CITE AS IN K. PRIBRAM, ED,
ORIGINS: BRAIN AND SELF-ORGANIZATION,
ERLBAUM, 1994.

Self-Organization: Reexamining the Basics and An Alternative to the Big Bang

Paul J. Werbos
Room 675, National Science Foundation[1]
Arlington, Virginia, USA 22230
pwerbos@note.nsf

This paper provides a basic, simplified overview of the field of self-organization, stressing those parts of the field -- like thermodynamics and nonlinear systems dynamics -- where there really does exist a coherent, unifying mathematical theory. At the same time, it points towards some basic unanswered questions in this field, and suggests a variety of heresies which merit further research. For example, the paper will conclude with a "caricature model" of cosmology, suggesting how life and order could have evolved in the universe without any need for a Big Bang or for the exogenous mechanisms used in some of the classical alternatives to the Big Bang. Another paper presented at this conference by Ilya Prigogine (discussed in more detail below) takes a position on this issue , which I regard as even stronger (if more tactful): it argues that the basic phenomenon of time-forwards evolution, which underlies life and order, can be deduced entirely from local microscopic effects, without any need to invoke assumptions such as the Big Bang, or special initial conditions, or the special kinds of field effects exemplified in my caricature model.

I have tried to make this paper as simplified as possible, for the sake of interdisciplinary cooperation, which will be crucial to progress in this field. However, I have also tried to include the key mathematical and logical points -- in simplified form -- which underlie the discussion, in the later part of some subsections; these details are essential, in my view, because the literature on this field has become so complex that many scientists become too accustomed to relying on blind faith or on assumptions justified by footnotes to other sources, which cite other sources, and so on... on trails which do not always support the strong beliefs of later authors.

One participant at this conference complained that this paper goes too far in stating (and appearing to support) the views of orthodox mathematical physics which, in his view, have become a great straitjacket to modern science, reducing everyone's openness to new empirical data. Lerner[1] has made similar points, which I do not fully agree with. Nevertheless, the pure empiricist should note that the mathematical heresies described in this paper -- if pursued by future researchers -- could loosen that straitjacket to a significant degree. The caricature model here is emphatically not a grand unified, comprehensive mathematical theory; it is intended purely as a minimalist starting point, designed to encourage the new experiments which would in fact be crucial to achieving a higher degree of precision in a realistic manner.

The paper will begin by discussing what self-organization is, in general terms. Next, it will describe certain basic types of system important to the field. Finally it will summarize classical views (and misconceptions), modern views, and some new heresies, in that order. All five sections will identify fundamental questions for future research. The technical appendix will provide some new ideas in classical stability of particular importance to control theory and artificial neural networks: i.e., new methods for constructing Liapunov functions. Definitions of a few key terms are collected at the end, as an aid to the nonspecialist.

[1]The views herein are those of the author, developed on personal time, and do not represent NSF in any way.

WHAT IS SELF-ORGANIZATION?

Self-organization is the study of the way in which systems made up of simpler elements, governed by simple dynamical principles, spontaneously develop organization or patterns or order at a higher level. Self-organization is one of the most important fundamental topics in all of science and engineering, because it provides a kind of glue to connect and unify our knowledge across different levels of aggregation. This section will discuss the role of self-organization as a key unifying component of basic research in general.

In the scientific establishment and in Congress, considerable fuss has been made about the role of "basic scientific research" and the billions of dollars being spent on basic research. Yet, when we get down to fundamentals, all of our science has been built on the efforts to answer only four truly basic questions:

1. What is reality? I.e., what are the underlying phenomena -- like matter and energy -- which define our cosmos, and how do they work?

2. What is the universe -- how big is it, where does it come from, and what is it evolving to?

3. What is life? Why and where does it exist, and how does it work?

4. What is mind or intelligence?

Only a tiny percentage of the funding for "basic" scientific research is directed towards systematic, strategic efforts to find new and deeper answers to these questions. The vast bulk of our efforts seem to be ruled by mental inertia, by people studying issues like ,"What is the 18th spectral line of the 17th isotope of the 88th element?" As Kuhn points out[2], there is a great need for such "normal science," but normal science by itself is not enough to find answers to these questions in an efficient manner.

There is an analogy here to artificial intelligence (AI): reasoning strategies based on forwards induction (like normal science) build up from what is currently known; backwards induction starts from long-term goals or questions, and works its way backwards from these questions, to find a chain of reasoning which can answer these questions. It is well-known in AI that forwards induction by itself works very poorly when the questions to be answered are highly complex and not well-structured in advance. For the four questions above, it is clear that backwards induction -- strategically motivated efforts -- will be crucial to any real progress in the future.

According to conventional wisdom, question number one -- concerned with physics -- is the only area where knowledge is truly unified and scientific (albeit incomplete). Only in physics do we have relatively simple and precise underlying principles, which are used to deduce a wide variety of predictions, applied to a very wide variety of experiments and to engineering. Yet even in physics, the phenomenon of mental inertia has become very common, perhaps as a side effect of the very success of the field in political organization, government-based funding and professional image-building. At last year's Radford conference, I presented a new heresy related to the Quantum Field Theory (QFT) foundations of physics[3-6], based on key concepts in self-organization discussed below.

Just in the last few years, neural network researchers have developed new mathematics which, in my view, begins to put question number four on a physics-like scientific basis[7,8]. More precisely, there is new mathematics which can encompass the central issues in building a brainlike intelligent system, in a unified manner, with a clear link to past and future empirical work on the brain. This was all extreme heresy when it was first proposed[7], but it is now well-established through engineering applications[9,10]. New government initiatives have been announced, to support the neuroscience-engineering collaborations required to follow through on the resulting opportunities.

Question number three is the one which calls directly for a theory of self-organization. It calls for a body of mathematical theory (and applications) which explains how life and order evolve, through the operation of simpler underlying laws of nature like those discussed by physicists.

No such unified body of theory exists at present, for the field of self-organization as a whole. Unified, precise concepts certainly do exist, but they do not begin to encompass the range of systems and issues which are necessary to really understanding phenomena such as life. If one were optimistic, one might say that the theory of self-organization today is comparable to the field of AI and neural networks thirty years ago: a heterogenous field, unified by a common question, composed of a few great continents of organized but limited designs, surrounded by lots of islands of suggestive examples and ideas which have yet to be unified into anything like a general theory.

The field of artificial life[11] is full of such islands, and is similar to the field of artificial neural networks thirty years ago; unfortunately, it is difficult to summarize that field, and I will not try to do so here. The fields of thermodynamics and nonlinear system dynamics -- great continents within the field of self-organization -- are far larger and far more unified than those islands; therefore, I will focus mainly on them. It is true that the islands may seem at first to be more relevant to the issue of life -- in all of its complexity -- than the continents now do; however, the continents provide the element of mathematical unity and generality which will be crucial to really understanding the phenomena discussed on the islands. For the truly creative mathematician, the absence of an all-encompassing theory here should not be discouraging; it suggests that question three -- unlike one and four -- is a good place to seek a fundamentally new paradigm -- if, in fact, a unified approach is at all possible here.

Several authors have suggested that the theory of self-organization or the theory of complexity might be the key to understanding intelligence in the brain. Certainly the phenomena of chaos and heat flow, etc., play an important role in the brain, but generic self-organization is not enough to yield intelligence directly. The phenomenon of intelligence requires a very special, very rare kind of underlying system dynamics -- like the dynamics of neurons -- which exists on earth only as the result of billions of years of natural selection. In any event, a generic theory of self-organization or of complexity does not yet exist. Usually, in the neural network field, the term "self-organization" is a loose synonym for "unsupervised learning" or for adaptivity in general.

Question number two above depends heavily on questions number one and number three. Considerable efforts have been made, in recent years, to unify physics and cosmology, in a mathematical way, by developing a unified theory of the Big Bang[12,13]. Observations from astronomy have played a significant role in fine-tuning the Big Bang models, but the assumption that the Big Bang must have happened relies heavily on an assumption about self-organization -- the assumption that the existence of life requires that something like a Big Bang must have happened. This paper will suggest an alternative to that assumption which, I hope, could stimulate more empirical research and a wider variety of mathematical theories.

TYPES OF SYSTEMS OR "RULES"

Self-organization, as defined above, studies how different types of underlying system dynamics -- the microscopic "rules" or "laws" of nature -- result in different patterns of order (like life) at a more macroscopic level. Before I can discuss the link from system rules to the resulting order, I must first discuss the most important types of system rules which have been studied:

1. Lagrangian systems, a type of conservative (energy conserving) system. Our universe itself is a Lagrangian system, according to all current credible theories of physics.

2. Dissipative systems, which are like conservative systems, but leak energy to the outside.

3. Open systems, which input energy (and perhaps matter) from the outside, as well as dissipating it.

4. Time-symmetric systems, whose "rules" look the same whether we look at them forwards or backwards in time.

Lagrangian Systems and Energy Conservation

Virtually every serious model in physics in this century has assumed that the universe is a Lagrangian system. This view achieved its strongest formulation in the field theory of Einstein. Quantum Field Theory (QFT) still assumes that the universe is a Lagrangian system, but there are some qualifications which I will discuss below.

For the sake of the mathematical reader, I will give a worked-out example here; however, I hope that the main ideas will be clear even without following the equations.

In Einstein's view, everything in the universe is made up of force fields -- fields whose intensity varies from point to point across three-dimensional space. Thus the state of the universe at any time is fully specified when we specify the intensity of each force field at each point in space. If there are n force fields, of intensity ϕ_1 through ϕ_n, we may consider these n numbers as a vector $\underline{\phi}$; thus the state of the universe at any time t is specified by specifying $\underline{\phi}(\underline{x},t)$ across all points $\underline{x}$ at time t.

Crudely speaking, when we say that the universe is a Lagrangian system, we are saying that the universe itself acts like a utility maximizer. We are saying that the universe seems to have a kind of utility function, $\mathfrak{L}$, and that it tries to maximize the sum of $\mathfrak{L}$ across all future time. In physics, $\mathfrak{L}$ is called a Lagrangian function.

More precisely, the value of $\mathfrak{L}$ at any point $\underline{x}$ is assumed to be a function of $\underline{\phi}(\underline{x},t)$ and of the derivatives of $\underline{\phi}$ at the point $\underline{x}$ at time t. When the current state of the universe is specified, at the current time t_0, then the universe will choose its later states ($\underline{\phi}(\underline{x},t)$ for $t>t_0$) so as to maximize or minimize the sum of $\mathfrak{L}$ over all future time, across all space.

These ideas by themselves are not enough to specify a theory of how the universe works. We still need to guess how large n is and, more importantly, come up with a theory for what the function $\mathfrak{L}$ is. However, once we specify these two pieces of information, we can deduce everything else mathematically. Furthermore, if the universe is a Lagrangian system, this implies that there is another function of $\underline{\phi}$, called $\mathcal{H}$ -- the mass-energy density, which is conserved over time (i.e., its total value across all space does not change from one time to the next). For some possible choices of $\mathfrak{L}$ -- but not all choices -- the function $\mathcal{H}$ will be positive-definite; this basically means that $\mathcal{H}$ can never be negative at any point in space. (When $\mathcal{H}$ is positive-definite, we can rule out the possibility that $\mathcal{H}$ will grow larger and larger, without limit, in one region of space, as it grows more and more negative in another.)

A common kind of Lagrangian, with n=1, in a one-dimensional space x, would be something like:

$$\mathfrak{L} = \tfrac{1}{2}\left(\frac{\partial\phi}{\partial t}\right)^2 - \tfrac{1}{2}\left(\frac{\partial\phi}{\partial x}\right)^2 + f(\phi) \quad , \tag{1}$$

which physicists would write in a more condensed notation:

$$\mathfrak{L} = \tfrac{1}{2}(\partial_t\phi)^2 - \tfrac{1}{2}(\partial_x\phi)^2 + f(\phi) \tag{2}$$

To work out the dynamic laws of the universe, as implied by any particular Lagrangian, we simply plug that Lagrangian into the Lagrange-Euler equation, a generalized equation which applies to all Lagrangian systems (explained further in [14] and [15]):

$$\textit{For all } i\text{:} \qquad \partial_t\left(\frac{\delta\mathfrak{L}}{\delta(\partial_t\phi_i)}\right) + \partial_x\left(\frac{\delta\mathfrak{L}}{\delta(\partial_x\phi_i)}\right) = \frac{\delta\mathfrak{L}}{\delta\phi_i} \tag{3}$$

The $\delta\mathfrak{L}/\delta...$ derivatives represent the derivatives of the function $\mathfrak{L}$ with respect to the things which appear in the function $\mathfrak{L}$. Thus in equation 2, the derivative of $\mathfrak{L}$ with respect to $\partial_t\phi$ is simply $\partial_t\phi$. When we plug equation 2 into equation 3, we arrive at the following equation, which specifies how ϕ changes over time in the theory specified by equation 2:

$$\partial_t(\partial_t\phi) - \partial_x(\partial_x\phi) = f'(\phi) \tag{4}$$

Equation 4 is a classical wave equation, the kind of equation used to describe radio waves, sound waves, etc.

Likewise, to figure out the formula for the conserved mass-energy, we plug in the $\mathfrak{L}$ of our choice into the Hamiltonian equation (for the case n=1):

$$\mathcal{H} = \dot{\phi}\left(\frac{\delta\mathfrak{L}}{\delta\dot{\phi}}\right) - \mathfrak{L} \quad , \tag{5}$$

which in the example of equation 2 yields:

$$\mathcal{H} = \tfrac{1}{2}\dot{\phi}^2 + \tfrac{1}{2}(\partial_x\phi)^2 - f(\phi) \tag{6}$$

Note that $\mathcal{H}$ is positive definite in this example, if the function f is positive definite. The quantity which is constant or conserved over time is simply the sum, at each time, over all space, of $\mathcal{H}$:

$$\textit{Total Mass-Energy} = H(t) = \int \mathcal{H}(x,t)\, dx \tag{7}$$

Conventional physical theories generally assume that $\mathcal{H}$ is positive definite, but there are notions of "false vacuum"[16] which begin to put that assumption into question; generally, the mathematics is easier when we assume that $\mathcal{H}$ is positive definite, but there is no apriori reason to assume that it must be. On the other hand, there is very good empirical evidence that mass-energy is in fact conserved. I am unaware of any theorem which says that energy-conserving systems must be Lagrangian systems, but with continuous field theories I am not aware of any workable alternatives.

Perhaps the most important true Lagrangian field theory is the "already unified" field theory of John Wheeler[17], which unified Maxwell's Laws and Einstein's general theory of relativity. More recent related concepts have been developed by Penrose[18] and others[19].

In QFT, the search for the true Lagrangian of the universe still continues, but there is more than just a Lagrange-Euler equation involved. In fact, there are three or four different versions of what we should do with the Lagrangian and of what it means[13]. Generally speaking, most of these versions require us to assume the existence of extra spatial dimensions, and additional "quantization" and "regularization" assumptions. At last year's conference, however, I showed[3] that we can replicate all the main features of QFT, starting from a purely Einsteinian Lagrangian theory, without such additional assumptions, so long as one is very careful not to throw in assumptions about time-forwards causality which do not emerge from the Lagrangian theory itself. Most physicists would recognize this claim as extreme heresy; the discussion of time-symmetric systems below will try to explain the basics of how it makes sense nevertheless.

Dissipative Systems

Even if the universe as a whole is a conservative, Lagrangian system, the systems which exist within our universe can input or output energy.

Classically, a dissipative system is just like a conservative system, except that there are additional terms in the dynamic equations which insure that energy will always be dissipated away, out of the system, until the system reaches some kind of minimum energy equilibrium. For example, consider the damped pendulum shown in Figure 1.

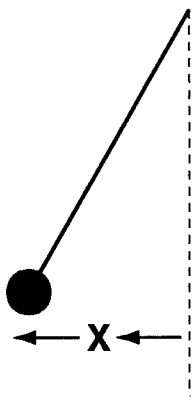

Figure 1. The Damped Pendulum

In the figure, x represents the gap between the pendulum and the center line, and $\dot{x}=v$ is the velocity of the pendulum. So long as x is small, the pendulum is governed by the equation:

$$\ddot{x} = \dot{v} = -ax - bv \tag{8}$$

The term "-bv" represents the effect of friction. Without the friction term, this would be a classical, Lagrangian system. In fact, it would be an example of a "harmonic oscillator," a simple model system used over and over again throughout physics. However, when we add the friction term to this conservative system, we guarantee that velocity and energy will always be dissipated away, until the velocity reaches zero. Friction guarantees that the system will move towards a definite, stable equilibrium point -- the unique point where x=v=0.

In the literature on chaos, the term "dissipative" has acquired a slightly different meanings, which is somewhat less precise[20]. To understand that alternative definition, one must first understand the concept of phase space, which is discussed in the next major section.

Open Systems

Open systems are like dissipative systems, but more general. They may input and output both matter and energy from the outside. As an example, consider the Continuously Stirred Tank Reactor (CSTR), shown in Figure 2.

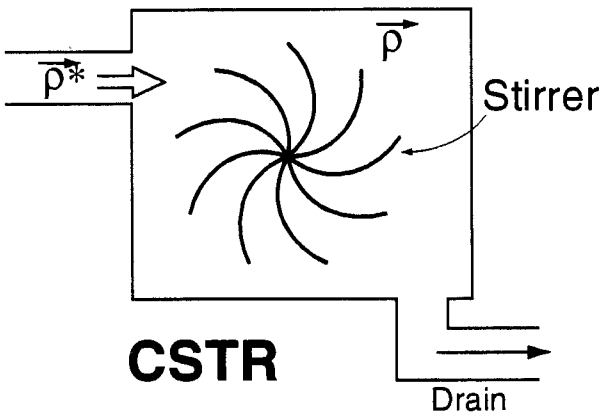

Figure 2. The Continuously Stirred Tank Reactor (CSTR)

The CSTR example plays a fundamental, all-pervasive role in chemical engineering. McAvoy has compared it to the role of the harmonic oscillator in physics. (See chapter 10 of [9] for examples of artificial neural networks used to control both CSTR models and real-world chemical plants.)

To specify the state of a CSTR at any time, we simply need to specify the concentrations of the various chemicals in the tank. Suppose that there are n chemicals, c_1 through c_N. Let ρ_i represent the density or concentration of the chemical c_i. To specify the state of the tank at time t, we need to know the N numbers, $\rho_1(t)$, $\rho_2(t)$, ... , $\rho_N(t)$, which we can think of as a vector, $\underline{\rho}(t)$.

The CSTR receives an input flow of these chemicals from the outside, indicated by the pipe on the upper left of Figure 2. The input stream is usually assumed to be a constant flow, going at a rate c (controlled by some valve). The input stream contains the same chemicals as the tank, and the concentrations of chemicals in the input stream forms another vector, $\underline{\rho}^*(t)$. The volume of fluid going out the drain is usually equal to the inflow rate, but the composition of what goes out the drain is simply the same as what is in the tank ($\underline{\rho}(t)$). The process of stirring insures a uniform concentration of all chemicals throughout the tank, so that we don't have to worry about concentrations varying from point to point in the tank (thereby complicating and confusing the mathematics).

Open systems like Figure 2 can only exist as a subset of our universe. Even then, most thermodynamicists would say that the influx of matter and energy can only continue for a finite time. However, the notion of a CSTR with a sustained influx and perfect stirring is still a mathematically interesting and tractable problem, close enough to real systems that it has enormous practical value.

Years ago, many people became excited by the idea that open systems (like the earth, if the sun would shine

forever) could develop and sustain life forever, even if closed systems could not. They concluded that the study of open systems might be crucial to our understanding of life, even if life itself -- like the sun -- must eventually die out in our universe.

Time-Symmetric Systems

The flow of time is a central issue -- perhaps the central issue -- in the study of self-organization. Several versions of the idea exist in science: time as a coordinate axis, no different from coordinates in space; time as the dimension which defines causality; time as the driving gradient along which life evolves; and time as a ruler of human thought and action. A clearer understanding of self-organization -- and even of physics itself [3-6] -- requires a deeper understanding of the relations between these various aspects of time.

A good place to start is by understanding the concept of time-symmetric systems.

In physics, a system is called time-symmetric (or T-symmetric) if the rules which govern that system look exactly the same when viewed in the forwards time direction or in the backwards time direction. In other words, if we captured the evolution of the system forwards in time on a movie film, and then played the movie backwards, the backwards movie should seem to be governed by exactly the same rules; there would be no difference between forwards or backwards.

In mathematical terms, we can test for time-symmetry by plugging in "-t" instead of "t" wherever "t" appears (explicitly or implicitly) in the system equations, and looking to see if the new version of the equations is the same as the old version. For example, consider the simple wave equation back in equation 4:

$$\partial_t(\partial_t\phi) = \partial_x(\partial_x\phi) + f'(\phi) \tag{9}$$

When we replace "t" by "-t," the only term which might be affected is the one on the left, which does contain "t." Indeed:

$$\partial_{-t} = \frac{\partial}{\partial(-t)} = -\frac{\partial}{\partial t}\ , \tag{10}$$

which would reverse the sign of our "∂_t" operator. But there are two ∂_t operators on the left in equation 9; multiplying the leftmost term by two minus signs, we end up with the term unchanged. Thus equation 9 is totally and perfectly T-symmetric.

On the other hand, consider equation 8, the equation for a damped pendulum, which we may write as:

$$\partial_t(\partial_t x) = -ax - b(\partial_t x) \tag{11}$$

The term on the far right -- the friction term -- contains only one "∂_t"; thus when time is reversed, we end up with the equation:

$$\partial_t(\partial_t x) = -ax + b(\partial_t x) \tag{12}$$

Equation 12 is very different in its behavior from equation 11. This is what we should expect; if we take a movie of a pendulum slowly winding down to a stop, we know that we would see something quite different (physically impossible, in everyday terms!) when we run the movie backwards.

Based on these two examples, one might guess that Lagrangian systems are time-symmetric while dissipative systems are not. But the first part of this guess would be very wrong. It is easy enough to postulate a Lagrangian which is asymmetric with respect to time, by including terms based on first-order time derivatives. High energy physicists use such terms routinely, in describing fields called fermion fields or spinor fields[14,19,20]. Today's grand unified theories of physics (including the superstring theories) generally are T-symmetric; however, physicists have known for decades that there is a class of nuclear reactions -- the "superweak interactions"[21] -- which do not fit the T-symmetric models, and suggest a likelihood of T-symmetry violations. (More precisely, the existing forms of QFT -- which assume CPT asymmetry, as I will discuss below -- requires that T symmetry must be violated somewhere, in order to explain the experiments already done.) The superweak effects appear extremely small at

present, but they may be the tip of a huge iceberg. In system dynamics, it is well-known that small feedback terms -- which appear inconsequential over small time intervals -- may exert large and decisive cumulative effects on the global state of a large system; the superweak effects, tiny though they are, might be a reflection of such small but decisive terms. (After the presentation of this paper at Radford, one participant stated that Weinberg of Harvard -- a major developer of unified theories -- may be developing a new model to cover superweak interactions; however, no information was available yet on the details. In any event, more empirical information would be essential to really understanding these interactions.) Even the existing theory predicts symmetry violations 100 times greater, in experiments now being set up for B mesons, than in the experiments reported so far.

High energy physics generally does require "CPT symmetry." To test for CPT symmetry, we switch t with -t, and we also switch **x** with -**x** and reverse the charges of all particles/fields. There are no known violations of CPT symmetry in any physical experiment at this time. Some physicists have assumed that CPT symmetry must be true, apriori, because they don't know how to make the mathematics of QFT work without absolute and perfect CPT symmetry. (Likewise, the whole elaborate apparatus of superstring theory rests entirely on trying to solve other problems in the mathematics -- without any empirical support whatsoever for any of the complexity.) Einsteinian field theory does not require CPT symmetry; therefore, the mathematics in [3-6] should make it possible to relax the assumption of CPT symmetry, if there were ever any empirical reasons to do. Still, even in my new approach, CPT symmetry currently appears to be a very natural and compelling assumption.

Many researchers believe that the central problem in self-organization is to explain why time seems to run forwards, in such an absolute asymmetric way, while at the same time being so absolutely symmetric at the microscopic level. Most people explain this by assuming that macroscopic causality (and life) is a temporary aberration, due to unusual starting conditions in our universe -- the Big Bang. The later part of this paper will suggest some alternative possibilities, related in part to recent work by Prigogine[23] as summarized by Prigogine at this conference.

Time-symmetry is also the key to my new reformulation of QFT. Because QFT describes the underlying, microscopic dynamics of our universe, and does not yet account for superweak interactions, there is no need or justification for assuming anything but T-symmetry as yet at that level.

In past decades, dozens of serious scientists -- including Einstein, Wiener, Von Neumann and DeBroglie -- made strenuous efforts to "explain" QFT as the statistical outcome of a deeper, underlying theory -- a theory of the universe as an Einsteinian Lagrangian system. Those efforts failed, and cast doubt on the whole idea. (However, some of DeBroglie's ideas may yet prove useful in the future.) Von Neumann even proved some theorems suggesting that such an explanation might be impossible. Actually, Von Neumann only proved that it would be impossible to replicate a range of predictions from QFT which had not been tested experimentally. But Bell, Shimony and others designed an actual experiment -- performed several times since 1974 or so [13,24] -- which could never be reconciled with what they called "local causal hidden variable theories."

Bell's theorem actually provides the key needed to overcome these difficulties. The key to Bell's Theorem was the assumption of time-forwards causality. DeBeauregard[13] has shown that QFT itself can only explain Bell's Theorem because of the way it assumes time-symmetry in the flow of causality in the experimental system. In retrospect, this whole episode in physics seems amazingly clumsy. Top scientists, in trying to compute the statistical implications of time-symmetric Lagrangians and dynamics, have generally added an assumption of time-forwards causality -- taken from personal intuition rather than the Lagrangians -- in order to simplify their calculations. The apparent failure of Einsteinian theories was not due to the theories themselves -- theories which in no way implied time-forwards causality -- but to an extraneous assumption which was thrown in when calculating the supposed predictions of these theories. The essence of my reformulation is simply to account for this problem, by going back to calculate the statistics in a more consistent manner. This is still far from trivial, but it does replicate all the main features of QFT, including renormalization, Fock space effects, interference, and so on, to an uncanny degree. (For more details and other issues, see [3-6].) This is not to say that I have constructed a grand unified theory on this basis; rather, I have provided a mathematical starting point which future researchers may use to develop such a theory. This, in turn, still leaves open the puzzle of causality at the macroscopic level.

Prigogine's new paper[23] suggests an alternative formalism for calculating what QFT calculates, which also starts from a purely Einsteinian Lagrangian basis, and provides another way to eliminate the need for apriori quantization assumptions and the like. It is conceivable that the two formalisms, though quite different, are mathematically equivalent, in much the same way that the Schrodinger and Heisenberg formalisms are now known to be equivalent; however, this has yet to be proven.

CLASSICAL CONCEPTS OF SELF-ORGANIZATION

Starting from the wide array of possible systems, discussed in the previous section, how can we deduce how these systems will actually behave at the macroscopic level? What kinds of life or order will they give rise to?

Prior to the mid-1970's, most scientists focused their attention on two kinds of behavior which can emerge, in equilibrium, in a dynamical system: simple, static, stable equilibrium, and states of total disorder or randomness. When behavior of this sort was proven to emerge in certain important systems, many scientists extrapolated this result, and made statements which appeared to suggest that all possible systems - including our own universe, whose dynamic laws are still unknown -- must culminate in a lifeless state, either totally static (like a Big Crunch) or totally random (like a Heat Death). At this conference, Prigogine provided some very strong warnings about such extrapolation or speculation.

This section will discuss some of the key concepts which emerged from this early research, and some of its extensions (and extrapolations). First it will illustrate the basic concepts related to stable equilibrium -- concepts are still the mainstay of control engineering today. Next it will describe the basic mathematics of particle "scattering" (i.e., collisions between particles) and statistics based on scattering. Finally, it will discuss a few extensions and extrapolations.

Static Stable Equilibrium and Related Concepts

The pendulum with friction, shown in Figure 1 above, is a classic example of a system which moves towards a simple, static equilibrium. No matter what state the pendulum start out in, it will move towards the state where x=v=0. It will never reach that state exactly, but as time goes on, it will move closer and closer.

The pendulum is a very simple kind of system, which can be described by only two variables, x and v, which vary continuously with time. Systems like this can be totally described by a phase plane picture, like the one shown in Figure 3.

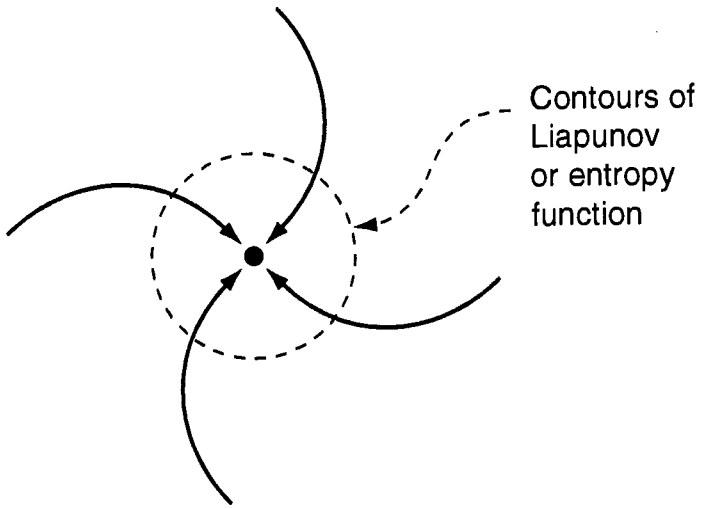

Figure 3. Phase Plane Picture For a Simple Static Equilibrium

In Figure 3, each point in the plane represents a possible state of the system; the two coordinates of the point are simply the values of x and v in that state. The solid curves with arrowheads represent how the system evolves over time, starting from any point on any of the curves. In this example, all arrows lead inwards towards a single point -- the static, stable equilibrium point.

This same notion of a point equilibrium can be applied to more complex systems as well. The phase-plane picture of a system defined by n variables (for n>2) cannot fit on a two-dimensional piece of paper, but we certainly

can think about what it looks like. Still, the difficulty of drawing or imagining phase-plane pictures in more than two dimensions was one of the reasons why it took so many years before people developed more advanced concepts in system dynamics. (Difficulties in visualizing time-symmetric causality have also been a problem.)

Even today, the vast bulk of research in control engineering (even with neural networks!) focuses on the effort to achieve a stable point equilibrium, of this classical variety. Yet several researchers [20,25] have shown that other approaches to control are possible, based on ideas like stable regions or points revisited on a regular schedule. Advanced neural net control designs, based on optimization methods[7,9], should be able to learn such unconventional control strategies, when such strategies can improve engineering performance over time.

Figure 3 illustrates another key concept from classical theory: the concept of a Liapunov function. Even for a simple pendulum, the Liapunov function is a function of two variables, x and v; therefore, the only way to show it on paper is by drawing its contour lines, like the dashed circle surrounding the stable equilibrium. You can think of Figure 3 as a kind of map, where each contour line (only one shown) indicate how "high" each point is, in terms of the Liapunov function.

But what is a Liapunov function? A Liapunov function may be defined as any function such that the solid curves always point downhill, except at the stable point, which is the lowest point on the map. (The lines need not point directly downhill; they need only point in some direction which goes down, which reduces the level of the Liapunov function.) Once we know that a system has this kind of Liapunov function, we know that it has a stable point equilibrium.

The usual entropy functions of thermodynamics may be viewed as a special case of Liapunov functions (with their sign reversed): the entropy functions always increase, until the system reaches equilibrium.
Energy, by contrast, is usually not a Liapunov function, except in special cases, like our simplified model of a pendulum with friction.

Even when systems do have a stable point equilibrium, it can be difficult to find a Liapunov function, to prove stability. Even today, most people rely mainly on cleverness and guessing here. (It reminds me of how people solved algebraic equations prior to Newton's method!) Certain neural network designs -- the adaptive critic methods [7,9] -- offer the hope of adapting or numerically locating Liapunov functions, because they combine to a mathematical function (the Jacobian function J) which is known to be a Liapunov function for a very broad class of systems; however, a great deal of work would be required to turn this into a working computer-based tool.
Some mathematical details of this possibility are discussed in the technical Appendix of this paper

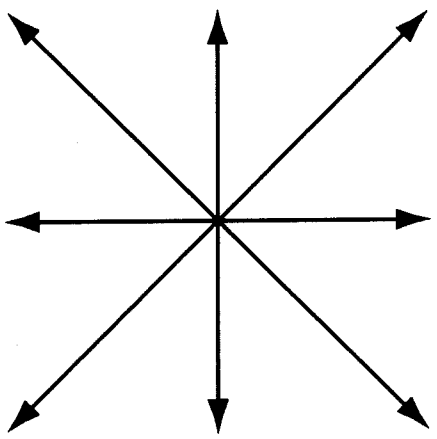

Figure 4. Phase Plane Picture of an Unstable System

Figure 4 illustrates the classical opposite to Figure 3: an unstable, divergent system which simply "blows up."

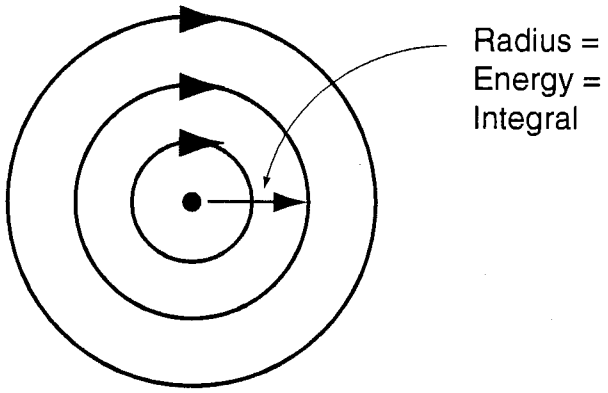

Figure 5. Phase Plane Picture of a Conservative System

Figure 5 illustrates a third classical picture: the conservative system (like the pendulum without friction). Energy-conserving systems cannot approach a universal, stable point equilibrium, because states of one energy level cannot evolve into a state at a different energy level. States of higher energy can dissipate their extra energy, as in the pendulum, in a dissipative system only. (If all possible states had the same "energy," it would be meaningless to say that energy is conserved.) Therefore, when energy or any other "integral" (conserved quantity) is conserved, the regions of phase space representing different energy levels are totally disconnected from each other. There could be stable points or oscillations within a given energy region, but no curves moving towards that point from states of higher or lower energy. In Figure 5, the circles each represent different energy levels; for example, they could represent the swings of a frictionless pendulum, starting off at different distances x from the center.

Scattering: the Classical Model

The classical scattering model begins from ideas about collisions between particles in free space, but it ends up being a fundamental tool in modern chemistry.

In classical, pre-Einsteinian physics, people often thought of the universe as a collection of particles -- simple particles or atoms or molecules -- moving around in free space. In this view, the laws of physics boiled down to predicting what will happen when different types of particle collide with each other, with what probability. Curiously enough, modern QFT has largely returned to this view[13,26]. Modern QFT focuses most of its attention on calculating scattering probabilities (or cross-sections), for different types of particles. These calculations are very complex, because they involve complex field effects in the zone of interaction between particles, but the final result is simply a set of scattering probabilities which are all we observe at the macroscopic level. (Several aspects of QFT which go beyond this will be discussed below.)

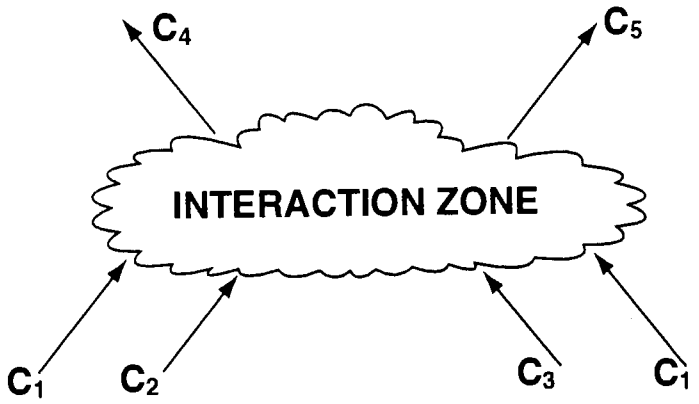

Figure 6. The Concept of Scattering

The notion of scattering or particle collision is illustrated by the example in Figure 6. From the underlying laws of physics or chemistry, we may know that 2 particles of chemical number 1 (c_1), one particle of c_2 and one particle of c_3 may collide and recombine to form particles c_4 and c_5. When the four particles come together, we may know that there is a certain probability of this happening. (Other types of collision may also be possible -- such as c_1 and c_2 colliding alone -- but let us just consider one type of collision at a time, for now.) Our problem, in thermodynamics, is to figure out what kinds of results these microscopic collisions will lead to, for the universe as a whole.

As a simple approximation, we may assume that these particles are distributed all across space, in a uniform, random sort of way, such that we don't have to worry about what happens in any specific location. We simply want to know how many of these particles there will be, for each type of particle, per volume of space. More precisely, we want to know, for each possible particle type, c_i, what will be its density, $\rho_i(t)$, measured in terms of particles per unit of space, at each time t, starting from a known initial density.

To figure this out, look back at Figure 6, and use some intuition. If the density of ρ_1 and ρ_2 is held constant, but the density of ρ_3 is cut in half, then the rate of four-way collisions will be cut in half. Likewise, cutting ρ_2 in half would also cut the frequency in half. Cutting ρ_1 in half would cut the frequency by a factor of 4, because two c_1's are needed for each collision. Putting this together, we would expect that the frequency of collisions (the rate of reaction) would simply be proportional to:

$$\rho_1 * \rho_2 * \rho_3 * \rho_1 = \rho_1^2 \rho_2 \rho_3 \tag{13}$$

Now if the underlying laws of the universe are T-symmetric -- as discussed at length above -- we must assume that this kind of collision could happen just as easily in reverse time, whenever a c_4 particle collides with a c_5 particle in free space. The rate of reverse collisions would be proportional to $\rho_4\rho_5$, and there must be the same likelihood of a backwards reaction as a forwards reaction, allowing for what the densities are. (As a practical matter, in chemistry, elementary reactions are symmetric only with respect to an equilibrium mixture, which will be explained in the discussion of equation 24 below.)

Translating this example into formal language, we would say that the chemicals c_1 through c_5 obey the following reaction equation:

$$2c_1 + c_2 * c_3 \leftrightarrow c_4 + c_5 \ , \tag{14}$$

where the double-headed arrow represents a two-way reaction. We would say that ρ_1, for example, obeys the dynamical equations:

$$\dot{\rho}_1 = -2R \tag{15}$$

$$R = k(\rho_1^2 \rho_2 \rho_3 - \rho_4 \rho_5) \ , \tag{16}$$

where R is the rate of the reaction, at the current time t, and where k is a rate constant. Equation 15 basically says that the reaction uses up two particles of c_1 per collision, and equation 16 is based on equation 13 and our discussion of the reverse collisions. It is interesting that equations 15 and 16 are not themselves time-symmetric, even though they describe statistics of a time-symmetric reaction. (One might hope, at first, that equation 15 -- the conventional kind of equation used in chemistry -- explains very easily how time-forwards statistics can emerge spontaneously in a system whose underlying physics is time-symmetric; however, this is not the case. Our derivation of equations 15 and 16 implicitly assumed that: (1) the densities ρ may start out, at some initial time t, at values different from their long-term equilibrium values; (2) time-forwards causality applies forwards from that time t. The next section of this paper will discus this general issue in more depth.)

More generally, we may consider reaction equations:

$$m_1 c_1 + \ldots + m_N c_N \leftrightharpoons n_1 c_1 + \ldots + n_N c_N \ , \tag{17}$$

where there are N particle types in existence in the universe, and where m_i will typically be zero for most particle types in any particular reaction. (In other words, most reactions involve only a handful of the particle types.) For each such reaction, the rate of reaction will equal the forward collision rate minus the backwards rate, such that:

$$R = k\left(\prod_{i=1}^{N} \rho_i^{m_i} - \prod_{i=1}^{N} \rho_i^{n_i}\right) \ , \tag{18}$$

where the large pi refers to a product. Equation 18 is simply the obvious generalization of equation 16. The net change in each ρ_i based on this reaction will simply be:

$$\dot{\rho}_i = (n_i - m_i)R \tag{19}$$

For a realistic system of particles or chemicals, we must assume that several reactions, like equation 17, are taking place in parallel. In that case, the total value of $\dot{\rho}_i$ will simply equal the sum of $(n_i - m_i)R$, for each reaction, summed over all the reaction types. Systems of this sort are easy to simulate on a personal computer, even in a spreadsheet.

Is it possible for such a simple, approximate model of particle statistics to yield interesting dynamic effects? For example, if we assumed a universe with very strange sorts of particles, such as negative energy particles, could one generate strange reaction equations leading to interesting dynamics here? Would the existence of negative energy states blow up the universe (as some have feared)?

What if we assumed some variation across space, such as two reaction chambers connected by some kind of diffusion of particles? Such a system could be represented by creating two variables for each chemical c_i, ρ_i^A and ρ_i^B , the densities in the two reaction chambers A and B; we can assume the usual reactions within each reaction chamber, and simply add N simple reaction equations to represent the diffusion process:

$$\rho_i^A \leftrightharpoons \rho_i^B \qquad \text{for } i=1,N \tag{20}$$

In general, none of these interesting complications changes the basic picture. Simple reaction systems like those above always have a stable point equilibrium. This can be seen by considering the simple Liapunov function:

$$L = \sum_i (\rho_i \log \rho_i - \rho_i) \tag{21}$$

It is a straightforward exercise in calculus to prove that this is in fact a Liapunov function for all these reaction systems; one need only calculate $\dot{L}$, plugging in equations 18 and 19 into the result, to verify that $\dot{L}$ is always negative, except at equilibrium points, where it is zero. (Recall that ρ_i can never be negative in these systems.) It is also a straightforward exercise to see that L has a global minimum when ρ_i equals 1, for any chemical c_i. Note that this global minimum does not even depend on the details of the reaction equations! This model provides very powerful support for the view that realistic universes do in fact approach a stable equilibrium, in a very strong way.

In order to approach reality, of course, we must account for some additional complexities here -- though they do not change the basic picture. The two main complications of importance to the classical picture are: (1) the role of mass-energy; (2) the existence of multiple energy levels for each particle type. Readers with less interest in the mathematical details can move on to the next subsection, without losing the basic idea.

Mass-energy does not appear explicitly in equation 17, but it is represented implicitly. If the mass-energy of each particle, c_i, is E_i, then the conservation of mass-energy effectively requires that total mass-energy be unchanged by the reaction:

$$\sum_{i=1}^{N} m_i E_i = \sum_{i=1}^{N} n_i E_i \quad , \tag{22}$$

for every reaction in the system. In vector notation, this may be expressed equivalently as:

$$\boldsymbol{E} \cdot (\boldsymbol{n} - \boldsymbol{m}) = 0 \tag{23}$$

In fact, any vector $\underline{E}$ which obeys equation 23 for all reactions will automatically be a conserved integral of this system. When a system of reactions does obey one or more conservation laws, then the system will not evolve towards the global minimum of the Liapunov function unless, by chance, the system starts out at the same energy level as that global minimum belongs to. The system will always evolve towards the unique, stable equilibrium within the set of points that have the same energy etc. If there is a set of integrals, $\underline{E}^{(1)}$ through $\underline{E}^{(M)}$, then it is straightforward to prove that the equilibrium points (points where $\dot{\underline{\rho}}=0$) are defined by:

$$\rho_i = \exp(-\theta_1 E_i^{(1)}) * \exp(-\theta_2 E_i^{(2)}) * \ldots * \exp(-\theta_M E_i^{(M)}) \quad , \tag{24}$$

where the constants θ_1 through θ_M can be any real numbers (positive or negative), and can be thought of as something like the "temperature" of the particular state. (The θ's must be the same, of course, for all the different chemicals in any one particular state.) For every possible state of the system, $\underline{\rho}$, whether an equilibrium state or not, there will always exist a "temperature level" ($\underline{\theta}$) which matches the energy levels of that state.

The discussion above assumed a finite number of particle types. In actuality, in physics, scattering probabilities are calculated as a function of particle type, velocity and spin. If we think of each c_i as representing a particular combination of particle type, velocity and spin, then the equations above still make sense, but they assume an infinite number of combinations c_i, because the speed of a particle can be anything between zero and the speed of light. This, in turn, limits the possible values of the temperature parameters θ in equation 24. For example, if a system starts out with a finite energy density at time 0, and if the energy values vary from 0 to $+\infty$, then it is impossible for the temperature to be negative, because a negative temperature would represent an infinite energy density. Systems like this still approach a strong equilibrium, despite the infinite number of combinations c_i, in essentially finite time. However, if we allow energy to vary between $-\infty$ and $+\infty$, we could generate an unstable kind of system in which higher energy levels become more and more populated as time goes on, even starting from a very simple initial state. Most physicists would consider this kind of situation highly nonphysical and implausible, but one never can be certain apriori. (Such situations could cause a breakdown in the assumption of spatial uniformity.)

As a practical matter, when there are infinite possible energy levels for each kind of particle, it is usually more convenient to write the reaction equations in terms of simple particle types, and to account for the energy-level effects indirectly in the dynamic equations. In chemistry, for example, one typically adds entropy terms to equations like equation 19. Because of such effects, many chemical reactions in the real world tend to go in one direction, at ordinary temperatures and pressures; however, this does not violate the underlying principle of time-symmetry, when all the various flows are accounted for.

Extensions and Extrapolations of the Classical Ideas

The scattering model described above is central to classical thermodynamics, but it is not, of course, the whole thing.

Most classical textbooks actually stress the work of Carnot, who proved very strong and rigorous limits to the efficiency of heat engines -- mechanical systems based on expansion and compression and chemical reactions like combustion. Many, many people confuse the narrow but rigorous work of Carnot with the more general Second Law of thermodynamics. For example, with small heat engines like automobile engines, Carnot's laws (applied to realistic operating temperatures for such engines) yield a limit of about 38% efficiency, even under optimal operating conditions (which are rare in actual driving). A few years ago, I saw a satire in one of the major energy newsletter, in which Senator Kennedy was portrayed as trying to pass an amendment to Carnot's Laws, to permit higher efficiency in automobiles. In actuality, small fuel cells -- an alternative way to get energy out of hydrocarbons -- have shown to get efficiency close to 100%, far more than Carnot's limits permit, because they are not heat engines. In testimony before Marilyn Lloyd's committee in the House, in the summer of 1993, Philip Haley of General Motors testified that 60% efficiency has been achieved in a certain class of fuel cells, which does appear suitable for automotive use, and is also capable of high efficiency under a wider variety of operating conditions, at 80-100 degrees Celsius. (Designers of coal-fired powerplants usually understand the difference between Carnot's Laws and the Second Law; however, even some world-class automotive experts confuse "T delta S" energy losses -- related to the Second Law -- with the much larger heat losses explained by Carnot. The need for high temperatures even to reach 38% efficiency in heat engines leads to serious problems with NOx emissions -- due to the combustion, in effect, of nitrogen in the air -- which is a crucial form of air pollution.)

The Second Law in its general form has been a source of great confusion, as noted by Prigogine at this conference. Many people have drawn the conclusions that: (1) all possible systems -- including our universe (whose dynamic laws are still unknown!) -- will converge to a maximum of entropy; (2) entropy is a measure of disorder, which can be maximized only when all life or order ceases. It is obvious that these conclusions go too far, because it is easy enough to draw up hypothetical, model systems which do not converge to a simple state of disorder. However, I will try to give at least some of the flavor of what has led to such sweeping extrapolations.

After the great success of Carnot and of the scattering model, it was natural for mathematicians to try to generalize their results to all possible systems. Among the results of this effort were new theorems with the following sort of flavor: Suppose that we try to be more rigorous, and study the dynamics of probability distributions for states of the universe, instead of just distributions for individual particle types. Suppose that we happen to live in a universe which converges to a unique probability distribution $P_0(\{\underline{\Phi}(\underline{x})\})$. (In the formal literature, P_0 is usually presented as an "invariant measure.") The function P_0 indicates the equilibrium probability for each possible state of the universe, $\{\underline{\Phi}(\underline{x})\}$; I have inserted curly brackets { } around $\underline{\Phi}(\underline{x})$, to remind you that a state of the universe is defined only when we know the set of values of $\underline{\Phi}$ across all points in space. After we know the function P_0, we can define the following entropy function:

$$S = \sum_{\Phi} \left(\log Pr(\{\Phi(x)\},t)) P_0(\{\Phi(x)\}) \right) , \qquad (25)$$

where the sum is taken over all possible states of the universe, and where Pr represents the current probability distribution for possible states of the universe, at the current time t. Under certain assumptions, it can be proven that S is a Liapunov function (with sign reversed); in other words, it can be proven that S will increase until it reaches its maximum. This is just a fancy way of saying that the actual probability distribution will gradually approach the equilibrium probability distribution.

Theorems of this sort are essentially just tautologies. They assume the existence of an equilibrium probability P_0. The theorems become more powerful if we add the additional assumption that, in equilibrium, the values of the field at any point in space will be statistically independent of the values at any other point. (When we say that two numbers are statistically independent of each other, this is like saying that they are not correlated with each other in any way.) This assumption tells us that:

$$\begin{aligned} P_0(\{\Phi(x)\}) &= P_0(\Phi(x_1),\Phi(x_2), \ldots ,\Phi(x_i),\ldots) \qquad \textit{(all points } x\textit{)} \\ &= P_0(\Phi(x_1))P_0(\Phi(x_2))\ldots P_0(\Phi(x_i))\ldots \end{aligned} \qquad (26)$$

If we take the logarithm of both sides of this equation, we can see that it lets us write the entropy, in effect, as something like:

$$\log Pr(\{\phi(x)\}) = \sum_x \log Pr(\phi(x)) \quad , \tag{27}$$

the sum of entropy across all points in space. This is called a local entropy function. If we assume statistical independence, then, we can deduce a local entropy function, which in turn tells us that the equilibrium state will be "disorderly" in the sense that there will be no order or correlation connecting different points in space. But once again, this is only a tautology! There is absolutely no guarantee that entropy functions, as defined in equation 25, must be local in the general case.

As a practical matter, many physicists have proven more restrictive theorems about entropy, of greater substantive importance than these general theorems. However, many efforts were also made in the 1970's to develop local entropy functions for open systems. Based in part on the hope that such efforts could be or had been completely successful, many scientists concluded that interesting patterns of order -- such as chemical oscillations -- would be impossible even in open systems[27]. This was a very curious hope, since it would appear to rule out the possibility of life on earth even in past history, let alone the long-term future. This is one more example of the need for all scientists -- even the very best scientists -- to work somewhat harder than they normally do to escape from the confines of "groupthink"[7,28] -- a pervasive problem which is not unique to science.

MODERN CONCEPTS OF SELF-ORGANIZATION

Since the late 1970's, the field of self-organization has gone through something of a revolution. A few key elements of the revolution have been:

1. Prigogine's theory of "dissipative structures" -- a rigorous theory which allowed for the possibility of chemical oscillations.

2. A vast expansion in the study of chemical oscillations, both empirical and theoretical.

3. The modern theory of nonlinear systems dynamics, including "chaos theory" and the like.

4. A greater appreciation of spontaneous symmetry breaking (SSB), which leads to forms of order or orientation in systems which appear, at first, to preclude such orientation.

5. A deeper understanding of solitons or solitary waves, ordered states which can emerge even in classical, Lagrangian systems.

Prigogine's theory was undoubtedly motivated in part by empirical work -- work starting from Belousev and Zhabotinsky in Russia -- actually demonstrating chemical oscillations. Prigogine's theory won him both the Nobel Prize, and a general perception that he is the leading thermodynamicist alive today. For a reasonable, popular account of Prigogine's theory and related ideas, see [29].

This section will discuss points 2 through 5 above, in that order, with comments both on the Big Bang and on Prigogine's most recent work, particularly in the section on SSB.

Chemical Oscillations

Work on chemical oscillations began with physical experiments showing how simple, closed systems -- starting out in a bland state far away from equilibrium -- can demonstrate beautiful patterns of oscillation and order on the way to their eventual equilibrium. Later on, open systems were developed, based on CSTRs (defined above), which demonstrated sustained oscillations. Some of these systems demonstrated beautiful color patterns, and were so easy to set up that they were suitable for high school chemistry labs. Several papers on these topics -- complete with big,

impressive color photographs -- appeared in Scientific American in the 1970's and 1980's. The early oscillators were essentially discovered by accident, but Epstein et al[27] later developed a more systematic procedure for finding or designing such systems. Later research has found all kinds of fascinating phenomena in these systems, such as complex spatial patterns [30] and chaos[31].

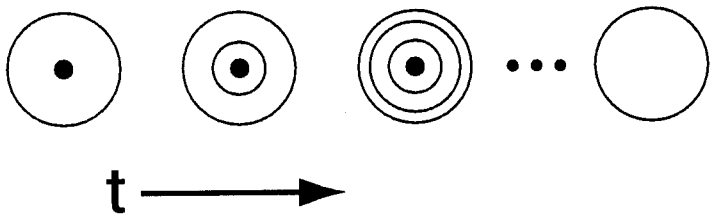

Figure 7. Schematic Picture of the Belousev-Zhabotinsky (BZ) Reaction

Figure 7 illustrates the classical BZ reaction. It shows a round bowl, viewed from overhead, at different times t. Before the reaction starts, one pours and mixes a certain set of chemicals into the bowl. Except for the choice of chemicals, there is no special order or pattern. Then, the reaction starts to take place, typically around a nucleus (like the dark circle in the middle of the leftmost picture). Waves of reaction spread out from the nucleus, as in Figure 7. The waves are red and blue, and usually form much more interesting patterns than this simple black-and-white drawing can show. Finally, after a time, the reaction goes to completion, and the contents of the bowl settle down into a static, pattern-free equilibrium (as in the far right).

The BZ reaction is today's paradigm for the nature of life and order in our universe. The Big Bang provides the initial reaction chamber and chemicals; life evolves as a set of patterns in the intermediate stage; and, finally, everything settles down into eternal death. Epstein, among others, has pointed out that chemical oscillations and cycles are truly fundamental to the functioning of the cells of living systems.

As a practical matter, even if we accept this paradigm, it is a major challenge to develop the details and the mathematics to the point where they describe life in our universe in a more detailed, less metaphorical way. Life on earth is not like a BZ reaction vessel, because it depends primarily on a relatively steady influx of energy from the sun. Thus the CSTR paradigm is a more relevant description, in practice, of what has happened here for the past few billion years. Also, the mathematics of the CSTR is far more tractable, in a sense, because the patterns and the order can exist as long-term phenomena.

In order to make this paper more self-contained, I was hoping to include a simple, simulated example, implemented as a spreadsheet on a personal computer, so that almost anyone could replicate it and study it more carefully. My real goal (which has been achieved) was to show that stable oscillations could be achieved in a simple energy-conserving, closed system based on reactions which are not all time-symmetric; however, in order to achieve this goal, I planned to start out from the "easier" task of generating oscillations in a truly classical system, based on equations 18 and 19, with additional terms to represent an influx of new material, as in the classic CSTR model. More precisely, I tried to develop a hypothetical system based on three chemical types (N=3) and two symmetric reactions, such that each chemical c_i would be governed by the equation:

$$\dot{\rho}_i = (n_i^{(1)} - m_i^{(1)})R_1 + (n_i^{(2)} - m_i^{(2)})R_2 + c(\rho_i^* - \rho_i) \quad , \tag{28}$$

where $n_i^{(1)}$ and $m_i^{(1)}$ are the coefficients n_i and m_i for the first chemical reaction, where $n_i^{(2)}$ and $m_i^{(2)}$ are the coefficients for the second reaction, where the reaction rates R_1 and R_2 are each calculated using equation 18, where $c\rho_i^*$ is the influx of chemical c_i coming in from the input pipe, and where $c\rho_i$ is the flow of chemical c_i going out down the drain pipe. To locate suitable values for the parameters n_i, m_i, k, c and $\underline{\rho}^*$, I planned to use Epstein's systematic design procedure.

This plan did not work. In fact, it is very difficult -- perhaps even impossible -- to generate oscillations in a very simple, totally symmetric system, even with the addition of a steady influx term. The oscillations observed empirically in chemical systems are consistent with modern physics, which is time-symmetric at a fundamental level, for the ordinary forces which dominate thse reactions; however, these empirical oscillations are far more complex, underneath, than equations 18 and 28. For example, Gyorgyi et al [32] have proposed an 80-reaction mechanism to describe what is really going on in the BZ reaction. The classic model of Boissonade and DeKepper can be expressed as a simple two-equation system[33], which has often been used in simulations; however, its relation to equations 18 and 28 is far from obvious, at least to the mathematician, even when the assumption of symmetry is relaxed. The more complex Oregonator system[34] can be reduced to a three-equation system, but it still involves 5 one-way reactions in addition to nontrivial influx and outflux assumptions. In more complex systems, phenomena like oscillation and chaos are far easier to find (even if they were misinterpreted as mere instability in the premodern literature) [35].

In the end, I was able to simulate oscillations on a spreadsheet -- surprisingly violent oscillations -- by simulating the one-way reactions as follows (based on equation 18 but without the backwards reaction term):

$$2\rho_p + \rho_h \rightarrow 3\rho_p \tag{29}$$

$$\rho_h + \rho_f \rightarrow 2\rho_h \tag{30}$$

$$\rho_c \rightarrow \rho_f \ , \tag{31}$$

with rate parameters (k) of 1, .45, and .045, respectively, and with initial conditions ρ_f=0.5, ρ_h=0.3 and ρ_p=0.2. The letters f, h and p stand for flora, herbivore and predator -- reflecting the inspiration provided by classical population biology [35]. Whatever the limitations of this example, the reader should be able to simulate it very easily. (In fact, my children enjoyed playing with this simulation on the spreadsheet.) It is easy to see that this system conserves a very simple measure of "mass-energy":

$$E = \rho_f + \rho_h + \rho_p \tag{32}$$

Similar examples have appeared in the past, but I am not personally aware of any with precisely the same characteristics; on the other hand, I am not a chemist.

The remainder of this subsection will give the technical details of my experience here. The nonmathematical reader could skip to the next section without loss.

The first step in Epstein's procedure is to locate chemical systems capable of bistability -- systems capable of two stable equilibria, for certain values of the flow and rate parameters. Epstein also asks that we focus on autocatalytic systems -- systems, for example, where some of the reactions have $n_i > m_i > 0$ for some chemical c_i. To implement this in the simplest way possible, I began by looking at a CSTR system with only two chemicals and one reaction:

$$3c_1 + 4c_2 \leftrightarrow 4c_1 + 3c_2 \tag{33}$$

To keep the system at one degree of freedom, without any loss of generality in describing the long-term dynamics, I restricted the CSTR influx and the initial values to:

$$E = \rho_1 + \rho_2 = \rho_1^* + \rho_2^* = 1 \tag{34}$$

To find the flow parameters which produce bistability, I simply set up a graph in Quattro with the curves on it: (1) the graph of ρ_1 production from the chemical reaction (equation 33) as a function of ρ_1, recalling the ρ_2=1-ρ_1; (2) the graph of net ρ_1 consumed as a function of ρ_1, based on the influx and outflux. Since the latter was just a straight line, it was easy enough to adjust its slope (c) and its intercept (ρ_1^*) to make it cross the other curve in three places. Still, equation 33 is disturbingly complex, and the resulting bistability too brittle to fit well with Epstein's later steps.

As an alternative, I then found a nicer bistability in the system:

$$c_1 + 2c_2 \leftrightarrow 2c_1 + c_2 \tag{35}$$

$$c_1 + c_2 \leftrightarrow 2c_3 \tag{36}$$

When I use a rate k of 1 for both equations, and $\rho_1^*=0$ and $\rho_2^*=.999$, I found that there would be bistability for a range of c (the influx rate) from .08 to .25. When I repeated this analysis for different rates k for equation 35, then the range of bistability for c would also change. The two curves (high c and low c), plotted as a function of k, seemed to converge very nicely, as in the "X" plots of Epstein. But, at the point of intersection, they did not lead to oscillation; instead, they led to a loss of bistability (for large k), or to the bottom of the physically acceptable range (k=0).

In order to understand these difficulties, and improve my chances of finding oscillations, I then looked more closely at the literature on nonlinear system dynamics. Bar-Eli[36], citing [37], discusses four different ways in which oscillations could show up suddenly, as system parameters are changed. The two most basic (and common) involve "Hopf bifurcations," which are discussed more completely by Abed and DeClaris[25], who cite [38] as a primary source. (See also [39], who cites [40]. See [41] for a more recent and colorful source based on the extensive experience of Japanese electronic engineers.)

To understand these bifurcations, one simply cannot avoid some use of matrix analysis. One must consider the properties of the Jacobian matrix (which should not be confused with the Jacobi function), defined by:

$$J_{i,j} = \frac{\partial \dot{\rho}_i}{\partial \rho_j} \tag{37}$$

In ecology, this matrix has sometimes been called the "community matrix" [35], because it describes the incremental impact of species number j on the rate of growth or decline of species number i.

To look for a Hopf bifurcation, one simply looks for a stable equilibrium point where the Jacobian contains two imaginary eigenvalues, $i\lambda$ and $-i\lambda$. For a simple system with two degrees of freedom, this requires that the trace of the Jacobian be zero. (Even in higher dimensions, the trace of J equals the divergence of the flow, a quantity of some importance.) With the simple Boissonade and DeKepper model, it is very easy to solve for such a bifurcation, using exact algebra; this, in turn, suggests that the Epstein mechanism may be a way of locating that type of oscillation, where it exists. (This is interesting, insofar as Bar-Eli states that most chemists tend to expect a more complex type of bifurcation.) To analyze equations 28, 35 and 36, one can also try to solve for such a point, solving for the equilibrium values of ρ_1 and ρ_2 and the values of ρ^{1*} and ρ_2^* required to zero the trace, as a function of k, c, etc.; however, across a wide range of c and k, the required values for $\underline{r}$ and $\underline{\rho}^*$ are unacceptable (<0). The attempt to solve directly for Hopf bifurcations is a very powerful computational tool in locating oscillations, but in this case it appears to suggest that oscillations cannot exist -- a suggestion supported by a very large number of simulations. By contrast, for equations 29-31, it is easy to solve algebraically for an equilibrium point with the trace of J equal to zero; that was how I picked the values .45 and .045 (out of a range of equally good parameters). The very first simulation led to strong oscillations.

Two other diagnostic tools were very useful in these runs: (1) the determinant of the Jacobian; and (2) the <u>derivatives</u> of the trace and determinant with respect to all the adjustable parameters, calculated by generalized backpropagation [7]. A Hopf bifurcation typically does <u>not</u> lead to oscillations when the determinant goes to zero along with the trace (though counterexamples can be designed, dependent on extreme bad luck). When only the determinant goes to zero, the system can become very unpredictable; it may bifurcate into two nearby branches, or it may "jump" a finite distance to a totally different solution, or any number of other things may happen. When the traces approaches zero, in a two-degree-of-freedom system, and the determinant does not, one may be sure that the eigenvalues of J have large imaginary components; typically, this implies the existence of strong but damped oscillations on the stable side of the Hopf bifurcation.

Nonlinear System Dynamics and Chaos

The field of nonlinear system dynamics has advanced far beyond the simple idea of stable equilibrium points versus total disorder. As an example, the concept of chaos has become very widely known [42], though commonly misunderstood.

The concept of attractor is fundamental to this field. A stable equilibrium point is one example of an attractor. Instead of settling down to one point, some systems settle down to some kind of fixed oscillation, or "limit cycle, like some of the systems discussed in the previous subsection. In phase space, these systems settle down into a kind of elliptical orbit (or a warped elliptical orbit), travelling around and around forever. (In simple spreadsheets, it is easy to plot part of the phase plane diagram, by using the XY plot option on the simulated values of the system variables.) Whenever the system starts out somewhere in the neighborhood of that orbit, it moves steadily closer and closer to the orbit.

Limit cycles are a second example of an attractor. Attractors may also be higher dimensional spaces, like the surface of a doughnut. In recent years, people have discovered stable attractors of fractal dimension -- something like 2½-dimensional spaces -- even for very reasonable-looking dynamic systems. These are called strange attractors. Strange attractors are not the same thing as chaos, though they often exist in association with chaos [20].

The concept of chaos refers to systems which are very stable on one level, but very unstable at another. A chaotic system always converges to a stable attractor, which usually has a small number of dimensions. Within the attractor itself, however, the system is essentially divergent. If you look at two different starting points on the attractor, very close to each other, you will see them diverging further and further away from each other with increasing time. This is called "sensitive dependence on initial conditions." For a readable but serious introduction to chaos and its applications, see [43]. For a deeper, physics-oriented survey, stressing the phenomenon of chaos in conservative systems, see [44].

Conventional studies of chaos focus on low-dimensional attractors. Between the realm of conventional chaos and the realm of total disorder, there is another realm which I think of as "turbulence." The physicist Kadanoff has done substantial work on turbulence, using both supercomputer simulations and experiments with liquid helium. Intuitively, if a physical system exists in a volume which is R atoms by R atoms by R atoms, then conventional chaos would imply that it settles down to a system with k degrees of freedom (i.e., a k-dimensional attractor); disorder implies that it retains kR^3 degrees of freedom, because all the atoms are independent of each other; turbulence would imply that it settles down to an attractor with something like $kR^{½}$ degrees of freedom, or kR, or something of the sort. Turbulent systems have very many degrees of freedom, even when they settle down, but they are still very far from total disorder.

The concept of self-organized criticality, discussed by Per Bak at this conference, can be seen as an example of "turbulence" in this sense. Bak's model of evolution seems highly abstract, but its predictions are very interesting. In particular, his model suggests that the extinction of the dinosaurs might have been due to something like a breakthrough in the evolution of other species, instead of a cosmic catastrophe. There is a fascinating match between his theory and some early suggestions by George Gaylord Simpson that the evolution of the neocortex, among small mammals, might have had a huge impact upon the biosphere. I would propose that we should look more carefully at the possibility that the evolution of the neocortex might have been the real cause of the extinction of the dinosaurs.

The study of turbulent dynamics has been limited by the extreme complexity of the mathematics. It is hoped that the techniques given in [4,5] (starting out from concepts given in [45]) will someday be of use here.

In general, the search for dynamical systems which generate chaos or turbulence is critical to our understanding of life. Living systems are clearly not static equilibria, nor are they states of total disorder; they are clearly a form of chaos or turbulence. Even today, most people believe that chaos or turbulence are possible only in open systems; that belief, in turn, reinforces the belief in the Big Bang.

Spontaneous Symmetry Breaking (SSB): The Iron Rod

Spontaneous symmetry breaking (SSB) is one of the most important forms of self-organization. People have known about SSB for a very long time, but only recently have they begun to appreciate its real importance. Weinberg and Salam have used SSB to unify our understanding of electricity and magnetism with a large part of nuclear reactions [21], and their work is now part of the "standard model" of physics. SSB is also central to the "grand unified theories"[46] of mainstream physics. This paper will argue that SSB might be even more important still than is realized at present; for example, it may play a crucial role in permitting life to exist in this universe.

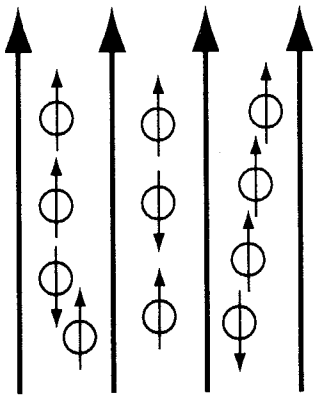

Figure 8. The Iron Magnet: An Example of SSB

Figure 8 describes the magnetization of an iron rod, one of the two classical examples of SSB. When an iron rod is hot enough, it will have no magnetism,, at the macroscopic level. Each atom in the rod is a tiny magnet, but the atoms bounce around at high velocity, constantly colliding and changing their direction, so that they all point in different directions at random. When the rod cools down, however, a kind of bandwagon effect takes over. Different atoms start to line up with their neighbors, until the rod as a whole develops a magnetic orientation (the large arrows in the figure). Even after the rod is magnetized, the movement of the atoms will shift some of them to the opposite orientation for short periods of time, as shown in the figure.

Notice what is going on here. The underlying dynamic laws which govern the individual atoms and the magnetic fields are totally symmetric with respect to space. There is no difference at all between the "up" direction and the "down" direction, so far as these laws are concerned. The microscopic laws are up/down-symmetric. Nevertheless, the macroscopic system which obeys these laws still develops a strong orientation with respect to space. It develops a strong a persistent orientation towards the "up" direction, in this example. A system which is totally symmetric at the microscopic level becomes totally asymmetric and oriented at the macroscopic level.

Could this same phenomenon explain the macroscopic orientation of our universe in time, despite the underlying time symmetry? This theoretical possibility should not be rejected out of hand. At this conference, Prigogine cited some new theorems he and Petrosky have proven [23] suggesting very strongly that a wide range of universes would in fact spontaneously develop a time orientation -- a form of SSB. Whether these theorems really have that effect in our particular universe or not, the possibilities here need to be studied very carefully. If they really do lead to a spontaneous breaking of time-symmetry, as claimed, then they would totally eliminate the need for a Big Bang as a mechanism to explain life and order. (Other motivations for the Big Bang will be discussed in the next major section of this paper.)

An interesting variation of this possibility is that CPT symmetry might be broken by SSB in our universe (or in large regions of it), while asymmetry with respect to T might be part of the underlying laws of physics. The CPT breaking would mainly be a matter of making us see lots of matter but very little antimatter in our universe Starting from there, the T asymmetry could generate the real "arrow of time" that underlies life as we know it -- as I will discuss further in the final section of this paper.

Note that the kind of system shown in Figure 8 does not fit the classical statistical model of random scattering across space. The macroscopic field provides a kind of global interaction effect. This may not be a necessary consequence of SSB, but it still is an important possibility. When QFT is used to analyze systems like the iron rod, one generally uses solid-state physics or tools borrowed from solid-state physics, rather than the pure scattering methods stressed in high-energy physics. The implications of this have yet to be fully understood.

SSB may also be directly relevant to social and biological phenomena. For example, the English language is a very useful tool in communication, but no one believes that every last detail of English grammar and spelling represents some kind of preordained optimal solution. The evolution of language, like the evolution of magnetism, depends heavily on local units lining up with (conforming to) their neighbors and creating a bandwagon effect [7].

It is even conceivable that the unique roles of DNA and protein molecules on earth is due to SSB to some extent; thus, it is conceivable that a different set of molecules could have emerged from evolution, if the initial conditions on earth had been slightly different. We will never know the answers to questions like this is our research is based only on empirical observation of the life we have now on one planet. A number of researchers [11,47] have begun to exploit very different sources of information, in trying to understand the dynamics of life; however, all of this work is only in its infancy, and new efforts to unify the various strands of thought could be very useful.

Solitons and Solitary Waves

Solitary waves were first discovered by empirical observation, by people staring at water waves in a canal in Scotland more than a century ago [48]. Over several decades, mathematicians specializing in the study of water waves gradually developed a basic understanding of this phenomenon [49]. Only in the last two decades have people begun to appreciate their widespread importance for engineering [50] and biology. As an example, the propagation of nerve impulses along the axon is usually modeled as a solitary wave. Hameroff, in the previous Radford conference, described how solitons propagating along microtubules inside the nerve cell may also play a crucial role, such as the implementation of backpropagation. (See [7] for a brief summary of the literature and an explanation of the importance of this point.)

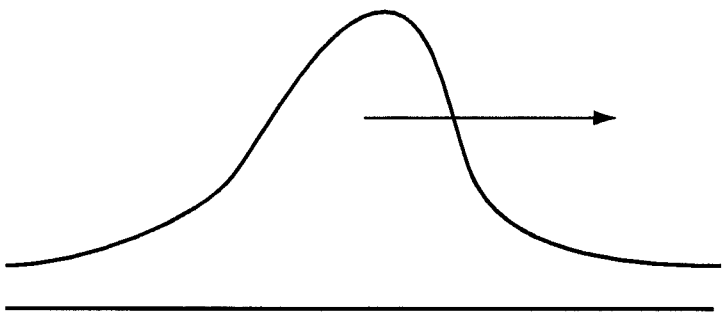

Figure 9. Picture of a Solitary Wave

Figure 9 illustrates the basic idea. In a linear system, an isolated wave like the one shown in Figure 9 would quickly disperse, giving rise to a totally disordered kind of equilibrium. But in a nonlinear Lagrangian system, such a wave could sometimes continue moving along indefinitely, without any change in shape. In certain very special systems, these solitary waves will maintain the exact same shape and speed even after they collide with each other (although the interaction zone would be more complex, of course); mathematicians have developed a very elaborate theory of solitons [51], of solitary waves in that sort of special system (Physicists often use the word "soliton" more loosely, to refer to the broad class of solitary waves.) Solitary waves are generally defined as stable patterns which retain their shape exactly, but may move around at a constant velocity.

In general, solitary waves represent a very interesting kind of order which emerges even in a classical Lagrangian system.

There are some very difficult mathematical challenges in trying to discover whether any given dynamical system can generate solitary waves or solitons. One curious method seems to mimic the Big Bang ideas: build a computer model of the system; start the system out with a huge energy impulse at one point in space; then watch what comes flying out of that initial explosion. But this is only a computational tool; it has nothing to do with astrophysics.

HERESIES: SOME CRITICAL OPPORTUNITIES FOR THE FUTURE

Every major new development in the field of self-organization has started out as a heresy. This section will describe some new opportunities or challenges of that sort. In particular, it will make three basic points, in order, in three subsections:

1. That patterns, order and even life may be possible in time-symmetric, closed Lagrangian systems, even without any spontaneous symmetry breaking (SSB). The word "life" has often been defined in a way which would make this a logical contradiction, but I will suggest that a broader view of life is more appropriate.

2. That life as we know it -- which is inherently asymmetric with respect to time -- may be possible in asymmetric, energy-conserving systems (which could, in fact, be regions within a greater time-symmetric cosmos affected by SSB).

3. That one can construct an image of our own universe -- a "caricature model" -- which provides an alternative to the Big Bang, and opens up empirical questions that can actually follow through on point 2.

The first two hypotheses call for mathematical and computer research, because they are hypotheses about what is possible, in general, for well-specified mathematical universes. The third is mainly an empirical issue; the most pragmatic researcher may want to jump directly to the final section. Nevertheless, a full development of the field of self-organization would require an ability to answer all of these questions, and to unify the mathematical and empirical strands of research.

This paper does not in any way question the importance of research into open systems in biology. To understand the evolution of life on earth as we know it, the open-systems approach is clearly the right starting point, because it reflects the primary role of light from the sun in powering the process of natural selection. But this leaves open the question of how the earth fits into the larger universe, and the possibility of life as we do not know it. Insights from a more general theory might well improve our understanding of classical open systems as well -- if such a theory can ever be developed.

Could Life Exist in Time-Symmetric Lagrangian Universes?

Could chaos, turbulence or life exist on a sustained basis in a conservative, time-symmetric Lagrangian universe?

The issue of chaos is a good starting point here, because it is easier to analyze than life as such, and life itself may be seen as a special case or extension of chaos. Also, the study of chaos in time-symmetric systems is a good warmup exercise, in overcoming the natural tendency to always assume time-forwards causality, unconsciously, even where that assumption is inappropriate.

Conventional wisdom suggests that chaos can occur only in dissipative systems, not in conservative systems. However, one of the first known examples of chaos was the simple system made up off three idealized planets moving around under the influence of each other's gravitational fields. Numerous similar examples have been studied since then. (See [44], especially the footnote on page 7.)

Chaos can occur far more easily when we account for the fact that fields vary over space and not just over time. In [6], I analyzed the issue of chaos in time-symmetric Lagrangian universes, accounting for variations over space, from a mathematical point of view. At least in theory, there seems to be strong reason to expect the existence of "chaotic solitons" or "chaoitons," patterns which remain stable and localized in space, like classical solitons, but which cycle through a complex attractor, like chaotic systems. In effect, chaoitons are to solitary waves what chaotic systems and the like are to stable point attractors. In the usual formulations of chaos theory, which do not account for fields varying over space as such, it is more difficult to generate chaos in a conservative system, because there is no way for higher-energy states to dissipate energy, as they would have to do in order to settle down into a lower-energy attractor; however, when there are spatial dimensions, energy can simply move away to other parts of space, on out to infinity. Stability based on this principle can apply at the same time to the forwards and backwards time directions. For near-term mathematical research issues, see [6]. In [4,5], I proposed that the neutron and proton might be chaoitons.

Is there a possibility, then, of life itself, in addition to chaos, in such a universe? Conventional authors have

sometimes even defined life as a time-forwards process based on DNA molecules and amino acids. Such definitions would tell us that partially silicon-based life, for example, is impossible simply by definition; however, semantic exercises of this sort are simply a way of covering up our fundamental ignorance about wha might evolve on other planets, under different physical conditions. The issues of what might evolve in a truly time-symmetric environment are part of the same intellectual challenge here.

Usually, when we think of natural selection, or we think about small but critical feedback effects, we think of a time-forwards process. We may think of living systems as a kind of attractor for a simple dynamical system, moving forwards in time.

On the other hand, recall that Lagrangian systems are a very special type of dynamical system. They are based on some kind of optimization process. This may be an important clue to understanding their macroscopic behavior. There is an analogy here to a phenomenon in economics, called "turnpike theorems." Certain economists have spent considerable effort trying to describe the optimal path for highly complex systems. For a wide variety of initial conditions and terminal conditions, they find that the same intermediate states can emerge as part of the optimal solution. (The analogy they give is to turnpikes: regardless of where you start driving from and where you want to go, if the distance is long enough, the fastest route will probably take you to the turnpike. This analogy is not as good for biology as for economics, but the mathematical insight is still valid.) If we interpret natural selection as this kind of emergence or crystallization process -- even in the time-forwards case -- then we should expect that the phenomenon could also occur in a time-symmetric world, or even in a time-symmetric niche in our own universe (if such could exist).

It is conceivable that the phenomenon of intelligence could also be understood better in this approach, and extended to the time-symmetric case. Intelligence as we know it seems to be best understood as a kind of approximate optimization system[7]. If the universe is somehow "trying" to optimize something, we should not be surprised that it creates subsystems which explicitly calculate optimal paths. (This may sound a bit too anthropomorphic, but -- as in the idea of Lagrangian systems itself -- the anthropomorphic metaphor can be useful, if we remember its limitations.) Intuitively, the solution to a constrained optimization problem (as in the Bryson and Ho formulation of Lagrange-Euler dynamics[15]) may involve the use of a distributed control architecture, where intelligent living systems serve as nodes in the system. It is interesting but difficult to imagine what intelligence in a time-symmetric universe might look like; it would certainly not be made up of the usual time-forwards Turing computers or model neurons, but it might have some connection to ideas about quantum computing which have begun to emerge in recent years.

Life in a time-symmetric universe might be similar to life as we know it from a mathematical viewpoint, but it would have very serious differences on a concrete level. For example, the processes of death and birth look very different when viewed in reverse time. What would substitute for birth and death? Would life itself appear to form a kind of vast web, across space and time, with phenomena like "loose ends" replacing both life and birth? Would it generate local causality gradients, on its own, at least along strands of the web which go between its periphery and the "center" (or central skeleton)? Would the phenomenon of learning follow these local causality gradients? Clearly, we would need to do a lot of mathematical work before being able to simulate such possibilities. Furthermore, we would have to be very careful to avoid inserting the assumption of time-forwards causality, unintentionally, into such simulations.

Many mathematicians have stated that the effort to prove Fermat's Last Theorem has led to very important and valuable advances in mathematics, far more important than the theorem itself. In a similar way, the effort to understand these kinds of phenomena may be very useful, even though our own universe does not appear to be completely time-symmetric. If, in fact, our universe is a hybrid or intermediate case, which is not causally time-symmetric but not perfectly time-forwards either, then we may need to understand both the time-symmetric and time-forwards extremes before we can understand the hybrid.

Could Life Persist in a Closed, Asymmetric Universe?

To explain the kind of life that we see on earth -- life which is clearly asymmetric -- we clearly have to assume that this life evolved in an asymmetric environment, either because of asymmetric laws of physics or because of SSB (which creates the appearance of asymmetry). Thus the question which is relevant for our kind of life within the universe as a whole is as follows: in a closed Lagrangian universe, which is not time-symmetric, could chaos and life evolve and continue indefinitely?

As in the time-symmetric case, a logical place to start is to display simple patterns of order -- oscillations,

and then later chaos -- in this kind of system. In the section on chemical oscillations above, I have shown a simple example of this phenomenon. That example is truly a baby-like starting point, but it is at least a significant step in showing that energy-conserving closed classical systems can generate interesting patterns. Energy conservation in scattering equations does not necessarily imply that such equations could result from a Lagrangian system, or that the quantity conserved in scattering would match the Hamiltonian of the Lagrangian system. (In fact, the empirical data for our universe really involves energy conservation and the like, more than the Lagrangian property as such.) Clearly, there is more work to be done, to develop and understand more complex examples. The simple model of oscillation in equations 29-31 reminds me of the popularized version of neural networks found in [52]: that work was hopelessly too elementary to be a model of intelligence, but it was a crucial starting point for the community, in coming to appreciate the more complex and relevant theory which had given rise to it[7]. In this case, however, I am hoping that the reader, rather than myself, will uncover the deeper aspects of this issue.

Intuitively, the real issue here is as follows: could a simple term or terms in the dynamic equations of our universe play the same role, for the universe as a whole, that uniform radiation from the sun does for the earth? Personally, I would expect the answer to be "yes," but the challenge here is to prove the answer, and to develop its implications. The most important task, however, is the empirical task, which the next section will address.

An Alternative to the Big Bang

This section will provide a simple "caricature" model, which is intended to encourage the empirical work and further theorizing which will be needed before anyone can develop a final verdict on the Big Bang theory. The emphasis is on the word "caricature," rather than the word "model."

The deepest and strongest reason to believe the Big Bang theory has been the belief that no other mechanism was available to explain order and life in our universe. Up until now, this paper has questioned that belief at a theoretical level. The caricature model gives us a more concrete example of how we could get order and life in the long-term, without a Big Bang. The caricature model is not intended as an all-encompassing theory of everything, but only as a minimal example of some important mechanisms we need to study empirically. In all likelihood, empirical studies will eventually lead us to something much more complicated than the caricature model itself.

What is the empirical evidence for and against the Big Bang, if we truly ignore the issue of explaining life and order (which has very much colored people's interpretation of other evidence)?

The first classical piece of evidence is the widespread redshift from distant galaxies and galaxy-like objects. It is well-known that the colors of any object turn red, when that object moves away from us. Decades ago, Hubble showed that objects further away from us tend to be shifted more towards the red; this seems to suggest a general pattern of expansion or explosion in the universe. On the other hand, more careful and more recent work by Arp and others[53] has shown that the pattern of redshifts is far more complex. There are other sources of very high redshifts, which are well established by the best experts in optical phenomena [54-7], but generally rejected by astronomers because of their apriori commitment to the classical explanation. Furthermore, there are anomalies even in the usual redshift patterns[58]. By and large, normal redshifts do tend to be proportional to the distance from earth, but the caricature model will suggest an alternative possibility for the cause.

The strongest piece of evidence for the Big Bang is said to be the low level microwave radiation reaching the earth from all parts of the sky. This microwave radiation follows the usual distribution of energies which one would expect from any thermodynamic system in equilibrium at the temperature of 3 degrees Kelvin. Actually, all of the well-worked-out versions of the Big Bang theory require that the radiation not be uniform. In 1992, there was great public excitement when NASA's COBE experiment reported variations on the order of 0.000001 from one part of the sky to another; however, this was still far less than the variation expected by Big Bang theories at the time[1]. Given the pressure of this data, some theorists have been creative enough to add some additional mathematical assumptions (epicycles?), which have brought some of the Big-Bang models back into consistency with the data. Significant interactions and fields in deep, intergalactic space are an alternative way to generate low-level radiation, which would naturally approach a black-body equilibrium over many billions of years, and would naturally tend to be more uniform in its point of origin; the high degree of observed uniformity is a more natural and unavoidable prediction in this alternative class of theory.

Lerner has recently published a book[1] describing a wide range of problems with the Big Bang theory, which have not received sufficient attention. Lerner's accounts of global field effects -- related in a sense to SSB -- are extremely interesting. Lerner does hurl some strong and unjustified criticism at mathematicians and mystics and Republicans -- criticisms which may offend many readers for good reason -- but this does not lessen the seriousness

of the questions he raises. The work of Arp, Hoyle, Lerner and others has not disproved the Big Bang theory, but it does indicate enough doubt that we are justified in trying to develop and test some alternative possibilities.

The caricature model is a minimal alternative, based on seven reaction equations and five types of "particle." Unlike the simplified system of equations 18-19, this system assumes that each type of "particle" requires one additional number to characterize it -- a velocity or energy level, or a particle number. The five types of particle are "Hydrogen" (H), "Helium" (He), neutrino (ν), photon (γ) and scavenger (s). Nothing is known or assumed about the detailed properties of s -- whether it is really only one particle type or more, whether it has positive mass-energy, or whether it would allow the possibility of supporting life in deep space. (An interesting option is that s might be a kind of "hole" in a "false vacuum," which could have bounded energy and therefore avoid some of the difficulties mentioned in connection with equation 24 above.) Likewise, there is no assumption (or requirement) that there be any oscillation or chaos in the mix of particle types at this level. In conventional QFT, some of these possibilities -- while plausible -- would suggest some important possibilities for instability over large regions of the universe[16]; it is not yet known whether the alternative formulations discussed above would support these suggestions.

In the model, a star is simply an H particle with a large particle number. A dead star or dark matter is an He particle with a large particle number. Other elements, like carbon, are not treated explicitly, because we already know a lot about the formation of planets and so on, starting from stars and hydrogen in space; the challenge here is to model the continued existence of active stars.

The caricature model begins with three chemical-like reactions, representing (in caricature-like fashion) the well-known effects of gravity and nuclear fusion (and related galaxy-forming effects):

$$H(n) + H(m) \leftrightarrow H(n+m) + \gamma \tag{38}$$

$$He(n) + He(m) \leftrightarrow He(n+m) + \gamma \tag{39}$$

$$H \leftrightarrow He + \gamma + \nu \tag{40}$$

These are all time-symmetric reactions, as is well-known from physics.

The model requires two additional reactions, to eat dark matter (dead stars, etc.) and put new hydrogen into deep space:

$$H(n) \ (+s) \leftrightarrow H(m) \ (+s) \tag{41}$$

$$He(n) \ (+s) \leftrightarrow He(m) \ (+s) \tag{42}$$

Note the testable hypothesis that protons can decay, with some very low probability which is nevertheless significant on astronomical time scales. Actually, many grand unified theories share this properties, often requiring a more dramatic decay. Dr. Per Anders Hansson, a consultant to the British scientific establishment, has reported informally that experiments to measure such decay rates may now be feasible, in light of modern high-precision nanotechnology[22]. Note that equations of this general sort would not create hydrogen so much as redistribute it. After time symmetry is broken, by later equations, these two equations by themselves result in a net consumption of energy from the s particles in forwards time.

Finally, to generate the crucial time-asymmetry -- the driving force -- the model would include two one-way equations:

$$s + \gamma \rightarrow s + \gamma \qquad (\gamma \textit{ of lower energy}) \tag{43}$$

$$s + \nu \rightarrow s + \nu \qquad (\nu \textit{ of lower energy}) \tag{44}$$

Equation 43 -- which is assumed to occur at a very low rate, just enough to explain the normal distance-related

redshifts of astronomy without attributing them to a Doppler effect. Equation 44 -- a somewhat stronger reaction -- would be used to explain the "missing neutrino" problem -- a fundamental, unexplained problem in today's astronomy. The "cost" of equation 44, in terms of Occam's Razor, is zero, because any complete theory of physics has to include something new to account for the missing neutrinos. Given that neutrinos are associated with weak nuclear reactions, which in turn are relatively close to the superweak interactions, it is very reasonable to have a moderate time-asymmetry in that particular equation. Equations 41 to 44 all stand in for a wide range of possible alternatives, and all suggest a variety of related experiments under different conditions. For example, high-precision experiments have been attempted (but not far enough to be conclusive) on distance-related redshift effects, even on earth, though a larger distance base might allow higher precision measurements. High precision neutrino detectors, such as the KARMEN project at Rutherford Appleton [22], might be useful as part of new experiments related to equation 44.

Some physicists have proposed that the missing neutrinos from the sun might be explained by assuming that neutrinos spontaneously mutate into other forms of neutrinos. Lately, I have heard some scientists argue that the total neutrino count from the sun is too low, ruling out this theory, while others think it may still be plausible. A time-asymmetric mutation of neutrinos could be part of a more detailed system implementing the basic ideas here, but that is only one of many, many possibilities.

Even equation 43 may not be quite so speculative as it appears at first. Even in conventional QFT, physical photons are not considered identical to the pure, idealized "bare photons" which appear in the usual Lagrangians; rather, they are held to be a mixture of bare photons, electron/positron pairs, and so on. This results from "renormalization effects" in the mathematics, effects which have been well-proven experimentally[14,26]. (See [4,5] for a more physical interpretation of these effects.) Furthermore, in today's standard model of physics, photons and neutrinos are very closely related to each other; these close relations should be accounted for in any correct, complete renormalization[21]. If one revised the equations for the neutrino, in a crude, direct way, to replicate the high observed rate of energy loss between the earth and the sun, one should also go back and recalculate the propagation equations for the photon. Even a very tiny correction might be enough to replicate the kind of red shift we see from distant spiral galaxies; after all, the observed energy loss per meter of distance travelled by the neutrino is many orders of magnitude larger than the usual red shifts. Nevertheless, I have not yet done these calculations, and the results may well be highly sensitive to one's choice between equally plausible starting points. Furthermore, the close link between photons and neutrinos also suggests that unknown phenomena which affect neutrinos directly might well affects photons directly in any case, simply as a matter of symmetry.

To understand this system, it is crucial once again to break free of the usual time-forwards way of thinking. It is crucial that a small time-forwards feedback effect, provided by the last two equations, can nonetheless lead to large-scale implications.

Equation 44 is the key mechanism which generates the arrow of time, in this model. By gobbling up neutrinos emerging from stars in forwards time, it prevents the phenomenon of fusion from occurring in the reverse time direction. This is what makes stars form and "burn" in forwards time only. That in turn is responsible for life as we know it on earth, in this model.

References

[1] E.J.Lerner, *The Big Bang Never Happened*, Expanded Edition. Random House, 1992.

[2] T.S.Kuhn, *The Structure of Scientific Revolutions*, U. of Chicago Press, 1962.

[3] P.Werbos, "Bell's theorem: the forgotten loophole and how to exploit it," in [13].

[4] P.Werbos, "Chaotic solitons and the foundations of physics: a potential revolution," *Applied Mathematics and Computation*, Vol. 56, p.289-340, July 1993.

[5] P.Werbos, "Quantum theory, computing and chaotic solitons," *IEICE Transactions on Fundamentals*, Vol.E76-A, No.5, p.689-694, May 1993.

[6] P.Werbos, "Chaotic solitons in conservative Systems: can they exist?," *Chaos, Solitons and Fractals*, Vol. 3, No.3, p.321-326, 1993.

[7] P.Werbos, *The Roots of Backpropagation: From Ordered Derivatives to Neural Networks and Political Forecasting*, Wiley, 1993.

[8] P.Werbos, "The brain as a neurocontroller: some new hypotheses and experimental possibilities," updated version, in this volume.

[9] D.White and D.Sofge, eds, *Handbook of Intelligent Control: Neural, Fuzzy and Adaptive Approaches*, Van Nostrand, 1992.
[10] P.Werbos, "Elastic fuzzy logic: a better fit to neurocontrol and true intelligence", *Journal of Intelligent and Fuzzy Systems*, Vol.1, No.4, 1993.
[11] C.Langton, ed., *Artificial Life* (Santa Fe Volume VI), Addison-Wesley, 1989.
[12] S.Hawking, *A Brief History of Time: From the Big Bang to Black Holes*,Bantam,1988.
[13] M.Kafatos, ed., *Bell's Theorem, Quantum Theory and Conceptions of the Universe*. Kluwer, 1989.
[14] F.Mandl, *Introduction to Quantum Field Theory*, Wiley, 1959.
[15] A.Bryson and Y.Ho, *Applied Optimal Control: Optimization, Estimation and Control*, Hemisphere, 1975.
[16] S.Coleman and F.DeLuccia, "Gravitational effects on and of vacuum decay," *Phys. Rev. D*, Vol.21, p.3305-15, June 1980.
[17] R.Adler, M.Bazin and M.Schiffer, *Introduction to General Relativity*, McGraw-Hill,1965.
[18] R.Penrose and W.Rindler, *Spinors and Space-Time*, Cambridge U.Press, Corrected Edition, 1987.
[19] M.Carmelli, *Classical Fields: General Relativity and Gauge Theory*, Wiley, 1982.
[20] C.Grebogi, E.Ott and J.Yorke, "Chaos, strange attractors and fractal basin boundaries in nonlinear dynamics," *Science*, Vol. 238, p.632-637, 30 October 1987
[21] J.Taylor, *Gauge Theories of Weak Interactions*, Cambridge U. Press, 1976.
[22] Private communication, Dr. Per Anders Hansson, Dec. 8, 1993.
[23] T.Petrosky and I.Prigogine, "Quantum chaos, complex spectral representations and time-symmetry breaking," *Chaos, Solitons and Fractals*, forthcoming.
[24] J.Horgan, "Quantum philosophy," *Scientific American*, Vol.267(1), p.94-104, July 1992.
[25] E.Abed and N.DeClaris, "Analytical and geometric problems in nonlinear system modeling and stability," in R. Kalman, Marchuk, and Viterbi, eds., *Recent Advances in Communication and Control Theory*, New York: Optimization Software Inc., 1987.
[26] C.Itzykson and Zuber, *Quantum Field Theory*, McGraw-Hill, 1980.
[27] I.Epstein, K.Kustin, P.De Kepper and M.Orban, "Oscillating chemical reactions," *Scientific American*, 248(3), p.112-123, 1983.
[28] I.Janis, *Victims of Groupthink*, Houghton Mifflin, 1973.
[29] I.Prigogine and I.Stengers, *Order Out of Chaos*, Bantam, 1984.
[30] Zs. Nagy-Ungvarai, J.Tyson, S.Muller, L.Watson and B.Hess, "Experimental study of spiral waves in the Ce-catalyzed Belousev-Zhabotinsky reaction," *J. Phys.Chem.*, Vol. 94, p.8677-8682, 1990.
[31] Z.Noszticzius, W.McCormick and H.Swinney, "Use of bifurcation diagrams as fingerprints of chemical mechanisms," *J. Phys. Chem.*, Vol. 93, p.2796-2800, 1989.
[32] L.Gyorgyi, T.Turanyi and R.Field, "Mechanistic details of the oscillatory Belousev-Zhabotinsky reaction," *J. Phys.Chem.*, Vol. 94, p.7162-7170, 1990.
[33] C.Hocker and I.Epstein, "Analysis of a four-variable model of coupled chemical oscillators," *J.Chem.Phys.*, Vol. 90 (6), p.3071-3080, March 1989.
[34] H.Krug,L.Pohlmann, and L.Kuhnert, "Analysis of the modified complete Oregonator accounting for oxygen sensitivity and photosensitivity of Belousev-Zhabotinsky systems," *J.Phys.Chem.*, Vol. 94, p.4862-4866, 1990.
[35] R.May, *Stability and Complexity in Model Ecosystems*, Princeton U. Press, Second Edition, 1974.
[36] K.Bar-Eli and M.Brons, "Period lengthening near the end of oscillations in chemical systems," *J. Phys. Chem.*, Vol. 94, p.7170-7177, 1990.
[37] A.Andronov, A.Vitt and S.E.Khaikin, *Theory of Oscillators*, Pergamon, New York, 1966.
[38] J.Casti, *Nonlinear System Theory*, Academic Press, Orlando, Florida, 1985.
[39] W.Zhang, *Synergetic Economics: Time and Change in Nonlinear Economics*, Springer, 1991.
[40] S.Chow and J.Hale, *Methods of Bifurcation Theory*, Springer, 1982.
[41] T.Matsumoto, M.Komuro, H.Kokubu and R.Tokunaga, *Bifurcations: Sights, Sounds and Mathematics*, Springer-Verlag, 1993.
[42] J.Gleick, *Chaos: The Making of a New Science*, Penguin, 1988.
[43] T.Mullin, ed., *The Nature of Chaos*, Oxford U. Press, 1993.
[44] A.J.Lichtenberg and M.A.Lieberman, *Regular and Chaotic Dynamics*, Second Edition, Springer-Verlag, 1992.
[45] G.Nicolis and I.Prigogine, *Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order Through Fluctuations*, Wiley, 1977.

[46] J.Wess and Bagger, *Supersymmetry and Supergravity*, Princeton U. Press, 1983
[47] L.Margulis and L.Olendzenski, eds, *Environmental Evolution: Effects of the Origin and Evolution of Life on Planet Earth*, MIT Press, 1992.
[48] Scott Russell, J., 1844. *Report on Waves*. Brit Assoc. Rep.
[49] G.Whitham, *Linear and Nonlinear Waves*, Wiley, 1974
[50] A.Scott, F.Chu and D.McLaughlin, "The soliton, a new concept in applied science," *Proceedings of the IEEE*, Vol. 61(10), 1973
[51] G.Eilenberger, *Solitons*, Springer-Verlag, 1983.
[52] D.Rumelhart and J.McClelland, eds, *Parallel Distributed Processing*, Volume I, MIT Press, 1986.
[53] H.Arp, G.Burbidge, F.Hoyle, J.Narlikar and N.Wickramashinge, "The extragalactic universe: an alternative view," *Nature*, Vol.346, p.807-12, Aug. 1990.
[54] E.Wolf, *The Red-Shift Controversy and a New Mechanism For Generating Frequency Shifts of Spectral Lines*, Technical Bulletin of the National Physical Laboratory, New Delhi, India, October 1991.
[55] E.Wolf, "Towards spectroscopy of partially coherent sources," in R. Inguva, ed., *Recent Developments in Quantum Optics*, Plenum Press, New York, 1993.
[56] E.Wolf, "Influence of source correlations on spectra of radiated fields," in J.W. Goodman, ed., *International Trends in Optics*, Academic Press, 1991.
[57] D. James and E.Wolf, "A class of scattering media which generate Doppler-like frequency shifts of spectral lines," submitted to *Physical Review Letters*, 1993.
[58] H.Arp, "The Hubble relation -- differences between galaxy types SB and SC," Astrophysics and Space Science, Vol. 167(2), p.183-219, 1990.
[59] R.K.Brayton and C.H.Tong, "Constructive stability and asymptotic stability of dynamical systems," *IEEE Transactions on Circuits and Systems*, Vol. 27, November 1980.
[60] A.N.Michel, N.R.Sarabudla and R.K.Miller, "Stability analysis of complex dynamical systems: some computational methods," *Circuits, Systems, Signal Processing* (Birkhauser Boston), Vol. 1, No. 2, 1982.
[61] A.N.Michel, B.H.Nam and Vijay Vittal, "Computer generated Lyapunov functions for interconnected systems: improved results with applications to power systems," *IEEE Transactions on Circuits and Systems*, Vol. 31, no,2, February 1984.
[62] A.N.Michel and R.K.Miller, "Stability analysis of discrete-time interconnected systems via computer-generated Lyapunov functions with applications to digital filters," *IEEE Transactions on Circuits and Systems*, Vol. 32, No.8, August 1985.
[63] A.Barto, R.Sutton and C.Anderson,"Neuronlike elements that can solve difficult learning control problems," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 13, No. 5, 1983.
[64] P.Werbos, "Consistency of HDP applied to a simple reinforcement learning problem," *Neural Networks*, Vol. 3, October 1990.
[65] G.W.Whitehead, *Homotopy Theory*, MIT Press, 1966.
[66] P.Werbos, "Supervised learning," *WCNN93 Proceedings*, Erlbaum, 1993.
[67] W.Miller, R.Sutton and P.Werbos, eds.,*Neural Networks for Control*, MIT Press, 1990.

APPENDIX. NEW METHODS FOR THE AUTOMATIC CONSTRUCTION OF LIAPUNOV FUNCTIONS

This technical appendix will present some new ideas in stability theory, mentioned briefly in the section on classical approaches to self-organization. It will address two closely related ideas in classical control theory, in order: (1) Given a nonlinear dynamical system, $\dot{\underline{x}}=\underline{f}(\underline{x})$, with an equilibrium point $\underline{x}_0$, to find a Liapunov function which proves stability about $\underline{x}_0$, either globally or over a large region, if the process is in fact stable; (2) Given a nonlinear dynamical system, $\dot{\underline{x}}=\underline{f}(\underline{x},\underline{u})$, influenced by control variables $\underline{u}$, to find a controller $\underline{u}=A(\underline{x})$ which results in global (or regional) stability about a desired point $\underline{x}_0$. (The new ideas were designed to handle more general problems[9], but this appendix will address only these two problems.)

More precisely, this appendix will discuss new approaches to the automatic construction of a Liapunov function (and controller) to solve these problems. With very few exceptions [59-62], the classical approaches to these problems all require that a human being simply guess both the Liapunov function and the controller -- something

which works very well in the linear case, but only rarely in the nonlinear case. There is an analogy here to the state of the art in solving algebraic equations before Newton: people depended very heavily on using massive amounts of ingenuity to locate closed-form solutions to algebraic equations. Nowadays, closed form solutions are still very useful when they are available, but for practical applications, we routinely rely on equation-solving systems which use methods of successive approximation based on computers. My goal here is to encourage the same kind of alternative approach available to classical stability analysis and control.

A few authors in the control field have occasionally suggested that they have fixed "general" Liapunov functions for "general" nonlinear controllers. However, global stability in the general case is not such a trivial issue. Stability is often easy for simple regulator systems, which require no real intelligence or planning in the controller. For a simple test problem, which helps illustrate the limits of controllers which do not incorporate a deep understanding of dynamics, consider the bioreactor test problem in Appendix A of [67].

In classical control theory, there is essentially only one established method for the automatic construction of Liapunov functions: the method of Brayton and Tong[59], generalized somewhat by Michel et al[60,61]. (Zhubov and others have described techniques for constructing Liapunov functions which are not "automatic" in the sense that they do not lead to specific computer algorithms; for example, Zhubov requires the solution of PDE, which do enter into my discussion below, but we do not have an exact procedure to solve nonlinear PDE in the general case[60].) Brayton's method has been useful in a few practical applications[61,62], but it is limited in several important ways:

1. As one would expect, we can only prove stability for systems which are in fact stable -- which have a Liapunov function (unknown), $V(\underline{x})$, with $V(\underline{x}_0)=0$. It is also necessary that the system be <u>smooth</u> in some sense, related to the Lipshitz condition. One might formalize this by requiring that there exists a characteristic length λ such that $|\underline{x}-\underline{y}|<\lambda$ implies that $|\log V(\underline{x})/V(\underline{y})|<K$, for some K.
2. The computational cost of the method is essentially proportional to adr^{d-1}, where a is a measure of the cost of some matrix comparison operations, d is the dimensionality of the vector $\underline{x}$, and r is the ratio between the size (diameter) of the region under study and the characteristic length λ.
3. The method makes a "worst case" assumption about the interactions of the Jacobians at different points $\underline{x}$ as the process evolves over time.
4. So far as I know, there is no complete rigorous theory which formally verifies stability after a finite number of points have been explored.

The first of these limitations is essentially unavoidable. The second -- the high computational cost -- is important in practice, and has been stressed very heavily by Michel et al [61], who have found ways to stretch the method a little in a few special cases. The fourth limitation could probably be overcome relatively easily, through further theoretical work. The third limitation, however, has not received enough attention: i.e., <u>when</u> the Brayton method proves stability, one may be sure that a process is stable, but there is no reason at all to believe that the method would in fact locate a Liapunov function for "most" stable well-behaved nonlinear system.

To explain this point, I need to explain the essence of how the Brayton method works. Essentially, Brayton argues that a dynamical system looks like a linear system, $\dot{\underline{x}}=J(\underline{x})\underline{x}$, where $J(\underline{x})$ is the Jacobian matrix at the point $\underline{x}$. In effect, the dynamical system generates $\underline{x}(t+dt)$ by multiplying $(I+J(\underline{x})dt)$ by $\underline{x}$. If we can proves that long products like $(I+J_1dt)(I+J_2dt)...(I+J_ndt)\underline{x}$ always stay within some fixed, bounded region, for <u>any</u> matrix J_i which equals $J(\underline{y})$ for some point $\underline{y}$ in that region, then we may be sure that the process cannot possibly blow up. In effect, this method allows for the possibility that the system encounters <u>any</u> possible sequence of Jacobian matrices, as it moves through state space. This is very much a worst-case assumption. There are many stable systems which achieve stability because the dynamics in one region -- which appear quite unstable -- automatically push the system into other regions, which are stable. (In the simulations on chemical systems described in this paper, I encountered this phenomenon over and over again, to my great frustration; this happened, for example, when I modified the coefficients in equation 30 to match those of equation 29.) Brayton's method would not demonstrate the global stability of such systems.

This appendix will discuss several related approaches to constructing Liapunov functions. The simplest approaches try to overcome the third limitation above, while maintaining the status quo on the three other limitations. The more radical approaches address the issue of computational cost as well. All of these approaches depend heavily on the use of <u>approximate dynamic programming</u> or <u>adaptive critics</u>, which are explained at length in [9].

The link between constructive Liapunov functions and dynamic programming can be understood at two levels -- intuitively, or formally.

Intuitively, the only way to find a Liapunov function in the general case is to start from the neighborhood of the equilibrium point (where $V(\underline{x})\approx 0$) and working backwards along the orbits of the process, setting higher values for V as one works one's way back. One would follow the usual dynamic programming approach of backwards movement from the end point, and incrementing the V function as one increments the J function in dynamic programming. (Henceforth, when I refer to J, I will be referring to the Jacobi function J of dynamic programming, rather than the Jacobian matrix J.)

More formally, the Hamilton-Jacobi-Bellman equation[15] reduces to the following equation in this case:

$$\frac{dJ(\underline{x}(t))}{dt} = U(\underline{x}(t)) \quad , \tag{45}$$

where we may require that U be positive definite and that it approach zero as $\underline{x}$ approaches $\underline{x}_0$. This requirement on J is equivalent to the requirement that J be a Liapunov function! In fact, we have a two-way equivalence here. Thus finding a Liapunov function is equivalent to solving a dynamic programming problem.

There is no way to actually solve the Bellman equation exactly for a generalized, continuous nonlinear system. Therefore, to develop general methods to construct Liapunov functions, we must use general-purpose approximations to dynamic programming instead. The most effective approximations to dynamic programming[9] make use of "critic networks" or "critic functions," which try to approximate the function $J(\underline{x})$ or to approximate its gradient, $\underline{\lambda}(\underline{x})$. More precisely, these methods require that the user specify a critic function, $\hat{J}(\underline{x},W)$ or $\hat{\underline{\lambda}}(\underline{x},W)$, which is capable of approximating a wide variety of possible functions J or $\underline{\lambda}$, depending on the value of the weights or parameters W; these methods then provide ways of adapting or estimating the set of weights W. Many choices for the function $\hat{J}$ have been used in the literature; most papers describe $\hat{J}$ as an "artificial neural network," but this is often little more than a semantic convention.

To use these methods in constructing a Liapunov function, one can proceed as follows. First, instead of specifying a single function $\hat{J}(\underline{x},W)$, specify a sequence of functions $\hat{J}_1(\underline{x},W_1)$, $\hat{J}_2(\underline{x},W_2),\ldots,\hat{J}_i(\underline{x},W_i),\ldots\hat{J}_\infty$, of ever-increasing complexity, with the property that the minimum approximation error (i.e. the minimum over W_i of $|J-\hat{J}|$) goes to zero as i goes to infinity. Pick an arbitrary "utility" function U such as:

$$U(\underline{x}) = (\underline{x} - \underline{x}_0)^T Q (\underline{x} - \underline{x}_0) \quad , \tag{46}$$

where Q is some positive-definite matrix. (Q could simply be set to I, or to any other "natural metric" for the system.) For any given i and cutoff U_{min}, we can use adaptive critic methods to try to solve the Bellman equation using the critic function $\hat{J}_i$ for the set of points $\underline{x}$ such that $U(\underline{x})>U_{min}$; to define the boundary conditions -- the values of J where $U=U_{min}$ -- one may use the simple linearized approximation:

$$V(\underline{x}) = (\underline{x} - \underline{x}_0)^T W (\underline{x} - \underline{x}_0) \tag{47}$$

$$WA + A^T W = Q \, , \tag{48}$$

where A is the Jacobian matrix at the point $\underline{x}_0$, and where equation 48 -- the Liapunov equation -- must be solved to find the matrix W. (Note that equation 48 is just a linear equation in the elements of W.)

This last procedure only gives us an approximation to the relevant dynamic programming problem, and to a J which is a true Liapunov function; however, a good enough approximation to a Liapunov function should also be a Liapunov function and, by increasing i and decreasing U_{min}, we are guaranteed to have an approximation as good as we like.

To try out this procedure in a simple case, such as a two-dimensional system, one might choose $\hat{J}$ to be a simple lookup table. In other words, one may simply slice up the relevant region of two-dimensional space into a set of square cells, and simply assign a single value of $\hat{J}$ to approximate J in each square. Approximations to dynamic programming of this sort have been very popular in the past -- generally on an ad hoc basis -- and they are also the basis of the original Barto, Sutton and Anderson algorithm[63].

Unfortunately, the square or rectangular approaches will not work here. As an example, consider the following simple process, to be analyzed in (x,y) coordinates, which I define for convenience in polar coordinates (with θ defined in radians):

$$\dot{\theta} = 1 \tag{49}$$

$$\dot{r} = -kr \ , \tag{50}$$

where k is 0.01 or 0.001. This process has a stable equilibrium at the origin. However, no matter how small the squares, the points within any square generally flow into three other squares. Assuming the worst (as is required in formal proofs of stability), knowledge of the flows from square to square still permits the possibility of divergence. Formally speaking, there exist flows from squares of lower $\hat{J}$ to higher $\hat{J}$, no matter how small the squares, for any reasonable method of assigning values of $\hat{J}$. (Again, the formal proof that such approximations cannot work really depends on the fact that the square-to-square flows appear to permit divergence;
one need only consider the problem of how to assign $\hat{J}$ values along the squares in such a divergent path, to see that a Liapunov function in this scheme is impossible.)

To make the original approach work, we need to use critic networks which allow a good approximation to the gradient of J, not just to J itself. This can be done quite easily in the two-dimensional case, by dividing up the plane into triangles, rather than squares. For example, one can easily divide the plane up into equilateral triangles (in the usual hexagonal sort of tiling), or one can simply split each of the square tiles in the usual arrangement into two triangles; one an simply draw in the diagonal line from upper-left to lower-right, within each square. One an try to estimate a value for J explicitly at each vertex, and use linear interpolations within each triangle to approximate J. For systems which obey a kind of proportional Lipshitz condition in the derivatives of J, we can be certain that such an approximation scheme allows us to get arbitrarily lose to the true derivatives as the triangles become smaller. If J is a well-behaved Liapunov function, for which the direction of flow is never orthogonal to the gradient of J (i.e., there exists an angle θ_{min} such that the angle between these vectors never exceeds $(\pi/2)-\theta_{min}$), this guarantees that the approximation to J will itself be a Liapunov function, eventually, as the triangles become smaller and smaller. To generalize this to n dimensions, one need only arrive at a tiling scheme made up of solids with n+1 vertices. Strictly speaking, the procedure here is also equivalent to approximating J as a weighted sum of continuous basis functions; for the split square tiling, for example, the equivalent basis functions are functions which look like triangular wedges in cross-section, with square-shaped contours rotated 45 degrees from the original squares.

It is one thing to prove that a Liapunov function exists, within a set of approximating functions. It is another thing to provide an explicit learning rule, and to prove that this learning rule will eventually find the relevant Liapunov function.

The obvious learning rule to use here is Heuristic Dynamic Programming (HDP), which I first proposed in print in 1977, for another application, and which is discussed at length in [9]. (The original Barto, Sutton and Anderson "method of temporal differences"[63] is the special case of HDP where $\hat{J}$ is a lookup table and U equals zero throughout.) In this approach, we would somehow sample points $\underline{x}$ throughout the region of interest; for the sake of efficiency, we would probably want to sample just the enters of each triangle, and begin sampling at the enter, and gradually work our way outwards. (Even random sampling, however, should eventually converge.)

At each sampled point $\underline{x}$, we can call on the system model available in the computer to evaluate $\underline{f}(\underline{x})$. We an find some h small enough that $\underline{x}+h\underline{f}(\underline{x})$ is still within the triangle, but perhaps lose to the edge. Following the procedure in [9], we would then adjust the weights W such that $\hat{J}$ approximates $\hat{J}(\underline{x}+h\underline{f}(\underline{x}),W)+hU(\underline{x})$. In this case, the "weights" are simply the estimates for J at the vertices of the triangle; I would conjecture that convergence is guaranteed if we limit ourselves to adjusting the estimated value of J on that vertex which is furthest "upstream", based on $\underline{f}$. This approach is essentially a special case of the total gradient approach described in [64], which was inconsistent in the general case, with a learning rate of 1. I would further conjecture that the partial gradient approach, with an appropriate choice of learning rates, would also converge efficiently to the right answer. The challenge to future research is either to prove these conjectures, or provide minor variations which would allow such a proof. In a similar way, one could apply DHP [9] -- a more powerful method -- to the same kind of problem. One alloyed try to estimate $\underline{\lambda}$ at each vertex, and again use linear interpolation in-between. This would have the advantage that it allows an exact fit to the usual quadrati Liapunov approximation in the innermost triangle.

The methods described above still have unacceptable costs, as n grows very large. Thus it is crucial to

develop more radical approaches, as I will discuss below. Before doing so, I would like to comment on the n-dimensional generalization of the triangle-tiling approach, if only for the sake of curiosity and esthetic completeness.

The main problem in specifying this generalization is to specify a way to slice up or "tile" n-dimensional space into regular locks which have only n+1 vertices each. There are many, many ways to do this. One might expect that excellent methods might be found in the literature on crystallography or on the interface between geometry and topology[65]. However, those are large and difficult bodies of literature; for now, I find it easier just to spell out two relatively simple techniques. Both techniques are n-dimensional generalizations of the simple idea of tiling the plane into squares, and then splitting the squares in half to generate triangles.

In the first alternative, we begin by choosing some unit of length, and then splitting up the region of interest into n-dimensional hypercubes of that size. Then, we split each hypercube into pieces, as follows. If the cube is a unit cube, based on the origin, define each piece as the convex hull of the following n+1 points: $\underline{\mathbf{0}}$, $\underline{\mathbf{e}}_i$ (for some i between 1 and n), $\underline{\mathbf{e}}_i + \underline{\mathbf{e}}_j$ (for the same i, for some j, not equal to i, between 1 and n), $\underline{\mathbf{e}}_i + \underline{\mathbf{e}}_j + \underline{\mathbf{e}}_k$ (for the same i and j, for a new k not equal to i or j, and still between 1 and n), ... , $\underline{\mathbf{1}}$. Because there are n! choices for the sequence of integers i,j,k,... , there will be n! different pieces. A simple change of coordinate axes (translation and change of scale) extends this procedure to any arbitrary hypercube.

A second alternative -- suggested by Elizabeth Werbos, my 13-year-old daughter -- would probably have better computational properties, because the pieces are less "gerrymandered" in appearance. More tiles are required per cube, but, because the tiles are more compact, one could probably afford to use larger cubes. In the three-dimensional case, she proposes that we first split the cube into six pieces; for each of the six faces of the cube, we cut out the pyramid stretching from that that face to the center of the cube. Then we can slice each pyramid into four pieces, by splitting the square face along its two diagonals (and extending each slice up to the top of the pyramid). The overall effect is to split the cube into 24 pieces, each containing a perfect corner vertex, linked to edges of similar size (0.5 and 0.7). The n-dimensional generalization is straightforward: i.e., slice the n-dimensional hypercube into 2n pieces, one from each face; further split each piece into 2(n-2) subpieces, by slicing each face from its center to its faces; subdivide each of the latter faces into 2(n-4) pieces, and so on.

The methods described above are very broad, and require much more development. Nevertheless, all of that effort would be merely a first step towards what is needed in practice: methods with much lower computational cost. The methods described above would essentially cost something on the order of r^d, which is similar to the preexisting methods. If the only available information about the function $\underline{\mathbf{f}}$, aside from our ability to evaluate it at selected points, is a conventional Lipshitz kind of assumption, then it is intrinsically impossible to prove stability at less than a very large computational cost, as d and r grow large. However, the work of Barron at Yale has shown that true neural networks, such as multilayer perceptrons, have the ability to approximate general functions with far fewer hidden nodes, as the dimensionality of the problem grows, than simple lookup or basis-function networks. Halbert White, of UCSD, has shown that MLPs can also approximate derivatives to an accuracy as great as one might like. The obvious approach, then, is to use HDP or DHP (or even GDHP) on an MLP or on new kinds of hybrid networks [65] instead of a basis function network. Using the approaches given in [9], one can use these methods to derive a controller, as well as prove the stability of the result. (Note, however, that the full working versions of DHP, GDHP and the related Gradient-Assisted Learning method, which directly minimizes errors in the derivatives of output, are all covered by a pending patent at BehavHeuristics, Inc., of College Park, Maryland.)
This approach would presumably yield a Liapunov function at a much smaller computational cost; however, the challenge to research lies in proving rigorously that one has in fact found a Liapunov function after one has.

How can one rise to this challenge? One approach would be to use symbolic methods like those used in AI to prove the relevant inequalities. When the critic networks are taken from a relatively simple, restricted set of functions (as with particular types of neural networks), such inequalities may be much easier to handle than they are in the general case.

Another approach is simulation. Even today, the "proofs" of stability in practical systems typically have two parts: (1) rigorous proofs for linear approximate models of the underlying nonlinear plant; (2) lots and lots of simple simulations of entire time-paths, based on more realistic nonlinear models. Narendra, in a 1990 NSF workshop, gave an example of a system which is "provably stable" in this conventional approach. The linear system -- based on the kinds of approximations commonly used in industry today -- was provably stable, assuming he validity of the approximations, based on theorems which Narendra himself originated over a decade ago. However, in realistic simulations of the nonlinear plant, it actually could blow up. He demonstrated a neural network controller directed developed on the nonlinear model, for which he then had no proof of stability, but which never blew up the plant, over many, many simulations.

Given the limitations of this conventional approach to stability analysis, neural networks could possibly make a significant improvement even when an exact approach is not possible. For example, one can use neural networks or hybrid networks[66] to approximate the nonlinear behavior of the plan as accurately as possible. Given a neural network plant model, symbolic reasoning about Liapunov functions may be much easier than it would be if any general algebraic expression were permitted. This is far better than simply using a linear model!

Alternatively, one might simply use these methods as a way to improve the numerical efficiency of the simulation process. One can sample states $\underline{x}$ at random, to prove that the Liapunov property is maintained near all points, instead of having to sample complete trajectories of the system. (Brute force simulation is numerically similar to the old Widrow blackjack design, which is well understood to be less efficient than the more modern critic designs.) As with drug testing, the goal may be to develop a kind of statistical bound on the degree of risk, based on some kind of optimized sampling.

This work, if pursued further, offers the hope of building workable general-purpose computer programs able to develop provably stable controllers for general, realistic nonlinear plants whenever stable control is in fact possible.

GLOSSARY OF A FEW BASIC TERMS

Attractor: Many dynamical systems gradually settle down into a fixed, stable equilibrium state or to a stable set of possible states. Each state may be thought of as a point in phase space. An attractor is simply a connected set of points in phase space which the system is attracted to (i.e. converges towards). In many practical applications, a system may have more than one "attractor," when the attractor points are not all connected to each other. There is no consensus on the exact details of the definition. (See [44], page 461.)

Bifurcation: A phenomenon which sometimes occurs when the underlying parameters which define a dynamical system are changed. (For example, we can experiment with different values of the parameter c in equation 28.) A small change in such a parameter can sometimes lead to an abrupt change in the system dynamics, such as a shift from having a simple static equilibrium to having two different stable equilibria; this is a basic form of bifurcation. More examples are given and cited in the 2½ pages which follow equation 32.

Causality: An intuitive concept -- like temperature (see definition) -- which, unlike temperature, has yet to be understood as a purely derived effect from the underlying system dynamics. See the discussion which follows after equation 12. The puzzle of causality, in different forms, permeates the rest of the paper.

Chaos: Chaos refers to the situation where a dynamical system converges to a stable attractor, but behaves in an unpredictable, divergent sort of way within the attractor. See the next subsection after equation 37. See also the definitions of oscillation and strange attractor.

Conservative system: A system which contains one or more "integrals" -- nontrivial quantities which, like energy, remain constant as the system evolves over time. See Figure 5.

CPT symmetry: See the discussion which follows equation 12.

Critical point: An equilibrium point, a point where $\dot{\underline{x}}=\underline{f}(\underline{x})$. Also sometimes called a "singular point." Papers using this terminology often refer as well to the "stable manifold" of such a point (the set of points which flow in towards the point) and the "unstable manifold" of the point (the set of points which points near to the equilibrium flow into).

Dissipation: See Figure 1.

Divergence: See Figure 4.

Energy: In modern thinking, the word "energy" is normally used as a shorthand for "mass-energy," as defined in equation 5 -- a quantity which is absolutely conserved, a quantity whose total value across all space does not change over time. Even in pre-Einsteinian thinking, equation 5 was used as the standard definition of energy. Under certain

conditions (as in Figure 1), physical energy may serve as a Liapunov function for approximate models of simple physical systems embedded with a larger conservative universe.

Entropy: A particular type of Liapunov function, normally applied to probability distribution. See the discussion of equation 25. (In information theory, "entropy" refers to a measure of information content or noise -- a different concept, but based on the same equation.) Equation 21 might also be considered an entropy function

Integral: See the definition of conservative system.

Jacobian: Because Jacobi was a very creative mathematician, his name is associated with at least two very different concepts -- the Jacobian matrix (defined in equation 37), and the Jacobian function J (exemplified in equation 45, and discussed further in [8]).

Lagrangian function or system: See the discussion of equation 1.

Liapunov function: A function used to prove the stability of a dynamical system. For any stable dynamical system, a Liapunov function is a function of the system variables which always decreases over time, as the system variables evolve over time, except at the stable equilibrium point (or set of points), where the function reaches its minimum value. See Figures 3 and 4.

Oscillation: A regular, periodic behavior over time -- neither static (as in stable point equilibria) nor chaotic. As the parameters of a system are changed (see the definition of "bifurcation"), there is a kind of standard progression from static equilibrium, to oscillation, to bifurcations within the oscillation, to chaos, to turbulence, and, finally, to thermodynamic disorder. In actuality, there are several different routes to chaos[20], and many variations on this progression.

Phase plane: See Figure 4.

QFT: Quantum field theory, a refined version of quantum mechanics developed in the 1950's by (primarily by Feynman and Schwinger), which is currently viewed as the foundation for all of physics. QFT has yet to have an operational, empirical impact on gravitational physics (though quantum gravity is a major field of research); however, it has had a pervasive impact on the other domains of physics, except at the level of very gross engineering approximations.

Soliton: See the discussion of Figure 9.

SSB: See the discussion of Figure 8.

Strange attractor: A kind of attractor (see the definition of attractor). Ordinary attractors are simple, compact geometric shapes, such as a point, an ellipse, or a surface. Strange attractors are nonordinary attractors -- attractors which typically have fractal dimensions. Strange attractors are typically associated with chaos, but there are exceptions [20].

Temperature: In everyday life, we think of temperature as a physical sensation in our bodies. But in physics, temperature is derived as a secondary quantity, useful in describing the dynamics of a system which does not "know" about temperature in the usual, intuitive sense. Temperature is simply the parameter θ, in equation 24, which does with energy. (In other words, if $\underline{E}^{(1)}$ represents energy, then ρ_1 is temperature.)

Time and time-symmetry: See the discussion surrounding equation 9.