledge which is constructed and conjectured in a rational ways was « scientific ». The phrase « in a rational way » in this proposition used to be a priori clear. On the contrary, a body of knwoledge which is contaminated by subjective biases was not « scientific ». This naive and happy demarcation criteria cannot hold anymore. Both the historical and methodological analysis of science have shown that.

§ 2. From the perspective of historical research, or rather historiography, the naive demarcationistic view had, as its background, a single-minded historical attitude. It can be called a progressionism of history or Whiggian interpretation of history. This attitude seems to consist of two parts ; scientific knowledge is regarded as universal and common in every period of time and all through history on the one hand, and history of science is regarded as the history of a gradual accumulation of such ever-unchanged scientific knowledge.

For instance, as is well-known, N. Copernicus proposed his helio-centric model in 1543. And today we believe that a helio-centric model is « scientific ». Then the progressionists say that Copernican helio-centric model is identical with ours. Thus, the progressionnists' view of history can be illustrated as in Fig. I.

Figure 1

Progressionist's Image of History of Science



However, if we take a closer look, we can easily find that the way of reasoning in helio-centric model of Copernican theory is almost completely different from the one in our helio-centric model. Copernican model was deeply influenced by the Neoplatonic philosophy which he first met at the University of Cracow. Of course, Copernicus, as a Catholic priest, knew well that Platonism or Neoplatonism retained various dangerous factors to the orthodoxy of Christianity. As St. Augustine had already pointed out, in Neoplatonism one could see an inclination to magic, or rather paganistic mysticism. But as a child of Renaissance, he, Copernicus did not hesitate not only to accept Neoplatonism but also to try to reconcile it with Christianity. Thus, he believed an animism of heavenly bodies on the one hand, and also believed the Creation of the world by God on the other. It was this half-animistic and half-orthodoxical philosophy that led Copernicus to this helio-centric model. So the conceptual framework in which Copernicus propounded his helio-centric model is quite far from ours, and in that sense, Copernican model is neither identical with ours, nor « scientific » as in today's sense of the word. In general, each seemingly « scientific » theory of the past was constructed within the particular conceptual framework of its particular historical period, and we could not overlook the point.

It is in this manner that historiography based on the « pluralism in historical periods », which seems to share the same ideas of « ethnomethodology » in sociology, and of the « theory of cultural equivalence » in cultural anthropology, has finally begun to exert a strong influence in the area of history of science.

The belief that science is always based on a rational methodology became imperiled as it became clear that the knowledge and theories referred to as « scientific » today were discovered or conceived under conceptual contexts that were far from « scientific », and were therefore devoid of the justification that science was supported and put into practice at all times through one rational and ordinary method.

At this critical stage, the line of thought advanced by K. Popper, I. Lakatos and so on bolstered the rationalist position. Popper belongs to the demarcationists, who always put the difference of « scientific » from « unscientific ». He proposed « falsifiability » as the rational criterion for the demarcation ; that is, when a theory satisfies this criterion, i.e. the theory can beforehand define the conditions of its falsification, it should be considered to be « scientific ». But Popper admits that actual history does not always follow the principle of falsifiability, which is to say that in individual cases of tangible and actual theory-change, falsifiability does not function at all. He brings into discussion « the question of rights » in the circumstances in which we have chosen a theory as « scientific » and have rejected ones of the past as « unscientific », Popper claims that all through the time of the history of science, in hindsight, it is dominated by one consistent rationality, falsifiability, and we can always reconstruct history in a rational way according to the principle of falsifiability.

The most radical counter-argument to this rationalistic position is made by P. Feyerabend. He negates rationalism, stating that « anything goes » would be the only possible rational principle that rationalists can find in history. Moreover, he contends that the today's « success » of our natural science is not the result or the outcome of rationality according to methodological standards that science is supposed to have but, rather, it has been brought out because modern society has systematically supported modern science exclusively, both economically and morally.

Such anti-rationalism, most pointedly represented by Feyerabend, also rejects the myth of a universal rationality in the analysis of the methodology of knowledge. As in the famous « paradigm theory » of T. Kuhn, the existence of a universal methodology that works consistently throughout the history of science is denied. An explanation of the various processes of creation, selection, rejection and replacement of knowledge as anarchic or chaotic processes has made its appearance in the field of methodology (or rather, non-methodology). As a corollary, attention is also focused on the question of incommensurability, that is incomparability, incommunicability or intranslatability alleged between tho theoretical systems which seemingly have logical or syntactical continuity such as classical mechanics and relativistic mechanics.

It can therefore be said that the question of rationality/irrationality that surrounds science, or more widely, general knowledge, has given rise to varied arguments from several overlapping areas.

§ 3. Insofar as the preceding arguments are concerned, I personally adhere to the irrationalist or anti-rationalist side. For instance, I believe that incommensurability, said to exist between classical mechanics and relativistic mechanics or between classical and quantum mechanics, does exist. It is also clear that the rational reconstruction of history is not a proper method to grasp actual history.

However, if the discussion is halted at this point, we would have great obstacles to tackle with. For instance, let me take up the question of incommensurability. To contend that incommensurability exists between two theoretical systems A and B destroys the very concept of incommensurability since in order to admit the existence of the incommensurability of A and B, one has to be able to « compare » A and B and yet to contend that incommensurability exists between A and B, would in the strict sense of the word, imply that it is impossible even to compare the two. In order to solve this contradiction, a logical base to claim incommensurability becomes necessary.

Among the possible solutions of this problem, I want to propose the holistic approach to knowledge. As long as we separate each theoretical system and look upon it as an isolated system, the comparison of a number of theoretical systems becomes impossibles, based on the argument of incommensurability which does not escape contradiction.

However, by placing such theoretical systems in a broader context, we seem to find a common basis for comparison that assures us of claiming incommensurability among the theoretical systems. By giving a concrete form to this symbolic proposition of placing theoretical systems in a broader context, I propose a triad structure of man, i.e. : of community, language and world.

a)The Multi-Layered Nature of Human Existence

It is possible to determine the form of the existence of a human being on the basis of the community in which he lives. When this human being belongs to one community of experts in a certain scientific area, that individual is influenced by the common language used in that scientific community — a sophisticated « theoretical language ». He thus comes to articulate or organize the world according to the language of his own community. Such a theoretical language, once it has been established, is closed and self-sufficient. As a result, when this person is put into alternative theoretical system, he either understands alternative theory in the light of his own theoretical system or does not understand it at all. In other words, here incommensurability arises for him between his own theoretical system and alternative systems. However, an individual does not always live as an expert in a single area. This person can also be an ordinary individual — as a member of a wider community using a common, ordinary language — or be a semi-expert in another specialized area. That is to say, a human being has a multi-layered structure that fluctuates dynamically among varied layers. And at the more basic layer(s) he can compare his own theory with other alternatives. This structure allows us to claim incommensurability without the contradiction mentioned above.

b) The Multi-Layered Nature of Language Systems

Together with the above multi-layered nature of man, language systems used in various communities also show a multi-layered character. For instance, the use of « weight » in ordinary language and the use of « mass » in the theoretical language system of elementary physics have a short of correspondence. And there is also a correspondence between the use of « mass » in the theoretical language of elementary physics and the use of « mass » in the more theoretically advanced particle theory. The semantics of « weight » in ordinary language and the one of « mass » in the theoretical language of elementary physics are by no means identical, and yet they correspond each other to the extent that the role of each in its respective language system can be said to be almost « identical ». It is therefore possible to imagine a multi-layered structure where highly theoretical language systems are based on more basic language systems.

c) Multi-Layered Nature of the « World »

In thinking about the « world » amid this multi-layered nature of man and language, we can also reflect on the world's multi-layered character. If the premise is that the « world » for us is articulated by language systems, it becomes almost evident from the above that this « world » is multi-layered. And the pictures of the world corresponding to each layer are « superposed » in quantum mechanical sense.

§ 4. Should we extract a fixed, isolated theoretical system from this multi-layered man-language-world triad as proposed above, we may find that incommensurability exists, but no matter how sophisticated the theory may be, it does not mean that it is always fixed, isolated and closed. Corresponding to the manner in which the existence of man always fluctuates from one layer to another dynamically, the language system also varies within a multi-layered structure from time to time, as does the « world ». In the sense that such movement among layers is not caused by permanently unchanged and universal cause-effect relationships, we cannot claim total « rationality » in our knowledge, but because it takes place in such a structured environment we can avoid reaching the conclusion that our knowledge is of a totally « irrational nature ».

This proposed model is not intended to serve simply as a means to solve the contradiction of the arguments of incommensurability ; I believe it will serve as a bridge between scientific theories and our daily lives. I also believe it will provide an answer to the questions of what constitutes theoretical knowledge, in what kind of structure simultaneous discovery occures, and what is the relationship between common thought and sophisticated theory, as well as numerous other problems surrounding scientific knowledge.

This proposed model denies the existence of an absolute, universal rationalism concerning knowledge. It seeks to find the possibility of a new rationality that could be called a « non-irrational structure » of knowledge or simply « soft rationality » of knowledge which allows us to admit the pluralism of rationality.

To conclude, the history of manking might ever be a travel to seek another rationality in the « non-irrational structure ». And I believe that this conclusion encourages us to sail out into the chaotic ocean of knowledge searching for a new rationality.

Economic and Social Effects of Factory Automation

Shuhei Aida
University of Tokyo

1. TREND OF AUTOMATION TECHNOLOGY

Rapid advancement in microelectronics has brought about great innovation of automation technologies, affecting individual workers as well as creating problems in social, economic and even in cultural contexts. Both physical and psychological maladjustment to machines among individual workers and groups of workers with different cultural backgrounds has caused growing concern. Expansion of factory automation (FA), in particular, has reduced satisfaction with jobs and dignity of labor among people at workplaces and divided skilled workers from unskilled workers. These are some examples of growing gaps caused by the expansion of FA, which are being observed internationally both in industrially advanced and developing countries. Accurate understanding of these problems and their settlement are most needed for future development of automation technologies.

Automation is primarily aimed at setting people free from painstaking and insignificant jobs as well as equally distributing benefits and wealth produced by modern technologies. Automation in the past were concerned with mass produced by modern technologies. Automation in the past were concerned with mass production of limited varieties of goods, thus lowering production costs and enabling a great majority of consumers to obtain lower-priced but better quality products. We can say it stemmed from the principle of « equalization ». But this equalization has been steadily getting linked with « efficiency » with the economic development and technological innovation as automation has shifted its importance in factory thinking to efficiency combining people,

goods and money. This has further pushed human alienation, raising interest in the development of « humane-oriented » automation technologies.

This has been brought about as a natural consequence of our high technology age and therein lies the importance of looking for research subjects matching new philosophy of technologies. It is evident that dissemination of industrial robots will have a great impact on society as a whole as robots have come to possess the ability which could replace human workers in many job situations.

Historically, mechanization of human labor became a social problem for the first time in the early 19th century when British workers revolted in what was known as the Luddite Movement. In the movement, workers in the British weaving industry tried to prevent new machines from coming into their workplaces by destroying them in a demonstration of their objection to the mechanization of weaving machines. But such a violent anti-machine movement gave way as the principle of equalization gained popular support as mechanization was getting spread in general society, giving birth to industrially advanced countries as we see them today. Automation technologies have penetrated into manufacturing industries as well as agriculture and civil engineering in many countries except developing countries.

Against this historical background, I try to look into factory automation (FA) in social, cultural and individual contexts, and identify the meaning of the deployment of automation technologies and its development goals. I will further discuss human nature, both as individuals and in groups, and social and cultural standards in relation to maladjustment of automation technologies, to assess farreaching effects produced by automation technologies from a more global point of view, positively as well as negatively.

Deployment of automation technologies, particularly in a new age of information, is not confined to factories alone. The situation prompts us to recognize the importance of discussing the process of feedback, in a wider social and economic scope involving consumers, as the basis of automation technologies. Simply stated, it means reorganization of FA by introducing high technologies in response to what consumers want or desire, representing a departure from automation technologies in the past which were organized and operated solely in response to intentions of manufacturers. The auto industry is a special case in point where engineers should look at technologies with the awareness that consumers are designers as well. We are now entering into a new age of rapid progress of automation technologies in almost every aspect of industrial activities, and I am trying to look into how this should be organized to fit an expanded social and economic framework, presenting an outline of our goals to make this adjustable to general society, then consider the optimality of automation technologies.

2. SOCIAL EFFECTS OF AUTOMATION

Although automation technologies created some negative reactions in the past, they have been accepted in industrially developed countries

because they have enhanced both productivity for employers and industrial safety for workers. The introduction of automation technologies has brought about a notable change in the decline of workforce required on the « manufacturing stage » of the entire production process. In contrast, this requires more workforce in market research, designing and other work on the « preparation stage », and distribution, sales and marketing on the « support stage ». As shown in Figure 1, jobs of the support and preparation stages require workers with more sophisticated skills than previously needed to conduct jobs of the manufacturing stage. It is quite normal, therefore, that professionally conscious workers should demand retraining for their new jobs and participation in a new automation technology society.

There are, no wonder, big differences emerging with the progress of automation technologies in retraining of workers and redistribution of human resources between Japan — a country of better job security and corporate education — and the United States and European countries where there is a general lack of such benefits for workers. As clearly demonstrated by a relatively higher rate of unemployment in the United States and Europe, redistribution of workers in an intra-company situation does not go smoothly in most cases in the United States and Europe as it does in Japan because of the general absence of corporate education and the structure of U.S. and European labor unions which are organized according to workers' jobs, not their corporate membership.

There was a time when the market expanded steadily as productivity went up, and when this happened on a global scale, there was market stability and the world economy grew up and up. But as the background of the current trade dispute clearly shows, technological innovation and market expansion do not go hand in hand any longer, instead they produce negative elements, both visible and invisible. The new age of automation technology requires us to understand all these problems and settle them in an expanded macroscopic framework.

It is commonly known that « human egineering » came into being as a new field of science about a bit more than 40 years ago, combining subjects of technology and psychology. This was designed to give birth to experimental science for the purpose of inter-adaptability of mechanical and human functions on the basis of examinations of the size, strength and speed of a machine, and human eyesight and power of hearing and memory.

The new science takes various human functions into consideration in designing automation technologies, thus making human backup system for such technologies more effective and contributing to their greater activation.

This can be seen in aspects of both hardware and software, and is credited with increasing the potential of automation. Technological innovation, particularly in the fields of information and communication, has created new feedback loops between humans and machines at various levels. We should be well aware in designing automation technologies that « systems are based on humans » and « humans are based on systems ».

In other words, it is vital to look at automation technologies in cultural contexts as « social engineering » is concerned with. Every possible effort has been made in industrially advanced countries to « educate people » in a bid to spread automation. Generally speaking, automation technologies are more likely to be accepted in culturally advanced countries, finding their way into many fields of industrial activities. Sophisticated automation technologies, even if forced upon a society of primitive agricultural economy, could not take root there in the absence of knowledge and skills to use them. In short, whether automation technologies will succeed or not depends on a particular culture where they are introduced.

American and European engineers often look in surprise at the high level of productivity of Japanese workers and better quality of Japanese products. But when they do so, it seems to me, Americans and Europeans tend to miss the social backgrounbd contributing to the higher productivity and better quality products. Industrial efforts such as corporate education and welfare measure are made in Japan in order to get workers involved in business management, and this has been associated with activation of industries through factory automation (FA).

In an obvious reflection of growing U.S. and European interest in this aspect of Japanese industrial relations, there were reports and lectures by guest speakers on this particular topic at a general meeting of the International Federation of Automatic Control (IFAC) which was held recently. I would say this was an indiction that Americans and Europeans are beginning to understand and positively evaluate, at least to some degree, the consensus-based Japanese corporate decision-making process which regulate Japanese group behavior and activate Japanese industries.

In introducing automation technologies, we should know what technological and economic resources we have gathered spontaneously in our own companies and examine how these can be utilized. In order to use such resources, we should look at the entire process of production as mentioned earlier with the aid of philosophy fitting factory automation. Automation technologies should be introduced into the production process, paying attention not only to the process itself but also to steps coming before as well as after the process.

Factory automation in the auto industry, incorporating every available high technology, has undergone sophistication to the extent that it has become a symbol of this kind of technological systems. This has come from the fact that automobiles are directly linked to consumers who are gettting more quality conscious and thus contributing in turn to the output of better products in some ways or others. Various forms of social enforcement such as product safety required as transportation vehicles has also stepped up the factory automation in the auto industry which has taken advantage of high technologies. The most important thing, after all, is the concept of global automation systems as shown in Chart 2 and the idea is being put into action in the auto industry. The auto industry goes beyond any particular local region, national bounda-

ries and cultural barriers and major research subjects remain to be undertaken in this nature of universality.

When we look at automobiles — products whose value is internationally recognized — it is vital to examine social effects of technology from a more global point of view as well as every technological innovation and its ripple effects, not to mention efficiency in the entire process of the industry as a whole from the preparation and manufacturing all through to the support stages.

3. CHANGEOVER OF PRODUCTION LABOR, AND SKILLS AND KNOWLEDGE

Advanced FA technologies are being completed as systems which can be introduced into the auto industry as well as every other industry. Computers at the command of a limited number of operators through displays control a variety of machine sections of a factory and undertake pieces of jobs one after another. Operators at all times monitor commands to computers and how the computers put them into action, adjust the parameter of the whole systems, and find out malfunctions and deviations from what have been predicted. And in the event of abnormality, operators analyze what has caused the situation and turn the systems into manual operation and control the production process on the basis of their apt experiences. Only very rarely is it possible to predict what ability is required of operators to find out errors and other forms of abnormality and how they should respond to them. Operators could be thrown into a succession of tensions at any time, without any warning, after being subjected to long hours of boredom. They are both in charge of similar kinds of machine operation in a smooth succession and pieces of operation unrelated to each other, emerging out of context. In this sense, what operators are doing resembles the role of business executives. In fact, it is no coincidence that pilots of jumbo jet liners are called « flight managers ».

With this in mind, let us take a look at supervisory control systems and they come out as follows. « Local computers » through a process of repeated simulations arrive at the best possible command for a robot. The computers are also used to process data, alter control commands and correct displays. When it comes to communication between local computers and terminal equipment at each stage, the time lag and frequency limit needed for such communication become important problems. The interface between each satellite computer, and sensors and terminal equipment at every stage is automatically controlled by computers. But these do not come out in a chain. Programs are set to conduct control actions at every step and there are loops enabling operators to manually control as required.

The existence of these systems helps us understand how fast production labor is changing over and technologies and knowledge are being subjected to major innovation. It is important here again as demonstrated by Figure 1 to pay attention to redistribution of human resources.

4. FA AND NEW CONCEPT OF SYSTEMS

We have to return to Figure 1 when we look at factory automation in the auto and other Japanese industries from macroscopic viewpoint. With the diversification of parts and spread of automated production processes in the auto industry, jobs at the preparation stage are increasing both in terms of quality and quantity. Attempts to push the efficiency at the production stage adversely affect jobs at the preparation and support stages. The production process using what is commonly known as the « Just-in-Time » formula has brought about delivery of parts according to a timetable and helped to get rid of storehouses in the factory compound. But it should be reminded that those auto parts are now being carried by trucks and the « moving storehouses » dominate on freeways. They often converge on parking places simply to kill time for delivery of their loads right on the timetable. This also means that the most effective production process in a particular company situation affect public transportation systems and causes many social problems. Looking at the problem from a viewpoint of Japanese industries as a whole, the production process using the « Just-in-Time » formula means a change in the form of storing parts and products. With this in mind, we should be aware of the importance to draw up future prospects of factory automation in an expanded framework in accordance with Figure 1. In doing so, it is vital for us to examine the structure of automobiles themselves, and to promote reviewing the way of assembling automobiles and delivering them, not to mention the number of parts.

A review of the contents of jobs on the preparation and support stages enables them to encourage feedback among themselves to a certain degree and also to streamline the production process further. Needless to say, it is important for the auto industry to carry on research and development projects to improve the efficiency of automobiles, but in view of its influences on general society, what is really important for the industry is to evaluate its far-reaching effects on society and pursue new research subjects within the global systems.

Japanese automakers have succeeded in improving the quality of their products by taking advantage of the fruit of advanced technologies. However, when we look at the magnitude of impact of the auto industry on Japanese industries as a whole, the industry's initiative to make its advanced system technologies available publicity will provide means and ways of their survival and future prosperity. The automobile industry should promote factory automation in an expanded social framework with the help of basic philosophy as part of an attempt to get out the situation in which moving parts storehouses (trucks) more or less occupy freeways as its products roll on there, too.

Factory automation in a new age of information aims at establishing production and distribution systems of the auto industry toward the 21st century, getting hold of the trend given in Figure 1 at various levels through the way of thinking emerging when factory automation is placed in a global perspective.

In discussing artificial intelligence (AI), a popular topic these days, we should look at it not only from the viewpoint of robot technology alone but also as a research subject of promoting intelligent substances in a more expanded framework. A good case in point will be the establishment of intelligent terminals in local regions as part of production and distribution systems with the aim of more active information transmission, comprehensive activation of « matters-energy-information », and then greater efficiency and prosperity of the automobile industry. In that sense, we can say activation of the automobile industry, as it goes along with the spread of factory automation, comes from technology philosophy through « new concepts of system thinking ».

BIBLIOGRAPHY

T.B. SHERIDAN, T. VAMOS and S. AIDA, « *Adapting automation to man, culture and society* », Automatica, Vol. 19, N° 6 (Nov. 1983), pp. 605-612.

S. AIDA, ed., « *The humane use of human ideas — Discoveries project and Eco-technology* », Pergamon Press, Oxford, 1983.

S. AIDA, « *Eco-technology ; between culture and technology* », Information Resource Management, ERICSSON, Nov. 1984.

S. AIDA, « *A concept of eco-technology ; technological approaches to Complexity* », The Science and Praxis of Complexity, The United Nations University, 1985.

S. AIDA et al., « *Culture and Automation* », Proc. of IFAC Congress 1984, Pergamon Press, Oxford, 1985.

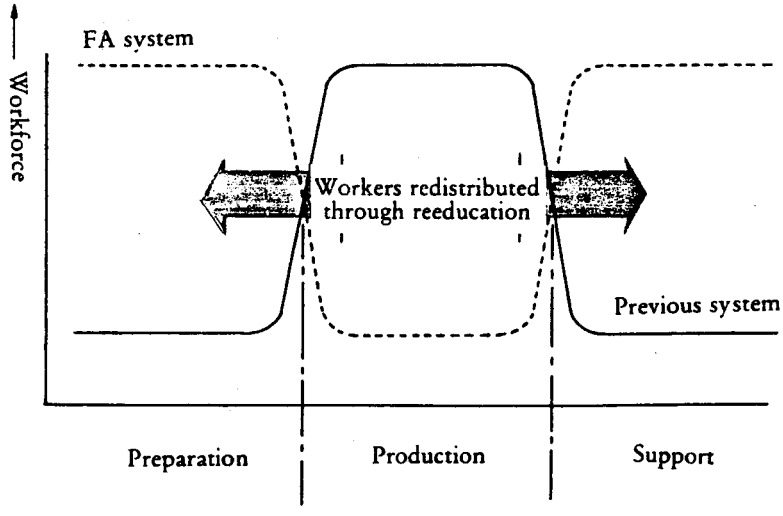


Figure 1
Shift of workforce in preparation
Production and support stages

Notes on the Limits to Certainty : Risk, Uncertainty and Economic Value

Orio Giarini
The Risk Institute

1. SOME PREMISES :

— the use and abuse of natural sciences as legitimation for social sciences and political ideologies ;

— the problem of demarcation in economics : the notion of value.

Deterministic science has been largely referred to and used by social (and political) theories (and ideologies) as a legitimizing reference for their own goals. This very ambiguous crossfertilization has given good and bad results.

The level of uncertainty and indetermination of social sciences has remained so great and evident, that it has rather produced the idea of inconciliability between « hard » (and very often assumed as precise, deterministic) sciences and « soft » sciences. The modern scientific revolution makes more and more apparent the fact that even « hard » sciences do not produce *definitive* truth, but *better* truth.

The notion of uncertainty should help to stimulate the improvment of social sciences and should not be manipulated to reevaluate processes or tendencies of thought which priviledge inaccuracy, sectarism, etc. The possibility of a continuum between hard and soft science through the notion of real time and uncertainty is probably a great ambition, but one should care *not* to find in the new scientific developments a legitimation for new social, economic, or political theories (as it *has always* been done in the past), but just the contrary. All paths of research must find in *their own* level of uncertainty and relative ignorance, the way to grow and to develop, without using the reference to other outside paradigms, as the reference authority for a (pseudo-rational) legitimation.

It is important in the case of economics, to show how many of its basic postulates are justified by and refer to deterministic scientific paradigms of last century. From there on, on the ground of a more open possibility of thinking the economic problems, we can start striving for a better economics. This does not mean that we cannot be stimulated by analogies in other sectors or sciences, but the origin of the stimulus cannot be confused with the pretention that the quest is « proved ».

The reductionism of current economic models and theories is also understandable because of the problem of demarcation of economics and its ambition to be or to become a specific, respected science (not to be diluted into something too vague and general). The notion of value and equilibrium is a strong-hold in this sense. But we feel that it is also essential to be sure that the theory of value correspond to the problems to be solved.

2. *ECONOMIC THEORY AND THE INDUSTRIAL REVOLUTION :*

The problem of measurement of economic value and the « objective » function of the « invisible hand ». :

— the equilibrium of supply and demand and the assumption of certainty ;

— monetarization of the economy and the development of technology (the development of capitalism).

The yardstick for the measurement of value in classical and neo-classical economics, is provided by the price of a goods exchanged, where supply and demand meet in an instant time dimension. This system is conceived as « general equilibrium » (as a « Newtonian » system), and it is based on the assumption of certainty. It is the « invisible hand » which, for Adam Smith, regulates this equilibrium.

The importance of the « invisible hand » lies primarily in the fact that it suggests the idea that price is an « objective » social measurement. This aspect is much more important than the interpretation of the notion of the « invisible hand » as a fundamental plea for economic liberalism. It is the social « objectivity » side of it which really matters and opens the door to the « scientific » pretensions of economics in all « political » directions. It is not really an accident that Marx was to a great extent a follower of Adam Smith (the « invisible hand » depends for Marx on another « objectivist » concept : the class struggle. The search for objectivity has the *same* philosophical foundation).

It is essential to note that this measurement of « production factors », is a measurement of flows and not of a stock, where it is assumed that the two coincide because of generalized scarcity. Productive work and industrialization are also considered as the same process.

Thanks to the tool provided by quantifiable (priced) values, economists have hoped to become as « scientific » as researchers in the natural sciences, who can better determine the systems or models they analyze

through the measurement of specific factors (heat, speed, weight, inertia, resistance, wave-length, etc.).

If priced values are the key to economic analysis, then economics will inevitably focus on *monetarized* production and consumption systems. Although classical economists still wrote extensively about non-monetarized use values, they did abandon with time this notion, on the one hand because they worked more and more on the assumption that what is « free » is not scarce and therefore place it outside the economists' pale (which, in our view, is largely wrong), and on the other because they saw no way of measuring them properly, given the positivistic-deterministic cultural conditions of their time.

It is indeed obvious that the notion of priced or monetarized values is necessarily linked to the situation of an industrializing economy where exchanges and production follow the monetarized pattern. It was precisely from the birth and development of the industrial revolution that the wide diffusion of the monetarized system took place. One should note the great confusion that usually prevails about the studies written on money (its origin and use) and the question of the *transition* of economic activities from a non-monetarized to a monetarized pattern. A key question today is that of the *interdependence between* the two systems.

The reason for the decisive growth of monetarization during the industrial revolution is closely linked to another fundamental factor : the development of modern technology. The more « modern » the technology, the more costly it is, which means that more « capital » (accumulated money) is required. In order to accumulate money efficiently, the economic system must be increasingly monetarized. During the eighteenth century the « capital requirements » in England were, as Adam Smith said, much higher than ever before in history (up to a maximum of 5-6 % of sales...). This percentage reached 12-15 % in the last century and 25 and 30 % in our time. The diffusion of monetarization is a key social factor allowing to solve the logistic problems to put new industrial technologies to use.

Industrialization therefore is essentially the history of technological development *and* capital requirements, of specialisation and of monetarization. A society which is not monetarization-prone has much greater difficulty in using and benefitting from modern technology, as this has evolved in the Western World.

3. THE PROBLEMS ECONOMIC THEORY HAD TO SOLVE DURING THE PERIOD OF THE INDUSTRIAL REVOLUTION :

— the priority given to develop tools and material products in a monetarized environment to fight against scarcity ;

— facing bottlenecks : from a supply-oriented notion of value to a demand-based notion of value, within perfect equilibrium ;

— the service economy as a system ;

— the utilization value.

« *The first theories — the first tentative solution problems — and the first problems must somehow have arisen together* » (from Karl POPPER « Unended Quest », Fontana Collins, Glasgow, 1977, p. 133).

It is important, in our view, not to underestimate the fact that *economics* itself is a *consequence* of the birth of the industrial revolution.

The problem at the origin of Economics was the scarcity of resources and the need to improve (by choice) and ameliorate (through social and technical tools) their output.

A theory of economics, as tentative solution to solve the problems of scarcity, has been, therefore quite understandably — from the first coherent theoretical system of Adam Smith up to the end of last century — supply (i.e. production) oriented.

From the beginning of the Industrial Revolution, new technolgies have enormously increased the capacity to use and exploit resources and to induce a powerful explosion in production. The power of technology (and the capacity of society to use it) has been in particular accelerated by the exploitation of scientific knowledge : a process which started in the middle of last century and came to full swing at the end of the second World War (since when and for a quarter of a century a historical unique period of growth has taken place in the industrialized countries).

This powerful extension of the potential of production has little by little switched the nature of the problems of economic development towards consumption and demand. Bottlenecks in the way to develop the wealth of nations, have been identified more and more from the end of last century and fully from the 1930s as dependant on solving the problem from the demand side. Economic theories (from Keynesianism to monetarists) in conjunction with the social evolution (i.e. the acceptance of labour unions and of their strategies for wage increases), have contributed to eliminate the bottlenecks of demand.

In theory and in practice, corroborated by the success of the twenty-five great years of growth, demand has become the key problem-theory. Consumption is considered as the priority factor in stimulating and regulating production. But the general equilibrium frame has remained the same : first, supply was by definition equal to demand, now it is demand which is equal by definition to supply.

The implicit assumption in most current economic models is that science, technology and societal adaptability are developing at such a speed, that supply (production) will remain elastic (responsive) enough to the stimulus of demand. Economists still often tend today to consider implicitly that technological development is an economic endogenous factor (of supply).

Supply, in particular after the 1960s, has become more and more rigid for reasons which are :

— quantitative (the phenomenon of the diminishing returns of technology for all the sectors which have been the back-bone of the traditional industrial revolution), and

— qualitative (in the contemporary economic process the notion of wealth production has to be more and more extended to the notion of wealth utilisation : the service functions in *any* type of economic activities have become the predominant economic element in terms of costs). The main shift of interest is from mere production of tools and products, to the systemic conditions under which they are usable and they are functioning.

The new problem for economics is now linked to a new type of rigidities at the supply level.

These new rigidities are different from those of the first period of the industrial revolution because they have less to do with bottlenecks and rigidities of material production (of tools, machines, final products), but much more with the functioning of all service functions (research, organization, storage, maintenance, protection and security, quality control, users' education, recycling, waste disposal, etc.), which condition not only production itself, but in particular its usability and utilization.

In parallel, the traditional service functions and activities include more and more tools and industrialized processes so that the distinction between what is a manufacturing and what is service activity is less and less apparent.

If the shift, in terms of the key problems' center in economics today, is from the mere production and existence of a tool or a product to the complex system which makes these tool and products effectively usable, the economic theory and analysis have to be enlarged by the introduction of the notion of real time (and of uncertainty which is necessarily linked to it).

In a situation in which the utilization of products is the real indicator of economic value, it is their life *duration* which becomes a focal problem and all the services which make their utilization possible.

We submit to consider the following notion of utilisation value in time :

P = Production phase
U = Utilization phase
W = Waste (treatment and recycling) phase

While : — U depends on products *and* services, and on monetarized as well as on non-monetarized factors ;

— Optimization of resources depends on costs of P + U + W, with respect to U (time and frequency of utilisation) ;

— SP is the selling point (the reference value for « equilibrium » economics) ;

— P, U, W have a real time dimension (duration).

Measurements in this case should probably be oriented more and more on :

— variations of stocks
— indicators of results,

where the value added measurements, taking into account also their gross and net variants, become indicators of specific subsystems, and not of *the* entire economic system.

4. *CONTEMPORARY ECONOMIC THEORY AS AN INCOMPLETE THEORY TO COPE WITH THE NEW* SERVICE *ECONOMY :*

— the real time factor between supply and demand ;

— the economy as a dynamic disequilibrium system.

« *I shall say nothing more that values emerge together with problems* » (p. 193). « *The reality of time and change is the crux of realism* » (p. 129) from Karl POPPER « Unended Quest » op. cit.

There can no longer be any « scientific » justification for considering a state of equilibrium in economics (as referred to the Newtonian model) as the necessary premise of economic analysis. In some cases, equilibrium might be desirable, but economic progress could well depend much more on specific and desirable states *of non-equilibrium* in cases where the isolated industrial system opens up to a multiplicity of new functions and interactions typical of the service economy. The key economic question of the future, rather than be « how shall we achieve a perfect (certain) equilibrium ? », might well become « how shall we create or stimulate productive non-equilibrium situations ? », situations which, contrary to « Newtonian » philosophy, have a real time dimension.

Today, those economists interested in the notion of uncertainty, still see it as a problem of imperfect information in view of a general equilibrium situation (where information would be perfect), or as a problem concerning only exogenous factors... On the contrary, introducing the notion of real time, information can never be perfect and it is precisely for that, that there can be responsibility and development. In this sense, social and economic systems are clearly disequilibrium systems. The problem is then to identify those disequilibria which are more acceptable (or less unacceptable) than others. The vision of equilibrium is corresponding simply to a static picture taken by a photographic tool, which, by necessity, shows an *ex post* situation. On the contrary, in real life, we have to live with *ex ante* images or hypotheses which are, by necessity, more or less blurred, and more or less uncertain. But it is this uncertainty which allows to have a goal.

Furthermore, introducing the notion of real time confronting economic supply and demand (service-based production and consumption in modern terms), is quite a radical alternative to the view of the economic process as based on timeless equilibrium.

It recognizes the fact that any decision to produce is inevitably taken in a situation of (greater or smaller) uncertainty, with regard to the moment when the outcome will be available and presented to the « market ». In a dynamic view of the economic process, it should be recognized that any decision to produce is taken *ex ante* the traditional moment of the

economic equilibrium and that any price analysis is always made *ex post*. *Ex post* analysis can be considered as (tautologically) certain.

This implies that any economic activity or endeavour is based in fact on some unknown and uncertain factors or possibilities. Even more, many competitive redundant production processes are continuously being proposed, of which only a part will reach the point of sale and / or the moment of utilization.

There are many uncertain productions which are then *selected* by demand. And there is a great difference between an equilibrium process (between supply and demand) and the process in which demand has not an equilibrium, but a selection function.

Something parallel to the above point has been written by Karl Popper in his refutation of induction and defense of empiricism :

« *There is no induction : we never argue* from *facts* to *theories, unless by the way of refutation or « falsification ». This view of science can be selective, as Darwinian. By contrast, theories of method which assert that we proceed by induction which stress verification (rather than falsification) are typically Lamarckian : they stress* instruction *by the environment, rather than* selection by *the environment* » (p. 86) from Karl POPPER « Unended Quest » op. cit.

In a similar way, current neo-classical demand-based economics sees demand as giving instruction to the economy, how to do things : in this way, a fundamentally deterministic philosophy still appears as permeating social sciences and economics in particular. By contrast, even if a process of selection can provide some hints and information, as to the way it will work, it will always be a simple hypothesis as it normally is in practice, which only will be verified empirically, *later*, by facts. But at the same time, a zone of uncertainty will always be preserved, because of the fundamental impossibility of forecasting a fully predictable environment if real time, evolution and dynamics are accepted as connotations of real life.

In my opinion then, we are now in a situation for which a static equilibrium theory of economics doesn't help to solve our major problems. Our hypothesis is that we are in a situation in which economic theories are too incomplete to be really efficient. This therefore means that they have to incorporate more problems and less incomplete hypothesis. This excludes also the idea that, by stressing the importance of supply, we would go somewhat backwards to the older economic thinking. Time dimension gives a much wider sense to the production function than it had in classical economics.

Whereas economic theories could swing in the past from supply to demand as priority, and be considered as workable instruments, we now need not only to reconsider the importance of the supply side, but also the fact that the *selection function* of demand is an absolute necessary complement to the production function.

Going back to the quotation of Karl Popper, we could say by analogy, that an economic system has to produce on the basis of hypothesis (and

may be even dreams or any other process stimulating action and initiative) : this is the first essential step ; then the demand process must be also as efficient as possible in his selective function (including criterias having to do with the best use of material and human resources and societal values).

5. *KEY ECONOMIC PROBLEMS IN CONTEMPORARY « SERVICE » ECONOMY :*

— risk and uncertainty in the service economy ;

— employement and productive activities.

The service economy as a system and the notion of risk :

As we have already proposed, in the service economy, the notion of system becomes essential. A system produces results or economic value when it functions properly. The very notion of functioning, requires to consider real time and the dynamics of real life. Whenever real time is taken into consideration, the degree of uncertainty and of probability which conditions any human action aiming at a result, becomes a central issue.

A system which is working to obtain some future results, is by definition in a situation of uncertainty, whereas of course, different degrees of risk, uncertainty or even indetermination characterize different situations. But in any case, risk and uncertainty are not simply a matter of choice : they are part of the human condition.

Rationality is therefore not so much a problem of avoiding risks and eliminating uncertainty, but of adapting to risks and uncertainty, accepted as given situations.

Furthermore, the very systemic nature of modern economic systems and the higher and higher degree of technological developments, require a deeper and deeper economic understanding and control of the system's *vulnerability*.

The notion of vulnerability is generally misunderstood. It might seem to be a paradox to say that vulnerability increases with the increase of the quality and performance of modern technology. In fact, the higher level of performance of most technological advances has as a consequence to reduce the margin of errors that the system can withstand to be able to continue to operate. Accidents and mismanagement must then happen and they do happen more and more rarely, but their effects have more and more costly systemic consequences. The opening of a door to the outside from a running automobile does not necessarily lead to a catastrophe. If it is an aeroplane, it does. This means that the notion of system functioning and of vulnerability control becomes a key economic function where the contribution of the economist and of the engineer must integrate. In a similar way, one can talk of vulnerability management considering the problems of social security and savings for the individuals.

As a consequence of having to do with systems, the notion of risk, the management of vulnerability and uncertainty becomes a key connotation of the service economy. The notion of risk itself has to be considered as referring to two complementary but substantially very different situations :

a) The so-called entrepreneurial risks where an action is taken with the hope of a positive result. The dynamic development of a service economy system — which is potentially growing more and more complex and diverse — is probably providing more and more space to the initiative and freedom of action of the individuals.

b) Any type of action is submitted to accidents or interruptions due to events outside the possibility of control of those involved in the functioning of the system itself. This is the so-called pure risk.

Evidence of the system functioning of the new service economy and of the relevance ot the pure risk is given by the fact that during the 70's, the world over, the real rate of growth of insurance activities has continued to be of the order of six percent per year (as in the previous two decades). Following the predictions of the Engel's law, this rate of growth should have fallen at least if not even lower than the level of the general economic growth. But the difference was that in the meantime the necessity of greater investment in system functioning had increased with regards to previous historical economic crisis periods.

It is within the framework of these notions of systemic developments and of risk that the problems related to the environment acquire their full economic relevance and significance, in the present situation.

Strategies for employment and the development of productive labour in the Service Economy :

These last notes concern the major problem economics should try to solve now and in the future : the better use of resources and in particular of human resources. Dr. Aurelio Peccei said several times, that there is something shocking in the fact of the obvious underemployment in qualitative and quantitative terms of the human potential : something must be very incomplete in the economic system if human power is too often described and considered as a problem and not as a potential for wealth production in the largest sense, for which adequate conditions for employment should be developed.

This, I think, is simply a question of realisitc idealism and should be far on top of economic and social priorities.

From an economic point of view, there are two different ways of approaching the employment problem :

a) To regard ourselves as merely facing a period of adjustment and adaptation of the industrial revolution which is reaching a new state of balance without nevertheless deeply altering its connotations. Here, the employment problem principally takes the form of a problem of investing in order to create enough jobs to absorb the available labour.

b) A different approach is to take the attitude that we are no longer living in an industrial society and that the old mechanisms can no longer provide adequate solutions. This means that the rules for production and the distribution of wealth, and those governing the mobilisation of productive forces are probably subject to a different logic from that of the industrial revolution. This is why we consider of the utmost importance to evaluate as far as possible the hypothesis of the service economy.

It is also obvious that the two approaches are essentially complementary. In any case, given the seriousness of the employment problem, every possibility of finding new, additional and better solutions should not be dismissed.

Considering the change towards the service economy, we would like here to introduce two lines of possible thought and action.

Productive labour and employment :

Considering the realm of economics as going beyond the monetarized economy, the problem of employment will probably require to reopen the old discussion of what is productive labour in a *service economy*.

This means considering the relevance of non-monetarized activities as far as they are economically wealth producing and the desirability of stimulating the best compromise (an optimisation) between the contribution of both monetarized and non-monetarized labour.

The social advantage of an economic theory considering both monetarized and non-monetarized activities as value producing could be to give more prestige to the non-monetarized sector and make it easier to develop it as a *complement* to the monetarized one. In turn, this would favour the present trend to more and more part-time monetarized jobs (of which non-monetarized productions are a complement). The *quality* of activities and jobs available in the service economy is also a key factor and probably represent a big chance.

Even more so that part-time jobs could become more than they are today a consistent complement for :

a) people after the present « retirement » age ;
b) for young people, to better couple education and practical experience ;
c) for women or men with a family obligation and any other group of people who wish to work for money, but do it only part-time.

The useful life of products and services and employment :

It is important, in the service economy, to define the value of the use of goods and services, i.e. what they bring as an addition to welfare and wealth during their useful life, taking into account all the costs which have been entailed (during production, use and recycling or waste treatment). In general, therefore, the idea of an instantaneous economic balance could be replaced by one of an optimum real time by a calcula-

tion of the costs/profits which should be made during the period of use (real values) of products and services. This reasoning is valid of course only insofar as technology, as is the case now in many sectors, no longer has the capacity to accelerate obsolescence in all sectors, as in the traditional period of the industrial revolution.

It must also be pointed out that, the shorter the useful life of products, the greater the energy consumption at the manufacturing stage (and afterwards, at the waste-disposal stage). Likewise, the longer the useful life of products, the more labour will be needed to cope with maintenance and repair. The cost of handling waste will in consequence be reduced.

Where greater capital-intensity no longer produces wealth (considering the notion of utilisation value), the use of labour becomes more and more competitive and this can be stimulated through a production structure which puts a true value on useful life and balances it over a period of time.

Stressing the useful life of products also means providing room to mobilize non-monetarized work. In fact, much maintenance, repair work and restoration depend in any case on it.

In both cases and probably in many others, it is the perception and analysis of the economy on the basis of a theory of the service economics, which might produce the mobilisation of will and action.

Hard criticism and verification of these hypothesis is urgently needed.

Toward a Nonequilibrium Systems Theory of History

Ervin Laszlo
General Evolution Research Group
Italy

Are the laws of nature applicable to the human world, in particular to the world of history ? Can society be conceptualized as a complex nonequilibrium system of which the basic dynamics are isomorphic with the dynamics of nonequilibrium systems in physics and biology ?

These problems have been debated for centuries and continue in our day. Many contemporary historians profess scepticism about the applicability of laws discovered in the natural sciences to society ; they consider history a record of « historical facts » rather than the exemplification of universal processes. The laws of nature are said to be inappropriate and inapplicable to the subject matter of history. History deals with the unique, the individual, and the concrete ; it also deals with events infused with human will and purpose. By contrast the hard sciences such as physics deal with the universal, the general and the abstract, and their subject matter is untouched by conscious purposes. The physical sciences postulate laws to predict and if possible control phenomena ; but prediction is impossible in the social sciences and control is undesirable.

The split between the natural sciences and the humanities has centuries-old roots in Western civilization. These cannot and need not be analyzed here. We should note, however, that inasmuch as the split prevents historians and social scientists from exploring the application of the concepts and theories of nonequilibrium systems — Prigogine's « dissipative structures » — to the development of human societies it is injurious to understanding in their field. Moreover whatever justification such reticence may have had in the past, it has none today. The developmental laws still resisted by historians are deterministic and mechanistic,

much like the laws of Newtonian physics. They apply to clockworks, not to organisms. But these laws are trascended even in the physical sciences. The evolutionary laws propounded by nonequilibrium thermodynamics and the related sciences of complexity describe how complex nonlinear systems maintain themselves and how they transform in a flow of energy in states far from equilibrium. They allow for the emergence of order and novelty from randomness and even chaos. They permit multiple choices ; they account for human freedom. Thus they are entirely appropriate to grasp large-scale patterns in the functioning and development of human societies. Indeed, the principles of the new sciences of nonequilibrium systems have little if anything in common with the scientific laws to which concerned humanists rightly object.

The proposition that human society evolves according to the same recently discovered evolutionary laws as other natural and biologic systems is inherently reasonable, but it is not one that can be answered by abstract logical analysis. In order to decide its validity we must first create a corresponding theory of history, and then confront the theory with historical records and empirical observations.

This is a major project that can be explored here only in a preliminary and tentative fashion. We shall proceed by sketching out the basic concepts of an evolutionary system theory of society, and exploring them both in regard to the overall *direction* of historical development, and with respect to its dominant *mode*.

Arguments for treating society as an entity radically different from natural systems have no *a priori* validity. A society is a dynamic structure composed of groups of human beings in specific relations. The social system, much as an organic population, a clade or an ecosystem, maintains itself or changes independently of the particular destinies of its individual members. Human beings pass through it in cycles of birth, maturation and death ; society persists, changes or decays according to processes that take place on its own, typically societal level. The fact that the individual members of human societies are conscious, language and tool-using persons introduces a good deal of noise into the system but it does not negate whatever dynamic properties it has of its own. All forms of reductionism can be firmly rejected. Society as a system is not reducible to an atom, a cell, an organism or an ecosystem, even if the same basic and general laws apply to all of them. Society is also not reducible to its parts or components. Just as a human being is « more than » the sum of the cells in his body, so a human society is « more than » the sum of its individual members. Each system derives its structural and dynamic characteristics from the kind of interactions and interrelations that obtain among its parts, rather than from the aggregate presence of its parts.

In the context of dissipative structures theory human society is a complex system far from equilibrium. It is composed of human beings and their manifold relations and maintained in the flow of energy in the biosphere. The elements of this system are individual human beings organized in various kinship, social, economic, political and cultural groupings such as families, clans, tribes, villages, provinces, cities, states,

nation-states politically federated states and multinational communities. The characteristics of human societies are determined by the way the elements are interrelated within the total social system. Since the members of society are conscious and symbolically intercommunicating human beings who code and store information in the form of culture, human societies are not just social, but specifically socio-cultural systems. They are embedded in the larger sphere of intersocietal relations spanning the subcontinents, the continents, and the globe itself. There are no absolute boundaries defining where one society ends and another begins ; any conceptualization of « a » society is necessarily an idealization. It is, however, one that is often necessary for purposes of analysis.

Having defined the principal parameters of human sociocultural systems in reference to the theory-paradigm of nonequilibrium systems, we can begin to apply our concepts to the world of history. The first and perhaps most fundamental question concerns the direction of historical development. Is the evolution of sociocultural systems directional, and if so, is it reversible or irreversible ?

HISTORY'S ARROW OF TIME

The question whether or not there is an overall direction in history, and whether or not such a direction, if it exists, is truly irreversible, has plagued historical debates for centuries. It is clear that, on first glance, any given instance of social change appears reversible. It does not appear possible to line up all such changes along a single axis, showing that they tend in the same direction. However, the sum total of social changes and transformations could nevertheless manifest an overall directionality. This is evident if we match a so-called « primitive » society with a « modern » one. The structure of Stone-Age tribes, for example, is considerably less complex than the structure of nation-states. More importantly perhaps, Stone Age societies constitute systems with a limited capacity to access and store energy. Modern societies are much further from thermodynamic equilibrium than traditional kinship systems : they access, store and utilize enormous quanta of energies in the various subsystems of production, consumption, services and administration.

It is clearly an oversimplification to compare merely Stone Age tribes and modern nation-states. But these two types of social systems represent the two temporal poles of historical development. Recorded history begins with Stone Age tribes and extends, to date, to modern nation-states. Whatever processes occurred in between, and whatever phases they may have conduced to, they were intermediary between these two. Even a cursory examination of the patterns of history bears out the hypothesis that the direction of history is validly given by the contrast between primitive and modern societies. Although in the intervening epochs there were numerous fluctuations and temporary reversals, on the whole capacities for the access, storage and use of energy have increased, and structural complexities have grown. The processes of convergent evolution, while still in early phases, are manifest within the span of recorded history. We witness their effects in the growing coordination of indivi-

dual and group functions within increasingly large, complex and energy intensive social, economic and political structures.

Yet, the problem of directionality in historical development can not be resolved in reference to contrasts between Stone Age tribes and modern societies. We would need to explore the course of recorded history in depth, following in the footsteps of Splenger's account of the life-cycle of civilizations, Toynbee's interpretation of history as challenge and response, Lévi-Strauss' formulation of invariant social structures and their transformations, Kroeber's and Kluckhohn's discovery of cross-cultural universals, and Benedict's conceptualization of universal culture patterns. Our analysis could turn out to be even more fruitful than these, since it would link the historical development of human societies with evolutionary processes in nature. However, this task must be left to historians. In this first attempt to formulate a scientifically based theory of history we must content ourselves with sketching the main elements and applying it to some major issues and problems of our time.

What, then, can we say as a first approximation about the overall direction of social change in the course of recorded history ? Although all generalizations have pitfalls, and generalizations in history are especially treacherous, we can nevertheless identify valid points of reference. We can note, to begin with, that since the dawn of history human societies have undergone major changes and transformations, and have also shown remarkable powers of persistence. Conservatism and innovation are both characteristic of human social systems ; jointly they dictate the tempo and the mode of historical development. On the whole, the traditions, belief systems, mainstream values and established institutions of society acted as forces of conservation, while the invention and application of new technologies — both the « software » of interpersonal communication and the « hardware » of new tools and instruments — were sources of change and transformation.

Stone Age societies were extremely enduring. They were stabilized in their immediate environment by a complex system of rites, rituals and taboos, and corresponding myths ans belief systems. Scores of millenia had elapsed before emerging technologies changed the ways of life of tradition-bound societies. The mastery of fire, the invention of the wheel, and the domestication of some varieties of plants and animals triggered slow but basic changes in the lifeways and belief systems of our Paleolithic forebears. The « Neolithic Revolution » was the first major evolutionary transformation in recorded history. It spread slowly from what is now the Middle East to the rest of the inhabited parts of the planet. Humanity has never again known societies as enduring as Paleolithic kinship-based nomadic tribes.

Technology — in the broad sense in which it includes all human activities that extend the powers of the body to interact with nature and the powers of the mind to intercommunicate with others — was always a major social change agent. We see its effects throughout recorded history. The invention of agriculture — i.e., the initial domestication of some varieties of plants and a few species of animals — was one of the

first technology-impelled transformations. It settled nomadic tribes in permanent abodes and made for growing specialization and division of labor as not everyone was needed to produce food for the community. Another technological breakthrough with vast consequences came with the invention of writing. It made possible the transmission of knowledge and experience, and it vastly accelerated the spread of new technologies and the corresponding values and ways of life.

In the span of the last millennium societal changes have begun to accelerate. They have been catalyzed by technological innovations and spread by armed conquests. The archaic empires could maintain themselves for centuries — some for millennia — in favorable natural environments, river-valleys for the most part (such as the Nile, the Ganges, the Tigris-Euphrates, and the Huang-Ho). But empires invented not only the means to produce food and administer far-flung territories ; they also invented the means to conquer other peoples by supplementing the power of human muscle with lances, arrows and swords. Empires could set out to vanquish and to subdue, and to spread their dominance over ever vaster distances and greater populations. The conquests of Alexander, of the Persians and of the Romans, followed by the Mongolians and the Turks, were as many giant fluctuations, destabilizing traditional social structures and imposing new technologies, new systems of beliefs and new institutions in place of the old.

Technology-induced societal change became vertiginous when, in the 19th century, traditional crafts entered into alliance with modern experimental science. Modern science transformed traditional crafts into powerful tools of production and communication, and buffeted the institutions to which it was held in bondage throughout the Middle Ages. New relations between man and man and man and nature destabilized established institutions and reformed traditional value and belief systems. Conditions approaching anarchy were created — the turbulences accompanying the « First Industrial Revolution ». But they did not spell the end of order and civilization ; on the contrary, they proved to be the furnace in which the institutions and values of the contemporary age were forged.

The industrial revolution displaced traditional occupations and opened up new ones. The villages lost ambitious young people and impoverished peasants to new urban and industrial centers, where the former could occasionally rise on the ladder of status and wealth while the latter were forced into bondage to the new « captains of industry ». In time better institutions emerged, some on the capitalist and others on the socialist pattern. Following periods of often painful adjustment, they restabilized medieval princedoms, kingdoms and tsardoms as modern industrial societies.

Hardly consolidated, the industrial revolution was rocked by two world wars. These cataclysmic events, in addition to loosening established orders and institutions, also promoted the discovery and application of several new technologies that soon moved from the military to the civilian sphere. The new technologies of information, automation and communication heralded another transformation : the « Second Industrial

Revolution ». Its shock waves are buffeting our values and institutions ; they are likely to destabilize numerous contemporary societies.

A factor common to all instances of technology-triggered social change is the increase in society's energy efficiency. With each technological « revolution » more energies are accessed, stored and used than in the preceding epoch. The structures of society have grown correspondingly larger and more complex. Technological change is known to be irreversible : whatever the nature of a technological revolution, it is always from the hoe to the plough, and not the other way around. Only those technological innovations are adopted and handed down which represent an improvement in the efficiency of some procedure. Improvement generally means greater efficiency in the use of energy, materials or information. It means greater speed, less investment of time and money, and operation on a larger scale. This holds true throughout the range of technological innovations, from the invention of the wheel to the invention of the silicon chip. Thus technology is the barb that keeps the wheel of social change from spinning backwards.

Technological innovation is the societal equivalent of improved energy-processing structure in nonequilibrium systems. Just as the structure of various natural systems becomes organized over time to access, store and use increasing amounts of energy in the environment, so human societies learn to access, store and use increasing amounts of increasingly diverse forms of energy (including various matter-energies and information) through the improvement of their technologies. Thereby societies grow larger in size, develop more intricate relations among their diverse subsystems, and create more massive and flexible modes of intercommunication among them. History's arrow of time leads, in a statistically irreversible progression, from the nomadic hunter-gatherers of the Paleolithic to the techno-industrial information societies of today.

BIFURCATIONS IN HISTORY

Clearly, the above simplified account of the dominant dynamics of societal change cannot do justice to the complexities of historical development. It can, however, highlight two important factors of historical change : the statistically irreversible increase of free energy in society (with the corresponding increase of structural complexity) ; and the alternation of comparatively long epochs of stasis with increasingly frequent periods of revolutionary transformation. The former testifies to the general direction of historical development ; the latter to its dominant mode.

The discontinuous and nonlinear mode of historical development is a consequence of the fact that human societies are open systems far from thermodynamic equilibrium. Human societies, as other varieties of nonequilibrium systems, maintain themselves in a flow of energy (as well as of matter and information) by means of auto-and cross-catalytic cycles. The cycles ensure the replication and replenishment of all societal structures and subsystems including human beings, natural resources, information and energies. Within certain limits the auto-and cross-catalytic cycles, functioning through established institutions and practices,

buffer out disturbances and maintain the status quo. However, major perturbations — such as those introduced by wars and technological revolutions — destabilize the cycles and replace the overall determinism introduced by their orderly functioning with a large degree of indeterminism. Governments fall, systems of law and order are challenged, new movements and ideas surface. Society encounters a point of bifurcation.

As a rule, randomness and chaos do not reign for long. No society can exist in the absence of some sort of order, however ad hoc it may be. The nature of the new order that emerges in the turbulence of bifurcation points is indeterminate with respect to the overall direction of historical development at any given moment in history. However, the sum of social transformations in the course of time does tend, with a statistical probability, toward high-energy societies with great structural complexity and low entropy. The general irreversibility of technological innovation overrides the indeterminacy of individual points of bifurcation and « drives » the processes of history in the observed direction — from primitive tribes to modern techno-industrial states.

History provides a rich terrain for the exploration of points of bifurcation in the development of societies. Bifurcations in society are in the nature of sudden breakdowns and transformations, hence revolutions. However, they were not always perceived as such. Bifurcations that were due to the destabilizing effects of new technologies were usually perceived as instances of societal *evolution* ; the traumas and crises were taken as the price of progress. On the other hand bifurcations triggered by the destabilizing effects of wars and foreign conquests were frequently perceived as revolutions or counterrevolutions, i.e. radical (and not always desirable) cases of societal mutation.

Of course, even the radical and rapid transformations that were perceived as revolutions did not mean complete breaks with the past ; historians have often pointed to elements of continuity. In the case of the French revolution of 1789, for example, Tocqueville found that it accomplished many goals pursued by French kings since Louis XI, and Acton announced that it brought the political structure of French society into line with long established traditions in its social structure. But however they are perceived by contemporaries and by historians, revolutions do change the established institutions of society, adapting them to the stresses which brought down the previously dominant regimes. The fact that the adaptations are functional from the viewpoint of social stability does not mean that they are humanistic and morally desirable : revolutions can establish dictatorial regimes and reigns of terror as well as democratic and enlightened forms of social order. Perhaps the clearest examples are furnished by the sudden societal convulsions of the 20th century : they created social and political systems as diverse as Communism in the Soviet Union, National Socialism in Germany, Maoism in China, Castroism in Cuba and Islamic fundamentalism in Iran, among others. These revolutions are dissimilar in many particulars ; no attempt should be made to draw moral and ideological comparisons between them. Yet they all exhibit the basic dynamic of historical bifurcations. Whatever the specific causes of instability in a society, and whatever the

ultimate outcome of the order which resolves the ensuing anarchy and chaos, the dynamic pattern repeats over and over again in the widest spectrum of historical circumstances. In these relatively sudden transformations — whether they are due to the destabilizing effects of technological innovation or to the more dramatic effects of wars and conquests — the institutions of society adjust to previously intolerable stresses. The outcomes are always unpredictable and often surprising. The factors of chance are well illustrated in the events in Russia and Germany ; neither the rise of Lenin nor that of Hitler was predicated on any discernible « historical necessity » at their particular time and place. Similarly, there was nothing necessary and predictable about the rise of Mao in China, Castro in Cuba and Khomeini in Iran. Yet many of these outcomes, surprising as they may be, are not inconsistent with the general directionality of historical development. Apart from all moral and ideological issues, it is clear that both the Soviet Communist and the German Nazi society mastered the new technologies of the early 20th century much more effectively than their immediate predecessors, and created high-energy, and structurally complex societies. China is now moving in the same direction, while the future shape of Cuba and Iran is as yet unclear. That many of the regimes established in the wake of revolutions proved to be temporary testifies to the fact that, although they were functional solutions of the crisis at the time, they were not final answers to the problems which confront societies today and in the future.

This brief review of patterns of history in light of theories and concepts developed in nonequilibrium thermodynamics and the related sciences of complexity cannot do more than demonstrate the inherent cogency of such an endeavor — and the need for historians and other social scientists to pursue it in detail. The invariance of the general laws of nonequilibrium systems throughout the realms of complexity holds out fresh hope that the day may not be far when we can overcome at last the centuries-old split between the « hard » sciences and the humanities, and gain a more integrated, illuminating and at the same time dependable insight into the nature of change and persistence in the systems that concern us the most — the communities and societies of human beings(*).

(*) A detailed exploration and exposition of a general theory of nonequilibrium systems is given in Ervin Laszlo, *Evolution* (Boston : Shamghala New Science 1987 ; in Italian : Milano : Fettrinelli, 1986 ; in German : Wien, Zürich, Munchen : Europa verlag, 1987).

Language, Logic, and Energy : A Thermodynamic Sketch of Greek Democracy

Robert Artigiani
History Department
U.S. Naval Academy — Annapolis, Maryland

As the branch of physics best able to explain living processes, thermodynamics provides a natural bridge between the sciences and the humanities. It is not surprising, therefore, to find thermodynamics borrowing the concept of historical time from the humanities to describe evolution in nature[34]. The humanities should be equally able to borrow from thermodynamics to make evolution scientific in history. Ervin Laszlo has begun the process[25]. Using the law of minimum entropy that Prigogine has shown to be associated with structures driven far from equilibrium[32], Laszlo finds that history records a continuing increase in social complexity. His attempt to track the growth of social complexity from the stone age to the present is thoroughly convincing. Students of society could make use of this theme to organize their data in a more scientifically sophisticated way. But they have been trained to be suspicious of all broad ranging theories[30]. Moreover, since Dilthey, historians, especially, have grown used to resisting the influence of science on their methods and categories[10]. Still, they might be persuaded to accept the basic conclusion that societies evolve toward more complex structures by a thermodynamic model. Although historians would cavil about the way complexity was defined and the extent to which it could be measured, they should be willing to accept it as a simple quantitative discriminator.

Conceding that societies do evolve toward more complex structures is more important than it seems at first. The evolution of complexity implies that for more complex societies to sustain themselves their internal systems of order must be more efficient, their networks of sen-

sors and feedback loops more extensive, and their regulating agents more numerous. From that it is possible to propose an unexpected conclusion. More complex societies that sustain themselves by increasing their internal efficiencies, their knowledge of the environment, the number of their allowed responses to environmental changes, and the options available to their agents are, it can be claimed, more democratic, liberal structures. Increasing complexity demands, on the social level, the active participation of more members of a society in its governance, since more actions must constantly be taken to counter greater tendencies toward disorder within the structure. Democracy is obviously the best way to increase this participation. But for democratic agents to affect the behavior of their social system they must be able to chose actions that are decisive. They must be autonomous and responsible. Thus, if thermodynamics can show that societies are becoming more complex it must, at the same time, be demonstrating that there is « progress » in history defined as an increase in the rights and significance of individual human beings. In this way, the simple quantitative determinant of thermodynamics leads to qualitative historical conclusions : Grant but increasing complexity and time will create a more human world.

Certainly these conclusions must be held tentatively, for research into the relationship between thermodynamics and history is only beginning. It must be agreed further that at this stage of the inquiry other conclusions might equally well be drawn. Finally, the fact that societies do not have to change in the direction of greater complexity must be honestly recognized. But philosophically consequential, even optimistic, conclusions are the kinds of claims that, if nothing else, will dramatize the potential for using twentieth-century scientific models for social and historical analysis. Within the context of the limitations mentioned above, a degree of boldness does seem appropriate.

Perhaps the most reasonable place to explore the potential of thermodynamics from this perspective is ancient Greece, the place where equal participation in government — « isonomia » — was first practiced. The civilization of Greece emerged from an anarchic period, called the Greek Middle Ages, that lasted from about 1500 to 900 B.C. In that time, various earlier societies, some quite sophisticated, were swamped by invading nomads and virtually destroyed. These nomads, gloriously recalled in Homer's epics, established brutal and barbarous norms of behavior that embarassed their descendents, including Homer. Centuries of sedantic living, commercial contacts with more advanced societies, literacy, and other factors, eventually domesticated the Greeks, at least to a point. But the aristocrats' independence and anarchic behavior was preserved by a monopoly of military power. No one could tame these ancient heroes without force, and as long as the aristocrats alone were able to purchase bronze armor and horses they were unassailable. Society may have been organized on a primitive basis, but little change could be expected without the introduction of some new flux that permeated the whole mix, creating new patterns of relationship in a manner comparable to that of a Bénard Instability.

That flux appeared around 700 B.C. with the introduction of iron technology. Iron made weapons and armor cheap enough for average citizens to possess. Horses, of course, remained outside the range of their purchasing power. But, given the rocky, broken terrain of ancient Greece, horses provided, at best, a limited advantage. Against a trained, disciplined, and determined infantry cavalry was almost always defeated. Arming the citizenry, therefore, meant distributing power downward through the ranks of society, for armed citizens were no longer obliged to defer to the decisions of the aristocrats. Within fifty years these new citizen-soldiers had eroded the power of the old leadership. Consequently, cities could no longer depend on their Homeric aristocracies for defense. Now Homeric aritocracies have many liabilities for a society, but since they are eager to win glory, gain prizes, and display areté they are always willing to fight — although, like Achilles, their independence and temperament often lead to whimsically unpredictable military behavior. Citizens, by contrast, have to develop new motivations and reasons to ensure their fighting spirit. In particular, the Greeks found that, instead of launching themselves into war spontaneously, citizens had to be talked into it. They persuaded themselves of the need to fight in their assemblies when they argued out the merits of various policies. Once persuaded of the rightness of a policy, however, the citizen-soldiers willingly supported it. In the city speech gained preeminence « over all other instruments of power »[37]. Thus, the military effects of iron technology are called the « Hoplite Revolution », because the creation of new infantry units made up of citizen-soldiers produced the first democratization of society.

In the Hoplite revolution a « phase change » took place. The Greeks reconstructed their society by creating new relationships that redefined citizens as free, equal, and active agents of the polis. The polis, the body of active citizens whom Aristotle considered an extended family, created a public space in which « nomos » — law and reason — prevailed over tribal, family, selfish, or private interests. The polis established a political language in which law prevailed over will, for since force was now distributed throughout the ranks of society only logic was persuasive and all were equally able to speak[20]. Thus the polis formalized the taming of the ancient heroes. That taming, begun in military tactics that suppressed individual quests for personal glory, was extended to every aspect of life, from women's dress through mourning rituals to criminal law[1]. In all these cases, citizens became equal with one another before both the law and the critical scrutiny of their peers. Yet the aristocratic goal of areté was not abandoned. It was redirected towards oratorical contests whose fierce competitiveness itself threatened social stability. New values were needed to secure society against the pressures of political discord. It was in this context that Greeks discovered « sophrosonye », moderation, and the « Golden Mean ». These new values controlled human action in ways that assured the endurance of the city-state by encouraging individuals to govern their own behavior. But when men govern their own behavior in ways calculated to sustain new relationships they have created a new social structure, and when a new structure

survives only by simultaneously changing both its relationships and constituent parts it is the product of a « phase change » in the simplest evolutionary sense[19].

Substituting citizen-soldiers for Homeric heroes made Greek society more complex by equalizing and democratizing it. A democratized society turned out to be amazingly effective militarily, as the Persians learned at Thermopylae and Salamis. Citizens soldiers who had persuaded themselves of the need for war — or its advantage — fought with a relentless determination. At Thermopylae every Spartan died rather than retreat ; at Salamis, the Athenians took to their ships and continued the war even after their city had been destroyed. But even in less dramatic encounters Hoplite deployments called for « shock tactics » that pitted ranks of soldiers in face to face confrontations so difficult to maintain that generations of later military commanders have despaired of ever bringing their troops to such a state of aggressive readiness[17]. Only « philia », the binding identification of the soldier with his unit and the citizen with his state, could produce that degree of spirited commitment[1]. At the same time, of course, the training required of Hoplite phalanxes to produce the skillfull coordination of maneuvres even in the most desperate circumstances, obliged them to drill regularly and often. Drill, the regularized repetition of bodily motions involving the large muscles, encouraged the tendency to sublimate the archaic love of self into a passionate attachment to the military unit[27]. Thus the identification of the soldier with his phalanx anticipated the identification of the citizen with his city. The Hoplite phalanx served as the « nucleation » around which the new social structure took shape. At the same time, the ability of the phalanx to transform raw recruits into disciplined soldiers through drill meant that the phalanx, as well as public debate, acted as a « Brusselator » for the replication of new generations of citizen-soldiers through the catalytic action of its values.

Only democracy can account for the military tenacity of the Hoplites, for they were not whipped into battle as Herodotus tells us the Persians were. On the contrary, it was the logic of persuasive arguments in which they had publically participated that controlled the Hoplites. They were creatures of the polis for whom, as Damaratus told Xerxes, « law is the master ». Their values are commemorated by the marker at Thermopylae, which reminds travelers that the Greeks « lie here having been persuaded to obey the lawful command of the Spartans ». Military tenacity is one measure of the increased human energy released by democratization. That released energy was so great it produced the Golden Age — the outpouring of artistic, scientific, philosophic, and cultural creativity for which Greece is eternally remembered. Interacting with a new technology, Greek cities evolved into more complex structures farther from equilibrium.

Farther from equilibrium, however, the problem then became one of assimilating this vast amount of released energy, for another consequence of it was to escalate frictions between city-states. They became expansive and soon found themselves in constant conflict. As Herodotus lamented,

These very Greeks are wont to wage wars against one another in the most foolish ways, through sheer perversity and doltishness. For no sooner is war proclaimed than they search out the smoothest and fairest plain that is to be found in all the land, and there assemble and fight ; whence it comes to pass that even the conquerors depart with great loss : I say nothing of the conquered for they are destroyed altogether. Now surely, as they are all of one speech, they ought to interchange heralds and messengers, and make up their differences by any means other than battle...

The results were disastrous, for the very same era that produced the fullest flowering of classical culture also produced the Pelopennesian Wars. Those wars, in turn, so weakened Greek civilization that it fell easy prey to empire builders from Macedon and Rome. The tragedy was that Greek cities could not integrate themselves, could not join their separate cultures into a larger, even more complex unity farther yet from equilibrium.

The failure to achieve unity on a national scale suggests that it would have required even greater energy than that released by democracy. There are other forms of human energy that can be tapped, of course, such as slavery, and the Greeks exploited them to varying degrees. But slave energy tends to dampen a society's capacity for growth, as the Spartans demonstrated. The mere presence of slavery discourages efforts to develop labor saving devices, for instance, while preserving the institution inhibits exploration of many avenues of human growth. Only non-human energy would have succeeded in supporting a complex Greek form of democracy above the level of the city-state. Non-human energy, of course, must come from external, environmental sources. Greek engineers apparently had the technical skills necessary to exploit such energies. Their mines, temples, clocks, gadgets, and military machines testify to a high level of technological mastery[31]. But they did not develop techniques for using water or wind or even for harnassing animal power efficiently as labor saving devices[18]. Nor did the machines they developed — such as Hero's vacuum and water pumps — alleviate in any way the labor of men[23]. Thus the Greeks could not increase the energy flow through the system and expand it to a larger scale.

One reason the Greeks were unable to mobilize the greater energy resources needed to stabilize a complex society on an enlarged scale was their failure to develop science, the only language sufficiently rigorous to control the behavior of natural phenomena. The Greeks, of course, are credited with having « discovered » science, and to the extent that they spoke of natural phenomena in natural terms that is correct[26]. But historians of science have long recognized that traditional social values inhibiting actual experimental and instrumental analyses of nature prevented Greek philosophers from becoming « scientists » in the modern sense[14]. Abjuring instruments and refusing systematic experiments left the Greeks with natural descriptions that, despite their originality and inventiveness, remained undisciplined speculations.

The absence of discipline can be explained partly as a hold over from the Achillean style of the Homeric aristocracy. It also followed from

the Greek reluctance to test their ideas against the hard realities of observation and experiment. What is particularly perplexing about this is that the close relationship between democracy and technology is so easy to recognize. According to the Greeks, « sophrosonye » — self-restraint — was the key to self-government. Sophrosonye is what gave the personality maturity and self-control, and for self-government in the political sense to be possible, people had to be self-governing in the personal sense. Now the practice of a technical craft teaches personal self-government at least as effectively as the military training that produces a phalanx, where the survival of all depends on the discipline of each.

Regretably, it was rarely the Greek craftsman who speculated about science. More often, it was upper class, educated men who probed the workings of nature. Their search, however, was restricted by their social attitudes : shirking direct, practical contact with nature — think of Euclid throwing a penny to the man who asked what his drawings were worth — these intellectual Achilles demonstrated their areté by devoting themselves to the beautiful but useless. In addition, limited direct contacts with nature revealed to the Greeks that sense data, even when corrected by the few instruments available to them, was highly unreliable. As a result, the language they used to describe nature was devoid of any enriching contact with actual phenomena. Their theories referred to words, not things, to logic, not experiment. They were mind-spun in the purest sense.

Without the test of observational verification the language of science was provocative but not compelling. Greek speculations could inspire awe or disgust, depending on whether the audience was attuned to Democritus or Parmenides. But it could never oblige a listener to consent. Thus its ability to control the energy released by democratization was limited. At the same time, the insistance that science must be pursued for purely esthetic purposes meant that it had no capacity to explore or exploit environmental energy sources. Together, these failures account, on the intellectual level, for the tragedy of Greek politics — and warn us that if we are to do any better we must integrate science into the very fabric of our own society.

As a science, thermodynamics, of course, was first developed by engineers, men with a direct and continuing contact with nature[4]. That it contradicted the basic elements of Newton's timeless mechanisms did not trouble them enough to force a repudiation of its demonstrated practical validity. Following their lead has enabled contemporary theorists to redefine Newtonian science in ways that reintegrate human beings into the picture of nature and redefine human accomplishments historically[33]. It may not be accidental that a branch of knowledge founded by engineers is producing breakthroughs in both our knowledge of objective reality and our perceptions of ourselves. Engineering gives us knowledge of nature in a direct way, sometimes generating information that was not otherwise available. But engineering — i.e., the active transformation of the world through the use of tools — has immediate human consequences as well. It is not just that the use of tools necessarily affects the way people live, however. That makes engineering a branch of

ethics[6]. More to the point here is that the use of tools shapes the values people have. A society like that of ancient Greece, which relied on human rather than natural energy, is more likely to have difficulty altering its values than a technological society, where the introduction of new tools means the constant alteration of the established value system and of the environment against which values are tested. As a result, the failure to develop techniques for exploiting environmental energy meant that Greeks were unable to preserve democracy while reorganizing their society on an expanded scale. It also meant that they remained, like Socrates and Aristotle, totally dedicated to the idea that the city-state was the only vehicle for achieving the civilized style of life they considered truly human. In the wake of self-destructive wars, Greek conservatives were able to resist change because, without technological innovation, there was no foundation on which to generate new values.

The claim that technology generates values must sound strange. We are much more accustomed to hearing that technology undermines all that is holy. In a sense, the claims are compatible, for if technology does generate new values then it must simultaneously overthrow old ones. But that compatibility does not explain the way technology generates values. Nor do arguments like Mesthesne's, which simply state that new tools allow us to do new things, tell us why we come to value new things[28]. To understand the process fully we have to turn to ideas like James Feibleman's, which begin by denying the fundamental dichtomy between body and mind that has plagued Western philosophy as well as Western science for three centuries[15,16].

Recognizing that the mind has a body as well as vice versa enables Feibleman to relate what the body does to what the mind thinks and feels. He concludes, quite simply, that the mind comes to value what the body does. Now nothing influences what the body does more directly than technology, since the mere use of a tool requires the body to move in ways prescribed by the form of the tool. Repeated motions like those associated with tool use establish temporal patterns — structures or forms. It was the sudden appreciation of the potency of forms that inspired Plato, whose frequent use of artisanal metaphors suggests at least a dim awareness of the linkage between technology and values[5]. Temporal patterns derived from tool use are similar in kind to those induced by more ancient procedures, such as ritual[7], or to the creation of Philia through regularized drill. Tool use structures the mind to find value in what the behavioral pattern enframes. A tool makes us value what the tool enables us to do, and the loop is closed by the fact that the act of using what the tool makes is what makes the tool valuable. Culturally, tool use functions like a cellular automaton, creating rules of the game, symbolized as values, that come to regulate behavior[13].

Moreover, tool use requires communication, for future generations must be taught how to make and use the tools that experience has found valuable. Tools are themselves systems of communication, for they incorporate in tangible form the minds of their creators and direct the behaviors of their users. They are very similar to works of art or DNA molecules in this regard. By catalyzing the actions of those who use them, tools

may also be said to function as energy multipliers, very similar in their effects to the multipliers with which economists are familiar. It is obvious that a society with a heavy dependence on technology must then have many vehicles for communicating energy as well as ideas.

Languages and values are just more abstract, and perhaps more efficient, energy multipliers. But languages and values that are closely linked to technology would be more effective in communicating both energy and information. The link to technology would produce redundancy in language that would ensure accuracy of communication because technology would keep the message sender in close contact with the environment. Errors would be quickly spotted and falsehoods quickly revealed. Technology would also contribute to the usable energy available to a social structure by increasing the number of communications and their accuracy, by catalyzing action, and by directly influencing behavior, as well as by harnessing non-human sources of energy directly. In this way technology would also help make more complex structures stable even when they were farther from equilibrium. Thus, technology would add to the benefits begun by democracy, for its improvement of communication and increase of usable energy would make complex human systems on a larger scale possible. Finally, by improving technology societies would alter behavior, create new forms, and produce new values, thus making the stability of complex social structures possible.

Greek resistance to developments of this sort is measured by Plato's philosophical use of « forms ». He seems to have confused the source of value, the process by which an object, idea, or behavior is « enframed », with value itself. But his purpose was perfectly straightforward : by the use of universal forms he hoped to prevent evolutionary change. He used examples from the experience of artisans to betray their politics[39]. Making values perfect and eternal, Plato thought, would make society forever static. He thus recognized the role of values in catalyzing actions calculated to replicate the status-quo. Of course, he failed. But in failing Plato may have betrayed his own expectations even more thoroughly than he had betrayed democratic politics, for his ideas became, in the hands of Christian philosophers, justifications for the complete separation of values from the established state. When they transfered allegiances from the city of man to the city of God, Christians removed the cultural brakes from the dynamics of nature and allowed for a social transformation even greater than that accomplished by the Hoplite Revolution. With the obstacles to change removed, new forces for development flowed in. The barbarian hordes that invaded an inadequately defended Roman Empire, in particular, brought with them new tools for the exploitation of nature that would, eventually, provide the base for a genuinely industrial society[5]. At that point, democracy and complexity could combine on an enlarged scale. But then, of course, more advanced technologies require that even more responsibilities be placed on individual human beings. To handle them, people must be better educated. Better educated people, in turn, have had the chance to explore and develop their own individuality. Thus, new technologies that sustain more complex structures do so by introducing new energy sources and

values in ways that support the claim that thermodynamics does, in fact, reveal « progress » in history.

Roman history may appear to contradict the thesis put forward in these remarks. Certainly, the Romans managed to organize society on a scale never approximated by the Greeks. Yet they shared with the Greeks an extended experience with democracy, an economy based on slavery, and a pronounced hostility to innovation. But bigger does not mean more complex. Thus, because the Romans expanded the scale of society without making it more technologically sophisticated, they did not make it more complex. Rather, after the Roman failure to contain the internal tensions of their democracy plunged the Mediterranean world into a centuries long civil war, they resolved their social disorders by substituting authoritarian government for democracy. They sacrified freedom for peace, autonomy for prosperity, and equality for hierarchy.

But when power collected in the hands of a single man, who often had total control over the lives of his subjects, the rest of the population became simply objects of his will. The dreadful uniformity that resulted is recorded in the gigantic stone statues of emperors like Constantine, who made themselves the single archetype and unique focus of all Roman life. To that must be added the policies, so admired by Gibbon, of emperors like Hadrian, who set fixed boundaries to the Roman state. While this action appealed to the rationalism of eighteenth century philosophical historians, Hadrian's Wall had the effect of making Rome a closed system. The internal entropy of the structure began immediately to increase. Population declined, trade decreased, the bureaucracy deteriorated, and the army compensated for its lack of effectiveness against external foes by turning to violence at home. Diocletian's attempt to save the state by increasing taxes merely recognized the lack of usable energy available. To escape the pressure from Rome, many simply fled into the countryside, there establishing lattifundia that supplied their needs at the expense of those of society. Diocletian's response, a policy of freezing the social positions of people in their towns and castes, completed the process by which a closed system crystallizes. Although Christians, began to make some progress in the state after Constantine, they lacked the skills necessary to revive it and, as St. Augustine's refusal to defend Hippo against the Vandals shows, the Christian subjects of an emperor who did enveryone's thinking for them remained unpersuaded that their salvation could be found in the worldly city.

The history of Rome therefore suggests that the passage of time is not tantamount to « progress ». Progress may be possible but it is not necessary. It occurs only when the creative acts of individual human beings are able to overcome the limitations of their time and place by the imaginative use of new resources. The process is difficult and the shifts dramatic. But if the results are an increase in structural complexity then it seems likely that those results were sustained by the liberation of people. The tragedy of ancient society is that both Greece and Rome, when they reached the bifurcation point at which progress is possible, chose instead the thermodynamic branch that led to equilibrium rather than evolution.

BIBLIOGRAPHY

(1) A.W.H. ADKINS, *Moral values and political behavior in Ancient Greece*, New York, Norton, 1972.

(2) T.W. AFRICA, *Science and the state in Greece and Rome*, New York, Wiley, 1968.

(3) R.J. BRAIDWOOD, *Prehistoric men*, Chicago, National History Museum, 1961.

(4) S. BRUSH, *The kind of motion we call heat*, Amsterdam, Elsevier, 1976.

(5) R.A. BUCHANAN, *History and industrial civilization*, New York, St. Martin's, 1979.

(6) M. BUNGE, « *Philosophical inputs and outputs of technology* » in G. Bugliarello and D.B. Dover (eds.), *History and philosophy of technology*, Chicago, University of Chicago, 1973.

(7) J. CAMPBELL, *The way of the animal powers*, San Francisco, Harper and Row, 1983.

(8) V. CSÁNYI, *General theory of evolution*, Budapest, Studia Biologica Hungarica, 1982.

(9) M. DAUMAS, *A history of technology and invention*, New York, Crown, 1969.

(10) W.D. DILTHEY, *Pattern and meaning in history*, New York, Harper, 1962.

(11) J. EARLEY, « *Self-organization and agency* », *Process Studies*, vol. ll, n° 4, winter 1981, pp. 242-58.

(12) L. EDELSTEIN, « *Motives and incentives for science in antiquity* », in A.C. Crombie (ed.) *Scientific Change*, New York, Basic Books, 1963.

(13) M. EIGEN and E.R. WINKLER, *Laws of the game*, New York, Harper, 1981.

(14) B. FARRINGTON, *Greek science*, Baltimore, Penguin, 1944.

(15) J.K. FEIBLEMAN, *Technology and reality*, Boston, Nijhoff, 1982.

(16) « *The philosophy of tools* », *Social Forces*, March 1967, vol. 45, n° 3, pp. 329-37.

(17) ARTHER FERRILL, *The origins of war from the Stone Age to Alexander*, London, Thames and Hudson, 1985.

(18) M.I. FINLEY, *The ancient economy*, Berkeley, University of California, 1973.

(19) S.W. FOX, « *The origins and nature of proto-life* » in W.H. Heidcamp, *XIII Nobel conference*, Baltimore, University Park, 1978.

(20) W.G. FORREST, *The emergence of Greece Democracy*, New York, Mc Graw-Hill, 1979.

(21) J. GRIFFEN, *Homer*, New York, Hill and Wang, 1980.

(22) C.H. KAHN, *Anaximander and the origins of Greek cosmology*, New York, Columbia, 1960.

(23) J.G. LANDELS, *Engineering in the ancient world*, London, Chatto and Windus, 1978.

(24) E. LASZLO, *Introduction to systems theory*, New York, Gordon and Beach, 1972.

(25) , *Evolution*, Unpublished Mss, 1985.

(26) G.E.R. LLOYD, *Early Greek science*, New York, Norton, 1970.

(27) W.H. McNEILL, *The pursuit of power*, Chicago, University of Chicago, 1982.

(28) E. MESTHENE, *Technology and social change*, Indianapolis, Bobbs-Merrill, 1967.

(29) R. NISBET, *History of the idea of progress*, New York, Basic Books, 1980.

(30) K.R. POPPER, *The open society and its enemies*, 2 vols., New York, Harper, 1962.

(31) D.K. DESOLLA PRICE, *Gears from the Greeks*, Philadelphia, American Philosophical Society, 1974.

(32) I. PRIGOGINE, *From being to becoming*, San Francisco, Freeman, 1980.

(33) and I. STENGERS, « *The New Alliance* », *Scientia*, 1977, vol. 112.

(34) , *Order out of chaos*, New York, Basic Books, 1984.

(35) W. SCHADERWALDT, « *The concepts of nature and technique according to the Greeks* », in P.T. Durbin (ed.), *Research in Philosophy and Technology*, Greenwich, Cn., JAI, 1979.

(36) E. SCHRÖDINGER, *Nature and the Greeks*, Cambridge, Cambridge University, 1954.

(37) J.P. VERNANT, *The origins of Greek thought*, Ithaca, Cornell University, 1982.

(38) L. WHITE, *Medieval technology and social change*, London, Oxford, 1962.

(39) E.M. and N. WOOD, *Class ideology and ancient political theory*, New York, Oxford, 1978.

Human Conduct and Natural Laws : Laws of Simple and Complex Systems

Bruce J. West
Division of Applied Nonlinear Problems — La Jolla Institute, Lajolla, Ca. (USA)

Jonas Salk
The Salk Institute for Biological Studies — San Diego, Ca. (USA)

1. INTRODUCTION TO COMPLEXITY

Poincáre once pointed out that if instead of the limited number of chemical elements that exist, there were thousands or millions of elements distributed uniformly, some plentiful and others scarce, then each pebble in its composition would be unique. Because of so large a number of elements, no matter how much one learned about a particular pebble it would not be possible to generalize that knowledge to the vast number of others. However, since the number of elements is relatively small we now know how these various configurations constitute what we see, feel and smell. Their modest number has enabled us to formulate general principles of both order and behavior. Examples of such general principles are the conservation laws and symmetry principles of physics and the laws of motion that govern physical and biological phenomena. It is the intent of this communication to examine the relationship between the *laws of nature* and human conduct and in so doing to determine whether or not *laws of human nature* exists in the same way as we understand the laws of nature to exist. In this pursuit many scientists, both past and present, have maintained the hope that, just as for chemical species, the number of « elements » of human conduct is of manageable size. The assumption is that the rich variety of human responses is due to the vast array of combinations and interactions among these fundamental elements. It is often further assumed that the repertoire of res-

ponses is somehow implicitly contained within these elements and that complicated activities result from a superposition of a large number of these more fundamental responses. This view is limited in scope and herein we propose an alternative perspective, one in which new qualities emerge that are not necessarily a predictable consequence of the elementary responses.

Human beings are the product of a long process of evolution, a process that continues with the same inevitability today that it did for our distant ancestors. The concept of evolution as first introduced into biology in the nineteenth century, needs now to be expanded to encompass the evolutionary emergence of all that is part of natural order. Thus, in the course of cosmic evolution there emerged patterns and structures of great diversity and complexity in the physical, the living and the human realms. As the more complex formed from the less complex, new properties and new functions appeared that previously had not existed. In this way all of the relationships in nature and all of the laws of nature blossomed including those described by the homologies and isomorphisms observed in the underlying and often unseen microstructures and the readily apprehended macrostructrures elaborated at all levels of complexity[1]. Thus, in the course of the process of evolution the substrates appeared for the organized bodies of knowledge of the different levels of complexity as diagramed in Figure 1.

2. COMPLEXITY, UNCERTAINTY AND ORDER

If a science of human nature is to exist with any degree of power to predict and to influence human development and evolution, it will be necessary to extend the methods of approach of natural science beyond their present limits so as to provide a scientific method for studies of human phenomena. In this regard a distinction must be made between a science of simple systems (linear) and a science of complex systems (nonlinear). Examples of simple systems are the planets in their orbits, light being refracted into a rainbow, and the mechanical workings of an internal combustion engine. Among complex systems there are some that are organized and others that are disorganized. Examples of complex systems that are disorganized (random) are the positions and velocities of the molecules of air in a room, the sizes of the grains of sand on the beach, the heights of individuals in an arbitrary group, and the sound of a large group of people talking simultaneously. Similar examples of complex systems that are organized (nonlinear, deterministic) are cities, social institutions of all kinds, the development of phenotype from genotype, and the lifestyle of an individual within a social context. Organized complex systems possess attributes of simple systems (deterministic and therefore predictable) as well as those of disordered complex systems (uncertain and therefore unpredictable). This partial ordering of complex systems is a recurrent theme in human conduct[2].

In the human realm questions concerning health and disease in body and mind, of individuals as well as societies, require for answers a science of complex as well as of simple systems. The method of the science of simple systems has produced a rich harvest to which our technological

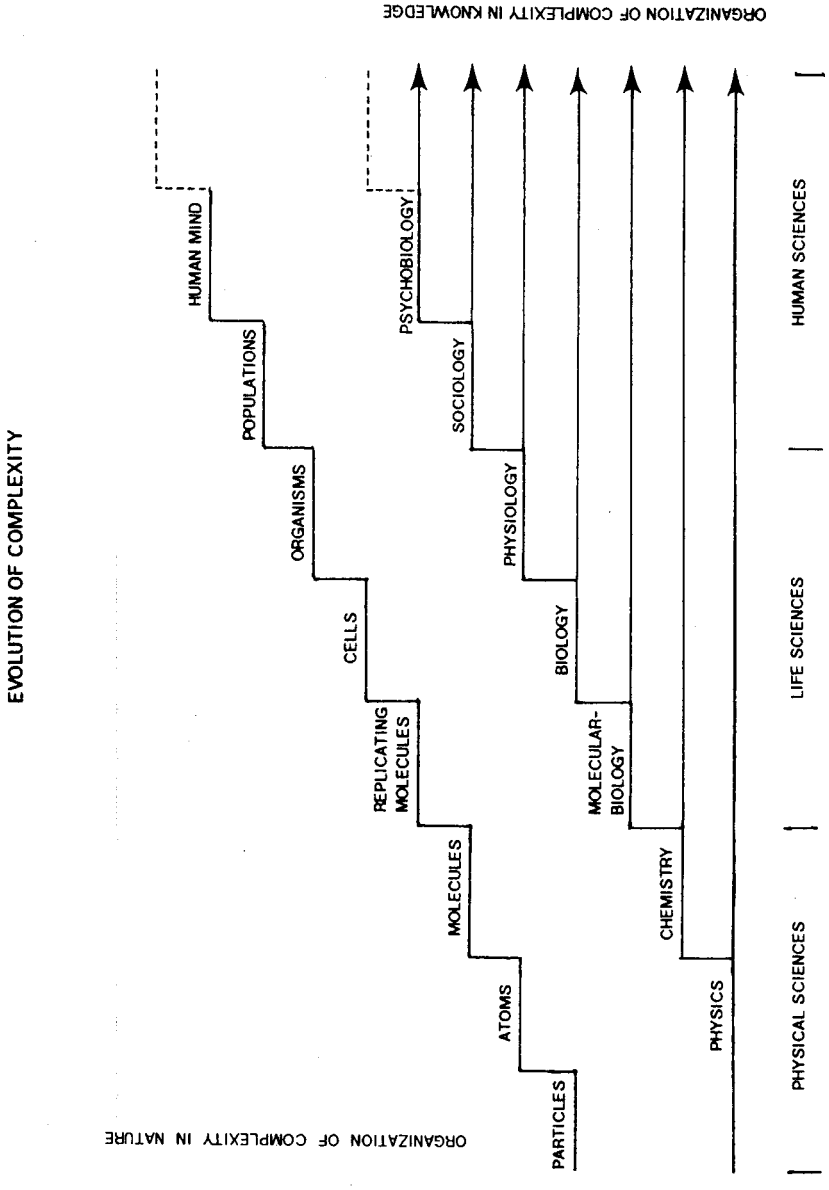


Figure 1

The taxonomic line of biological organization is depicted on the upper set of steps and the associated organization of information into disciplines is indicated on the lower set of steps.

society bears witness. The time has come, however, for a science of complex systems to make its contribution in solving the myriad of problems that have resisted the simple approaches.

The distinctions that exists between simple and complex systems can be illustrated in yet another way. A simple system is one that may be iso-

lated by separating a part from the whole. Complex systems, however, must be considered as a whole in which the whole is greater than the sum of its parts. An example of a complex system is a family, the properties of which are more than the simple summation of the effect of each individual family member. The phenomena of « family » is an emergent property of the group and is the resultant of the complex interactions among the family members. Any attempt to understand the meaning, or property, of family by isolating for study the individuals free of interaction with others members of the family group results in the loss of the emergent property, that of family which is the unit of interest.

Thus, simple systems, isolated from their context, can interact and induce effects which can be measured, i.e., that of individuals unrelated in any way to others. However, complex systems such as family include the context as well as the parts, and are the resultant of the relationships of a continuous series of events among strongly interacting components.

Whether a system is seen as simple or complex is often times determined by the observer rather than by the intrinsic properties of the system itself. For example advertisers or politicians interested in influencing the opinions of large numbers of individuals may consider a person as an isolatable component of a population and, therefore, as a simple system, since their interest is in a response to a yes or no question in terms of a purchase or a vote of each person as a singularity. A psychologist, on the other hand, sees individuals in their complexity and recognizes the individual as a part of a group or in a context. Each specialist focuses their interest on their own preferred level of complexity.

In the physical realm, relatively simple systems and disorganized complex systems predominate. In the living and the human realms partially organized complex systems predominate. In the living and the human realms partially organized complex systems predominate. Strongly interacting complex systems cannot be separated into their constituent elements : the process itself is the whole ; if any part is altered or modified in any way then the overall character of the phenomenon will change. Such systems are more readily characterized by their qualitative attributes than by arbitrary quantitative measures or indices. Such phenomena must be seen as a process, as a continuum, rather than as a series of discrete events because of the multiplicity of strong interactions among the elements.

The distinction between process and event, complex and simple, whole and parts, quality and quantity, suggests that if there exist natural laws that apply to complex systems (nonlinear) they will differ from those that apply to simple systems (linear). It is probable that the deterministic laws of the physical sciences are not applicable to human conduct. Nevertheless it is reasonable to expect that natural laws also exist in respect to human conduct. However, the laws that apply, within the constrained randomness of human activities, would differ from the mechanical laws that describe the eternal motions of the planets and stars. Patterns evident from human experience, as reflected in data from psychology, sociology, and other disciplines in the human sciences, suggest the

existence and operation of natural laws in the human realm as manifest in the distribution functions and intrinsic uncertainties that are different from the deterministic laws that apply in the physical sciences.

3. POWER LAWS AND HUMAN CONDUCT

One way to distinguish between the laws of the physical sciences and the laws of human conduct is through a comparison of the *symmetrical* bell-shaped frequency distribution curve characterizing the properties of simple systems, with its well-defined average and variance, and the *asymmetrical* frequency distribution curve with a long tail at one extreme to which averages and variances may not apply (cf. Figure 2). Gauss was the first of the physical scientists to appreciate the fundamental importance of the symmetrical bell-shaped curve in interpreting measurements. He emphasized the order rather than the variability embedded in a sequence of fluctuating measurements[3]. In non-quantum simple systems the variation in the physical measurements is related to the resolving power of the measuring instrument and can be made larger or smaller by the choice of the measuring apparatus ; thus, the variations of the measurements are extrinsic to the phenomena in question (simple systems). In the human realm, however, the variations are often intrinsic to the observed phenomena (complex systems).

The methods and approach of reductionist science are altogether appropriate and useful for the analysis of simple isolatable systems, whether physical or biological. However, for complex systems, or processes, the approach of a constructionist science for modeling is more appropriate. The latter would be more suitable for the development of a science of the integrated whole, its qualities, its properties and its performance as distinct from the science of added or subtracted parts.

The essence of the science of parts is expressed in the lyric that describes a limb, « ... the foot-bone is connected to the leg-bone, the leg-bone is connected to the knee-bone, the knee-bone is connected to the thigh-bone... » This lyric reflects the nature of most physical and biological phenomena involving the addition of identifiable parts to one another. The law of conservation of energy has this characteristic, since there is no gain or loss of energy in the universe ; if one doubles the size of a homogeneous system then its energy is similarly doubled.

However, for emergent phenomena and functional systems, or a science of the whole, Franklin's proverb is apropos : « For want of a nail a shoe was lost, for want of a shoe a horse was lost, for want of a horse a rider was lost, for want of a rider a battle was lost... » This lyric reveals that a complex process is contingent upon the actualization of a series of antecedent processes ; the failure of any one element impairs the culmination of the process as in the loss of a kingdom for want of a nail.

In the domain of complex human phenomena the overall process, culminating in an event, is contingent on the realization of a number of subprocesses or sequences of events rather than on simple addition that characterizes simple systems. The failure of any one subevent inhibits the culmination of the ultimate event. Thus, a new form of thinking and a

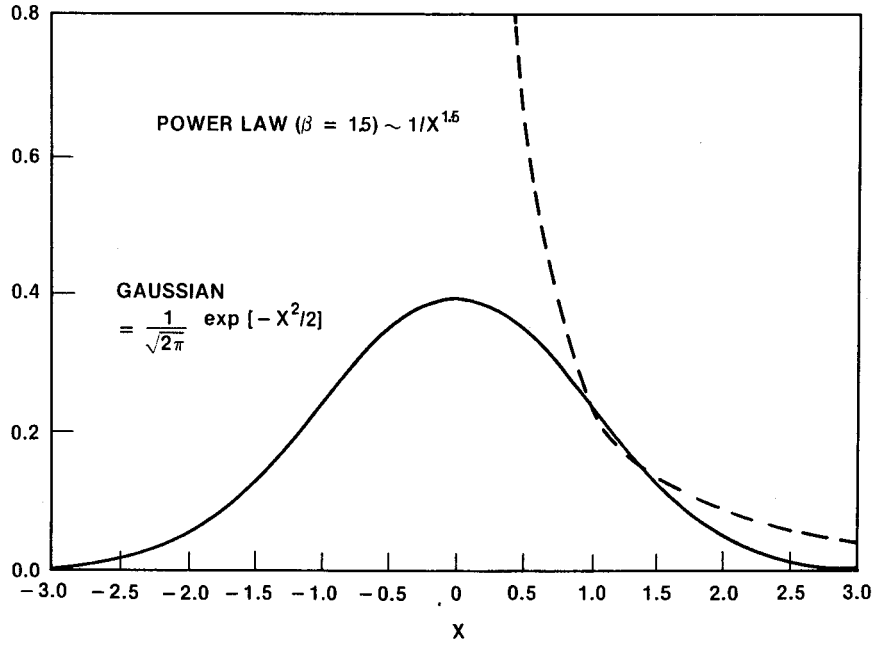


Figure 2

The Gaussian distribution for unit variance (—) is contrased with the power-law distribution (-.-) having a power-law index ß = 1.5. Zipf's law falls off more slowly with x than the indicated power law, Pareto's law is approximately of this form and Lotka's law falls off even more rapidly with x.

new methodology is needed that will allow the investigation of complex human systems in a way that takes into account the strong interconnectedness of events and process.

In further illustration of the foregoing is the example of two individuals who invest money in the stock market. The first invests $1000 in company 1 and the second invests $100 in company 2. Both make a profit of $100. The first investor receives a 10 % return on his investment and is satisfied. The second receives a 100 % return and is ecstatic. Thus, their response is proportional to the relative profits of 10 % and 100 % even though both investors make the same absolute profit ($100). The sensitivity of human response to relative change rather than to absolute levels seems to be far-reaching and may well have a psychophysiological basis ; this is the subject of psychophysics in which the subjective response to objective stimuli is measured[4]. It has been shown experimentally that an individual's response is proportional to the degree of change perceived rather than to the level of intensity of the applied stimulus[*].

Not only in psychophysics do such power-law relations exist. In linguistics the pattern of frequency with which a word appears in written language is not altogether random but follows a well-defined distribution. The distribution pattern is not the symmetrical bell-shaped curve of Gauss, but rather one that is asymmetric and decreases inversely with the word number, creating long-tailed distribution curves that can be

characterized by a single power-law index which differs from language to language (cf. Eq. (4))[***]. The pattern of frequency with which a word appears in the written language, when properly normalized to the available vocabulary, reveals a power-law relationship. Equation (4) in linguistics is known as Zipf's law[5] and the power-law index has been experimentally determined to have the value for most languages in the range $1<ß<1.2$. In exceptional cases ß may reach 1.6, as, for example, in children's talk[6].

In economics the law that relates the number of people having a given income to that income level is another example of a power-law relationship of the form of Eq.(4). Pareto, a nineteenth century social economist discovered that the power-law index ß is approximately 1.5 for a large number of western societies[7]. Thus, the achievement of a particular income level is the result of a process involving a series of contingencies, i.e., one must have attained a certain education, one must favorably impress the boss, one must be at the right place at the right time, etc. If any one of these conditions is not met then the probability of achieving a given income level is greatly reduced.

Another example of a power-law relationship, is the pattern of publication of scientific papers. Lotka demonstrated that the number of papers published by scientists also follows a power law with an index of two[8]. Again to publish a paper one must first have an idea, then the technical ability to express the idea, the time to develop it, the money for the publication costs and so on. Thus this activity too can be viewed as a contingency process.

Thus, the « laws » governing each of these processes have a common form, that of a power law with a power-law index that differs from process to process. As we have repeatedly stated, the quantitative and qualitative measures of partially ordered complex systems are distinct from those of simple systems or disorganized complex systems. The power-law distribution is one example of such a distinct measure and a study into its implications regarding the properties of the underlying process has only recently been initiated[3].

4. CONCLUSION

In conclusion, it is of interest to observe the misunderstanding, contradiction and conflicts that can occur when traditional models are applied in inappropriate contexts. There is the account of the poet Gœthe who, having made important contributions in his taxonomic studies in anatomy and botany, undertook an investigation into the physical nature of white light. Even though he was in agreement with Sir Isaac Newton on the observations they both made of optical phenomena, Gœthe disagreed with Newton's explanations and his general theory of optics. Goethe's apologist, Helmholtz, an accomplished nineteenth century physicist and physiologist, was of the opinion that the nature of Gœthe's genius as a poet and biologist would necessarily lead him astray in the physical sciences[9]. Helmholtz recognized that Gœthe's rejection of Newton's theory stemmed from the desire to find in the phenomenon

itself the direct explanation of an observation. Gœthe believed that one can derive an understanding of white light from the study of the « sunlight in the garden » rather than in the laboratory. Thus, it was the reductionist paradigm of the nineteenth century physics that Gœthe refuted in his rejection of Newton's theory of optics.

However, Helmholtz, assuming that the reductionist view was superior for the study of all phenomena, regarded the attempt by Gœthe the poet to consider a physical phenomena in its wholeness as a noble but misguided idea. We now recognize that Helmholtz's view was as much in error as was Gœthe's in assuming the universal applicability of the reductionist principles in the scientific study of natural and of human phenomena. This issue is not resolvable in either/or terms ; a both/and resolution is more fitting. The reductionist approach is applicable to events and to parts isolated from the whole and the integrative approach is applicable to processes, to the whole that is greater than the sum of the parts.

There is need for a mutually respectful view of each school of thought as well as by the epistemologists of science and the epistemologists of human knowledge and human experience. Through their merging we will come to recognize the existence of the laws of nature that determine human conduct and human evolution. When we do so we will then develop the capacity to better understand human nature and perhaps be able intentionally, to facilitate its development and evolution, and to influence the emergence of human values that will recognize the role of wisdom in nature and in human nature and to test the human capacity to choose and to act wisely for survival and continued evolution[10].

(*) In the language of mathematics this is expressed as the logarithm of the subjective response as proportional to the logarithm of the objective stimulus, yielding a power-law (nonlinear) relationship between stimulus and response. The implications of this experimental result in determining the overall response of an individual (R) to a sequence of N stimuli (S_j,j=1,...,N) is as follows :

$$\log R = \alpha_1 \log S_1 + \alpha_2 \log S_2 + \alpha_3 \log S_3 + \ldots + \alpha_n \log S_n \qquad (1)$$

when α_1 through α_n are a set of response cœfficients. Using the fact that the sum of logarithms is the logarithm of products, this series may be rewritten as

$$\log R = \log [S_1^{\alpha_1} S_2^{\alpha_2} \ldots S_N^{\alpha_N}]$$

and by inverting the logarithm we observe that

$$R = S_1^{\alpha_1} S_2^{\alpha_2} \ldots S_N^{\alpha_N} . \qquad (2)$$

Thus in this idealized case even if only one stimulus S_j is absent, say $S_q = 0$, then the response R is not realized, i.e., as in the lyric, for want of a nail the kingdom is lost.

(**) Assume that the response function R is a measure of the reaction of a population to a set of stimuli $\{ S_j \}$. For simplicity we interpret the stimuli to be a string of the same stimulus repeated N times and $\alpha_1 = \alpha_2 = \ldots \alpha_N = \alpha/N$ to be the average response to each of the repetitions. Then the response function can be written

$$R = (S \cdot S \ldots S)^{\alpha/N} = S^{\alpha} \qquad (3)$$

$$N - \textit{factors}$$

which for $\alpha = -\beta$ is of the power-law form

$$R \propto 1/S^{\beta}. \qquad (4)$$

REFERENCES

(1) J. Salk, *Anatomy of Reality.*

(2) B.J. West and J. Salk, « Complexity, Organization and Uncertainty », to appear in Eur. J. of Op. Res.

(3) B.J. West, An Essay of the Importance of Being Nonlinear, Lect. Notes in Biomathematics *62*, Springer-Verlag, Berlin (1985).

(4) F.S. Roberts, *Measurements Theory*, Ency. of Math. *7*, Addison-Wesley, Reading, Mass. (1979).

(5) G.K. Zipf, *Human Behavior and the Principle of Least-Effort*, Addison-Wesley, Cambridge, Mass. (1949).

(6) L. Brillouin, *Science and Information Theory*, Academic Press, New York (1962).

(7) V. Pareto, *Cours d'Economie Politique*, Lausanne (1897).

(8) A.J. Lotka, J. Wash. Acad. Sci. *16* 317 (1926).

(9) H. von Helmholtz, *Popular Scientific Lectures*, Longuans, Green and Comp. (1881) ; Engl. Trans. Dover, New York, N.Y. (1962).

(10) J. Salk, *Survival of the Wisest.*

New Science of Instability throws Light on Politics

Alvin Toffler
Author

Behind the clash and clatter of today's headlines lie hidden models of reality. Politicians and opinion makers say things like :

« Government spending must be cut because deficits lead to higher interest rates. » Or « spending money for arms reduces social welfare. » Or « trade between nations lowers the risk of war. » Or « more police equal less crime. ».

Right or wrong, all such claims are based on seldom-questioned assumptions about how things work in the real world. On the surface, statements like these have little or nothing to do with science. But all of them imply cause-and-effect relationships, and the very way we habitually think about cause and consequence has been strongly shaped by science. In fact, some of our most cherished political ideologies and attitudes are, at least in part, products of the scientific revolution that began with Newton three centuries ago.

Today science, itself, is caught up in a conceptual revolution. And I believe that certain new ways of thinking about change in the physical universe may soon begin altering the way we conceptualize change in social, economic and political systems as well.

What we have called science until now is being sharply challenged by what might be called post-industrial or « Third Wave » science. And it suggests that many of our assumptions about how things, people and social systems change are hopelessly simplistic and obsolete.

Consider our basic concepts about how the world works. They are rooted in the system of thought that came together in the 17th and 18th

centuries under the heading « Newtonianism ». They pictured a clock-like universe in which chance played no part, in which all the pieces fit together cogs. An emergent smokestack civilization enthusiastically embraced this machine model of reality. Newtonian mechanism even influenced the framers of the American Constitution to create a machine for governing, its checks and balances clicking like the parts of a steam engine.

American politicians still speak of railroading legislation, greasing the congressional gears, or steamrollering the opposition. Marxists routinely speak of « capturing the state machine. »

Today this Newtonian view is being challenged by what I would call « Post-industrial » or « Third Wave » science. The new insights focus not on the order and regularity so crucial to Newton and a machine world, but on upheaval, upset, instability — in fact, on the chaotic disorder which seems to characterize today's world. And no one has done more to pioneer the new science of instability than a Brussels-based physical chemist named Ilya Prigogine.

Prigogine is a preeminent theorist of disorder and discontinuity. He won the Nobel Prize for chemistry in 1977 for his work on the thermodynamics of nonequilibrium systems. Deeply interested in archeology, art, and history, he brings to science a remarkable, diverse intelligence. *Order Out of Chaos*, the book he wrote with Isabelle Stengers to outline his thinking, is a landmark. Contentious and mind-energizing, it is filled with flashing insights that subvert many of our most basic assumptions and suggest new ways to think about them.

The key to understanding Prigogine's theories — and their possible relevance to political and social change — is the recognition that when systems are pushed out of equilibrium, they behave in odd ways. They do not behave like machines at all.

They go « non-linear » — i.e., suddenly small inputs can produce giant changes, while massive forces can, sometimes, result in little or nothing. At the same time, positive feedback loops multiply, so you get self-reinforcing or self-feeding processes. Chance operates. Once-entrenched structures may fall apart. Or they may completely revolutionize themselves.

In short, Prigogine tells us, we should not expect a chemical system to follow the ordinary rules of classical science when it is streched or pushed into a far-from-equilibrium condition. The system goes haywire.

Consider, for example, Prigogine's examination of « chemical clocks », which illustrates the surprising self-reorganizations that can take place in complex systems in far from equilibrium conditions.

Imagine a million white ping-pong balls mixed at random with a million black ones, bouncing around chaotically in a tank with a glass window in it. Most of the time, the mass seen through the window would appear to be gray. But now and then, at irregular moments, the sample

seen through the glass might seem black or white, depending on the distribution of the balls at a given moment in the vicinity of the window.

Now imagine that suddenly the window goes all white, then all black, then all white again, and on, and on, changing its color completely at fixed intervals — like a clock ticking or a metronome.

Why do all the white balls and all the black ones suddenly organize themselves to change color in time with one another, Prigogine asks. Do they, in some unknown way, communicate with one another ?

By all the traditional rules, this should not happen at all. Yet, if we leave ping-pong behind and look at molecules in certain chemical reactions, we find that precisely such astonishing self-organizations or ordering can and do occur, despite classical physics and probability theories.

Now, while I mistrust automatic comparisons between physical systems and human affairs, it strikes me that this is a pretty fair description of what happens to bureaucracies, economies and even political systems when they, too, get pushed too far out of whack. They, too, behave in truly bizarre ways.

What was more bizare than inflation-ravaged Germany in the 1920^{s}. Or Mao's China in the 1960's and early 70's ? Or Pol Pot's Cambodia ? Or Lebanon today ?

I'm not suggesting that Prigoginian concepts can give us a science of social change or upheaval, but that it does suggest provide us with novel new analogies and metaphors.

Classical science focussed on stability. Even the arrival of the computer did not change that much at first. The early work on information theory emphasized « negative feedback » — process that tend to damp down change and bring systems back into equilibrium. Prigogine, by contrast, urges us to study how « positive feedback » promotes change and instability.

Imagine a situation in which a chemical or other reaction produces an enzyme whose presence then encourages further production of the same enzyme. Such situations are rare in inorganic chemistry. But in recent decades the molecular biologists have found that such loops are very stuff of life itself. Such processes help explain how we go from little lumps of DNA to complex living organisms.

Such positive feedback loops are found in society as well : « The rich get richer » is a description of an positive feedback process. So is or the electoral « bandwagon effect » that happens once a politician begins to win. So is « a run on the banks ». In the arms race, each defensive step by one side triggers an escalation which, in turn, triggers a further escalation.

According to Prigogine, positive feedback loops like these proliferate when a system is pushed beyond its equilibrial limits, which may help explain some of today's accelarating, out-of-control changes.

Although Prigogine is a physical chemist, he heads cross-disciplinary teams in Brussels and Austin that are busy exploring the implications of his ideas in fields as disparate as the social behavior of ant colonies, diffusion reactions in chemical systems and dissipative processes in quantum field theory. When his book appeared in France in 1979, it triggered a scientific free-for-all among prestigious intellectuals in fields as diverse as entomology and literary criticism.

The new science of instability speaks to a generation witnessing high-speed economic, social, technological and geopolitical changes. It serves us by shifting the attention of scientists and intellectuals to disorder, instability, diversity, disequilibrium, non-linear relationships, and the importance of time.

I find Prigogine's investigation of disorderly processes in the natural world very useful in thinking about how societies transform themselves. As one who has studied social change for many years, I have been, struck repeatedly by the inadequacy of our dominant social theories and ideologies, most of which are remnants from the Newtonian age.

Marxists assure us that social change is a reflection of class conflict. Evolutionists insist that societies change according to Darwinian laws, trying experiments and discarding failures. Toynbee wrote of challenges and response. Still others build grand theories on a single factor, so we get a welter of determinisms — energy is the key variable shaping societies, or culture predominates, or ecology, or demography, or whatever.

My own work has attempted to cross much disciplines. In *Future Shock, The Third Wave* and othert books, I focus on the changing interrelationships among multiple factors, from economics and politics to technology, communications, culture and psychology. But crossing traditional intellectual boundaries is a lonely work. It violates academic turfs. Beyond this, to focus, as I have, on discontinuities, upsets and reversals, rather than on continuity, and simple, straightline changes, is even harder and lonelier.

This is why Ilya Prigogine's work strikes me as a breath of intellectual oxygen. Prigogine combines sweeping synthesis with an emphasis on process. He and his colleagues in the so-called « Brussels school » may well represent the next revolution in science as it enters into a new dialogue not merely with nature, but with society as well. It may affect our social and political thinking because it brings with it a novel, comprehensive theory of change. And some day this model may help us understand the politics of unstable systems.

There is far more to the Prigoginian world view than I have so far suggested. The emergent science of instability holds that while some parts of the universe may operate like machines, these are closed systems, and closed systems, at best, form only a small part of the physical universe.

Most phenomena of interest to us are, in fact, open systems, exchanging energy or matter (and information) with their environment. Surely biological, political, economic and social systems are open, which means that the attempt to understand them in mechanistic terms is doomed to failure.

This tells us, moreover, that most of reality, instead of being orderly, stable, and equilibrial, is seething and bubbing with change, disorder, and process, and that chaos is not unusual or exceptional. It is a « normal » part of reality and it can give rise to structure and non-randomness. In fact, chaos may be the necessary precursor of a higher level order.

Prigogine argues that all systems contain subsystems, which are continually « fluctuating ». At times, a single fluctuation in one of them may become so powerful, as a result of positive feedback, that is shatters the entire preexisting organization.

At this revolutionary moment — called a « singular moment » or a «bifurcation point » — it is inherently impossible to determine in advance which direction change will take : whether the system will disintegrate into chaos or leap to a new, more differentiated, higher level of order or organization, called a « dissipative structure ». (Such physical or chemical structures are termed dissipative because, compared with the simpler structures they replace, they require more energy to sustain them).

Could this model help us to understand the political phenomenon of revolution ? And what about bifurcation points ? If ever one saw a society at the brink of bifurcation, hovering each day at the edge of either total break-up or fundamental restructure on a higher level, it is South Africa today. Is it about to become a Prigoginian « dissipative structure » ?

Another of Prigogine's findings is that when systems are thrown into a non-equilibrial condition, when internal instability is very high, they become supersensitive to outside forces. This might be a physical vector. But in social terms, might it also be a sudden flood of immigrants ? Or foreign propaganda beamed into country when it is on the edge of upheaval ? Or a sudden shiftin the currency ratios when an economy is at the breaking point ? If he is right, such events are likely to have far greater internal impact on an unstable society than on one resting comfortably in equilibrium.

And since small inputs can yield large non-linear effects, it should hardly surprise us when a seemingly small political act touches off a huge reaction.

The recent Daniloff case is instructive. The U.S. arrests a Soviet spy in New York. Not a major matter, as such things go. This provokes the Soviets to arrest Daniloff, escalating matters. Now, amplified by the media and other non-linear and positive feedback processes, the Daniloff case becomes part of global summitry, first as an obstacle to the holding of a U.S.-U.S.S.R. summit, then as a trigger for a strange new phenomenon — a so-called pre-summit summit. A man hands a reporter a small package in Moscow and superpower relations change as a result.

Or take the Greenpeace case in France not long ago. Here a tiny environmental organization decides to protest French nuclear testing in the Pacific. Compared to the French government, it is a mere gnat. Nonetheless, a decision is made in Paris to damage the Greenpeace ship

in order to prevent its sailing into the test zone. The ship is sunk. A man is killed. The effects cascade rapidly, and the French media, led by *Le Monde* amplify them into a Cabinet crisis. Before we know it, a tiny group with minimal influence in France forces a top-level reshuffle and vastly complicates France's relations with New Zealand and other Pacific countries. Small input — big effect.

Finally, what about Prigogine's contention that at various moments — when systems are at the bifurcation point — chance plays a key role ? In classical science, chance was an unwelcome outsider. It did not exist. In the new science, it is determinism's partner. And that makes the new science much closer to the realities we deal with in economics and politics. Ask any stock market investor or political pro about Lady Luck.

A group of terrorists in the tour ship *Achille Lauro* sit in a cabin polishing their guns. By a chance, just at that moment, a steward walks by and catches a glimpse of them — and before long an elderly and handicapped man is murdered, jets roar into the sky and the Italian Government is shaken.

The new world revealed by the science of instability is filled with upheavals and revolutions, with rise and fall and reorganization and just plain fortune. It is not a world of absolutes and eternals. It is a world in which time hastens us toward unknown — and sometimes unknowable — destinies. It is a world that refuses to oblige the political or religious fanatic whose dogmas are held to be timeless.

The agricultural revolution of 10.000 years ago launched a first wave of historic change and introduced new models of reality. In turn, the Newtonian system arose at a time when feudalism in Western Europe was crumbling, when the social system was far from equilibrium. Its model was embraced by a machine civilization. For these reasons, it makes sense to me to regard the Newtonian knowledge system as, itself, a « cultural dissipative structure » born of that massive social fluctuation we call the industrial revolution — the Second Wave of transformatory change.

Similarly, I believe that Prigogine's ideas are central to the latest revolution in science. They seem far more relevant to human affairs than the bloodless, cold machine model taught by classical science. And I cannot help but see these ideas in relationship to the demise of the Machine Age and the rise of what I have called a « Third Wave » of civilization.

Alvin Toffler is the author of « Future Shock » and « The Third Wave ».

Concluding Remarks

Edward W. Ploman
Vice-Rector of the University of United Nations (Tokyo)

The request to give some concluding remarks at the end ot this 1985 Discoveries Symposium is an honour but also caused me apprehension. I am not a specialist in mathematics, physics, chemistry, ecology or other fields of scientific inquiry represented here. However, there was the possibility that I might contribute through a different perspective. My work concerned with the development of international law has been focussed on new common resource realms such as outer space, ocean resources, international environmental issues, communications and information — all areas that have been opened up of been deeply influenced by advances in science and technology. Thus, this work involved both laws of nature and laws to order human conduct.

I will therefore not pretend to draw any conclusions, a task which would be daunting for anybody facing the rich and varied fare offered by this conference. I will rather offer reflections, some of them necessarily quite personal, on what we have discussed during these last few days — and also on what we have not done, or have neglected.

The title of this Symposium « Laws of nature and human conduct » is challenging, particularly through the need to build bridges between the two. In one perspective the issues implied in this title could be seen in terms of « transition in science » and « science in an age of transition ». Generally, we have devoted more attention to laws of nature than to human conduct. This proves our fascination with the transition in science and our difficulties in dealing with human conduct in an age of transition. In the long run, the transition in science will obviously affect our concepts about human nature and social conduct but we seem unclear about, and probably also afraid of, how the transition in science

will influence our images of human conduct — and thus of ourselves. As Jonas Salk put it in his inimitable way during the pause : we all have the same itch but we don't know where to scratch. It is against this background that I offer my reflections, focussed on three main points.

The first point concerns science in transition reflected in new images, new models and metaphors. These new concepts and images point to the depth of this transition and are most clearly expressed by *complexity* as one of the key concepts all through the Symposium. Complexity is clearly more difficult to deal with than simplicity ; complexity can therefore be seen to offend our « rage for order » but, suddenly it seems, complexity is found to exist everywhere, even in basic physical systems and, surprisingly, in pure mathematics. With complexity go other concepts such as non-linearity, non-equilibrium, randomness, open dynamical systems. Some of these concepts denoting new approaches to the description and investigation of reality were however only mentioned, but not further explored even though they would seem crucial in relation to images of human beings, to our comprehending their conduct and the conduct of human affairs.

Other, perhaps unexpected, key concepts have been ambiguity, even paradox. However, some reflection on our debates at this Symposium should make them seem less surprising : our discussions have also shown examples of deep ambiguity, sometimes reflected in an opaque veil between models and reality. To my mind, some of the models presented, ostensibly designed to provide a better understanding of economic behaviour, made the messiness and disarray of the current international economic scene seem conspicuous by its absence. Similarly, the concept of artificial intelligence seems to me profoundly ambiguous and reminds me — with all due respect — of a cartoon that appeared in an English journal some months ago. In this cartoon, a rather upper-class looking lady turns to her husband and says : « Oh, I see in the paper that there is something called artificial intelligence. Surely, then, there must be something called artificial stupidity. »

Unwarranted paradox ? May be, but certainly no worse than the paradoxes of our current global situation. One apparent paradox could be seen in terms of great capacity and equally great incapacity at the same time. We have the capacity to destroy life on earth and exhaust the carrying capacity of the planet but not to cope with these new powers. We have the capacity to feed people everywhere but seem incapable of gaining the fight against hunger and poverty. If we no longer equate rationality with certainty, we can also look in new ways at risk and risk-taking — but what about those who live at or under the poverty line and cannot afford the freedom of taking risks since failure would mean starvation ? All societies, be they strong or weak, are now exposed to forces beyond their control, and there is a growing realization that we do not seem to understand and have lost control over the global processes of change which affect all of us. Nations and societies have become more vulnerable, one cause being our incapacity to deal with the complexities and asymmetries of interdependence.

However, paradox apart, there seemed to be reflected in the Symposium together with the recognition of complexity, of indeterminacy and what I would see as the liberating effects of these concepts, a strong longing for certainty, a sometimes almost desperate search for a new rationality able to provide the same security as the old. But if the new rationality is taken to include randomness, fluctuation, risk one could ask : how deep is the acceptance of change, of transition, of the stochastic — and how rational is rational when applied to human conduct. Ambiguity is embedded in our experience of day-to-day living but still causes strong aversion in science, an aversion which paradoxically at this stage sometimes seems stronger in the social sciences than in the physical sciences. Perhaps, then, the West should learn from those cultures that have, cheerfully it seems, accepted paradox, and expressed contradiction in strong symbols such as the yin-yang.

I was intrigued by the fleeting reference to the possibilities offered by non-equilibrium constraints. I would tend to see this concept as a well-known phenomenon in art, particularly in music, a concept which deserves further elaboration and application — as would other such concepts that glimmered in the discussions.

I have in fact already started to move in on the second point which concerns the linkages between laws of nature and human conduct or, rather, ideas about human conduct.

With my background it would be natural to look at laws in two different ways :

— laws of nature as an expression of the dialogue between nature and human consciousness, in an attempt to capture what is, what « really happens out there » ;

— laws of man which concern both what is and what ought to be, and which thus concern not only human conduct but also human values which have hardly been, mentioned at all during the Symposium.

While Mr. Honda and others have emphasized the idea of « man at the centre », at this Symposium human beings figured explicitly in only one presentation : Peter Allen's stochastic and Cartesian fishermen.

Interestingly enough, the only other individuals mentioned were bees and ants.

There are, though, some bridges between the « new laws » of nature and human conduct, mainly seen in the applications of the emerging new scientific concepts and paradigms. Professor Prigogine and his colleagues have applied their ideas to regional planning in north Holland, to urban problems and more recently, in co-operation with the United Nations University, to the problems of fisheries management.

The flow between the two realms can also be seen the other way round : Wolfram's application of a simple rule over and over again to form unexpected computer patterns, reminded me of paintings by Vasarely and of the patterns in a Persian carpet.

A key linkage was of course provided by Holling's concept of man as a planetary phenomenon, and as a steward of this planet — which I should like to extend to our immediate environment which includes the solar system. It would not seem necessary to use our new-found scientific and technical power to pollute outer space with human folly.

Another example of linkage I can draw from my own field. A few legal scholars have pointed out that legal theory and practice should as closely as possible be based on up-to-date scientific knowledge. This requirement should be obvious when we try to define legal rules for outer space, ocean resources, the use of the radio frequency spectrum, acid rain or environmental conservation.

It should not be difficult to grasp the correspondence between the concept of eco-technology promoted by the Honda Foundation and the concept of eco-economy defined by René Passet. However, such linkages can cause their own ambiguities. Some economists have pointed to the linkages between economic theories and the evolving laws of physics ; thus

— classical liberal economics is seen as based on Newtonian physics

— Marxian economics is seen as based on the classical thermodynamics of closed systems

— which economic theory should correspond to the thermodynamics of open systems far from equilibrium and to the new complexity paradigms ?

Thirdly, I should like to combine the two perspectives : laws of nature and human conduct seen in combination whith science in transition and science in an age of transition.

Laws of nature : their meaning, whatever epistemological approach used, concerns what really happens in the dialogue between man and nature. Laws to regulate human conduct concern not only what is but also what ought to be. There would seem to be an abyss between the descriptive laws which science elaborates and the normative rules conceived by society and defined by the jurists. However, the need is for building bridges, to provide a dialogue between a changing science of nature and human conduct in an age of transition.

The combination of the two areas, particularly in the perspective of a reconceptualization, the emergence of a new rationality, thus of a new world view, a new Weltanschauung, puts formidable demands on learning. On learning and the development of the learning capacity not only of individuals but also of groups and institutions, even of communities, from villages to entire societies. This is one of the key elements in the concept of global learning as conceived by the United Nations University : global learning seen both as learning about global processes and issues and learning seen as a global process.

To learn how to generate and utilize knowledge represents only one aspect of global learning. Perhaps even more important is the learning required for coping with transition and change which is not only difficult

but also painful. And more, we have to learn how to cope with the ambiguities of change. I agree with Alvin Toffler when he speaks of instability as an inevitable effect of « the third wave ». It seems necessary, though, to add the idea of bifurcation : at certain points the instability can lead either in the direction of a new, open rationality or in the direction of a new fundamentalism.

To achieve a new rationality we would also have to learn how to cope with and manage complexity which can well be represented by the phenomena of and the new perceptions of interdependence, of countries as well as of issues.

The issues of complex interdependence lead us directly to the problems of governance, to what appears as a deep crisis of governance, every day reflected in the newspapers : terrorism depends on our failure to solve social problems and on our lack to govern the availability of weapons ; refugees and migrations now represent movements of people of a size unprecedented in modern times ; there is crisis in the world economy as much as in economic development theory. And these crises of governance hit us in a situation where normally facts are uncertain, opinions — even scientific ones — divided, values in dispute, stakes large and decisions urgent.

These problems are such that they obviously cannot be solved on a single-issue or single-discipline basis ; not even a single-culture approach would any longer be adequate.

Thus we need to learn from different cultures, to accept multiple perceptions and multiple perspectives. This, of course, has already happened in art but we need it in all our endeavours, of science and of governance. In terms of the issues discussed at this Symposium we also need to learn how to go through that door which is symbolized by the current reconceptualization, that door which, to use an expression coined by Professor Prigogine, is symbolized by the opening up of a new theoretical space.

V

Administrative Council

of

— The Honda Foundation
— The International Institutes
for Physics and Chemistry
Founded by E. Solvay.

Lists of

— Scientific Committee
of the Symposium
— Scientific Secretaries
of the Symposium
— Participants
— Observers.

Administrative Council of The International Institutes for Physics and Chemistry founded by E. Solvay

Solvay
Président de la Commission Administrative des Instituts Internationaux de Physique et de Chimie.

E. Bingen
Ondervoorzitter van de Raad van Beheer van de Internationale Instituten voor Fysica en Chemie.

I. Prigogine
Directeur des Instituts Internationaux de Physique et de Chimie.

J. Philippot
Adjunct-Directeur van de Raad van Beheer van de Internationale Instituten voor Fysica en Chemie.

J. Reignier
Penningmeester van de Raad van Beheer van de Internationale Instituten voor Fysica en Chemie.

G. Nicolis
Secrétaire du Conseil d'Administration des Instituts Internationaux de Physique et de Chimie.

R. Debever, P. Glansdorff, A. Jaumotte, J.M. Piret, G. Van Binst
Members.

J. Gehenieu, A. Bellemans
Scientific Secretaries.

Administrative Council of the Honda Foundation

S. Honda,
Founder of Honda Foundation and Honda Motors, Co., Ltd.

T. Shimoda
President of the Honda Foundation.

H. Kawashima
Vice-President of the Honda Foundation.

S. Aïda
Director of the Honda Foundation.

E. Tokura
Director of the Honda Foundation.

S. Shinomiya
Director of the Honda Foundation.

T. Harata
Director of the Honda Foundation.

K. Kikutake
Councillor of the Honda Foundation.

Y. Murakami
Councillor of the Honda Foundation.

K. Atsumi
Councillor of the Honda Foundation.

T. Furakawa
Councillor of the Honda Foundation.

T. Ueda
Managing Director of the Honda Foundation.

Scientific Committee

I. Prigogine *Université Libre de Bruxelles* *Service de Chimie Physique II*	Belgium
S. Aida *University of Electro-Communications* *Engineering Systems*	Japan
W. Barnett *The University of Texas at Austin* *Department of Economics*	USA
Sir Henry Chilver *Vice-Chancellor* *Cranfield Institute of Technology*	U.K.
U. Colombo *Presidente* *Comitato Nazionale dell'ENEA*	Italy
A. Jaumotte *Université Libre de Bruxelles* *Faculté des Sciences Appliquées*	Belgium
Sir Karl Popper	U.K.
J. Salk *The Salk Institute*	USA
A. Toffler	USA
A. van Cauwenberghe *Rijksuniversiteit Gent* *Laboratorium voor Regeltechniek*	Belgium

Scientific Secretaries

G. Nicolis, M. Sanglier, P. Allen, S. Pahaut
Université Libre de Bruxelles

List of Participants

P.M. Allen
Université Libre de Bruxelles
Service de Chimie Physique II — Belgium

S. Aida
University of Electro-Communications
Engineering Systems — Japan

B. Arthur
Stanford University
Population Studies — USA

R. Artigiani
U.S. Naval Academy
History Department — USA

K. Atsumi
University of Tokyo
Faculty of Medicine — Japan

W. Barnett
The University of Texas at Austin
Department of Economics — USA

L. Bellemin
Adm. Principal, CEE — Belgium

W.A. Brock
University of Wisconsin
Department of Economics — USA

A. Bruylants
Académie Royale de Belgique — Belgium

I. Burton
The International Federation of Institute of Advanced Studies (IFIAS) — France

E.R. Caianiello
University of Salerno
Dept. of Theor. Physics — Italy

G. Careri
Universita degli studi di Roma
Dipartimento di Fisica — Italy

A. van Cauwenberghe
Rijksuniversiteit Gent
Laboratorium voor Regeltechniek — Belgium

G.J. Chaitin *IBM, Thomas J. Watson Research Center*	USA
H. Chantrenne *Académie Royale de Belgique*	Belgium
Chen Ping *Center for Statistical Mechanics Robert Lee Moore Hall*	USA
H. Chestnut *Swiss Foundation, Inc.*	USA
Sir Henry Chilver *Vice-Chancellor* *Cranfield Institute of Technology*	U.K.
E. Clementi *IBM*	USA
J.F. Coales *C.B.E., C. Eng.*	U.K.
M. Conrad *Wayne State University* *Dept. of Computer Sciences*	USA
R.W. Crosby *Federal Aviation Administration* *Department of Transportation*	USA
A.R. Curran *Keene State College* *Department of Management*	USA
A. Danzin *Cabinet d'Etudes Scientifiques et Techniques*	France
J.L. Deneubourg *Université Libre de Bruxelles* *Service de Chimie Physique II*	Belgium
J. Dockery *Office Joint Chiefs of Staff*	USA
J. Earley *Georgetown University* *Chairman Dept. of Chemistry*	USA

W. Ebeling
Humboldt Universität
Sektion Physik — GDR

Fang Fu-kang
Vice President
Beijing Normal University — People's Republic of China

T. Furukawa
Univeristy of Tokyo
Faculty of Medicine — Japan

J. Géhéniau
Académie Royale de Belgique — Belgium

O. Giarini
The Risk Institute — Switzerland

P. Glansdorff
Académie Royale de Belgique — Belgium

L. Glass
McGill University
Department of Physiology — Canada

W. Grünsteidl
Director of CPMS
Corporate Strategic Planning — Netherlands

G. Hambraeus
Chairman
Royal Swedish Academy of Engineering — Sweden

Hao Bai-Lin
Institute of Theoretical Physics
Academia Sinica — People's Republic of China

J. Hartland
Council of Europe
Committee on Science and Technology — France

B. Hess
Max-Planck Institut für Ernährungsphysiologie — FRG

C. Holling
University of British Columbia
Institute of Animal Resource Ecology — Canada

M. Homès *Académie Royale de Belgique*	Belgium
G. Homès-Van Schoor *Académie Royale de Belgique*	Belgium
B.A. Huberman *Xerox Palo Alto Research Center*	USA
E. Illy	Italy
A. Jaumotte *Université Libre de Bruxelles* *Faculté des Sciences Appliquées*	Belgium
S. Kauffman *University of Pennsylvania* *Dpt of Biochemistry and Biophysics*	USA
D. Kondepudi *University of Texas* *Center for Statistical Mechanics*	USA
E. Laszlo	Italy
P. Ledoux *Académie Royale de Belgique*	Belgium
M.J. Lesser *PANCAN*	U.K.
J. Mac Glade *Bedford Institute of Oceanography Marine Fish Division*	Canada
P. Malgrange *Centre d'Etudes Prospectives d'Economie Mathématique* *Appliquée à la Planification*	France
C. Masson *Laboratoire de Neurobiologie sensorielle de l'Insecte (CNRS)*	France
F. Margulies *Honorary Secretary of IFAC*	Austria
P. Michot *Académie Royale de Belgique*	Belgium

M. Migeotte
Académie Royale de Belgique — Belgium

E. Mosekilde
Technical University of Denmark
Physics Laboratory III — Denmark

Y. Murakami
University of Tokyo
Faculty of Medicine — Japan

G. Nicolis
Université Libre de Bruxelles
Faculté des Sciences — Belgium

C. Nicolis
Institut d'Aéronomie Spatiale — Belgium

J.S. Nicolis
University of Patras
Commun. Systems, School of Engineering — Greece

S. Pahaut
Université Libre de Bruxelles
Service de Chimie Physique II — Belgium

R. Passet
CESTA — France

J. Pasteels
Université Libre de Bruxelles
Département de Biologie Animale — Belgium

A. Philippart
Ministère de l'Education Nationale — Belgium

E.W. Ploman
Vice Rector
The United Nations University — Japan

I. Prigogine
Université Libre de Bruxelles
Service de Chimie Physique II — Belgium

J.L. Proops
University of Keele
Department of Economics — U.K.

Ph. Roberts-Jones *Secrétaire Perpétuel de l'Académie Royale de Belgique*	Belgium
J. Salk *The Salk Institute*	USA
M. Sanglier *Université Libre de Bruxelles* *Service de Chimie Physique II*	Belgium
A. Schaff *Pension Research Institute*	Poland
E. Schneider *University of Maryland* *Chesapeake Biological Laboratory*	USA
T.B. Sheridan *Engineering and Applied Psychology MIT*	USA
K. Sigmund *Universität Wien* *Institut für Mathematik*	Austria
T.D. Spearman *European Science Foundation*	France
J. Subba Rao *Jawaharlal Nehru University* *School of Environment Sciences*	India
G. Subba Rao *All India Institute of Medical Science* *Department of Biophysics*	India
I. Stengers *Musée du Parc de la Villette*	France
R. Thomas *Université Libre de Bruxelles* *Laboratoire de Génétique*	Belgium
A. Toffler	USA
H. Toffler	USA

K. Tomita
University of Kyoto
Department of Physics Japan

R. Ulanowicz
University of Maryland
Ctr. for Environ. & Estuar. Studies USA

M.B. Velarde
Univ. Nacional De Educ. A Distancia
Dept. de Fisica Fundamental Spain

J. Voge
CNET, Bureau des affaires générales France

I. Wallerstein
Fernand Braudel Center USA

B. West
La Jolla Institute
Center for Stud. of Non Linear Dynamics USA

S. Wolfram
Institute for Advanced Studies
Princeton USA

E.R. Woodcock
Synetics Corporation USA

Sir David Zeidler
President Australian, Academy of Th. Sc.
Clunies Ross House Australia

Observers

Cl. d'Aspremont *Université Catholique de Louvain* *CORE*	Belgium
A. Babloyantz *Université Libre de Bruxelles* *Service de Chimie Physique II*	Belgium
T. Bernold *Gottlieb Duttweiler Institut*	Switzerland
J.P. Boon *Université Libre de Bruxelles* *Service de Chimie Physique II*	Belgium
F. Boon-Wolter *STRATEC*	Belgium
G. Engelen *Université Libre de Bruxelles* *Service de Chimie Physique II*	Belgium
C. George *Université Libre de Bruxelles* *Service de Chimie Physique II*	Belgium
A. Goldbeter *Université Libre de Bruxelles* *Service de Chimie Physique II*	Belgium
S. Goss *Université Libre de Bruxelles* *Service de Chimie Physique II*	Belgium
M. Ikonomov *SNCF*	France
M. Kerszberg *Institut für Festkorperforschung*	FRG
H. Nicholson *The University of Sheffield* *Department of Control Engineering*	U.K.
J.F. Richard *Université Catholique de Louvain* *CORE*	Belgium

J. Sinton
Gottlieb Duttweiler Institut Switzerland

G. Thines
Université Catholique de Louvain
Centre Albert Mic, Biol. du Comport. Belgium

• ABSTRACTS

PRIGOGINE Ilya — « A New Rationality ?», *Mondes en Développement (14), 54-55, 86 : pp. 19-39.*

The historian Al. Koyré has aptly described the western scientific revolution of the seventeenth century as a transition « From the closed world to the infinite universe ». Still, classical science concentrated its efforts mainly on stable, periodic motions. We now begin to understand the message of the second law of thermodynamics : we are living in a world of unstable dynamical systems, and it is this very instability which leads to stochasticity and irreversibility. Some of the main subjects of contemporary science, be it biology or cosmology, illustrate in a striking way the constructive role of irreversible processes. We must give up the myth of « complete knowledge » which has haunted western science for three centuries. Both in the hard sciences and in the so-called soft sciences, we have only a window knowledge of the world we want to describe.

CARERI Giorgio — « Natural Order, Natural Laws and Correlated Events », *Mondes en Développement (14), 54-55, 86 : pp. 43-48.*

The aim of this paper is to call attention on the development of the idea of natural law, from Galileo, Descartes and Newton up to the present time. It is shown that both is physics and biology the original formulation of deterministic law is now beeing replaced by a probabilistic one, better formulated in terms of correlated events. The parallel developments in the western philosophical through is briefly mentioned.

ALLEN Peter M., Mc GLADE Jacqueline M. — « Modelling Complexity : the Dynamics of Discovery and Exploitation in a Fisheries Example », *Mondes en Développement (14), 54-55, 86 : pp. 49-76.*

The basis of current models used in Fishery Management is briefly examined, and various shortcomings are discussed. Alternative, dynamic models are described which are based on the data available for the Nova Scotia Groundfish Fisheries. It is shown that human responses can amplify relatively small annual environmental fluctuations, leading to large, quasi- cyclic changes in catch and profit. In a detailed spatial model it is shown that stochastic behaviour on the part of some fishermen is necessary for the survival of the fishery, and that the efficiency and size of the industry depends very much on the information flows concerning catch. A very general discussion is given which shows how these ideas are important in our understanding of innovation and discovery in general terms.

HESS Benno, MARKUS M. — « Transitions between Biochemical Oscillations and Their Relevance to Biology », *Mondes en Développement (14), 54-55, 86 : pp. 81-90.*

The general dynamics of biochemical processes are reviewed using glycolysis as a model on the basis of theoretical as well as experimental investigations. This process, especially under periodic excitation, displays a large variety of time patterns such as periodic, quasiperiodic as well as chaotic oscillations, and acts as a time pattern converter. These time patterns may correspond to attractors coexisting in phase space, and show hysteresis properties. In addition, however, they also may display crisis phenomena which are transitions between chaotic behaviour toward an attractor of e.g. periodic oscillations. These autonomous transitions from chaos to other types of oscillations through crisis may be of fundamental importance in biochemical as well as biological processes and furthermore in any nonlinear and open dynamic system with the property of self-organization and time pattern display, and yield a new understanding of complex dynamic behaviour.

JENSEN Klaus S., MOSEKILDE Erik, HOLSTEIN-RATHLOU Niels-Henrik — « Self-Sustained Oscillations and Chaotic Behaviour in Kidney Pressure Regulation », *Mondes en Développement (14), 54-55, 86 : pp. 91-109.*

Recent experimental results have revealed an oscillatory behaviour of the proximal intratubular pressure in rat kidneys. For normal, healthy rats, these oscillations have

a well-defined frequency and amplitude. For genetically hypertensive rats, however, the oscillations are highly irregular.

This paper presents a dynamical model of the most prominent feedback mechanisms involved in nephron pressure- and flow regulation. With realistic parameter values, the model exhibits a limit cycle behaviour with amplitudes and frequencies in accordance with experimental findings. Variation of those parameters which are likely to differ between normal and hypertensive rats causes the model to develop through a cascade of period-doubling bifurcations to reach a chaotic regime.

CONRAD Michael — « Reversibility in the Light of Evolution », *Mondes en Développement (14), 54-55, 86 : pp. 111-121.*

Recent efforts to unify physics and biological evolution give new importance to questions about the origin of irreversibility and the physical meaning of probabilities (fluctuations, variability). Probabilities can be interpreted in dynamical terms (e.g. chaos), in terms of limits of computation, or in terms of measurement. We consider in particular the posssibility of measurement-embedding models since these appear to minimize the strain in the overall conceptual architecture of science. Breakdown of microscopic reversibility is necessary for probability to have an irreducible status, but whether probability in fact has this status may be undecidable.

CHAITIN Gregory J. — « Randomness and Gödel's Theorem », *Mondes en Développement (14), 54-55, 86 : pp. 125-128.*

Complexity, non-predictability and randomness not only occur in quantum mechanics and non-linear dynamics, they also occur in pure mathematics and shed new light on the limitations of the axiomatic method. In particular, we discuss a Diophantine equation exhibiting randomness, and how it yields a proof of Gödel's incompleteness theorem.

PASTEELS Jacques, DENEUBOURG Jean-Louis, GOSS Simon — « Transmission and Amplification of Information in a Changing Environment : the Case of Insect Societies », *Mondes en Développement (14), 54-55, 86 : pp. 129-156.*

Self-organization processes in insect social activities are discussed and illustrated by the ant's foraging behaviour.

From the interactions between stochastic events inherent to the behaviour of the individuals and amplifying mechanisms due to communication, collective coherent and efficient patterns emerge which are characterized by larger scales of time and magnitude than the corresponding individual scales.

Self-organization of this kind is related to the concept of mass communication (Wilson, 1971).

Mass recruitment allowing societies of numerous unspecialized workers to exploit large, ephemeral and unpredictible resources by chemical trails illustrates this process and is analyzed in detail. Mass recruitment is highly flexible. « Errors » during trail following are more frequent when the recruitment is low, i.e. at the beginning of the recruitment or when food sources are of poor quality. « Lost ants » patrolling the foraging ground allow new discoveries and the selection of the best resource is possible by the adjustment of recruitment rates to food quality. Efficiency in trail following increases non-linearly as the recruitment proceeds and the trail is reinforced by successful recruits. Two identical sources are exploited asymmetrically illustrating a bifurcation phenomenom, the adaptive value of which is discussed.

It is suggested that the level of accurracy in recruitment could be tuned to ecological constraints (e.g. food distribution in time and space) and biological constraints (e.g. colony size). Small societies use other recruitment techniques which are less flexible, but more accurate.

When resources are highly predictable or long-lasting, specialization of the workers is observed. It is suggested that task or territorial specialization is in part due to the acquisition of long-term memory, and that it is autocatalytic : the more an ant performs a task, the more it will continue to do so.

Specific patterns could then emerge from the interplay between : the number of participants competing for a specific task ; more or less stochastic events in the environment inducing different spatio-temporal distribution of tasks to perform, or different reinforcement patterns ; and the ants' learning capacity which could be tuned to environmental and biological constraints.

Both self-organization processes, mass communication and specialization by the acquisition of long-term memory, occur in other social activities and can cooperate to regulate social insect behaviour.

NICOLIS John S. — « A Study Program of Chaotic Dynamics Applied to Information Processing », *Mondes en Développement (14), 54-55, 86 : pp. 157-174.*

Information processing aims at category formation in the (human) cognitive system or the classification of the external (and a good deal of the system's « internal ») world in an hierarchy of abstracts patterns. These patterns in turn may be used as algorithms for simulating observed physical processes.

In this paper we examine how this abstraction process may be dynamically implemented using the concepts of « Chaotic strange attractors ».

HOGG Terry, HUBERMAN Bernardo A — « Order, Complexity, and Disorder », *Mondes en Développement (14), 54-55, 86 : pp. 175-184.*

A physical measure of the complexity of a system is introduced which is based on its diversity, while ignoring its detailed specification. It applies to hierarchical structures made of elementary parts and provides a precise, readily computable measure. Its applicability to ecosystems and other adaptive structures is discussed.

BROCK William A — « Testing for Deterministic Chaos in Economic Data », *Mondes en Développement (14), 54-55, 86 : pp. 191-207.*

This paper discusses tests on time series for the presence of low dimensional deterministic chaos. Empirical applications on U.S. business cycle data are reviewed. Two tasks are necessary to test a time series for (low dimensional) deterministic chaos : (1) dimension must be calculated and shown to be low relative to comparable (psuedo) random numbers ; (2) Lyapunov exponents and other measures of local instability must be calculated and show local instability relative to a useful standard. The evidence for low dimensional deterministic chaos in U.S. postwar II GNP data is weak.

EARLEY Joseph E. — « Aggregation, Actualities and Attractors », *Mondes en Développement (14), 54-55, 86 : pp. 225-228.*

Presence of an attractor makes it possible for an open dynamic system to interact with external others as an integrated whole. Coherence of an open dynamic system is detected only by external others that take part in interactions that have appropriate characteristics (e.g., whether attractors are detected in economic systems depends on details of « aggregation »). Long-standing problems of epistemology and ontology (illustrated by an issue between Foucault and Derrida) might be clarified by keeping both points in mind.

SCHUSTER Peter, SIGMUND Karl — « Evolutionary Game Dynamics », *Mondes en Développement (14), 54-55, 86 : pp. 229-236.*

Phenotypes can be viewed as competing strategies in an evolutionary game where the payoff is Darwinian fitness. The resulting game dynamics yields a global picture of the evolution of frequencies. Equilibria, limit cycles and chaotic attractors occur. Notions like evolutionary stability or permanence are appropriate to describe the principal asymptotic features. Game dynamics can be used to study fighting behaviour, sex ratio, parental investment or the evolution of cooperation. Similar equations occur in population genetics, mathematical ecology and chemical kinetics. The analysis of « games without rationality » helps to understand the effects of selection.

VOGE Jean — « The Information Economy and the Restructuring of Human Organizations », *Mondes en Développement (14), 54-55, 86 : pp. 237-244.*

The important information economy in the most industrialized countries shows an increase of complexity with jumping costs consuming profits brought by new technologies. There is, then, a stagnation of productivity. « Science of complexity » which studies the interactions between the structures and the operational economy — imputes this stagnation to the pyramidical structures of human organizations. In short a structural change is necessary that would bring along hope for a new productivity.

PROOPS John L.R. — « The Physical Input to Economics », *Mondes en Développement (14), 54-55, 86 : pp. 245-252.*

Physical science has been influential upon economics, mainly through the adoption by economics of the mechanical mode of reasoning. Recently the Second Law of Thermodynamics has also been recognised as having implications for economics, but neither mechanics nor classical thermodynamics offer a conceptual framework that encompasses technical progress and the evolution of economies that we observe. Modern far-from-equilibrium thermodynamics, however, does offer a framework for dealing explicitly with the irreversibility of time, and the implications of such time irreversibility for economic theory are explored.

METCALFE John S., GIBBONS Michael — « Technology, Variety and Competition », *Mondes en Développement (14), 54-55, 86 : pp. 253-266.*

In this paper a dynamic theory of the competitive process is developed around the three concepts of efficiency, fitness and creativity. The burden of the argument is that differences between firms in terms of efficiency, fitness and creativity determine the rate and direction of technological change within a population of firms. The paper is divided into three sections. In section I the dimensions of firm behaviour are discussed. In section II a model of the competitive process is presented and finally, in section III some analyses of selection processes in simple environments which focus upon the ways in which technologies compete for economic weight are discussed.

AIDA Shushei — « Economic and Social Effects of Factory Automation », *Mondes en Développement (14), 54-55, 86 : pp. 277-283.*

Automation is primarily aimed at setting people free from painstaking and insignificant jobs as well as equally distributing benefits and wealth produced by modern technologies. Automation in the past were concerned with mass produced by modern technologies.

Although automation technologies created some negative reactions in the past, they have been accepted in industrially developed countries because they have enhanced both productivity for employers and industrial safety for workers. In contrast, this requires more workforce in market research, designing and other work on the « preparation stage », and distribution, sales and marketing on the « support stage » instead of « manufacturing stage ».

A review of the contents of jobs on the preparation and support stages enables them to encourage feedback among themselves to a certain degree and also to streamline the production process further.

GIARINI Orio — « Notes on the Limits to Certainty : Risk, Uncertainty and Economic Value », *Mondes en Développement (14), 54-55, 86 : pp. 285-295.*

Economics, in order to exist as a specific science or discipline, had to find a body of theory with specific demarcation lines : this was made possible by the notion of value as it was first defined by Adam Smith.

Such notion of value is based on a general equilibrium theory, which assume certainty and has developed in the same general cultural environment which produced the Newtonian theory of the universe.

Furthermore, the first formalizations of an independent and autonomous discipline of economics were quite consistent with the Industrial Revolution which stimulated them.

The modern economy is essentially based now on service functions, which incorporate as part of a system the manufacturing processes which were the focal point of the Industrial Revolution. In this sense, traditional economic theory is now incomplete with regard to contemporary economic realities.

The « Service Economy » is more concerned by the « functioning » of systems than by the « existence » of products per se, and is closely related to such concepts as disequilibrium, real time (duration), risk and uncertainty. Employment, as one of the key economic problems, can find new solutions in the perspective of the « Service Economy ».

LASZLO Ervin — « Toward a Non-Equilibrium System Theory of History », *Mondes en Développement (14), 54-55, 86 : pp. 297-304.*

Laws of nature, as formulated in contemporary nonequilibrium thermodynamics are not mechanistic but possibilistic : they define constraints and ranges of possibilities and do not predetermine evolutionary trajectories. They apply to human societies the same as to natural systems. Demonstration of this proposition calls for a nonequilibrium system theory of history, testable against empirical records and observations. The outline of such theory shows that the general direction, as well as the basic mode, of the historical development of societies is isomorphic with the direction and mode of evolution of nonequilibrium systems in nature. The detailed exploration of the theory could demonstrate the applicability of general laws of system development to society, thereby paving the way for a unitary and consistent conceptual framework for the natural and the social sciences.

ARTIGIANI Robert — « Language, Logic and Energy : A Thermodynamic Sketch of Greek Democracy », *Mondes en Développement (14), 54-55, 86 : pp. 305-315.*

Rigorous analyses of the role of irreversibility have greatly enriched the scientific picture of nature by incorporating its historical dimensions. This new scientific picture, applied to social and cultural developments, could provide links between the « stories » historians tell to describe particular events and fundamental natural processes. Linking social developments to natural processes, whose regularities have been established across a wide spectrum of phenomena, could supply the long-sought « covering laws » that would make human history comprehensible as well as memorable. In nature, irreversible processes often reveal a tendency to evolve through the self-organization of more complex structures, quantitatively defined. Unexpectedly, applying quantitative discriminators to history suggests a basis for qualitative conclusions about its humanistic implications. The emergence of democracy in ancient Greece is offered as an example.

WEST Bruce J., SALK Jonas — « Human Conduct and Natural Laws : Laws of Simple and Complex Systems », *Mondes en Développement (14), 54-55, 86 : pp. 317-325.*

A distinction is made between systems that are simple (linear, deterministic and predictable) and those that are complex. The latter can be either disorganized (unpredictable) or organized (nonlinear and of limited predictability). Human conduct is seen as falling into the last category, that of partially ordered systems, from which the tentative conclusion is drawn that the laws of human nature are probabilistic rather than deterministic and predictable. Data in support of this view are discussed.

PLOMAN Edward W. — « Discoveries Symposium, Bruxelles, October 1985, Concluding Remarks », *Mondes en Développement (14), 54-55, 86 : pp. 333-337.*

The following remarks presented at the end of the Symposium represent some reflections on the discussions at the Symposium. The challenge implied in the title of the conference « Laws of nature and human conduct » points to the need to build bridges between the two aspects of the theme. The presentations and discussions were mainly concerned with the new laws of nature ; they proved the fascination with science in transition, reflected in new models, paradigms and models and also the uncertainty

in dealing with human conduct and human values which themselves will be influenced by the new scientific approaches. Mention is made of some linkages between laws of nature and human conduct, in areas such as law, and in the application of new concepts to real-life situations. The reconceptualization in science also puts new demands on learning and the learning capacity not only of individuals but also of groups and institutions, even of societies to cope with what now is recognized as a crisis in the governance of human affairs, linked to the complex interdependence both of countries and of issues.

LAWS OF NATURE AND HUMAN CONDUCT

Specificities and Unifying Themes, Discoveries 1985 Symposium,
Hosted by « the International Institutes for Physics and Chemistry Founded by E. Solvay »
and Supported by « the Honda Foundation »

Ilya PRIGOGINE	5	Laws of Nature and Human Conduct — Specificities and Unifying Themes
Jacques SOLVAY	11	Opening address
Takeso SHIMODA	15	Opening address

I. The Reconceptualisation of the Physical Sciences, in the Perspective of Complex Systems

Ilya PRIGOGINE	19	A New Rationality ?
Hao BAI-LIN	41	On the Concept of Chaos
Giorgio CARERI	43	Natural Order, Natural Laws and Correlated Events
Peter M. ALLEN, Jacqueline M. Mc GLADE	49	Modelling Complexity : the Dynamics of Discovery and Exploitation in a Fisheries Example
Werner EBELING	77	On Stochastic Models for Human Conduct in the Evolution of Science and Technology
Benno HESS, M. MARKUS	81	Transitions between Biochemical Oscillations and their Relevance to Biology
Klaus S. JENSEN, Erik MOSEKILDE, Niels-Henrik HOLSTEIN-RATHLOU	91	Self-Sustained Oscillations and Chaotic Behaviour in Kidney Pressure Regulation
Michael CONRAD	111	Reversibility in the Light of Evolution

II. Dynamical Systems, Information and Cognitive Sciences

Gregory J. CHAITIN	125	Randomness and Gödel's Theorem
Jacques M. PASTEELS, Jean-Louis DENEUBOURG, Simon GOSS	129	Transmission and Amplification of Information in a Changing Environment : the Case of Insect Societies
John S. NICOLIS	157	A Study Program of Chaotic Dynamics Applied to Information Processing
Terry HOGG, Bernardo A. HUBERMAN	175	Order, Complexity, and Disorder
Toshiyuki FURAKAWA	185	Intelligence Quotient of Artificial Intelligences

III. Dynamical Systems, Economic Theory and Games

William A. BROCK	191	Testing for Deterministic Chaos in Economic Data
William BARNETT, Ping CHEN	209	Economic Theory as a Generator of Measurable Attractors
Joseph E. EARLEY	225	Aggregation, Actualities and Attractors
Peter SCHUSTER, Karl SIGMUND	229	Evolutionary Game Dynamics
Jean VOGE	237	The Information Economy and the Restructuring of Human Organizations
John L.R. PROOPS	245	The Physical Input to Economics
Michael GIBBONS, John S. METCALFE	253	Technology, Variety and Competition

IV. Toward a New « Rationality »

Yoichiro P. MURAKAMI	269	Rationality and Irrationality in the Theory of Knowledge
Shushei AIDA	277	Economic and Social Effects of Factory Automation
Orio GIARINI	285	Notes on the Limits to Certainty : Risk, Uncertainty and Economic Value
Ervin LASZLO	297	Toward a Nonequilibrium Systems Theory of History
Robert ARTIGIANI	305	Language, Logic, and Energy : A Thermodynamic Sketch of Greek Democracy
Bruce J. WEST, Jonas SALK	317	Human Conduct and Natural Laws : Laws of Simple and Complex Systems
Alvin TOFFLER	327	New Science of Instability throws Light on Politics
Edward W. PLOMAN	333	Concluding Remarks

V. Administrative Council of the Honda Foundation.
Administrative Council of the Instituts Internationaux de Physique et de Chimie founded by E. Solvay

	List of
343	Scientific Committee of the Symposium
344	Scientific Secretaris of the Symposium
345	Participants
352	Observers

GORDES (Bruxelles)

DE CET OUVRAGE, IL A ETE TIRE
SUR LES PRESSES DE J. DIEU-BRICHART
A OTTIGNIES-LOUVAIN-LA-NEUVE,
1500 EXEMPLAIRES SUR PAPIER OFFSET SENLIS 80 gr.
CONSTITUANT L'EDITION ORIGINALE.

Imprimé en Belgique

Dépôt légal : D/1987/4027/1

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means without the prior written permission of the publisher.

© Copyright 1987 by G.O.R.D.E.S. / Task Force of Research Information and Study on Science.

G.O.R.D.E.S.
44, avenue Jeanne - C.P. 124
B-1050 Bruxelles

Prix : U.S.$ 80.00

Printed in Belgium.