The Dynamics of Biological Order

John D. Collier
jcollier@alum.mit.edu

The most casual observer notices that order, complexity and organization are found in biological organisms. The most striking evidence for evolution is the regular increase in these properties displayed in the fossil succession. Despite this, the core of conventional selectionist evolutionary theory avoids mentioning the order concepts altogether. Natural selection of fit traits merely requires that fitness promotes reproduction in a line of descent. The order concepts are required only to account for the promotion of reproduction, which is kept out of the core of the theory, and relegated to boundary conditions. This tactic is peculiar at best, since it deliberately relegates explanation of an obvious fact of biology to the periphery of evolutionary theory. Furthermore, as I will argue later, if the environment is the only cause of biological order, cyclic changes in environmental characteristics affecting fitness should result in cyclic changes in biological forms. This is a direct result of omitting order concepts from the core of evolutionary theory. A unified general theory of evolution must include order among its key concepts.

Rigorous, and preferably quantitative, definitions of order, complexity and organization would be helpful for formulating and comparing both specific evolutionary hypotheses and general evolutionary theories. These concepts are vague in ordinary usage, so our common notions must be replaced with more precise concepts. The precise and quantitative concept of entropy is widely thought to underlie order through its relation to information. The connection of these concepts to complexity and organization is less clear. Whatever the connection is, it isn't straight-forward: It isn't possible to have organization without order, but it is possible to have complex disorder. Despite this, the information required to specify any complex or any organized system is high. This suggests that entropy and information are not related in any simple way, contrary to the observations of both communications theorists (Shannon and Weaver, 1949) and measurement theorists (Brillouin, 1962). I believe that a theory of information coded in physical systems is required to resolve the inter-relationships of entropy, information, order, complexity and organization (Collier, 1986). Before explaining this theory, though, I would like to describe a related problem concerning biological order.

Most things in the world, if left alone, tend to disintegrate rather than organize themselves. Biological order, on the other hand, appears to originate spontaneously. This peculiarity led Schrodinger (1945) to describe life as negentropic. This leads to a paradox: Why should life be negentropic if the physical and chemical processes that underlie it are entropic? The usual response is that life depends on the existence of an entropy gradient around it, and that it gets its order at the expense of the surrounding environment. This answer is not entirely satisfactory, since although it shows the possibility of negentropic systems, it does not show why they come to exist. This requires an understanding of the dynamics of biological order.

Since the rise of the neo-Darwinian synthesis, and especially with the successes of molecular biology since 1953, the genetic code is the key to understanding the nature and evolution of biological order. Biological organisms themselves do not evolve; they live and die. Nor does the material substance of organisms evolve; it becomes a part of some organism, and then ceases to be a part of that organism. Not even the forms of organisms evolve; forms appear and disappear as the organisms that have them are born and die. What evolves is an historical sequence of forms, each causally related to its predecessor. The continuity between forms is provided by the information transmitted to successors through replication. The medium of transmission (largely, if not exclusively) is the genetic code. Fundamentally, biological evolution is the evolution of genetic encodings, the physical embodiment of the information responsible for biological organization.
The genetic code has both a material and a formal aspect, each having implications for the role of information in evolutionary theory. The material basis of the genetic code is chemical, and must obey the laws of chemistry and physics, including those of thermodynamics. It is possible to define the physical information content of a particular state of a chemical system in thermodynamic terms using the standard relationship between information and entropy developed for measurement theory by Szilard (1929) and Brillouin (1962). This technique has been widely applied in molecular biology and physiology, including in calculations of the entropies of some biological structures of molecular size and larger (Holzmuller, 1984).

Formally, the genetic code is the major means by which the form of biological organisms is transmitted both in the development of the organism and in reproduction. It provides a set of open-ended instructions similar to a computer program, which the zygote interprets to develop into a mature organism. On this computational model, these instructions are encoded information, mostly inherited from ancestors, much of which is passed on to offspring. Transmission of information by a code falls within the domain of communications theory, as originally developed by Shannon and Weaver (1949). Since both the formal and physical information increase during evolution, it appears that we need an information theory which encompasses both communications theory and measurement theory. Furthermore, this theory must account for the spontaneous increase of order, and allow an adequate definition of organization.

Previous attempts to resolve the paradox of biological order have concentrated on either the material or formal aspect of the problem, in line with the two reductionist tendencies in biology. Both tendencies can be traced to positivistic trends in the philosophy of science (Rosenberg, 1985: 21-25). The first, more commonly considered reductionist, attempts to reduce biology to the laws governing its physical and chemical components. The second, which is equally reductionist in spirit, though it is often represented as being opposed to reductionism, attempts to reduce biology to laws about the functions of biological entities, as understood in terms of adaptive value. Although I doubt that any biologist fully and consistently subscribes to either of these schools of thought, their opposition has tended to direct attention away from the advantages of synthesis. Before discussing the requirements for a synthetic approach to the puzzles about biological order, I would like to point out the difficulties, first with proposed materialist reductionist solutions, and then with formalist proposals.

The laws of thermodynamics apply to all natural systems, including biological systems. In addition, since biological systems exist under non-equilibrium conditions, we can assume that they obey the principles of non-equilibrium thermodynamics. Prigogine (1961; also Nicolis and Prigogine, 1977; Jantsch, 1981) has observed that non-equilibrium conditions permit the spontaneous development and maintenance of self-organizing systems called "dissipative structures", which maintain their organization, in spite of a general increase in entropy, by expelling matter and energy into the environment. An example is the vortices produced by heating a pot of water. If biological systems are dissipative structures, the paradox of biological order seems to be resolved. Biological order can spontaneously increase locally at the expense of increased entropy in the whole system.

There are two closely related difficulties with this resolution of the paradox. First, although it is true that biological organisms, considered as chemical systems, almost certainly fall within the scope of dissipative structures as described by Prigogine, it is not clear how the chemical order that results is physically related to the biological order in evolving genetic encodings. We don't know nearly enough about organisms as chemical systems to be able to tell whether their functioning as dissipative structures has anything to do with the creation or even maintenance of inherited order.

Second, although Prigogine's theory can account for the existence of increasingly complex chemistry in living systems (see Wicken, 1977, 1980), it is not clear how dissipative structures can account for the peculiarities of biological order, which, unlike the order of the simple dissipative structures described by Prigogine, is (a) encoded and (b) hereditary. If thermodynamics is to explain biological order, it must account for these two features. To do this at the chemical level would require an account of the origins and development of the genetic code in purely chemical terms, along with an account of the chemical origins and operations of the mechanisms of inheritance. These are currently intractable problems, and are likely to remain so for the
foreseeable future.

Since purely physico-chemical explanations of biological order are incomplete or (perhaps irremediably) intractable, functional aspects of genetic encoding must be considered. It is empirically well-established that the genetic code is the basis for hereditary transmission, via gene replication and transcription. The inheritance of characters through reproduction, then, is the primary function of the genetic code. Using communications theory (Shannon and Weaver, 1949), it is in principle possible to determine the information capacity of each segment of genetic material. If we know how proteins are encoded for transcription, how these proteins are involved in developmental pathways, and how these developmental pathways eventually lead to expressed characters, we can in principle determine the amount of structural, developmental and phenotypic information actually encoded. Even though we cannot carry this out in practice, it is often possible to make reliable comparisons of the relative complexity of various biological structures, if only on phenomenological grounds. This provides an estimate of the amount of information that must be transmitted through reproduction. Communications theory allows us to make qualitative and (in principle) quantitative measurements of biological order, but it cannot explain the origin of the order it measures. Although communications theory explicitly permits sources of information, it gives no account of these, since it deals only with information transmission. In its classic form it is restricted to cases in which information either remains constant or degrades.

In standard evolutionary theory genetic variation increases the capacity of the genetic encoding, while natural selection determines which variations are informative (functional). On this model, biological organization is all adaptive, and is attributable to the action of the forces of selection. Random genetic variants may be either useless "noise" or survivable "messages", selected for their adaptive value (Cooper, 1981; Mayr, 1982: 67-69). Biological order is a "reflection" of order in the environment in the sense that it is a consequence, via natural selection, of environmental order. Except inasmuch as it is affected by selection, the chemical level is entirely random with respect to biological order, and does not contribute to it. Although the contributions of the various characters of an organism to its fitness are the proximate causes of its reproduction rate, and might be thought to be the source of order in the organism, the degree to which a given character contributes to an organism's fitness depends ultimately entirely on the environment. The contributions of the characters of an organism to its fitness are contributions to its biological order only insofar as they are selected for.

The fundamental problem with this approach is that it requires that biological order varies solely according to the environment. If there were a circular succession of environments, the same gene might be first informative, then noise, and then again informative. This would be inconsistent with the observation that biological order steadily increases. Selection is not the source of irreversibility unless environmental conditions are irreversible. Although all environmental processes are irreversible, only some properties of these processes exert selective forces. Some properties of the environment do change irreversibly, but many evolutionarily significant conditions, such as climate and food supply, are cyclic. It seems improbable, therefore, that irreversible environmental changes are the source of evolutionary irreversibility. Although selection is an important factor in determining how biological order is expressed, it cannot be the only source of biological order, since it cannot account for the irreversibility of evolution.

The most direct response to this objection is to maintain that biological order and the irreversibility of evolution are independent of each other. Irreversibility, on this account, is due to the constant change in the genotype, so that if a need arises again for some capacity that has been lost, the capacity will be generated by a different genotype than the original (Mayr, 1982: 609). The change in the genotype is partly due to selection, and partly due to random changes (i.e., changes that are not ordered with respect to selection). Granted, the reply goes, selection does not contribute to irreversibility, but irreversibility can be explained by the random changes that do not contribute to biological order.

The answer to this response is equally direct. Explaining the irreversibility of evolution independently of biological order is impossible, since it is increases in biological order that need explanation. A more sophisticated selectionist response would be that environmental conditions affecting fitness are not linked, and cyclic changes in the environment affecting one capacity are accompanied by adaptations for other properties,
so that by the time the environment cycles around, the phenotype to be selected has changed. This mechanism can account for the non-repetition of biological forms, but it is doubtful if it can account for increasing order. If order originates entirely in the environment, and is imposed on organisms via selection for functional adaptations, the amount of order would be just that imposed by the environmental conditions. Changing environmental conditions could just as easily decrease the total order as increase it. Cyclic environmental changes would sometimes result in the cyclic creation and destruction of order, contradicting our intuitive observation that biological order increases. Biological order is not just the product of natural selection. If Darwinism, which attributes evolution to genetic variation and selection alone, is to be maintained, genetic variation must also contribute to biological order. But genetic variation depends on the physico-chemical structure of the genetic system. The (mythical) extreme functional reductionist, who maintains that all biological order can be explained by adaptive value, is wrong.

There are several constraints on genetic variation that introduce order. Although it is sometimes thought that selectionists must assume that genetic variation is random, this is not true. Darwinian selection requires only that genetic variation is not causally linked to selective forces. For selection to work, genetic variations must be physically possible and even likely, given the current genetic state. Two factors that constrain genetic variation, then, are physical likelihood and the current genotype. Since these two constraints apply throughout the whole history of a lineage, the potential for genetic variations in a population at a particular time is a result of the history of genetic variations of the lineage and the statistical laws governing variation. Both contribute to biological order. Since the evolutionary increase of biological order is unlikely to be an historical quirk, we can be pretty certain that it can be explained systematically. Although the particular history of a lineage depends on chance events, and is not amenable to systematic analysis, the statistical laws governing variation must be incorporated into an adequate general theory of biological order. The theory can safely ignore the particular chance events determining the unique character of the biological order within a particular lineage, since this is not part of what is to be explained.

Lionel Harrison (1987) has suggested that increases of biological order can be understood in terms of kinetic theory as the result of diffusion and self-catalysis. These mechanisms seem promising, and are worth further investigation. If the mechanisms of gene-kinetics can be understood entirely in terms of the underlying chemical processes, gene-kinetics can be reconciled with the laws governing these chemical processes, and the paradox of biological order can be resolved. Unfortunately, this sort of reduction isn't possible now. A different approach would be to show that, solely on its own principles, gene-kinetics spontaneously and irreversibly produces biological order. This would be to develop a statistical gene-mechanics, with a derivative "gene thermodynamics". Order would need to be suitably defined so that it would be statistically likely to increase in an arbitrary gene system. This requires the definition of macrostates of the genetic system, preferably in functional terms in order to avoid the intractabilities of genetic reduction.

It seems unlikely on information-theoretic grounds that a purely functional genetic statistical mechanics can resolve the paradox of biological order. Since biological order is measured by the information it bears, it obeys a law of information theory parallel to the second law of thermodynamics: The information transmitted by a channel cannot increase as it is transmitted through the channel; it can only decrease or remain the same. It seems that no matter how genetic information is re-ordered or varied, it cannot show an irreversible increase in content unless the additional information comes from outside the functional system (and hence from outside of the scope of a purely functional theory). This limitation applies generally to any theory of biological order that tries to deal with only the formal aspect. If we define macrostates internally in purely functional terms, the order of these states can only remain constant or degrade with time. This is true whether we take the functional system to be just the gene-reproduction system or whole eco-systems. Internalist accounts of biological order cannot explain irreversibility because they inherently deal with closed systems.

An adequate theory of biological order will give a functional definition of order and organization to avoid the problems of intractability and incompleteness, but will also be compatible with materialist accounts of order so that the system is not closed but can allow order to be incorporated from the outside. It is possible for such systems to increase their functional order by dissipating some of the information that flows through
them while incorporating a small part of it, similarly to the way Prigogine's dissipative structures self-organize at the expense of dissipating some of the energy and matter that flows through them. An increase in functional order involving the incorporation of new information in which the genetic state produced is more likely can explain a spontaneous increase in order. For this, we need a common representation of functional and statistical order.

The statistical entropy of a system is a logarithmic function of the proportion of microstates compatible with the macrostate of the system. This definition presumes a distinction between microstates and macrostates. If the statistical entropy is real, rather than being an artificial projection, both the microstates and macrostates must interact in regular ways with similar microstates and macrostates. Put differently, there must be causal laws governing the succession and interactions of both the microstates and macrostates. The key point is that the regularities we can use to identify the macrostate, though they are statistical regularities in the regularities that we use to identify the microstate, must be macroscopically detectable. For example, the lift of a kite flying is due to the pressure of individual air molecules, more of which hit one side than another. The integral structure of the kite, however, allows us to directly detect the macroscopic effect, the differential air pressure, and allows us to directly determine the macrostate of the kite without knowing its microstate.

Measurement theory is about the physical detection of the state of a system. A measurement gives the information needed to specify which state the system is in. This information equals the negentropy of the system, which is the difference between its maximal entropy (with all microstates equally probable) and its actual entropy. This difference can be measured with a device which is incapable of discriminating between the different microstates of the system, but which can distinguish the different macrostates. The amount of information obtainable with such a device is the macroscopic information of the system. The information required to specify the microstate of the system, given the macrostate, is the microinformation of the system. As macroinformation is lost through increases in entropy, the microinformation increases (Layzer, 1975). It might appear that microinformation can be used to create macroinformation, but this requires a macroscopic detector to receive the microinformation, that is, to convert microinformation into macro information. This can be done only at the cost of an equal or greater amount of macroinformation. Any device which violates this would be a "Maxwell's Demon", which violates the Second Law of Thermodynamics. There are no "free" observations. This was first explained by Szilard (1929), and generalized for physical systems by Gabor (Brillouin, 1961). The principle seems to be extendible to generalized statistical entropies, since the derivation has no dependency on non-information theoretic values (e.g. Planck's constant). Assuming this can be done, any system which has natural macrostates and microstates will behave like thermodynamical systems, for the same reasons.

In communications theory, information is carried by a code made up of elements which can combine according to various rules, which are the constraints of the system. A message is a chunk of code. Its syntactic information content, which is a measure of its capacity to carry semantic information (meaning) is a logarithmic function of its probability: The more improbable the message, the greater its information content. This is the same as for macrostates, but the underlying explanation leads to an important difference: A message in which the elements are randomly distributed has the highest syntactic information content. Just the opposite holds for macrostates, for which greater order means greater information. In this respect, message information is more like microinformation. So, in communications theory, information is equivalent to entropy, not negentropy. (Brillouin is wrong when he suggests that Shannon merely got the sign wrong.) The message entropy cannot be a physical entropy, however, since in some irreversible changes, such as the loss of information in a filter, the entropy decreases. The information of a message is a measure of its complexity, and increases as the order decreases. It is tempting to identify organization with complexity, but as I noted above, complexity without order is impossible, but organization without order is not. I think that the tendency to identify organization with syntactic information comes from its usual association with semantic information.

Brooks, Leblond and Cumming (1987) have invented a Hierarchical Information Theory (HIT), in which entropy is a measure of complexity and negentropy is a measure of order. Their work is based on Landsberg's (1984) observation that an expanding phase space allows order and entropy to increase
simultaneously (the original insight was due, I believe, to Layzer). The innovation of HIT is the recognition that order is hierarchically arranged according to degrees of redundancy, where the degree of a redundancy is the number of elements required to exhibit the redundant pattern. The same elements can simultaneously participate in order of varying degrees, as, for example, when the same letters make up words, sentences and paragraphs, which involve different degrees of grammatical order. Macrostates and microstates can be defined for a particular level by taking the macrostates to be determined by the order at that level, and the microstates to be all variants at level one which are compatible with that order. HIT comes to nothing, however, unless the orders of various degrees deviate from what would be expected if the total redundancy were evenly distributed, in which case we could more simply describe the system in terms of its total entropy. If there are laws which work on groups of elements of particular sizes, however, we would expect an uneven distribution of order. Then the question of the correlation or lack of correlation of order of different degrees (whether they share the same elements) becomes interesting. Given the operation of laws, a correlation between orders of different degrees requires a correlation of the laws which produce the order. I postulate, therefore, that organization is measured by the correlation of deviant or unexpected order of different degrees or ranges of degrees. Organization would then involve order directly, and complexity because of order at the lower degree which is not expressed at the higher degree. In a sense, HIT gives us too much, since only some degrees of order result directly from the operation of natural laws; other order is either carried over from lower levels, or is coincidental.

HIT is a useful system for representing hierarchical information, but it needs to be supplemented with a naturalistic account of information coding. This is provided by the theory of physical information systems (Collier, 1986). Messages are abstract entities, defined without any reference to their physical embodiment. They are also usually thought of as linear. The analogue to messages in physical information systems are arrays, which do not have these limitations. Arrays are composed of relatively stable physical elements which combine according to specific (generally statistical) laws. The laws of combination for arrays are equivalent to the constraints on the combination of the elements of messages. The major problem in applying the concept of macrostates in classical communications theory is that potential message length is not restricted. Any finite message can be completed in an infinite number of ways, so if a message is a macrostate, a complete specification of its microstate is impossible in principle. Real systems, however, are limited in size, even if this limit can be defined only statistically. For any given array, then, we can put a rough limit on the ways it can be completed. The more clearly this limit can be defined, the more well-defined the statistical entropy of the array relative to its possible completions. In a biological systems, the main basis of order is the DNA, which comes in chromosomes of relatively fixed length. DNA also combines according to well-defined rules, and interacts systematically with higher levels of organization such as cells, organisms and species. Biological systems, therefore, are ideal for analysis by hierarchical physical information theory.

Brooks and Wiley (1986) propose that evolution involves increases in entropy and information in an expanding phase space at various levels in the hierarchy of biological information. One fundamental mechanism of the theory is the expression of information from lower levels at higher levels. This expands the phase space at the higher level, increasing both entropy and information, complexity and order (Brooks, Leblond and Cumming, 1984). The correlation between the lower levels and higher levels implicit in this expression increases organization. On this model (see Collier 1986 for details), the species is a macrosystem forming an array bound together by reproduction and history. Its microstates are phenotypes, whose microstates in turn are genetic, and the genetic information is coded by molecules of DNA. Genetic information which is expressed phenotypically leads to variation in the species, increasing complexity. Some of this phenotypic information can become involved in reproduction, usually with the help of natural selection. This creates new order at the species level. Order at the species level can break down when reproductive cohesion fails, resulting in speciation.

This short description doesn't do the Brooks-Wiley theory justice, but suggests its general form. Evolution is driven by the entropy produced by the capture of randomly varying molecular information through its involvement in reproduction. This capture is possible because it increases the entropy of the species by
increasing its phase space. If this were not so, life would behave like a "Maxwell's Demon", and would be negentropic, as Schrödinger suggested. Once reproduction is established, this process can "bootstrap" itself to produce ever greater order, complexity and organization. The resources of hierarchical physical information systems theory allow any physical system which can be naturally described as bearing information to be described in terms of macrostates, microstates, entropy and information. These concepts can be used to explain order, complexity and organization. A natural consequence of this explanation is that order, complexity and organization will increase together spontaneously.

The Brooks-Wiley theory is a unified theory of evolution which accounts for the presence and evolution of biological order. It also implies some novel and testable hypotheses. I will mention only two. The first prediction is that evolution can occur independently of environmental selection. This prediction can be checked by determining whether all the variation between related populations in a given area can be accounted for by natural selection. There are already reasons to doubt this. The second prediction is that order will not be distributed randomly among degrees of redundancy, or concentrated at lower levels, but that jumps will occur at higher levels. Preliminary studies (Brooks, Cumming and Leblond, 1987) suggest this is so. If doubts remain about the applicability of the entropy concept outside the narrow domain of chemical thermodynamics, the explanatory power and predictive fruitfulness of the application of the concept to evolutionary biology should allay them.

References