TITLE: ALGORITHMIC INFORMATION CONTENT, CHURCH-TURING THESIS,
PHYSICAL ENTROPY, AND MAXWELL'S DEMON

AUTHOR(S): Wojciech H. Zurek, T-6

LA-UR--90-3702

DE91 001804

SUBMITTED TO: Proceedings of NATO/Advanced Study Institute on
Information Dynamics, June 1990, Irsee, Kaufbeuren,
Bavaria, West Germany

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

By acceptance of this article the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes.

The Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy

Oct. 15, 1990

Received by USIL

NOV 05 1990

Los Alamos Los Alamos National Laboratory
Los Alamos, New Mexico 87545

MASTER

W. H. Zurek

Theoretical Division, T-6, MS B288, Los Alamos National Laboratory, Los Alamos, New Mexico 87545 and the Santa Fe Institute, 1120 Canyon Road, Santa Fe, New Mexico 87501

Algorithmic Information Content, Church-Turing Thesis, Physical Entropy, and Maxwell's Demon

Measurements convert alternative possibilities of its potential outcomes into the definiteness of the "record" — data describing the actual outcome. The resulting decrease of statistical entropy has been, since the inception of the Maxwell's demon, regarded as a threat to the second law of thermodynamics. For, when the statistical entropy is employed as the measure of the useful work which can be extracted from the system, its decrease by the information gathering actions of the observer would lead one to believe that, at least from the observer's viewpoint, the second law can be violated.

I show that the decrease of ignorance does not necessarily lead to the lowering of disorder of the measured physical system. Measurements can only convert uncertainty (quantified by the statistical entropy) into randomness of the outcome (given by the algorithmic information content of the data). The ability to extract useful work is measured by *physical entropy*, which is equal to the sum of these two measures of disorder. So defined physical entropy is, on the average, constant in course of the measurements carried out by the observer on an equilibrium system.

1. INTRODUCTION

Algorithmic information content (also known as algorithmic randomness) of a physical entity is given by the size, in bits, of the most concise message (e.g., of the shortest program for a universal computer) which describes that entity with the requisite accuracy. Regular systems can be specified by means of concise descriptions. Therefore, algorithmic information content can be regarded as a measure of disorder.

Algorithmic randomness is defined without a recourse to probabilities. It provides an alternative to the usual ensemble measures of disorder: it quantifies randomness of the known features of the state of the physical system. I shall demonstrate that it is indispensable in formulating thermodynamics from the viewpoint of the *information gathering and using system* ("IGUS")—a Maxwell's demon-like entity capable of performing measurements and of modifying its strategies (for example, for extraction of useful work) on the basis of the outcomes of the measurements. Such an IGUS can be regarded as a "complex adaptive system." The aim of this paper is to review the concept of the algorithmic information content in the context of statistical mechanics and discuss its recently discovered physical applications.

2. OVERVIEW

Algorithmic randomness, an alternative measure of the information capacity of a specific physical or mathematical object, was independently introduced in the mid-60's by Solomonoff,²² Kolmogorov,¹³ and Chaitin.⁵ It is based on an intuitively appealing idea that the information content is equal to the size, in bits, of the shortest description. Formalization of this idea will be briefly described in the next section: In its development, it draws on the theory of algorithms and in the process makes use of the theory of computation,^{7,19} establishes a firm and useful connection with Shannon's theory of information,^{12,20} and benefits from its implications for coding.¹⁰

Applications of the algorithmic measures of the information content were initially mostly mathematical in nature. More recently, Bennett, in an influential paper,¹ has pointed out that the average algorithmic entropy of a thermodynamic ensemble has the same value as its statistical (ensemble) entropy and, consequently, one could attempt to build a consistent thermodynamics on an algorithmic foundation.

I have applied algorithmic randomness to the problem of measurement as seen by an observer.^{25,26,27}

Following a measurement, an observer (IGUS) is in possession of a specific record. From its intrinsic point of view, this record is quite definite. Therefore, further analysis of the measurement in terms of the ensemble language is pointless.

Rather, the observer must deal with the specific measurement outcomes and with their implications for extraction of useful work. In this "Maxwell's demon" context, algorithmic randomness assumes, in part, a function analogous to the Boltzmann-Gibbs-Shannon entropy: From the observer's point of view, the second law of thermodynamics must be formulated by taking into account both the remaining ignorance (measurements are typically far from exhaustive) and the randomness in the already available data. Thus, the *physical entropy*, the quantity which allows for the formulation of thermodynamics from the viewpoint of the observer, must consist of two contributions:

$$\left(\begin{array}{c} \text{physical} \\ \text{entropy} \end{array} \right) = \left(\begin{array}{c} \text{remaining} \\ \text{ignorance} \end{array} \right) + \left(\begin{array}{c} \text{algorithmic} \\ \text{randomness} \end{array} \right)$$

or

$$S(\rho) = H(\rho) + K(\rho), \quad (2.1)$$

where ρ is the density matrix of the system;

$$H(\rho) = -\text{Tr } \rho \log_2 \rho \quad (2.2)$$

is its BGS entropy, and $K(\rho)$ is the size of the shortest program capable of describing ρ :

$$K(g) = |p_g^*| \quad (2.3)$$

This recent proposal for physical entropy will be described in more detail in sections 3 and 4.

In section 4, I shall also discuss the importance of "compression" of the acquired data to their most concise form: thermodynamic efficiency of an IGUS-operated engine depends on its ability to find concise measurement descriptions. In turn, this efficiency can be regarded as a consequence of the IGUS's ability to "understand" or "model" the part of the Universe employed as the "engine" in terms of the regularities which can be regarded as analogs of physical laws. In this sense, "intellectual capabilities" of an IGUS are quite critical for its "success."

3. ALGORITHMIC RANDOMNESS AS A MEASURE OF PHYSICAL DISORDER

Consider a one-dimensional lattice of spins. I shall suppose that each spin can point either up (and designate this by "1") or down ("0"). A sequence of 0's and 1's with the number of digits equal to the number of spins in the system can be then regarded as a "description" of its state. Consider two such descriptions:

$$0101010101010101 \quad (3.1)$$

and

$$10110100101100010111. \quad (3.2)$$

The first system is "regular": It can be simply and concisely described as 10 "01's." There is no equally concise description of the second spin system. To reconstruct it, one would have to have a "verbatim" description (Eq. 3.2); there is no way to "compactify" this description into a more concise message.

The concept of *algorithmic information content* (known also as the *algorithmic randomness*, *algorithmic complexity*, or *algorithmic entropy*) captures this intuitive difference between the "regular" and "random" binary sequences.

Algorithmic information content of a binary sequence s is defined as the size of the minimal program, s_U^* , which computes sequence s on the universal computer U :

$$K_U(s) \equiv |s_U^*|. \quad (3.3)$$

Above, vertical lines indicate the size of the binary string in bits. It is important to note that this definition of the algorithmic information content makes it explicitly subjective in the sense that it is computer dependent; hence, the subscript U in Eq. (3.3). However, by the very definition of the universal computer, programs executable on U can be also executed (and will yield the same output) on any other universal computer U' , provided that they are preceded by a prefix $\tau_{UU'}$ which depends on U and U' , but not on the program. Hence, algorithmic randomness defined for two different computers will differ by at most the size of the prefix^[1]:

$$|K_U(s) - K_{U'}(s)| \leq |\tau_{UU'}|. \quad (3.4)$$

Such differences by a constant are usually ignored in mathematically motivated considerations, as they become relatively unimportant in the limit of very long strings. Moreover, as we shall see in section 4, differences of algorithmic complexities (rather than its absolute value) are relevant in computations of engine efficiencies. This further reduces—but does not eliminate—the thermodynamic importance of the subjectivity.

It is worth noting that a binary string, or a physical system, which appears to be random at first sight may nevertheless have a concise description. Binary representations of π , $\sqrt{2}$, etc. are examples of such apparently random but, in fact, algorithmically regular binary strings. This difficulty in assessing a system's algorithmic complexity is related to Gödel's undecidability.^{5,6,9,24,25} In spite of it, the distribution of the algorithmic randomness in physically or mathematically constructed ensembles, its value for whole classes of strings, and many of its more paradoxical (and intriguing) properties can be discussed in a rigorous manner. In particular, one can demonstrate that only exponentially few binary strings are algorithmically simple.^{5,6}

[1] Note that here we have used the vertical lines in two different ways: on the left-hand side they stand for the absolute value while on the right-hand they indicate the size of the binary string in bits. Only this second meaning will be employed below.

FIGURE 1 Turing machine T uses a set of instructions residing inside its "central unit" as well as the input it reads in by means of the "head" scanning the input tape to modify the content of the tape. A Universal Turing machine U can simulate any other Turing machine by reading on the input tape the "description" of T . In particular, a single-tape U can simulate operations of the modern computers, which can be modelled as "polyccephalic" Turing machines with access to several multidimensional tapes and other kinds of hardware. Such "modern" machines (one of which is illustrated above) may be more convenient to use, but their capabilities are limited to the same range of "computable" tasks as for the original, one-tape U . This universality justifies the importance attached to the universal computers. In particular, it limits the subjectivity of the algorithmic information content defined by means of the minimal program, Eq. (3.3).

The distribution of the number of strings of a fixed length as a function of their algorithmic randomness is given by:

$$\mathcal{N}(K) \simeq 2^K. \quad (3.5)$$

This implies that "cryptoregular" strings (i.e., algorithmically simple but "hard to decipher" binary sequences like those associated with π , $\sqrt{2}$, etc.) are rare and a typical binary string s is algorithmically random and cannot be compressed:

$$K(s) \simeq |s|. \quad (3.6)$$

In particular, a typical natural number n is algorithmically random. Hence, to the leading order,

$$K(n) \simeq \log_2 n. \quad (3.7)$$

As in the "usual" information theory, one can define algorithmic randomness of sets of strings. For example, *joint algorithmic randomness* of a pair (s, t) is given by the minimal program which outputs s followed by t on the output tape. $K(s, t)$ satisfies the commuting equality

$$K(s, t) \cong K(t, s) \quad (3.8)$$

as well as the usual inequality

$$K(s, t) \lesssim K(s) + K(t). \quad (3.9)$$

In the equations above, I have used "approximate" ($\cong, \lesssim$) rather than "exact" ($=, \leq$) signs to describe relationships; this is because most of the equations of the algorithmic information theory are subject to " $O(1)$ " corrections which, in most cases, can be traced to the subjectivity in the definition of the algorithmic information content associated with the choice of the universal computer U .

Conditional algorithmic information content $K(s | t)$ is given by the size of the smallest program q_{st}^* , which computes string s from t (or, equivalently, both s and t from t). By analogy with Shannon's information theory, one would expect a relation:

$$K(s | t) = K(s, t) - K(t). \quad (3.10)$$

This equality does indeed hold, but only with logarithmic corrections.

It is convenient, for the reasons which will soon become apparent, to impose certain requirements on the properties of minimal programs. In particular, it is convenient to demand that they be *self-delimiting*; that is, that they must contain within them the information about the size of the input—their own size. This demand brings algorithmic information theory into a close correspondence with coding theory¹⁰ where such codes are known as *prefix* or *prefix-free* codes and are used to guarantee the unique decodability of messages: It can be demonstrated that every uniquely decodable encoding can also be made self-delimiting without changing sizes of the words (programs) associated with different messages (outputs). This correspondence between communication theory and computation (which can be regarded as "decoding" of programs by means of universal computers) puts the powerful mathematical formalism of the coding theory at the disposal of the algorithmic information theory.

The most important consequence of the requirement of self-delimiting minimal programs is the ability to associate binary sequences with the probability that they will be generated by a randomly obtained program for the universal computer U : any random sequence of bits obtained, for instance, by coin flipping, can be used as an input to the computer U . To be a "legal" input program i , this bit sequence must be self-delimiting which, in practice, means that U will stop reading additional bits—stop asking for further flips of the coin—after a finite number of bits were supplied. Hence, the probability of generating this sequence is equal to:

$$p(i) = 2^{-|i|}. \quad (3.11)$$

FIGURE 2 Lexicographic tree establishes a correspondence between binary strings and natural numbers. An algorithm which computes a certain string s from the input can be regarded as a Turing machine which "decodes" the input. The prescription for decoding is contained in the description of the computer.

Probability associated with the output o will be given by the sum over all inputs which yields the given output o :

$$p(o) = \sum_{\{i|U(i)=o\}} 2^{-|i|} \quad (3.12)$$

Dominant contribution in this sum will come from the shortest program. Hence,

$$K(o) \simeq \log_2 p(o). \quad (3.13)$$

This connection between the probability that a given string will be generated by a random input and its algorithmic information content can be employed in proving that the average algorithmic randomness of a member of a simply described ("thermodynamic") ensemble is almost identical to its Boltzmann-Gibbs-Shannon entropy.^{1,4,26} The relevant double inequality

$$H(\{p(s_i)\}) \leq \sum p(s_i) K(s_i) < H(\{p(s_i)\}) + K(\{s_i, p(s_i)\}) + O(1) \quad (3.14)$$

has been demonstrated by Levin^{16,17} and Chaitin⁵ (see, also, Bennett³ for a more accessible presentation and Caves⁴ for a discussion in the physical context). Above, $K(s_i, \{p(s_i)\})$ is the size of the minimal description of the ensemble. Bennett¹ has

pointed out that, in the case when the BGS entropy is large compared with the size of the ensemble description:

$$H(\{p(s_i)\}) = - \sum_{(s_i)} p(s_i) \log_2 p(s_i) \gg K(s_i, \{p(s_i)\}), \quad (3.15)$$

one could base thermodynamic formalism on the average algorithmic randomness of the ensemble.

In a recent paper I have considered an application of algorithmic randomness to the situation in which an observer attempts to extract maximum useful work from the system on the basis of partial measurements. In the next section I shall discuss this situation which forces one to consider *physical entropy* defined as the sum of the remaining ignorance $H(\{p(s_i)\})$ and of the cost of storage of the available information $K(s_i, \{p(s_i)\})$.

The last quantity which can be defined in the algorithmic context is the algorithmic information distance given by the sum of the conditional information contents:

$$\Delta(s, t) \equiv K(s | t) + K(t | s). \quad (3.16)$$

Algorithmic information distance satisfies the requirements expected of a metric.²⁵

In addition to the "simple" distance defined by Eq. (3.16), one can consider several related quantities. For example,

$$K(s!t!u) \equiv K(s | t, u) + K(t | u, s) + K(u | s, t) \quad (3.17)$$

is also positive, reflexive, and satisfies the obvious generalization of triangle inequality. Hence, $K(s!t!u)$ and its further generalizations involving more strings can be regarded as direct extensions of $\Delta(s, t) = K(s!t)$.

It is sometimes useful to express distance as the difference between the joint and mutual information content

$$\Delta'(s, t) = K(s, t) - K(s : t), \quad (3.18)$$

where the mutual information is given by

$$K(s : t) = K(s, t) - (K(s) + K(t)). \quad (3.19)$$

The quantity Δ' defined by Eq. (3.18) differs from the "original" distance in Eq. (3.16) by logarithmic terms because of the similar logarithmic errors entering into Eq. (3.10). The advantage of employing Eq. (3.18) is its intuitive appeal: The distance between two binary strings is the information which they contain but do not share.

Mutual information can also be used to define algorithmic independence of two strings: s and t are independent when $K(s : t)$ is small; for example,

$$K(s : t) \ll \min(K(s), K(t)).$$

Information distance can also be defined for statistical (that is, BGS) entropy. In this case, Δ and Δ' coincide. Indeed, information distance was independently discovered in the domain of the Shannon's information theory by at least three authors before it was discussed (again without the benefit of knowledge of these references) by this author²⁵ in the algorithmic context.

4. MAXWELL'S DEMON AS AN ADAPTIVE COMPLEX SYSTEM

A. MAXWELL'S DEMON AND THE CHURCH-TURING THESIS

Smoluchowski²¹ (see also Feynman⁸ for an accessible up-to-date treatment) ruled out the possibility that a simple automated device (such as a trapdoor) can be used to violate the second law of thermodynamics; he has shown that the thermal fluctuations which inevitably occur whenever a trapdoor is coupled to the heat reservoir make it "jitter," which in turn makes the selection process it is supposed to implement completely unreliable. Thus, it cannot be sufficiently "selective" about its actions to create and maintain temperature or pressure differences in a closed system. Smoluchowski has, however, raised the question of whether an "intelligent being" employed as a "demon" could be more successful than the simple trapdoor. Until now, no definite answer to this question could be given. (Although Szilard²³ did write a celebrated and influential article, the title of which suggested that it would address the question of the second law and "intelligent beings." While intelligent beings never really enter into Szilard's considerations, it raises the issue of the relationship between thermodynamic entropy and information, and I shall make extensive use of its approach below.)

The reason for this state of affairs is easy to explain. After all, how can one analyze actions of an "intelligent being" within the realm of physics? In the first place, the very idea sounds "sacreligious." More importantly, a physically sensible definition of what is an "intelligent being" was not easy to come by, especially in the times of Szilard.

In the intervening years, a compelling mathematical definition of an intelligent being became available. It is suggested by the so-called Church's thesis (or the Church-Turing thesis, or the Church-Tarski-Turing thesis, see Hofstadter¹¹ and Penrose¹⁸ for an accessible discussion). In simple terms, it states that "What is human computable is universal computer computable." Thus, it equates information-processing capabilities of a human being (or of any intelligent being) with the "intellectual capacities" of a universal Turing machine. This may appear to be a gross oversimplification. Indeed, Church's thesis (CT) might be misunderstood as a claim that "all brains are created equal." For the time being, we shall only note that no counterexamples to CT have been discovered (but see Penrose¹⁸ for a critical assessment of the extent to which computers may be able to simulate "consciousness"). Moreover, while with Penrose and others one might be concerned about the ability to simulate consciousness or the validity of CT for the "right-hemisphere" functions, it is harder to find a reason to question CT with regard to the logical and mathematical operations usually associated with the "left hemisphere" of the brain. Below, in the thermodynamic considerations, I shall need only the uncontroversial "left hemisphere" part of the thesis.

B. SZILARD'S ENGINE AND THE COST OF ERASURE

The connection between the mathematical model of an "intelligent being" and thermodynamics goes back to the above-mentioned paper by Szilard.²³ In the analysis of the famous one-gas-particle engine (Figure 3), Szilard concluded that the second law could be indeed violated by a fairly simple "demon" unless the cost of measurements is no less than $k_B T$ per bit of acquired information. Further, essential clarification of the situation is due to the recent work by Bennett^{1,2} who, basing his discussion on the earlier considerations of Landauer^{14,15} on the costs of information erasure, concluded that it is the "resetting" of the measuring apparatus which is thermodynamically expensive and must be responsible for restoring the validity of the second law in Szilard's engine. (Indeed, this observation was anticipated, if only in a somewhat half-hearted manner, by Szilard.²³)

Algorithmic randomness proved essential in attempts to generalize this discussion of Maxwell's demon.^{4,25,26,27} The validity of the original experiment about the cost of erasure was limited to the context of Szilard's engine. In that case, the outcome of the measurement can always be described by a single bit. Hence, the gain of useful work in the course of the expansion is given by

$$\Delta W^+ = k_B T \log_2 \left[\frac{V}{V/2} \right] = k_B T \quad (4.1)$$

Note that above we are using Boltzmann constant k_B which differs from the usual one by a factor of $\ln 2$. This distinction reflects the difference between entropy measured in "bits" and "nats."

This gain of useful work is "paid for" by ΔW^- of the energy needed to restore the memory part of the "brain" of the "demon" to the "blank," "ready to measure" state,

$$\Delta W^- = -k_B T. \quad (4.2)$$

FIGURE 3 Szilard's engine employs one-molecule gas in contact with a heat bath at temperature T to extract $\sim T \ln 2$ of work per cycle (which is illustrated in a self-explanatory manner above). The measurement which establishes the location of the molecule is crucial. The importance of the cost of erasure for the proper accounting for the net energy gain is discussed in the text.

Therefore, the net gain of work is at best nil;

$$\Delta W = \Delta W^+ + \Delta W^- = 0. \quad (4.3)$$

It is, nevertheless, far from clear how to apply this "cost of erasure" argument to less idealized and more realistic situations.

One simple (although not very realistic) generalization is to consider a sequence of measurements on the Szilard's engine and to postpone the "erasure" indefinitely. This requires a demon with a significant memory size. One can then, as noted by Bennett,^{1,2} use Szilard's engine to extract $k_B T$ of work per cycle as long as there is "empty" tape. This is, of course, only an apparent violation of the second law since the empty tape can be regarded as a zero-entropy (and, hence, zero-temperature) reservoir. Consequently, an ideally efficient engine can, in accord with the second law and, in particular, with the Carnot efficiency formula, attain exactly $k_B T$ of work per cycle.

The cost of erasure does not have to be paid for as long as the "memory tape" is available. However, for this very reason, the process is not truly cyclic: the demon's memory is never restored to the initial "blank" state. The gain of useful work is paid for by the "clutter" in its "brain." If the outcomes of consecutive measurements are random, getting rid of this clutter would cost $k_B T$ per bit, and all the apparent gain of work would have to be "paid back" by the final costs of erasure.

C. COMPRESSIBILITY OF INFORMATION: SECOND LAW AND CODING THEORY

Consecutive "bits" in the demon's memory may have been inscribed with 0's and 1's in a regular fashion. For example, a situation of this kind would automatically arise in the "multicylinder" Szilard engine shown in Figure 4. There the recording can be made by simply writing "0" or "1" on the squares of the tape corresponding to the empty or full cylinders. The gain of useful work extractable from the n -cylinder engine in contact with the heat bath at temperature T is

$$\Delta W^+ = k_B T \log_2 n. \quad (4.4)$$

Each measurement results in filling up n blanks of the tape with 0's and 1's. Hence, the cost of erasure would be

$$\Delta W^- = k_B T \cdot n. \quad (4.5)$$

Again, one could postpone erasures indefinitely and just "dump" all of the "cluttered-up" tape into the garbage can. In the final count, however, the cost of erasure (linear in n) would outweigh the gain of useful work (which is logarithmic in n).

FIGURE 4 "Multicylinder" Szilard's engine. A single gas molecule is enclosed in a cylinder with n partitions, each of them with the same volume ΔV . The engine cycle begins with a measurement to determine which of the positions contains the molecule. Expansion following a measurement yields $\sim \ln(V/\Delta V)$ of useful work per cycle. The cost of erasure and, consequently, the efficiency of the engine depends on the manner in which the information is recorded. The most concise record (the size of which is given by the algorithmic information content) will result in the maximum efficiency. Thus, algorithmic information content inescapably enters the "demon's own" formulation of thermodynamics.

A more efficient multicylinder engine would be achieved by a fairly obvious but more economical encoding. Instead of making a "direct" record of the cylinders' content, one could "compress" such "verbose" messages and keep in the memory only the label of the "occupied" engine cell. It is now easy to verify that both ΔW^+ and ΔW^- are logarithmic in n (providing that one insists, as one must, on the unique decodability of the label). Hence, a computerized "demon" employing this second, more economical way of encoding measurement results would be able to come much closer to the "break even point" than was the case for the operating system of the multicylinder Szilard engine considered above.

The general question raised by this example concerns the "ultimate compressibility" of a set of measurement outcomes. In particular, in the context of the example considered before, one could imagine that there is a still more efficient encoding of the alternative outcomes which results in a still more concise average size of the record. Then the demon would be able to do more than just break even, as the cost of erasure would be only partially offset by the gain of useful work. In this sense, the ability to compress records representing measurement outcomes determines the efficiency of a thermodynamic process whenever information acquisition and processing are involved. Indeed, the second law could be violated unless there is a limit on the compressibility which relates the average size $\langle \Delta K \rangle$ of the

record with the decreased statistical entropy ΔH of the measured system via an inequality

$$\Delta H \leq \langle \Delta K \rangle. \quad (4.6)$$

For, unless this inequality holds, the gain of useful work which is equal to

$$\Delta W^+ = k_B T \Delta H \quad (4.7)$$

could exceed the cost of erasure which is given by

$$\langle \Delta W^- \rangle = -k_B T \langle \Delta K \rangle. \quad (4.8)$$

Hence, the net average gain of useful energy per cycle would be

$$\langle \Delta W \rangle = k_B T (\Delta H - \langle \Delta K \rangle). \quad (4.9)$$

The second law demands that $\langle \Delta W \rangle \leq 0$, which leads to the inequality (4.6).

Fortunately, this inequality is indeed respected: It is an immediate consequence of the left-hand side of the inequality in Eq. (3.14). Indeed, it follows from the first basic result of Shannon's theory of communication (the so-called noiseless channel coding theorem; see Shannon and Weaver,²⁰ Khintchin,¹² Hamming,¹⁰ and Caves⁴ for discussion): The average size of minimal "descriptions" needed to unambiguously describe measurement outcomes cannot be made smaller than the statistical entropy of the "source" of information (in our case, of the measured physical system). In this context, the second law can be regarded as a direct consequence of the Kraft inequality¹⁰ which plays a basic role in the coding theory and thus enters physics^{25,26}: Suppose that $\{K_i\}$ are sizes (in the number of bits) of distinct symbols (programs) $\{s_i\}$ which correspond to different signals (measurement outcomes). Then one can prove that in order for the encoding to be uniquely decodable, the following inequality must be obeyed:

$$\sum_i 2^{-K_i} \leq 1. \quad (4.10)$$

The inequality (4.6) follows from Kraft inequality (4.10) since it can be immediately rewritten as

$$\log_2 \sum_i \left(\frac{p_i(s_i) 2^{-K_i}}{p(s_i)} \right) \leq 0$$

where $p(s_i)$ are the probabilities corresponding to signals (states) s_i . Now, employing convexity of the logarithm, one can write

$$\sum_i p(s_i) \log \frac{1}{p(s_i)} \leq \sum_i p(s_i) K_i, \quad (4.11)$$

which establishes the desired result, Eq. (4.6).

Thus, the most efficient encoding corresponding to entropy decrease (information gain) ΔH will have to fill up, on the average, at least $\langle \Delta K \rangle \geq \Delta H$ bits in the memory of the "demon."

The second law cannot be violated because the average size of the record cannot be compressed to less than the corresponding gain of information. The success of an intelligent Maxwell's demon is ruled out as a consequence of a theorem which was proposed a century after the second law in a very different context of the theory of communication!

D. FROM IGNORANCE TO DISORDER: THE ROLE OF MEASUREMENTS

The ability of living organisms to perform measurements and "profit" by exploiting their outcomes can be analyzed in the algorithmic terms of the above discussion. Measurements decrease ignorance about the specific state of the system, but increase the size of the record necessary to encode the acquired information. This process is illustrated in Figure 5(a) for an equilibrium ensemble: The approximate equality of the increase of the size of the record and of the decrease of ignorance is guaranteed by the coding theorem. To approach the optimal coding efficiency $\langle \Delta K \rangle = \Delta H$, the demon can assign to different outcomes symbols with length K_i given by

$$K_i = \lceil \log_2 (1/p(s_i)) \rceil. \quad (4.12)$$

Here the symbol $\lceil a \rceil$ signifies the smallest natural number that is $\geq a$. It is easy to see that with this assignment, Kraft inequality [Eq. (4.11)] is automatically satisfied. Therefore, there exists a uniquely decodable "labeling" with the word length given by Eq. (4.13). Moreover,

$$K_i < \log_2 1/p(s_i) + 1.$$

Therefore,

$$\langle \Delta K \rangle < \Delta H + 1. \quad (4.13)$$

Thus, the encoding satisfying Eq. (4.12) (and known as Shannon-Fano coding¹⁰) is quite efficient in that it exceeds the absolute lower limit on the minimal average size of the description given by $\langle \Delta K \rangle = \Delta H$ by no more than one bit.

Indeed, Shannon-Fano coding can suggest a near-optimal "labeling" strategy for a demon dealing with an unknown ensemble: The demon could perform sufficiently many measurements to estimate probabilities $p(s_i)$ from the frequencies of the different outcomes; and then adopt short descriptions for the more frequent outcomes and longer descriptions for the rare ones in accordance with the Shannon-Fano prescription [Eq. (4.12)]. For a demon dealing with an equilibrium system (Figure 5(a)), it would guarantee an (almost) optimal performance: The demon would lose no more than $k_B T$ of useful work (corresponding to the "extra" bit) per measurement. [The demon could "break even" if the probabilities $p(s_i)$ were exactly equal to the inverse powers of 2.] Of course, no useful work could be extracted—Shannon's noiseless channel coding theorem guarantees the validity of the second law.

To restate the content of Figure 5(a) one can regard measurements as a way of turning statistical uncertainty into algorithmic randomness. Moreover, for equilibrium ensembles, the two quantities are, on the average, changing in such a way that their sum

$$S = H + K \quad (4.14)$$

FIGURE 5 The effect of measurements on (i) The Shannon entropy H_d in presence of the partial information—data d ; (ii) The algorithmic information content of the data $K(d)$; and (iii) The *physical entropy* $S_d \equiv H_d + K(d)$, which measures the net amount of work that can be extracted from the system given the information contained in the data d . (a) When the measurements are carried out on the equilibrium ensemble, the randomness in the data increases at the rate given by the decrease of ignorance. (b) For systems far from equilibrium the increase of randomness is smaller than the decrease of ignorance, which allows the observer to extract useful work and makes measurements energetically attractive.

remains approximately constant. Furthermore, as it was already discussed above, the amount of useful work which can be extracted by a “demon” from a system is given by

$$\Delta W = k_B T (\Delta H + \Delta K) = k_B T \Delta S. \quad (4.15)$$

Therefore, from the point of view of the observer (the demon) evaluating its ability to extract *net* useful work from the physical system, the hybrid quantity S plays the role of entropy. I shall therefore refer to S as *physical entropy*.

It is important to emphasize that the actual record r_i corresponding, in the demon’s memory, to the state s_i will rarely be minimal; the inequality

$$|r_i| > K(r_i) \quad (4.16)$$

will usually be satisfied. This is because the minimal programs cannot be found in a systematic manner. Moreover, there is usually no incentive to look for the truly minimal record. A reasonably short record, often with built-in redundancies to allow for error-correction, can be far more useful. Consequently, the actual gain of useful work will be given by

$$\Delta W \simeq k_B T (\Delta H + \Delta |r|). \quad (4.17)$$

One might be tempted even to use quantity $H + |r|$ as a measure of entropy. As M. Gell-Mann emphasizes, this would be incorrect: $H + |r|$ can be lowered by *computation* at no thermodynamic expense. Hence, if $H + |r|$ was regarded as entropy, the second law could be violated by reversible computation. Such violation is impossible with S where the record part of S is already as short as it can be. The second law is safe when formulated in terms of the physical entropy S .

Indeed, physical entropy S has the great advantage of removing the "illusion" that entropy decreases in the course of measurement. Of course, the ignorance (measured by H) does indeed decrease, but only at the expense of the increase of the minimal record size K . Hence, in measurements performed on equilibrium ensembles $S \equiv H + K$ is, on the average, constant (Figure 5(a)).

By contrast, measurements performed on far-from-equilibrium systems can result in a decrease of ignorance which is much larger than the resulting increase in the record size (Figure 5(b)). Fortunately, the Universe that we inhabit is precisely such a nonequilibrium environment: It pays to measure. Moreover, it is not really necessary to look for truly minimal programs: the possible gain of useful work outweighs even substantial inefficiencies in the record size optimization (as well as in other more "traditional" thermodynamic inefficiencies). Nevertheless, the ability to recognize in the measurement outcome the opportunity for such a gain of useful work is essential for the well-being of IGUS's.

ACKNOWLEDGMENTS

I would like to thank Charles Bennett, Carl Caves, Stirling Colgate, Doyne Farmer, Murray Gell-Mann, James Hartle, Rolf Landauer, Seth Lloyd, Bill Unruh, and John Wheeler for stimulating and enjoyable discussions on the subject of this paper. The warm hospitality of the Aspen Center for Physics, the Institute for Theoretical Physics in Santa Barbara, and the Santa Fe Institute is very much appreciated.

REFERENCES

1. Bennett, C. H. *Int. J. Theor. Phys.* **21** (1982):305-340.
2. Bennett, C. H. *Sci. Am.* **255**(11) (1987):108-117.
3. Bennett, C. H. In *The Universal Turing Machine—A Half-Century Survey*, edited by R. Herkin. Oxford: Oxford University Press, 1988.
4. Caves, C. M., in *Complexity, Entropy, and the Physics of Information*, W. H. Zurek, ed. (Addison-Wesley, 1990)
5. Chaitin, G. J. *J. ACM* **13** (1966):547-569.
6. Chaitin, G. J. *Sci. Am.* **232**(5) (1975):47-52.
7. Davis, M. *Computability and Unsolvability*. New York: Dover, 1973.
8. Feynman, R. P., R. B. Leighton, and M. Sands. *Feynman Lectures on Physics*, sect. 46, vol. 1. Reading, MA: Addison-Wesley, 1964.
9. Gödel, K. *Monat. Nachr. Mat. Phys.* **38** (1931):173-198.
10. Hamming, R. W. *Coding and Information Theory*. Englewood Cliffs: Prentice-Hall, 1987.
11. Hofstadter, D. *Gödel, Escher, Bach*. New York: Random House, 1979.
12. Khintchin, A. I. *Information Theory*. New York: Dover, 1957.
13. Kolmogorov, A. N. *Information Transmission* **1** (1965)3-11.
14. Landauer, R. *IBM J. Res. Dev.* **3** (1961):113-131.
15. Landauer, R. In *Signal Processing*, edited by S. Haykin. New York: Prentice-Hall, 1989, 18-47.
16. Levin, L. A. *Dokl. Akad. Nauk SSSR* **227** (1976).
17. Levin, L. A. *Sov. Math. Doklady* **17** (1976):522-526.
18. Penrose, R. *The Emperor's New Mind*. Oxford: Oxford University Press, 1989.
19. Rogers, H. *Theory of Recursive Functions and Effective Computability*. New York: McGraw-Hill, 1967.
20. Shannon, C. E., and W. Weaver. *The Mathematical Theory of Communication*. Urbana: Univ. of Illinois Press, 1949.
21. Smoluchowski, M. In *Vorträge über die Kinetische Theorie der Materie und der Elektrizität*. Leipzig: Teubner, 1914.
22. Solomonoff, R. J. *Info. & Control* **7** (1964):1-22.
23. Szilard, L. *Z. Phys.* **53** (1929):840-856.
24. Turing, A. M. *Proc. Lond. Math. Soc.* **42** (1936):230-265.
25. Zurek, W. H. *Nature* **341** (1989):119-124.
26. Zurek, W. H. *Phys. Rev. A* **40** (1989):4731-4751.
27. Zurek, W. H. In *Proceedings of the International Symposium on Quantum Mechanics*, edited by Y. Murayama. Tokyo: Physical Society of Japan, 1990.

END

DATE FILMED

12 / 03 / 90