CHARLES H. BENNETT
IBM Research, Yorktown Heights, NY 10598, April 1985

Dissipation, Information, Computational Complexity and the Definition of Organization

I address two questions belonging to an interdisciplinary area between statistical mechanics and the theory of computation:

1. What is the proper measure of intrinsic complexity to apply to states of a physical system?
2. What role does thermodynamic irreversibility play in enabling systems to evolve spontaneously toward states of high complexity?

I. INTRODUCTION

A fundamental problem for statistical mechanics is to explain why dissipative systems (those in which entropy is continually being produced and removed to the surroundings) tend to undergo "self-organization," a spontaneous increase of structural complexity, of which the most extreme example is the origin and evolution of life. The converse principle, namely that nothing very interesting is likely to happen in a system at thermal equilibrium, is reflected in the term "heat death." In the modern world view, thermodynamic driving forces, such as the temperature difference between the hot sun and the cold night sky, have taken over one of the functions

ISBN 0-201-15686-5 David Pines, Ed.

of God: they make matter transcend its clod-like nature and behave instead in dramatic and unforseen ways, for example molding itself into thunderstorms, people, and umbrellas.

The notion that dissipation begets self-organization has remained informal, and not susceptible to rigorous proof or refutation, largely through lack of an adequate mathematical definition of organization. Section II, after reviewing alternative definitions, proposes that organization be defined as "logical depth," a notion based on algorithmic information and computational time complexity. Informally, logical depth is the number of steps in the deductive or causal path connecting a thing with its plausible origin. The theory of computation is invoked to formalize this notion as the time required by a universal computer to compute the object in question from a program that could not itself have been computed from a more concise program.

Having settled on a definition of organization, we address briefly in section III the problem of characterizing the conditions (in particular, thermodynamic irreversibility) under which physical systems evolve toward states of high organization. We do not solve this problem, but rather suggest that it can be reduced to several other problems, some of which can already be regarded as solved, some of which are promising areas of research, and some of which are well-known unsolved problems in mathematics (notably the P=PSPACE question).

II. THE PROBLEM OF DEFINING ORGANIZATION

Just what is it that distinguishes an "organized" or "complex" structure like the human body from, say, a crystal or a gas? Candidates for a definition of organization can be divided into those based on function and those based on structure.

A. FUNCTIONAL DEFINITIONS

Living organisms are noted for their capacity for complex function in an appropriate environment, in particular the ability to grow, metabolize, reproduce, adapt, and mutate. While this functional characterization may be a good way to define "life," in distinction to nonliving phenomena that possess some but not all of life's attributes (e.g., a crystal's trivial growth; a flame's metabolism), it is not really a satisfactory way to define organization. We should still like to be able to call organized such functionally inert objects as a frozen human body, a printout of the human genome, or a car with a dead battery. In other words, what we need is not a definition of life or organism (probably inherently fuzzy concepts anyway), but rather a definition for the kind of structural complexity that in our world is chiefly found in living organisms and their artifacts, a kind that can be produced to a lesser degree by laboratory experiments in "self-organization," but which is absent from such structurally trivial objects as gases and crystals.

Another functional characterization of complexity, more mathematical in flavor than the lifelike properties mentioned above, is as the capacity for universal computation. A computationally universal system is one that can be programmed, through its initial conditions, to simulate any digital computation. For example, the computational universality of the well-known deterministic cellular automaton of Conway called the "game of life" implies that one can find an initial configuration that will evolve so as to turn a certain site on if and only if white has a winning strategy at chess, another initial configuration that will do so if and only if the millionth decimal digit of pi is a 7, and so on. On a grander scale, one can in principle find initial conditions enabling the Conway automaton to simulate any physical or chemical process that can be digitally simulated, even presumably the geological and biological evolution of the earth.

The property of computational universality was originally demonstrated for irreversible, noiseless systems such as Turing machines and deterministic cellular automata having little resemblance to the systems ordinarily studied in mechanics and statistical mechanics. Later, some reversible, deterministic systems (e.g., the hard sphere gas [Fredkin-Toffoli, 1982] with appropriate initial and boundary conditions, and Margolus' billiard ball cellular automaton [Margolus, 1984] which models this gas) have been shown to be computationally universal. Very recently [Gacs, 1983; Gacs-Reif, 1985], certain irreversible, noisy systems (probabilistic cellular automata in 1 and 3 dimensions with all local transition probabilities positive) have been shown to be universal. Computational universality, therefore, now appears to be a property that realistic physical systems can have; moreover, if a physical system does have that property, it is by definition capable of behavior as complex as any that can be digitally simulated.

However, computational universality is an unsuitable complexity measure for our purposes because it is a functional property of systems rather than a structural property of states. In other words, it does not distinguish between a system merely capable of complex behavior and one in which the complex behavior has actually occurred. The complexity measure we will ultimately advocate, called logical depth, is closely related to the notion of universal computation, but it allows complexity to increase as it intuitively should in the course of a "self-organizing" system's time development.

B. THERMODYNAMIC POTENTIALS

In spite of the well-known ability of dissipative systems to lower their entropy at the expense of their surroundings, flouting the spirit of the second law while they obey its letter, organization cannot be directly identified with thermodynamic potentials such as entropy or free energy: the human body is intermediate in entropy between a crystal and a gas; and a bottle of sterile nutrient solution has higher free energy, but lower subjective organization, than the bacterial culture it would turn into if inocculated with a single bacterium.

This difference in free energy means that, even without the seed bacterium, the transformation from nutrients to bacteria (albeit an improbable case of spontaneous biogenesis) is still vastly more improbable case of spontaneous biogenesis) is still vastly more probable than the reverse transformation, from bacteria to sterile, high free-energy nutrients. The situation is analogous to the crystallization of a long-lived supersaturated solution: although crystallization without the catalytic assistance of a seed crystal may be so slow as to be unobservable in practice, it is not thermodynamically forbidden, and is, in fact, overwhelmingly more probable than the reverse process.

Subjective organization seems to obey a "slow growth law" which states that, except by a lucky accident, organization cannot increase quickly in any deterministic or probabilistic process, but it can increase slowly. It is this law which forbids sterile nutrient from turning into bacteria in the laboratory, but allows a similar transformation over geological time. If the slow growth law is to be obeyed, the rapid multiplication of bacteria after inocculation must not represent much increase in organization, beyond that already present in the seed bacterium. This, in turn, means that subjective organization is not additive: 1 bacterium contains much more organization that 0 bacteria, but 2 sibling bacteria contain about the same amount as 1.

C. INFORMATION CONTENT

The apparent non-additivity of "organization" suggest another definition for it, namely as information content, an object's information content being the number of bits required to specify it uniquely. Clearly, two large message-like objects (e.g., DNA molecules), if they happen to be identical, do not together contain significantly more information than one alone.

This subsection will review various definitions of information, especially the algorithmic definition implied by the phrase "number of bits necessary to specify a structure uniquely." However, it should be pointed out that information in this sense, like entropy, leads to absurd conclusions when used as the measure of subjective organization: just as the human body is intermediate in entropy between a crystal and a gas, so the human genome is intermediate in information between a totally redundant sequence, e.g., AAAAA..., of near zero information content and a purely random sequence of maximal information content. Although information itself is a poor measure of organization, it will be discussed at some length because it underlies two of the more adequate organization measures to be discussed later, vis. mutual information and logical depth.

There is some uncertainty as to how the "information content" of biological molecules ought to be defined. The easiest definition is simply as the information *capacity* of the molecule, e.g., 2N bits for a DNA molecule of N nucleotides. This definition is not very useful, since it assigns all sequences of a given length the same information content.

In the classical formulation of Shannon, information is an essentially statistical property. The information content in bits of a message is defined as the negative base-2 logarithm of its probability of having been emitted by some source, and it is improper to treat information content as if it were a function of the message itself, without specifying the probability. This is rather awkward in a biological context, where one is frequently faced with a bare message, e.g., a DNA sequence, without any indication of its probability. The information capacity is equivalent to assuming a uniform probability distribution over all sequences. It would be more informative to define the information content of a sequence x as its -log probability in some physically specified distribution, such as an (equilibrium or nonequilibrium) statistical mechanical ensemble. However, this approach departs from the goal of making the definition of organization intrinsic to the sequence.

A third approach to defining information is as the number of bits necessary to uniquely describe an object in some absolute sense, rather than with respect to a particular probability distribution. This approach has been put on a firm mathematical basis by regarding the digital object x as the output of a universal computer (e.g., a universal Turing machine), and defining its algorithmic information content $H(x)$ as the number of bits in its "minimal algorithmic description" $x*$, where $x*$ is the smallest binary input string that causes the universal computer to produce exactly x as its output. Clearly this definition depends on the choice of universal computer, but this arbitrariness leads only to an additive $O(1)$ uncertainty (typically ± a few thousand bits) in the value of $H(x)$, because of the ability of universal machines to simulate one another. Algorithmic information theory also allows randomness to be defined for individual strings: a string is called "algorithmically random" if it is incompressible, i.e., if its minimal description is about the same size as the string itself. Algorithmic information is discussed further in the introductory article by Chaitin [1975], and in review articles by Zvonkin and Levin [1970] and Chaitin [1977].

The advantage of using a universal computer to regenerate the message is that, for sufficiently long messages, it subsumes all other more specialized schemes of effective description and data compression, e.g., the use of a dictionary of abbreviated encodings for frequently occurring subsequences. Any non-universal scheme of data compression fails to compress some sequences of obviously low information content. For example, the sequence consisting of the first million digits of pi, though it admits a concise algorithmic description, probably cannot be significantly compressed by abbreviating frequent sequences.

As noted above, information per se does not provide a good measure of organization, inasmuch as messages of maximal information content, such as those produced by coin tossing, are among the least organized subjectively. Typical organized objects, on the other hand, precisely because they are partially constrained and determined by the need to encode coherent function or meaning, contains less information than random sequences of the same length; and this information reflects not their organization, but their residual randomness.

For example, the information content of a genome, as defined above, represents the extent to which it is underdetermined by the constraint of viability. The existence of noncoding DNA, and the several percent differences between proteins performing apparently identical functions in different species, make it clear that a sizable fraction of the genetic coding capacity is given over to transmitting such "frozen accidents," evolutionary choices that might just as well have been made otherwise.

D. MUTUAL INFORMATION AND LONG-RANGE ORDER

A better way of applying information theory to the definition of organization is suggested by the nonadditivity of subjective organization. Subjectively organized objects generally have the property that their parts are correlated: two parts taken together typically require fewer bits to describe than the same two parts taken separately. This difference, the *mutual information* between the parts, is the algorithmic counterpart of the non-additivity of statistical or thermodynamic entropy between the two parts. In many contexts, e.g., communication through a noisy channel, the mutual information between a message and something else can be viewed as the "meaningful" part of the message's information, the rest being meaningless information or "noise."

A body is said to have long-range order if even arbitrarily remote parts of it are correlated. However, crystals have long-range order but are not subjectively very complex. Organization has more to do with the *amount* of long-range correlation, i.e., the number of bits of mutual information between remote parts of the body. Although we will ultimately recommend a different organization measure (logical depth), remote mutual information merits some discussion, because it is characteristically formed by nonequilibrium processes, and can apparently be present only in small amounts at thermal equilibrium. Notions similar to mutual information have been introduced in many discussions of biological organization, but often without clearly distinguishing among gross information content (i.e., accidental or arbitrary aspects of the object as a whole), mutual information (amount of correlation between parts that individually are accidental and arbitrary), and determined, non-accidental aspects of the object as a whole which, as argued above, are not information at all, but rather a form of redundancy.

If two cells are taken from opposite ends of a multicellular organism, they will have a large amount of mutual information, if for no other reason than the presence in each cell of the same genome with the same load of frozen accidents. As indicated earlier, it is reasonably certain that at least several percent of the coding capacity of natural genomes is used to transmit frozen accidents, and, hence, that the mutual information between parts of a higher organism is at least in the hundred megabit range. More generally, mutual information exists between remote parts of an organism (or a genome, or a book) because the parts contain evidence of a common, somewhat accidental history, and because they must function together in a way that imposes correlations between the parts without strictly determining the structure of any one part. An attractive feature of remote mutual information for physical systems is that it tends to a finite limit as the fineness of coarse graining is increased, unlike simple information or entropy in a classical system.

Since mutual information arises when an accident occurring in one place is replicated or propagated to another remote place, its creation is an almost unavoidable side effect of reproduction in a probabilistic environment. Another obvious connection between mutual information and biology is the growth of mutual information between an organism and its environment when the organism adapts or learns.

Further support for remote mutual information as an organization measure comes from the fact that systems stable at thermal equilibrium, even those with long-range order, exhibit much less of it than nonequilibrium systems. Correlations in systems at equilibrium are generally of two kinds: short-range correlations involving a large number of bits of information (e.g., the frozen-in correlations between adjacent lattice planes of an ice crystal, or the instantaneous correlations between atomic positions in adjacent regions of any solid or liquid), and long-range correlations involving only a few bits of information. These latter include correlations associated with conserved quantities in a canonical or microcanonical ensemble (e.g., if one half of a gas cylinder contains more than half the atoms, the other half will contain fewer than half of the atoms) and correlations associated with order parameters such as magnetization and crystal lattice orientation. In either case, the amount of mutual information due to long-range correlations is small: for example, in a gas of 10^{23} atoms, conservation of the number of atoms causes the entropy of the whole to be about $\log\sqrt{10^{23}} \approx 39$ bits less than the sum of the entropies of its halves. It may at first seem that a real-valued order parameter, such as phase or orientation of a crystal lattice, already represents an infinite amount of information; however, in an N-atom crystal, owing to thermal and zero-point fluctuations, the instantaneous microstate of the entire crystal suffices to determine such order parameters only to about log N bits precision; and, hence, the mutual information between remote regions of a macroscopic crystal amounts to only a few dozen bits.

Unfortunately, some subjectively not-very-organized objects also contain large amounts of remote mutual information. For example, consider an igneous rock or other polycrystalline solid formed under nonequilibrium conditions. Such solids, though not subjectively very “organized,” typically contain extended crystal defects such as dislocations and grain boundaries, which presumably carry many bits of information forward from the earlier-crystallized to the later-crystallized portions of the specimen, thus giving rise to the correlated frozen accidents that constitute mutual information. On a larger scale, terrestrial and planetary geological processes create large amounts of mutual information in the form of complementary fracture surfaces on widely separated rock fragments. Mutual information does not obey the slow growth law, since an ordinary piece of glass, after a few minutes of hammering and stirring, would be transformed into a three-dimensional jigsaw puzzle with more of it than any genome or book. Even larger amounts of mutual information could be produced by synthesizing a few grams of random, biologically meaningless DNA molecules, replicating them enzymatically, and stirring the resulting mixture to produce a sort of jigsaw-puzzle soup. Two spoonfuls of this soup would have macroscopically less than twice the entropy of one spoonful. In all these examples, the mutual information is formed by nonequilibrium processes and would decay if the system were allowed to approach a state of true thermal equilibrium, e.g., by annealing of the separated fracture surfaces. Remote mutual information is somewhat unsatisfying as a measure of organization because it depends on accidents, assigning low organization to some objects (such as the binary expansion of pi) which seem organized though they lack accidents, and high organization to other objects whose correlated accidents are of a rather trivial sort (random palindromes, broken glass).

E. SELF-SIMILARITY

A conspicuous feature of many nontrivial objects in nature and mathematics is the possession of a fractal or self-similar structure, in which a part of the object is identical to, or is described by the same statistics as, an appropriately scaled image of the whole. I feel that this often beautiful property is too specialized to be an intuitively satisfactory criterion of organization because it is absent from some intuitively organized objects, such as the decimal expansion of pi, and because, on the other hand, self-similar structures can be produced quickly, e.g., by deterministic cellular automata, in violation of the slow growth law. Even so, the frequent association of self-similarity with other forms of organization deserves comment. In some cases, self-similarity is a side-effect of computational universality, because a universal computer's ability to simulate other computers gives it, in particular, the ability to simulate itself. This makes the behavior of the computer on a subset of its input space (e.g., all inputs beginning with some prefix p that tells the computer to simulate itself) replicate its behavior on the whole input space.

F. LOGICAL DEPTH

The problem of defining organization is akin to that of defining the value of a message, as opposed to its information content. A typical sequence of coin tosses has high information content, but little message value; an ephemeris, giving the positions of the moon and planets every day for a hundred years, has no more information than the equations of motion and initial conditions from which it was calculated, but saves its owner the effort of recalculating these positions. The value of a message, thus, appears to reside not in its information (its absolutely unpredicatble parts), nor in its obvious redundancy (verbatim repetitions, unequal digit frequencies), but rather in what might be called its buried redundance—parts predictable only with difficulty, things the receiver could in principle have figured out without being told, but only at considerable cost in money, time or computation. In other words, the value of a message is the amount of mathematical or other work plausibly done by its originator, which its receiver is saved from having to repeat.

Of course, the receiver of a message does not know exactly how it originated; it might even have been produced by coin tossing. However, the receiver of an obviously non-random message, such as the first million bits of pi, would reject this "null" hypothesis on the grounds that it entails nearly a million bits worth of ad-hoc assumptions, and would favor an alternative hypothesis that the message originated from some mechanism for computing pi. The plausible work involved in creating a message, then, is the amount of work required to derive it from a hypothetical cause involving no unnecessary ad-hoc assumptions.

These ideas may be formalized in terms of algorithmic information theory: a message's most plausible cause is identified with its minimal algorithmic description, and its "logical depth," or plausible content of mathematical work, is (roughly speaking) identified with time required to compute the message from this minimal description. Formulating an adequately robust quantitative definition of depth is not quite this simple and, in particular, requires a properly weighted consideration of other descriptions besides the minimal one. When these refinements are introduced [cf Appendix], one obtains a definition of depth that is machine independent, and obeys the slow growth law, to within a polynomial depending on the universal machine. The essential idea remains that a deep object is one that is implausible except as the result of a long computation.

It is a common observation that the more concisely a message is encoded (e.g., to speed its transmission through a channel of limited bandwidth), the more random it looks and the harder it is to decode. This tendency is carried to its extreme in a message's minimal description, which looks almost completely random (if $x*$ had any significant regularity, that regularity could be exploited to encode the message still more concisely) and which, for a nontrivial (deep) message, requires as much work to decode as plausibly went into producing the message in the first place. The minimal description $x*$, thus, has all the information of the original message x, but none of its value.

Returning to the realm of physical phenomena, we advocate identifying subjective organization or complexity with logical depth, in other words, with the length of the logical chain connecting a phenomenon with a plausible hypothesis explaining it. The use of a universal computer frees the notion of depth from excessive dependence on particular physical processes (e.g., prebiotic chemistry) and allows an object to be called deep only if there is no shortcut path, physical or non-physical, to reconstruct it from a concise description. An object's logical depth may, therefore, be less than its chronological age. For example, old rocks typically contain physical evidence (e.g., isotope ratios) of the time elapsed since their solidification, but would not be called deep if the aging process could be recapitulated quickly in a computer simulation. Intuitively, this means that the rocks' plausible history, though long in time, was rather uneventful, and, therefore, does not deserve to be called long in a logical sense.

The relevance of logical depth to physical self-organization depends on the assumption that the time development of physical systems can be efficiently simulated by digital computation. This is a rather delicate question; if by simulation one means an exact integration of differential equations of motion, then no finite number of digital operations could simulate even one second of physical time development. Even when simulation is defined less restrictively (roughly, as an effective uniformly convergent approximation by rational numbers), Myhill [1971] showed that there is a computable differentiable function with a noncomputable solution. On the other hand, it remains plausible that realistic physical systems, which are subject throughout their time development to finite random influences (e.g., thermal and gravitational radiation) from an uncontrolled environment, can be efficiently approximated by digital simulation to within the errors induced by these influences. The evidence supporting this thesis is of the same sort, and as strong as, that supporting the empirically very successful master equation [van Kampen, 1962], which approximates the time development of a statistical mechanical system as a sequence of probabilistic transitions among its coarse-grained microstates.

Accepting the master equation viewpoint, the natural model of physical time development, at least in a system with short-ranged forces, would be a three-dimensional probabilistic cellular automaton. Such automata can be simulated in approximately linear time by a universal three-dimensional cellular automaton each of whose sites is equipped with a coin-toss mechanism; hence, time on such a universal automaton might be the most appropriate dynamic resource in terms of which to define depth. Usually we will be less specific, since other reasonable machine models (e.g., the universal Turing machines in terms of which algorithmic information theory is usually developed) can simulate probabilistic cellular automata, and one another, in polynomial time. We will assume conservatively that any t seconds in the time development of a realistic physical system with N degress of freedom can be simulated by probabilistic computation using time bounded by a polynomial in Nt.

Although time (machine cycles) is the complexity measure closest to the intuitive notion of computation work, memory (also called space or tape) is also important because it corresponds to a statistical mechanical system's number of particles or degrees of freedom. The maximum relevant time for a system with N degrees of freedom is of order $2^{O(N)}$, the Poincaré recurrence time; and the deepest state such a system could relax to would be one requiring time $2^{O(N)}$, but only memory N, to compute from a concise description.

Unfortunately, it is not known that any space-bounded physical system or computer can indeed produce objects of such great depth (exponential in N). This uncertainty stems from the famous open P=?PSPACE question in computational complexity theory, i.e., from the fact that it is not known whether there exist computable functions requiring exponentially more time to compute than space. In other words, though most complexity theorists suspect the contrary, it is possible that the outcome of every exponentially long computation or physical time evolution in a space-bounded system can be predicted or anticipated by a more efficient algorithm using only polynomial time.

A widely held contrary view among complexity theorists today, considerably stronger than the mere belief that P is not equal to PSPACE, is that there are "cryptographically strong" pseudorandom number generators [Blum-Micali, 1984; Levin, 1985], whose successive outputs, on an N-bit seed, satisfy all polynomial time (in N) tests of randomness. The existence of such generators implies that space-bounded universal computers, and, therefore, any physical systems that mimic such computers, can after all produce exponentially deep outputs.

If, on the other hand, it turns out that P=PSPACE, then exponentially deep N-bit strings can still be produced (by well-known "diagonal" method, the gist of which is to generate a complete list of all shallow N-bit strings and then output the first N-bit string not on the list), but the computations leading to these deep objects will require more than polynomial space during their intermediate stages.

It is worth noting that neither algorithmic information nor depth is an effectively computable property. This limitation follows from the most basic result of computability theory, the unsolvability of the halting problem, and reflects the fact that although we can prove a string nonrandom (by exhibiting a small program to compute it), we can not, in general, prove it random. A string that seems shallow and random might, in fact, be the output of some very slow-running, small program, which ultimately halts, but whose halting we have no means of predicting. This open-endedness is also a feature of the scientific method: a phenomenon that seems to occur randomly (e.g., pregnancy) may later turn out to have a cause so remote or unexpected as to have been overlooked at first. In other words, if the cause of a phenomenon is unknown, we can never be sure that we are not underestimating its depth and overestimating its randomness.

The uncomputability of depth is no hindrance in the present theoretical setting where we assume a known cause (e.g., a physical system's initial conditions and equations of motion) and try to prove theorems about the depth of its typical effects. Here, it is usually possible to set an upper bound on the depth of the effect by first showing that the system can be simulated by a universal computer within a time t and then invoking the slow growth rule to argue that such a computation, deterministic or probabilistic, is unlikely to have produced a result much deeper than t. On the other hand, proving lower bounds for depth, e.g., proving that a given deterministic or probabilistic cause certainly or probably leads to a deep effect, though always possible in principle, is more difficult, because it requires showing that no equally simple cause could have produced the same effect more quickly.

III. TOWARDS AN UNDERSTANDING OF THE NECESSARY AND SUFFICIENT CONDITIONS FOR SELF-ORGANIZATION

We have already pointed out a mathematical requirement, namely the conjectured inequality of the complexity classes P and PSPACE, necessary for a finite model system to evolve to a state of depth comparable to its Poincare time. In this section, we mention recent results in computation theory and statistical mechanics which may soon leads to a comprehensive understanding of other conditions necessary and sufficient for model systems to self-organize, i.e., to evolve deterministically or with high probability to a state's deep compared to the system's initial condition.

It is clear that universal computation, and, hence, self-organization, can occur *without dissipation* in reversible deterministic systems such as Fredkin and Toffoli's "billiard ball model" [1982], which consists of classical hard spheres moving on a plane with fixed obstacles (without loss of generality the array of obstacles may be taken to be spatially periodic); or in Margolus' billiard ball cellular automaton [1984] which discretely simulates this model. In these models, the initial condition must be low-entropy, because a reversible system cannot decrease its own entropy (the continuous billiard ball model, because of the dynamical instability of its collisions, in fact requires an initial condition with infinite negative entropy relative to the random hard sphere gas). Moreover, if the system is to preform a nontrivial computation, the initial condition must lack translational symmetry, because a deterministic system cannot break its own symmetries. It would suffice for the initial condition to be periodic except at a single site, which would serve as the origin for a depth-producing computation.

The systems just considered are noiseless. As indicated earlier, it is more realistic to imagine that a physical system is subject to environmental noise, and to treat its motion as random walk, rather than a deterministic trajectory, on the relevant discrete or continuous state space.

In general, such noisy systems require at least some dissipation to enable them to correct their errors and engage in a purposeful computation; the amount of dissipation depends on the noise's intensity and especially on its pervasiveness, i.e., on whether it is considered to affect all, or only some aspects of the system's structure and operation. At the low end of the pervasiveness spectrum are systems such as the clockwork computer of Bennett [1982], in which the noise causes only transitions forward and backward along the intended path of computation, not transitions from one computation into another, or transitions that degrade the structure of the hardware itself. In such systems, all errors are recoverable and the required dissipation tends to zero in the limit of zero speed. More pervasive noise can be found in the situation of error-correcting codes, where some unrecoverable errors occur but the decoding apparatus itself is considered perfectly reliable; and in proofreading enzyme systems [cf Bennett, 1979], where the decoding apparatus is unreliable but still structurally stable. These systems require finite dissipation even in the limit of zero speed. Von Neumann's [1952] classic construction of a reliable computer from unreliable parts is also of this sort: all gates were considered unreliable, but the wires connecting them were considered reliable and their complex interconnection pattern structurally stable. Only recently has decisive progress been made in understanding systems at the high end of the pervasiveness spectrum, in particular, "noisy" cellular automata (henceforth NCA) in which all local transition probabilities are strictly positive. For such an automaton, any two finitely differing configurations are mutually accessible.

An NCA may be synchronous or asynchronous, reversible or irreversible. The former distinction (i.e., between a random walk occurring in discrete time or continuous time) appears to have little qualitative effect on the computing powers of the automata, but the latter distinction is of major importance. In particular, irreversible NCA can function as reliable universal computers [Gacs, 1983; Gacs-Reif, 1985], and can do so robustly despite arbitrary small perturbations of their transition probabilities; while reversible NCA, for almost all choices of the transition probabilities, are ergodic, relaxing to a structurally simple state (the thermodynamic phase of lowest free energy) independent of the initial condition. Irreversibility enables NCA to be robustly nonergodic essentially by protecting them from the nucleation and growth of a unique phase of lowest free energy [Toom, 1980; Domany-Kinzel, 1984; Bennett-Grinstein, 1985].

(An NCA is considered reversible or nondissipative if its matrix of transition probabilities is of the "miscroscopically reversible" form DSD^{-1}, where D is diagonal and S symmetric. In that case, a movie of the system at equilibrium would look the same shown forwards as backwards and the stationary distribution can be represented (exactly for asynchronous automata, approximately for synchronous) as the Boltzmann exponential of a locally additive potential. On the other hand, if the local transition probabilities are not microscopically reversible, the stationary macrostate is dissipative (corresponding physically to a system whose environment continually removes entropy from it), a movie of the system would not look the same forwards as backwards, and the distribution of microstates, in general, cannot be approximated by the exponential of any locally additive potential. Asynchronous reversible NCA, otherwise known as generalized kinetic Ising models, are widely studied in statistical mechanics.)

The computationally universal NCA of Gacs and Gacs-Reif are still somewhat unsatisfactory because they require special initial conditions to behave in a nontrivial manner. A truly convincing case of self-organization would be an NCA with generic transition probabilities that would initiate a depth-producing computation from generic initial conditions (e.g., a random soup). Such an automaton has not been found, though Gacs believes it can be. If it is found, it will lend support to the philosophical doctrine that the observed complexity of our world represents an intrinsic propensity of nature, rather than an improbable accident requiring special initial conditions or special laws of nature, which we observe only because this same complexity is a necessary condition for our own existence.

APPENDIX: MATHEMATICAL CHARACTERIZATION OF DEPTH

Two rather different kinds of computing resources have been considered in the theory of computational complexity: static or definitional resources such as program size, and dynamic resources such as time and memory. Algorithmic information theory allows a static complexity or information content to be defined both for finite and for infinite objects, as the size in bits of the smallest program to computer the object on a standard universal computer. This minimal program has long been regarded as analogous to the most economical scientific theory able to explain a given body of experimental data. Dynamic complexity, on the other hand, is usually considered meaningful only for infinite objects such as functions or sets, since a finite object can always be computed or recognized in very little time by means of a table look-up or print program, which includes a verbatim copy of the object as part of the program.

In view of the philosophical significance of the minimal program, it would be natural to associate with each finite object the cost in dynamic resources of reconstructing it from its minimal program. A "deep" or dynamically complex object would then be one whose most plausible origin, via an effective process, entails a lengthy computation. (It should be emphasized that just as the plausibility of a scientific theory depends on the economy of its assumptions, not on the length of the deductive path connecting them with observed phenomena, so the plausibility of the minimal program, as an effective "explanation" of its output, does not depend on its cost of execution.) A qualitative definition of depth is quoted by Chaitin [1977], and related notions have been independently introduced by Adleman [1979] ("potential") and Levin [Levin and V'jugin, 1977] ("incomplete sequence").

In order for depth to be a useful concept, it ought to be reasonably machine-independent, as well as being stable in the sense that a trivial computation ought not to be able to produce a deep object from a shallow one. In order to achieve these ends, it is ncessary to define depth a little more subtly, introducing a significance parameter that takes account of the realtive plausibility of all programs that yield the given object as output, not merely the minimal program. Several slightly different definitions of depth are considered below; the one finally adopted calls an object "d-deep with b bits significance" if all self-delimiting programs to compute it in time d are algorithmically compressible (expressible as the output of programs smaller than themselves) by at least b bits. Intuitively this implies that the "null" hypothesis, that the object originated by an effective process of fewer than d steps, is less plausible than a sequence of coin tosses beginning with b consecutive tails.

The difficulty with defining depth as simply the run time of the minimal program arises in cases where the minimal program is only a few bits smaller than some much faster program, such as a print program, to compute the same output x. In this case, slight changes in x may induce arbitrarily large changes in the run time of the minimal program, by changing which of the two competing programs is minimal. This instability emphasizes the essential role of the quantity of buried redundancy, not as a measure of depth, but as a certifier of depth. In terms of the philosophy-of-science metaphor, an object whose minimal program is only a few bits smaller than its print program is like an observation that points to a nontrivial hypothesis, but with only a low level of statistical confidence.

We develop the theory of depth using a universal machine U, similar to that described in detail by Chaitin [1975B], which has two tapes, a program tape and work tape. The expression $U(s) = x$ will be used to indicate that the machine, started with the binary string s on its program tape and a blank work tape, embarks on a computation that halts after a finite number of steps, leaving the output x on the work tape. The number of steps (run time) is denoted $t(s)$. The work tape can also be used as an auxiliary input, with $U(s,w)$ denoting the output and $t(s,w)$ the run time of a computation beginning with s on the program tape and w on the work tape. In case the computation fails to halt, the functions U and t are considered to be undefined.

The program tape is treated in a special way [Gacs, 1974; Levin, 1974; Chaitin, 1975] in order to allow a natural relative weighting of programs of different lengths. The details of this treatment are described by Chaitin, but the essential feature is that the machine itself must decide how many bits to read off its program tape, without being guided by any special endmarker symbol. Another way of looking at this is to say that the expression $U(s,w) = x$ means that, if the machine were given w on its work tape and any *infinite* binary sequence beginning with s on its program tape, it would halt with the infinite program. This "self-delimiting" formalism allows the *algorithmic probability* of an output x to be defined in a natural way, as the sum of the negative binary exponentials of the lengths of all programs leading to that output:

$$Pu(x) = \sum_{\{s:U(s)=x\}} 2^{-|s|}$$

Here $|s|$ denotes the length of the binary string s, regarded as a self-delimiting program for the U machine. (Without the self-delimiting requirement, this sum would, in general, diverge.) An analogous conditional algorithmic probability, $Pu(x/w)$, may be defined for computations that begin with a string w on the work tape. This represents the probability that a program generated by coin tossing would transform string w into string x.

Besides being self-delimiting, the U machine must be *efficiently universal* in the sense of being able to simulate any other self-delimiting Turing machine with additive increase in program size and polynomial increase in time and space. That such machines exist is well known. The *minimal program* for a string x, denoted $x*$, is the least string p such that $U(p) = x$. The algorithmic information or entropy of a string $H(x)$ may be defined either as the size of its minimal program, or the negative base-two logarithm of its algorithmic probability, since it can be shown that the difference between these two quantities is bounded by a constant depending on U but independent of x (this is another advantage of the self-delimiting formalism). A string x is said to be compressible by b bits if its minimal program is b bits shorter than x. Regardless of how compressible their outputs may be, all minimal programs are incompressible to within an O(1) constant depending on the standard machine. (If they were not, i.e., if for some s, $x**$ were significantly shorter than $x*$, then $x*$ would be undercut in its fole as executing $x**$.) Finite strings, such as minimal programs, which are incompressible or nearly so are called *algorithmically random*. The above formulation in terms of halting, self-delimiting programs appears the most natural way of defining information content for discrete objects such as integers, binary strings, or Ising microstates.

To adequately characterize a finite string's depth, one must consider both the amount of redundancy and the depth of its burial. Several definitions are given below; the best appears to be to say that a string x is *(d,b)-deep*, or *d-deep with b bits significance*, if

i. every program to compute s in time $\leq d$ is compressible by at least b bits.

It can be shown that any (d, b)-deep string according to this definition is deep in two other, perhaps more intuitive senses:

ii. computations running in time $\leq d$ supply less than $1/2^{b+O(1)}$ of the string's algorithmic probability.

iii. the smallest program to compute x in time $\leq d$ is at least $b + O(1)$ bits larger than the minimal program $x*$.

Alternative 2), perhaps the most natural (because it fairly weights all computations leading to x) is very close to the chosen definition, since it can be shown (by a proof similar to that of Chaitin's [1975B] theorem 3.2) that any (d, b)-shallow string (one not (d, b)-deep) receives at least $1/2^{b+O(\log b)}$ of its algorithmic probability form programs running in time $\leq d$. Alternative 1) is favored because it satisfies a sharper slow growth law. Alternative 3), perhaps the most obvious, might seriously overestimate the depth of a string with a great many large fast programs, but no single, small, fast program. Whether such strings exist is not known; if they do exist, they should probably not be called deep, since they have a significant probability of being produced by small, fast-running probabilistic algorithms.

It is obviously desirable that depth obey the slow growth law, i.e., that no fast, simple, deterministic or probabilistic algorithm be able to transform a shallow object into a deep one. With the chosen definition of depth, it is easy to show that this is the case: for any strings w and x, if w is less than (d, b) deep, and the algorithmic probability for U to transform w (furnished as an auxiliary input on the work tape) into x within time t is at least 2^{-k}, then s can be no more than $(d + t + O(1),\ b + k + O(1))$-deep.

Similarly, depth can be shown to be reasonably machine-independent, in the sense that for any two, efficiently universal, self-delimiting machines, there exists a constant c and a polynomial p such that $(p(d), b + c)$ depth on either machine is a sufficient condition for (d, b) depth on the other.

One may well wonder whether, by defining some kind of weighted average run time, a string's depth may reasonably be expressed as a single number. This may, in fact, be done, at the cost of, in effect, imposing a somewhat arbitrary rate of exchange between the two conceptually very different quantities' run time and program size. Proceeding from alternative definition 2) above, one might try to define a string's average depth as the average run time of all computations contributing to its algorithmic probability. Unfortunately, this average diverges because it is dominated by programs that waste arbitrarily much time. To make the average depth of s depend chiefly on the fastest programs of any given size that compute s, it suffices to use the reciprocal mean reciprocal run time in place of a straight average. The *reciprocal mean reciprocal depth* of a string x is thus defined as

$$d_{rmr}(x) = \left[\sum_{\{s\,:\,U(s)=x\}} 2^{-|s|}\right] \Big/ \left[\sum_{\{s\,:\,U(s)=x\}} (2^{-|s|}/t(s))\right],$$

In this definition, the various computations that produce x act like parallel resistors, the fast computations in effect short-circuiting the slow ones. Although reciprocal mean reciprocal depth doesn't satisfy as sharp a slow growth law as two-parameter depth (multiplicative rather than additive error in the computation time), and doesn't allow strings to have depth more than exponential in their length (due to the short-circuiting of slower programs, no matter how small, by the print program), it does provide a simple quantitative measure of a strong's nontriviality.

An even rougher, qualitative distinction may be drawn between "deep" and "shallow" strings according to whether their reciprocal mean reciprocal depth is exponential or polynomial in the strings' length, or some other parameter under discussion. This rough dichotomy, in which all merely polynomially-deep strings are called shallow, is justified by the typically polynomial cost for one machine model to simulate another, and the consequent arbitrariness in the definition of computation time.

ACKNOWLEDGEMENTS

These ideas developed over many years with the help of countless discussions, especially with Gregory Chaitin, Rolf Landauer, Peter Gacs, Leonid Levin, Tom Toffoli, Norman Margolus, and Stephen Wolfram.