Propagating Organization: An Enquiry

May 3 , 2006 - Toronto draft 36BLSK

Stuart Kauffman
Institute for Biocomplexity and Informatics, University of Calgary

Robert K. Logan
Department of Physics, University of Toronto
Institute for Biocomplexity and Informatics, University of Calgary

Robert Este,
Institute for Biocomplexity and Informatics, University of Calgary

Randy Goebel
Department of Computer Science, University of Alberta

David Hobill
Department of Physics and Astronomy, University of Calgary

Ilya Shmulevich
Institute of Systems Biology, Seattle Washington

Abstract

It is obvious that information is both created and communicated within the process comprising biological evolution, and that the systematic reductionist analysis of component parts has provided much scientific insight about Darwin's grand theory. For example, the combined work of biologists, physicists, and chemists has provided our understanding of DNA as a

significant foundation for the "gemmules" that Darwin and his contemporaries sought as the explanation for the evolution of the biotic universe.

Here we seek a non-reductionist explanation for the synthesis, accumulation, and propagation of information, which we hope will provide some insight into both the biotic and abiotic universe, in terms of the basic work cycle energy, work, information creation and the propagation of organization. The idea is to combine ideas from biology, physics, and computer science, to formulate explanatory hypotheses on how information can be captured and rendered in the expected physical manifestation, which can then propagate the organization of process in the expected biological work cycle to create the diversity in our observable biosphere.

Our study is far from complete, but our progress places fundamental perspectives on information in context, to help focus the challenge of possible explanatory hypotheses and possibly the first steps to developing a theory of the organization of process. This includes the potential role of the analysis of information by Shannon and Kolmogorov, which are both found wanting with respect to characterizing the information requirements of evolution.

We argue that the proper and deep understanding of Schrödinger's intuition from his essay *What is Life* is that an aperiodic crystal contains a very large number of diverse constraints that are partially causal in guiding the huge diversity of events and processes which occur physically in cells. From this we shall arrive at a new formulation: constraints are information and information is constraints.

We further argue that information is a relative concept and cannot be defined as an invariant like the speed of light. Its definition depends on the context of the situation that it is describing or representing. Hence there is no conflict between the instructional or biotic information we have defined and Shannon or Kolmogorov information. Each has its own place and its own usefulness. We locate biotic (but not linguistic) semiosis, as a subcase of information as constraints. Finally we identify a number of other examples of the propagation of organization which includes language and culture and specific elements of culture, namely technology, governance and economies.

Our conclusions, to date, suggest a foundation which views information as the construction of constraints, which, in their physical manifestation, guide the processes of evolution to dynamically determine the fitness of organisms within the context of a biotic university.

> An organized being is then not a mere machine, for that has merely moving power, but it possesses in itself formative power of a self-propagating kind which it communicates to its materials though they have it not of themselves; it organizes them, in fact, and this cannot explained by the mere mechanical faculty of motion.
>
> Immanuel Kant - Critique of Judgement

Introduction

Our broad aim is to understand propagating organization as exemplified by the vast organization of the co-evolving biosphere. Our effort is a rather mysterious undertaking, for we entirely lack a theory of organization of process, yet the biosphere, from the inception of life to today manifestly propagates organization of process. Indeed, we believe that the evolving universe as a whole also manifests the propagation of organization. We shall focus most of our efforts on the biotic case, but undertake an initial extension of our analysis to the abiotic case as well.

The role of information in biology, what it "is," how it accumulates, and how it is "used," has not been directly addressed by mainstream biologists. Yet information is likely to be one of the key unifying concepts in the emerging field of systems biology. As part of the propagating organization within living cells, the cell operates as an information processing unit, receiving information from its environment, propagating that information through complex molecular networks, and using the information stored in its DNA and cell-molecular systems to mount the appropriate response. Indeed, biology is acquiring many characteristics of an information science (Hood and Galas 2003).

Although information is only a part of the propagating organization in cells which also involves work, constraints, the linking of exergonic and endergonic process, as well as other features, a deeper understanding of the role of information in biology can only help us gain new insights about the fundamental processes of life.

It is sometimes the case that science progresses by finding the concepts and language to “see that which is directly in front of us.” Such is the case with the present enquiry. We are persuaded that we are not wholly successful, but hope that we shall have at least started a far broader discussion.

Two predecessors to this article can be found in *Investigations* (Kauffman 2000), and “Emergence, Autonomous Agents, and Organization” (Kauffman and Clayton 2006). At its core, *Investigations* seeks to understand the physical nature of agency itself, and proposes that a molecular autonomous agent, able to act on its own behalf in an environment, is an autocatalytic system carrying out at least one thermodynamic work cycle. Much follows from this tentative definition, which implies that an autonomous agent is an open non-equilibrium chemical system, and finds general biotic importance in the fact that work cycles link spontaneous and non-spontaneous (exergonic and endergonic) processes. This linkage has built up the enormous complexity of the biosphere.

Further analysis reveals this work to be the constrained release of energy into a few degrees of freedom. But if one asks where the constraints themselves come from –as in the example of a cylinder and piston that confine the expansion of the working gas in the head of the cylinder to yield the translational motion of the piston, hence the release of energy into a few degrees of freedom–one finds that it typically takes work to construct the constraints[1]. Thus we arrive at the first surprise - it takes constraints on the release of energy for work to happen, but work for the constraints themselves to come into existence. This circle of work and constraint shall turn out to be part of the theory of propagating organization that we shall discuss.

Most importantly, contemporary cells, which are collectively autocatalytic and do work cycles, also carry out work to construct constraints on the release of energy. When released, this energy constitutes further work that drives non-spontaneous processes that builds structures, drives processes, and also builds further constraints on the release of energy, which when released can build still more such constraints. In short, cells carry out propagating work linking spontaneous processes, constraints, work, and non-

[1] Here we use the word "constraint" in a very general sense that includes "global constraints" (e.g. conservation of energy, symmetry conditions etc.,) and "local constraints" or boundary conditions (e.g. initial conditions, reflection or absorption at a spatial location).

spontaneous processes, and more broadly as we shall see, the propagating organization of process. In doing so, the cell carries out a set of interlocked tasks that achieve a closure of tasks whereby the cell literally builds a rough copy of itself. We know this, yet we have no clear way to say what we know. This closure of work, constraints, tasks, and information, as we shall see below, is a new form of matter, energy, information, and organization that constitutes the living state.

The new insight that we explore in this article is that the constraints that allow autonomous agents to channel free energy into work are connected to information: in fact, simply put, the constraints *are* the information, are partially causal in the diversity of what occurs in cells, and are part of the organization that is propagated.

In "Emergence, Autonomous Agents, and Organization" (Kauffman and Clayton 2006), the tentative definition of autonomous agent is extended to include construction of boundaries enclosing the agent, discrimination of "yuck" (meaning poison) or "yum" (meaning food), and at least one choice of action: flee (or not), approach (or not). Our language is teleological. We believe that autonomous agents constitute the minimal physical system to which teleological language rightly applies.

This article is organized as follows:

In Section 1 we discuss Darwinian adaptations and preadaptations, argue that the first implies that biology cannot be reduced to physics, while the second, stunningly, implies that the future evolution of the biosphere cannot be finitely prestated. Much follows from these surprising conclusions.

In Section 2 we discuss Shannon information and argue that it does not apply to the evolution of the biosphere. One reason is that due to Darwinian preadaptations, the ensemble of possibilities and their entropy cannot be calculated.

In Section 3 we discuss Kolmogorov information and argue that it does not apply to the ongoing evolution of the biosphere.

In Section 4 we begin with Schrödinger's famous statement that a periodic crystal cannot "say" a lot, while an aperiodic crystal can say a lot. We shall argue that the proper and deep understanding of Schrödinger's intuition is that an aperiodic crystal contains a very large number of diverse constraints that are partially causal in guiding the huge diversity of events and processes which occur physically in cells. From this we shall arrive at a new formulation: constraints are information and information is constraints. The first part of this twosome, constraints are information is, we believe, secure. The second part, information is constraints, may be more problematic.

In Section 5 we argue that information is a relative concept and cannot be defined as an invariant like the speed of light. Its definition depends on the context of the situation that it is describing or representing. Hence there is no conflict between the instructional or biotic information we have defined and Shannon or Kolmogorov information. Each has its own place and its own usefulness.

In Section 6 we shall place our definition of biotic information in the larger context in which information is "about" something, arguing that when an autonomous agent discriminates yuck or yum, the molecular signatures of yuck or yum are about yuck or yum, hence the rudiment of semiotics. We shall locate biotic (but not linguistic) semiosis, as a subcase of information as constraints.

In Section 7 we shall stress that constraints as information, and, derivatively, semiotic information, must have causal consequences for the autonomous agent. These consequences increase its fitness such that the information is assembled by natural selection into the ongoing evolution of the biosphere. Without this coupling to fitness, the information and its effects would not come to exist in the universe. Therefore we shall argue that natural selection constitutes the assembly machinery, when coupled with heritable variation, that literally assembles the propagating organization of matter, energy, constraint, work, and information. This constitutes the propagating organization in autonomous agents, whose coevolution drives the biosphere's progressive exploration of what we call the Adjacent Possible.

In Section 8 we attempt to extend our analysis to the abiotic universe. We find that our analysis that considers information as constraints is equivalent to the statement that information consists in boundary conditions and in global constraints. But, in classical and quantum physics, boundary conditions – like the cylinder and piston – are only partially causal for what occurs. Physicists often "put in by hand" the boundary conditions of a problem, such as the behavior of the cylinder piston working gas system. But in the unfolding of the biosphere or universe since the Big Bang, the very coming into existence of new boundary conditions - information we argue - is itself part of the full dynamics of the total system. We thus assume a context with information understood as boundary conditions on the release of energy that makes diverse processes happen. So we argue that in the proper union of matter, energy and information is precisely the union of matter, energy, and boundary conditions that, in an expanding and cooling universe, progressively break symmetries, invade the Adjacent Possible, and cause an increasing diversity of events, processes and structures come into existence. The evolution of the biosphere is but one case of this general process.

In Section 9 we make brief mention of conscious human agents, and the evolution of language, culture, technology, governance, and economies. Since consciousness is not understood, our purpose here is only to point to these further examples of propagating organization, not to propose an analysis of them.

Section 1. Darwinian Adaptations and Preadaptations.

Were one to have asked Darwin what the function of the heart is, he would presumably have responded that the function of the heart is to pump blood. But the heart has a wealth of other causal consequences, such as heart sounds. Heart sounds are not the function of the heart. That is, the causal consequence of the heart that matters, the virtue for which it was selected, was the pumping of blood. So the function of a part (or organ) of an organism is typically, if not always, a subset of its causal consequences. This has major implications. Among these, the function of a part (or organ) of an organism cannot be analysed except in the context of the whole organism in its selective environment. But further, this fact is just one of the reasons that biology cannot be reduced to physics. In Kauffman and Clayton (2006), it is argued that, if we grant the physicist a theory of everything, say string theory to cite one example, and the capacity to deduce upwards to all that occurs in the universe - an impossibility given throws of the quantum dice - the physicst could deduce all the causal features of the heart, *but would* *have no way to pick out the pumping of blood as the relevant causal property which is the function of the heart* and which is the property that gave rise to the evolutionary emergence of this organ.

To do so, the physicist would have to discuss whole organisms as causal agents in their own right, evolving under natural selection in changing environments. That is, the physicist would have to become a biologist and talk biology talk. Thus, biology cannot be reduced to physics, rather physics has to be lifted up to biology.

Darwin had many brilliant insights. Among these is what is now called a Darwinian preadaptation. Here the central concept is that a causal property of a part of an organism that is not of selective significance in the normal environment might become useful in a different environment, and hence become subject to selection. It is critical to point out, first that Darwinian preadaptations have occurred repeatedly in evolution, and second, that such an evolutionary step results in the emergence in the biosphere of a novel function. For example, lungs evolved from the swim bladders of certain early fish. The swim bladders, partially filled with water and partially with air, adjusted the height in the water column to establish neutral bouyancy of the fish. But the swim bladder, with air in it, was preadapted for use as a lung, and air breathing was a novel functionality that allowed life to conquer land.

We now raise a central question discussed in *Investigations*. Is it possible to say ahead of time what all possible Darwinian preadaptations are for human beings, or for the whole biota of the contemporary biosphere for that matter? The answer appears to be “No.” We cannot finitely prestate all possible Darwinian preadaptations. Part of the difficulty, or impossibility, in doing so is that we cannot even begin the task of prestating what all possible selective environments will be. That is, there appears to be no finitely stateable procedure which would allow us to enumerate all possible environments.

Part of the challenge is that the concept of such environments is systematically vague. It is not even clear how to begin on the project of listing all possible environments for all actual, let alone possible, organisms. While we do not know how to prove our claim, we believe it to be true and shall assume that it is.

We point out that the property or causal consequence which becomes the subject of a Darwinian preadaptation need not be a mutant property. It might be a normal feature of the organism, but normally of no selective significance until the new environment is encountered. Therefore, an attempt to enumerate the possible preadaptations by trying to count the number of mutations possible to a genome is irrelevant. Darwinian preadaptations cannot, in general, be prestated.

Much follows from the claim that we cannot finitely prestate all possible Darwinian preadaptations of all contemporary organisms. First, it means in a radical sense that we cannot predict the future evolution of the biosphere. We literally have no idea of what such preadaptations may be. Second, it means that a frequency interpretation of probability statements does not apply to possible probability statements about the evolution of the biosphere. In the normal frequency interpretation of probability, say that a fair coin will be heads about 5000 times out of 10,000 coin flips, one can finitely prestate all possible outcomes. This is not possible for the evolving biosphere. Third, and dramatically, the incapacity to say ahead of time what the relevant preadaptations will be means that we cannot write down a statable set of variables whose dynamics captures the evolution of the biosphere. But all our mathematical techniques in physics begin with a prestatement of the full set of variables and the configuration space of the system. This is true in Newtonian dynamics, statistical mechanics, general relativity and in quantum mechanics if one does not believe in hidden variables. If one believes in hidden variables then because they are hidden they cannot be prestated hence the caveat for quantum mechanics.

But we cannot prestate the configuration space of the biosphere. Now a classical physist might argue that, if we take the solar system, it is just a large classical 6N dimensional system where N is the number of particles in the solar system and the current biosphere is, with the rest of the solar system, a point in that vast space. Let us grant the move. Then, we rejoin, the physicist has no way to pick out the collective variables, the lungs and hearts and wings, and features of the environment that are the relevant causal variables for the ongoing evolution of the biosphere. Thus, again we see that we cannot write down causal laws with a prestated set of (collective) variables for the evolution of the biosphere.

Section 2. Shannon Information

Shannon (1948) information theory has been a brilliant mathematical construct. At its core, Shannon envisioned a Source with a set of messages, symbol strings, over which a well defined probability distribution might be attributed. Then he envisioned a (perhaps noisy) channel over which information is transmitted. He then envisioned a receiver and, importantly, a decoder. Shannon's move was to calculate the entropy of the set of messages at the Source.

The information that propagated down the channel and was received at the receiver removed uncertainty with respect to the entropy of the Source. This reduction of uncertainty, hence the lowering of the entropy of the Source, constitutes the amount of information transmitted. One interpretation, not given by Shannon himself who abjured to say what information "is," is that information is just the reduction in uncertainty at the receiver. This definition leaves open exactly what the claim might mean. It might be the reduction of uncertainty in a human receiver's mind, for example.

Importantly, and widely recognized, is the fact that Shannon information considers the amount of information, nominally in bits, but is devoid of semantics. There is no sense of what information is "about" in Shannon information.

Now we ask whether Shannon information applies to the evolution of the biosphere. We answer that it does not. In particular, Shannon information requires that a prestated probability distribution (frequency interpreted) be well stated concerning the message ensemble, from which its entropy can be computed. But if Darwinian preadaptations cannot be prestated, then the entropy calculation cannot be carried out ahead of time with respect to the distribution of features of organisms in the biosphere. This, we believe, is a sufficient condition to state that Shannon information does not describe the information content in the evolution of the biosphere.

There are further difficulties with Shannon information and the evolving biosphere. What might consititute the "Source"? Start at the origin of life, or the last common anscestor. What is the source of something like "messages" that are being transmitted in the process of evolution from that Source? The answer is entirely unclear. Further, what is the transmission channel? Contemporary terrestial life is based on DNA, RNA, and proteins via the genetic code. It is insufficient to state that the channel is the transmission of DNA from one generation to the next. Instead, one would have to say that the actual "channel" involves successive life cycles of whole organisms. For sexual organisms this involves the generation of the zygote, the development of the adult from that zygote, the pairing of that adult with a mate, and a further life cycle. Hence, part of one answer to what the "channel" might be is that the fertilized egg is a channel with the Shannon information to yield the subsequent adult. But it has turned out that even if all orientations of all molecules in the zygote were utilized, there is not enough information capacity to store the information to yield the adult. This move was countered by noting that, if anything, development is rather more like an algorithm than an information channel. In short, a channel to transmit Shannon information along life cycles does not exist, so again, Shannon information does not seem to apply to the biosphere.

It seems central to point out that the evolution of the biosphere is not the transmission of information down some channel from some source, but rather the persistent, ongoing, co-construction, via propagating organization, heritable variation, and natural selection, of the collective biosphere. Propagating organization requires work. It is important to note that Shannon ignored the work requirements to transmit "abstract" information.

One might be tempted to argue that a Shannon-like information theory could be applied to the vast set of selective events that have led the specific DNA sequences that are in contemporary organisms. But does this move work? Can we specify a finite ensemble of possible DNA sequences out of which the present DNA sequences have been derived? If we consider all DNA sequences longer than, say 1000 nucleotides, it would take vastly large repetitions of the history of the universe for the universe to construct one copy of each possibility. This cannot physically consitute the ensemble. Is the ensemble the set of DNA sequences that have been explored in the actual evolution of the biosphere, some accepted, most rejected? This approach initially seems promising, but has the obvious difficulty that we cannot specify the ensemble explored in 3.8 billion years, hence do not and cannot know the Shannon information content of the biosphere. A further difficulty with this approach is that it measures the information content of the biosphere as a function of the number of DNA sequences "tried" in evolution. But very different numbers of attempted mutations might have led to the same biosphere, hence quantitating the information of the biosphere by the number of attempted DNA mutations is not in direct correspondence to any specific biosphere.

We conclude that a Shannon Information content analysis of the information content of the evolving biosphere is not legitimate.

Section 3. Kolmogorov Information

Kolmogorov introduced a concept of information in which the information in, say, a symbol string, is the shortest program which would produce that string on a universal computer. From the viewpoint of our notion of algorithm, this shortest program is the Kolmogorov compression of a string. At first, Kolmogorov information as a measure of the information content of the biosphere seems promising. For, happily, Kolmogorov information does not need to specify a finitely prestatable ensemble of messages from which the Source entropy can be computed, hence avoids the difficulties mentioned above for Shannon-like information. We do note, however, that like Shannon, Kolmogorov does not state what information "is."

But, we argue, even under the assumption that a simulation of the evolution of the biosphere were possible, Kolmogorov information does not apply well to the evolving biosphere because diverse histories of life might lead to the same biosphere, yet have different Kolmogorov complexity. Further, there is no way to confirm which specific history of life occurred. Thus Kolmogorov information cannot uniquely specify the information content of this specific biosphere.

Section 4. Schrödinger's Aperiodic Crystal: "Instructional" Information as Constraint or Boundary Condition

In *What is Life*, Schrödinger (1992) is concerned with the order in organisms and hence the physical basis of the gene. He argues, based on X-ray mutation induction frequency, that the gene must have a few hundred to a few thousand atoms, and points out that statistical mechanical equilibria cannot account for the stability of the organism over generations. He then posits that quantum mechanics in the form of chemical bonds is the answer. Then he brilliantly points out that the order of life cannot be based on a periodic crystal, for such a crystal cannot say a lot, or carry much information. He places his bet on aperiodic crystals which can, in strong contrast, say a lot, or carry much information.

He was brilliantly right, and presaged DNA and the genetic code. Now we come to the critical issue. In just what sense can an aperiodic crystal "say a lot?" Schrödinger does not himself say more than suggesting the aperiodic crystal contains a microcode.

We believe Schrödinger was deeply correct, and that the proper and deep understanding of his intuition is precisely that an aperiodic solid crystal can contain a wide variety of microconstraints, or micro boundary conditions, that help cause a wide variety of different events to happen in the cell or organism. Therefore we starkly identify information, which we here call “instructional information” or “biotic information,” not with Shannon or Kolmogorov, but with constraints or boundary conditions, and the amount of information will be related to the diversity of constraints and the diversity of processes that they can partially cause to occur. By taking this step, we embed the concept of information in the ongoing processes of the biosphere, for they are causally relevant to that which happens in the unfolding of the biosphere.

We therefore conclude that constraints are information and, as we argue below, information is constraints which we term as instructional or biotic information to distinguish it from Shannon or Kolmogorov information. We use the term “instructional information” because of the instructional function this information performs and we sometimes call it “biotic information” because this is the domain it acts in, as opposed to human telecommunication or computer information systems where Shannon and Kolmogorov information operate. This step, identifying information as constraint or boundary condition, is perhaps the central step in our analysis. We believe it applies in the unfolding biosphere and the evolving universe, expanding and cooling and breaking symmetries, that we will discuss below.

Is this interpretation right? It certainly seems right. Precisely what the DNA molecule, an aperiodic solid, does, is to “specify” via the heterogeneity of its structural constraints on the behavior of RNA polymerase, the transcription of DNA into messenger RNA. Importantly, this constitutes the copying or propagating of information. Also, importantly, typically, the information contained in aperiodic solids requires complex solids, i.e., molecules, whose construction requires the linking of spontaneous and non-sponetneous, exergonic and endergonic, processes. These linkages are part of the work cycles that cells carry out as they propagate organization.

It is essential to note that the set of constraints in a contemporary cell is not merely the DNA and RNA, but lies also in the specific stereochemistry of a vast horde of specific molecular species. So, when an enzyme binds two substrates and holds them in proximity, lowering the potential energy barrier to their joining, the enzyme is acting as a constraint on the motion of the two substrates, hence as a catalyst. The working of a cell is, in part, a complex web of constraints, or boundary conditions, which partially direct or cause the events which happen. Importantly, the propagating organization in the cell is the structural union of constraints as instructional information, the constrained release of energy as work, the use of work in the construction of copies of information, the use of work in the construction of other structures, and the construction of further constraints as instructional information. This instructional information further constrains the further release of energy *in diverse specific ways,* all of which propagates organization of process that completes a closure of tasks whereby the cell reproduces.

Section 5. The Relativity of Information

In Sections 2 and 3 we have argued that the Shannon and Kolmogorov conceptions of information are not directly suited to describe the information of autonomous agents that propagate their organization. In Section 4 we have defined a new form of information, instructional or biotic information as the constraints that direct the flow of free energy to do work.

The reader may legitimately ask the question "isn't information just information?", i.e., an invariant like the speed of light. Our response to this question is no, and to then clarify what seems arbitrary about the definition of information. Instructional or biotic information is a useful definition for biotic systems just as Shannon information was useful for telecommunication channel engineering, and Kolmogorov information was useful for the study of information compression with respect to Turing machines.

The definition of information is relative and depends on the context in which it is to be considered. There appears to be no such thing as absolute information that is an invariant that applies to all circumstances. Just as Shannon defined information in such a way as to understand the engineering of telecommunication channels, our definition of instructional or biotic information best describes the interaction and evolution of biological systems and the propagation of organization. Information is a tool and as such it comes in different forms just as screwdrivers are not all the same. They come in different forms, slot, square, and Philips –depending in what screw environment they are to operate. We therefore would like to suggest that information is not an invariant but rather a quantity that is relative to the environment in which it operates. To drive home this point we will now examine the historic context in which Shannon (1948) information emerged.

Before delving into the origin of Shannon information we will first examine the relationship of information and materiality. Information is about material things and furthermore is instantiated in material things but is not material itself. Information is an abstraction we use to describe the behavior of material things and often is thought as something that controls, in the cybernetic sense, material things. So what do we mean when we say the constraints are information and information are the constraints as we did in Section 4.

“The constraints are the information” is a way to describe the limits on the behavior of an autonomous agent who acts on its own behalf but is nevertheless constrained by the internal logic that allows it to propagate its organization. This is consistent with Hayle’s (ibid., p. 72) description of the way information is regarded by information science: “It constructs information as the site of mastery and control over the material world.” She claims, and we concur, that information science treats information as separate from the material base in which it is instantiated. This suggests that there is nothing intrinsic about information but rather it is merely a description of or a metaphor for the complex patterns of behavior of material things. In fact, the key is to what degree information is a completely vivid description of the objects in question.

This understanding of the nature of information arises from Shannon’s (1948) original formulation of information, dating back to his original paper:

> The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages. The system must be designed to operate for each possible selection, not just the one that will actually be chosen since this is unknown at the time of design. If the number of messages in the set is finite then this number or

> any monotonic function of this number can be regarded as a measure of the information produced when one message is chosen from the set, all choices being equally likely.

A number of problems for biology emerge from this view of information. The first is that the number of possible messages is not finite because we are not able to prestate all possible preadaptations from which a particular message can be selected and therefore the Shannon measure breaks down. Another problem is that for Shannon the semantics or meaning of the message does not matter, whereas in biology the opposite is true. Biotic agents have purpose and hence meaning.

The third problem is that Shannon information is defined independent of the medium of its instantiation. This independence of the medium is at the heart of a strong AI approach in which it is claimed that human intelligence does not require a wet computer, the brain, to operate but can be instantiated onto a silicon-based computer. In the biosphere, however, one cannot separate the information from the material in which it is instantiated. The DNA is not a sign for something else it is the actual thing in itself, which regulates other genes, generates messenger RNA, which in turn control the production of proteins. Information on a computer or a telecommunication device can slide from one computer or device to another and then via a printer to paper and not really change, McLuhan's "the medium is the message" aside. This is not true of living things. The same genotype does not always produce the same phenotype.

According to the Shannon definition of information, a structured set of numbers like the set of even numbers has less information than a set of random numbers because one can predict the sequence of even numbers. By this argument, a random soup of organic chemicals has more information that a structured biotic agent. The biotic agent has more meaning than the soup, however.

The living organism with more structure and more organization has less Shannon information. This is counterintuitive to a biologist's understanding of a living organism. We therefore conclude that the use of Shannon information to describe a biotic system would not be valid. Shannon information for a biotic system is simply a category error.

A living organism has meaning because it is an autonomous agent acting on its own behalf. A random soup of organic chemicals has no meaning and no organization. We may therefore conclude the meaning of life is organization—organization that propagates.

Section 6. Semiosis as a Special Case of Constraint as Information.

We wish next to consider the minimal physical conditions for semiosis. We shall not concern ourselves with fully human linguistic symbols, but with the semiosis of our minimal molecular autonomous agent. Consider an agent that is confronted by molecules in its environment, which constitute "yuck" or "yum." To respond to these environmental features, the agent, assumed to be bounded (Kauffman and Clayton 2006), must also have yuck and yum receptors, capable in the simplest case of "recognizing" molecules of yuck or yum, and responding appropriately by avoiding yuck and eating yum. Assume such molecular machinery exists in the agent. They of course exist in prokaryotic and eukaryotic cells. We wish to say that the agent confronting yuck or yum receives information "about" yuck or yum. This appears to constitute the minimal physical system to which semiotic information might apply. And it is worth noting that the "meaning," or semiotic content of the yuck and yum molecules is built into the propagating organization of the cell. The cell, we want to say, has embodied knowledge and know-how with respect to the proper responses to yuck and yum, which was assembled for the agent and its descendants by heritable variation and natural selection.

The existence of yuck and yum as semiotic signs is a subcase of constraint as information. How does the agent detect yuck? A concrete case would be that a yuck molecule binds a yuck receptor, constraining the receptor's motions, which in turn acts as a constraint in unleashing a cell signaling cascade leading to motion away from yuck. Further, if yuck is present below a detection threshold, it will not be detected by the agent. Hence that threshold, and the receptor itself, act as a constraints partially determining the behavior of the agent in fleeing or not fleeing.

A contentious point is whether the communication of semiotic information requires the extension of "instructional" information to some wider sense that embraces the properties of external objects, like the "yuckiness" of the yuck. Interestingly, one can construct an underlying set theoretical interpretation for this yuck and yum semantics in two equivalent ways: The first posits a set of instances, and a set of properties to which each instance is assigned. The second posits a set of instances and detectors, or classifying operators, that classify "properties" of instances. Note that in the second case, those properties need not themselves be discussed because the detectors do the job. If the second stance is taken, then detectors, "yuck" and "not yuck,"suffice and no extension beyond instructional information is required. If the second stance suffices, we want to say not only that constraints are information but also that information is constraints. We recognize that this second step is arguable and do not analyze this issue further here.

Semiotic information can not itself embody "agentness," for it has no agency; but identified agents can be observed to respect the semiotic interpretation like yuck and yum. This inspectable behaviour provides the opportunity to attribute constraint-directed behaviour to the agent organism.

Another important point in this attempt to understand propagating organization is that the semiotic behaviour can identify a source of free energy, yum in this case, from which work can be extracted and propagate in the cell. This behaviour is part of a theory that unifies matter, energy, information and propagating organization.

We end this section with the description of a final interesting feature of the yum receptor. A wide variety of molecules might bind to the yum receptor with modest affinity, hence mimic true yum molecules. So the yum receptor can be "fooled." This might allow another agent to emit a poison that mimics the yum molecule, fools the receptor, and leads to the death of the agent. So evolves the biosphere. Now ask, can a Shannon channel be "fooled?" Clearly noise can be present in the channel. Due to noise a 1 value can replace a 0 value in the constrained sense of 1 and 0 as subsets of the physical carriers of 1 and 0. But the Shannon channel cannot be fooled: "fooling" is a semantic property of detectors, hence not present in a Shannon channel. Therfore, while one might be tempted to measure the amount of semiotic information using a Shannon-like approach, the fact that semiosis in an organism can be fooled suggests that a symbol based Shannon move is inappropriate.

We conclude that semiotic information in molecular agents such as organisms is a special case of information as constraint. For semiotic information to be "about" something, and to be extracted, it appears that a constraint must be present in one or more variables that are themselves causally derived from that which the information is about.

Like the threshold level of yum needed for detection, to use the information, the extracted semoitic information must do work on some system. That work might copy the information, for example into a record, or might construct constraints on the release of energy which is further work. Here, semiotic information becomes part of propagating organization.

We comment that in standard semiotic analyses with human agents and language, there are three elements to semiotic information, namely,

1. the subject of the information or the agent being informed;
2. the object of the information or what the information is about; and
3. the possibly arbitrary, sign or symbol referring to the object.
4. with J. Monod (?) we add that allosteric chemistry allows arbitrary molecules to cause events. If we wish to call such molecules "symbols" that "refer to" "yum," the standard semiotic analysis just noted applies to molecular autonomous agents.

Section 7. Heritable Variation and Natural Selection as Assembly Processes

We have now grounded biotic information as "instructional information" or constraint, or boundary condition, that partially causes subsequent events in the unfolding of the biosphere. In this view information is not an abstraction, but is causally efficacious in the biosphere and we argue below in the unfolding of the abiotic universe. And we have grounded semiotic information as information detected about external (to the agent) features of the environment about which it learns. These semiotic cases are also cases of constraints, or boundary conditions, detecting and categorizing inputs and partially causing subsequent events. We note again that we remain neutral for the moment about whether information needs to be extended beyond instructional information for a set theory analysis of the categorization of objects.

At the level of complex molecules, as noted above, the universe has not had time to create all possible versions. For example, the universe has not had time to create all proteins length 200, by about 10 to the 67th power repititions of the history of the universe.

Consider a simple set of organic molecules and all the reactions they can collectively undergo. Call the initial set of molecules the Actual. Now among the reactions that might happen, some may lead to molecular species that are not present in the initial Actual. Call these new molecular species the Adjacent Possible. They are the molecular species that are reachable in a single reaction step from the current actual. It is of fundamental importance that the biosphere has been evolving into the Adjacent Possible for 3.8 billion years, from an initial diversity of perhaps 1000 organic molecules to trillions. The biotic world advances into the adjacent possible in terms of molecules, morphologies, species, behaviors, and technologically from pressure flaked stones; it lurks in everything from the global economy to the computer, and the millions of products in the current global economy.

Once at a level of complexity sufficiently above the atom, the universe, the biosphere, and the technosphere can never exhaust the diversity of things and events that can happen. The evolving universe and biosphere advance persistently into the adjacent possible. This means that what comes to exist at these levels of complexity is typically unique in the universe.

Now consider a heritable variation which gives rise to a new constraint, physical biotic information, that helps cause a sequence of events in a molecular agent. If that heritble variation is to the selective benefit of the agent, the new constraint, the new biotic information, will be grafted into the organism, its progeny, and the ongoing evolution of the biosphere.

It is essential to note that in the absence of heritable variation, an increase in fitness, and natural selection, this new functionality would not come to exist in the universe: but lungs and flight have come to exist. The mechanisms of heritable variation and natural selection comprise an assembly process by which propagating organization is modified in normal Darwinian adaptations and preadaptations where new functionalities arise, and these modifications are built into the ongoing evolution of the biosphere.

It is clear then, that heritable variation and natural selection are sufficient mechanisms in the biosphere to build an expanding mesh of functionalities as the biosphere invades the adjacent possible. We will ask below whether similar processes can happen in the abiotic universe.

Section 8. The Evolution of the Abiotic Expanding Universe: Propagating Organization Diversifying Sources of Constraint, Free Energy, and Coupling of Spontaneous and Non-Spontaneous Processes

We here ask whether we can find generalizations of the above analysis of information, matter, energy, consraint, work, in the biosphere, in the abiotic expanding universe.

For some time, scholars have struggled to find the union of matter, energy, and information. Cases such as Maxwell's demon, the Bekenstein bound on the entropy of a black hole, and the holographic principle, all seem to be places in physics where matter, energy, and information come together. These cases merit attention, but we leave them unanalyzed, except for this comment:

For information to be united with matter and energy, information must be part of the physical unfolding of the universe. Thus, consider Maxwell's demon. It has been shown that the demon cannot "win" with respect to the Second Law of Thermodynamics for a closed equilibrium system (Kauffman 2000). However, in a non-equilibrium setting, the demon can win by making measurements that reduce the entropy of the measured system, with respect to the demon, *faster* than the most compressed record of the measured system grows, on average, in length. Now physicists usually end their argument with a claim rather like, "Then, in principle, work could be extracted." Such a statement is inadequate for a theory that unites matter, energy, and information. What is required is that, in the non-equilibrium setting, a displacement from equilibrium that is a source of free energy must be detected by at least one measurement; a physical system able to couple to that source of free energy must have come to exist and must actually extract free energy, and must release that energy in a constrained way to carry out actual work. Thereafter, this work may propagate.

If we conceive of an abiotic physical system able to carry out these processes of measurement and work extraction in the abiotic universe, it will have to be an abiotically derived system able to perform such measurements, recording the results, and employ the record of the measurements to extract actual work. Such a system will be a case of propagating organization with boundary conditions as constraints, including measurements in the record as constraints on the behavior of the system conditional on the recorded measurements, and the constrained release of energy in work. Whether the coming into existence in the universe of such a system is plausible abiotically is certainly open to question but may be worthy of consideration. Biotically, of course, such systems abound: sources of free energy from sunlight to prey are detected and coupled to work extraction. Records of sources of free energy in the form of food are seen in ant phermone trails. The measurement of a source of free energy and extracting that free energy typically involves thresholds and other constraints or boundary conditions. For example, ants will not follow a phermone trail if it is below a detection threshold, and the boundaries of the trail are boundary conditions on the ants' motions.

These considerations suggest that we take information to be constraint or its physical equivalent, boundary conditions that partially cause events, where the coming into existence of the constraint is itself part of propagating organization. If we do so, the issue starts to clarify in a simple way. It is fully familiar in physics that one must specify the laws, particles, the initial and boundary conditions, then calculate the behavior of the system in a defined state space. Now it is common, as noted, in physics, to "put in by hand" the boundary conditions, as in the cylinder and piston case. But in the evolving biosphere, itself part of the evolving universe, and in the evolving universe as a whole, new boundary conditions come into existence and partially determine the future unfolding of the biosphere or the universe. These evolving boundary conditions and constraints are part of the propagating organization of the universe.

We consider a single, but complex case in cosmic evolution. It is well known that molecular grains are found in interstellar space. These grains aggregate up to the scale of planetessimals. Now it is also well known that the grains have surfaces with complex molecular features on which complex chemistry appears to be occurring. The grains themselves act as constraints, or boundary conditions, that confine reacting substrates, hence may catalyze reactions, some of which may be endergonic, requiring, for example, photons. In some cases, the product molecules presumably are bound to the growing grain, thereby modifying the boundary conditions afforded by the grain, which in turn modifies the chemical reactions that can occur. Furthermore, the product molecules can be novel substrates – hence novel sources of free energy – which again allow novel chemical reactions to occur. In short, the grains appear to behave as constraints that can partially guide spontaneous or non-spontaneous processes, can, in addition, link spontaneous and non-spontaneous processes, can create new constraints enabling such processes and linked processes, and can create novel sources of free energy in the form of novel substrates able to enter into new chemical reactions.

Assume the above account is roughly correct. Then the growing grains appear to be cases in which matter, energy, and continuously evolving boundary conditions and novel sources of free energy *emerge*, and condition the future evolution of the grains. The grains are at levels of complexity sufficently above atoms so that what occurs is typically unique in the universe. It seems virtually sure that no two modest size grains are molecularly identical. Here we confront a union of matter, energy, and evolving and diversifying boundary conditions linking, for example, spontanous and non-spontanous processes, and providing diversifying sources of free energy, which alter the ever diversifying structures that come to exist in the evolving expanding universe.

If this approach has merit, it appears to afford a direct union of matter, energy, and information as constraint or boundary condition.

Section 9. Human Language, Culture, Technology, Economies and the Propagation of Organization

Human language, culture, technology, and economies occupy a special place in the biosphere. They are a product of human conceptual thought (Logan 2006 & in press) and represent emergent phenomena in which their organization is propagated. They differ from the cases considered in the previous sections in that they are abstract and symbolic and not materially instantiated as such, with the exception of technology. In the case of technology, it is the concepts and organization that goes into the creation of the physical tools that propagates, not the physical tool. Technologies and economies are actually a part of culture but we have listed them explicitly because they represent vivid examples of the propagation of organization.

We might mention that Dawkins (1989) and others have characterized the propagation of language and culture as the replication of memes, and that Christiansen (1994) and Deacon (1997) have likened language to an organism that Christiansen and Ellefson (2002) have described as “a kind of beneficial parasite – a nonobligate symbiant – that confers some selective advantage onto its human hosts without whom it cannot survive.” Language and culture differ from autonomous agents as defined by Kauffman (2000) which were the subject of Sections 1-7 because they do not perform any thermodynamic work cycle. They are organisms only in the metaphoric sense that they evolve and propagate their organization. They are after all "beneficial parasites" that derive their energy from their human hosts.

Because these phenomena emerge from human autonomous agents and as such represent a still higher level of organization and complexity than materially-based biotic systems like plants, animals, and humans and because they are symbolic and conceptual and not material, they deserve and will receive a separate treatment in a future paper. The purpose of mentioning them here and now is to bring to the attention of the reader the full sweep and scope of systems that propagate their organization.

Summary

We have traveled a new path in which we have discussed Darwinian adaptations and the non-reducibility of biology to physics, the mysterious Darwinian preadaptations which seem to preclude finite prestatement and lead to evolution where the state space cannot be prestated. This brings us to serious doubts about whether Shannon or Kolmogorov information directly apply to the evolution of the biosphere, and lead to Schrödinger's aperiodic crystal and the hypothesis that information is constraints and boundary conditions, to semiotic information and records, and to the realization that, in the biosphere, it is heritable variation and natural selection that build the intricate web of propagating organization. This provides the basis for considering a new union of matter, energy, information-constraint, work, in cells. This leads to questions about the abiotic universe, where infomation as boundary conditions affords a simple means to unite matter energy and information.

We have been led to doubt that Shannon and Kolmogorov information are physically instantiated, whereas the evolving universe and biosphere are.

We seek a new theory of propagating organization, the unfolding of Kant's statement at the outset of this article. We further seek a theory of the diversifying sources of free energy and constraints that are used to couple spontaneous and non-spontaneous processes into an ever expanding diversity of processes in the biosphere and universe. We do not believe our analysis is fully adequate, but believe it is a start.

References

Christiansen, Morten. 1994. Infinite languages finite minds: Connectionism, learning and linguistic structure. Unpublished doctoral dissertation, Centre for Cognitive Studies, University of Edinburgh UK.

Christiansen and Ellefson. 2002. Linguistic adaptation without linguistic constraints: The role of sequential learning in language evolution. In A. Wray (ed), The Transition to Language. Oxford: Oxford University Press, pp 335-58.

Dawkins, Richard. 1989 edition (originally published in 1976). The Selfish Gene. Oxford: Oxford University Press.

Deacon, Terrence. 1997. Symbolic Species: The Co-evolution of the Brain and Language. New York: W. W. Norton & Co.

Hayle, Katherine. 1999. "The Condition of Virtuality" In Peter Lunenfeld (ed), The Digital Dialectic. Cambridge MA: MIT Press.

Hood, L. and D. Galas. 2003. The digital code of DNA. Nature 421: 444-448.

Kauffman, Stuart. 2000. Investigations. Oxford: Oxford University Press.

Kauffman, Stuart and Philip Clayton. 2006. Emergence, autonomous agents, and organization. Biology and Philosophy.

Logan, Robert K. 2006. The extended mind model of the origin of language and culture. In Nathalie Gontier, Jean Paul Van Bendegem and Diederik Aerts (Eds). Evolutionary epistemology, language and culture. Dordrecht: Springer.

______in press. The Extended Mind: The Origin of language and Culture. Toronto: University of Toronto Press.

Schrödinger, Erwin. 1992. What is Life? Cambridge: Cambridge University Press.

Shannon, Claude E. "A mathematical theory of communication". Bell System Technical Journal, vol. 27, pp. 379-423 and 623-656, July and October, 1948

Shiryayev, A.N. (Editor). 1993. Selected Works of A.N. Kolmogorov: Volume III: Information Theory and the Theory of Algorithms (Mathematics and its Applications). New York: Kluwer Academic Publishing.