LEGIBILITY NOTICE

A major purpose of the Technical Information Center is to provide the broadest dissemination possible of information contained in DOE's Research and Development Reports to business, industry, the academic community, and federal, state and local governments.

Although a small portion of this report is not reproducible, it is being made available to expedite the availability of information on the research discussed herein.

TITLE **ALGORITHMIC RANDOMNESS, PHYSICAL ENTROPY, MEASUREMENTS,
AND THE SECOND LAW**

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

SUBMITTED TO **Proceedings of the Third International Symposium on
Foundations of Quantum Mechanics, 8/20/89-9/8/89,
Tokyo, Japan**

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, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent 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.

It is hereby certified that this report contains neither recommendations nor conclusions of the United States Government. It is the policy of the United States Government to purchase and distribute reprints for individuals at a special rate of \$1.00 per copy.

Approved for public release; distribution is unlimited. This report was prepared under the auspices of the Los Alamos National Laboratory, which is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-88.

10/23/89

ALGORITHMIC RANDOMNESS, PHYSICAL ENTROPY, MEASUREMENTS, AND THE SECOND LAW

W. H. Zurek

Theoretical Division, T-6, MS B288

Los Alamos National Laboratory

Los Alamos, New Mexico 87545 U.S.A.

Abstract

Algorithmic information content is equal to the size -- in the number of bits -- of the shortest program for a universal Turing machine which can reproduce (i.e. plot with the requisite accuracy) a state of a physical system. In contrast to the statistical Boltzmann-Gibbs-Shannon entropy, which measures ignorance, the algorithmic information content is a measure of the available information. It is defined without a recourse to probabilities and can be regarded as a measure of randomness of a definite microstate. I suggest that the physical entropy S -- that is, the quantity which determines the amount of the work ΔW which can be extracted in the cyclic isothermal expansion process through the equation $\Delta W = k_B T \Delta S$ -- is a sum of two contributions: (i) The missing information measured by the usual statistical entropy and (ii) The known randomness measured by the algorithmic information content. The sum of these two contributions is a "constant of motion" in the process of a dissipationless measurement on an equilibrium ensemble. This conservation under a measurement, which can be traced back to the noiseless coding theorem of Shannon, is necessary to rule out existence of a successful Maxwell's demon.

Proceedings of the Third International Symposium on

FOUNDATIONS OF QUANTUM MECHANICS

Y. Murayama, ed. (Physical Society of Japan, Tokyo, 1990)

I. Introduction

Statistical entropy of a system described by the density matrix ρ :

$$H(\rho) = -\text{Tr} \rho \log \rho \quad (1.1)$$

can be decreased in the process of measurement. Thus --- as it was originally pointed out by Maxwell --- an intelligent being capable of modifying its actions according to the outcomes of measurements it performs could in principle extract useful work from an equilibrium system. This threat to the second law is based on the assumption that the physical entropy $\mathcal{S}$ --- the quantity which determines the amount of useful work which can be extracted from the system through the equation:

$$\Delta W = k_B T \Delta \mathcal{S} \quad (1.2)$$

is identically equal to the statistical entropy $H(\rho)$, Eq. (1.1).

I show that this danger to the second law can be avoided when the algorithmic contribution to physical entropy $\mathcal{S}$ is recognized: $\mathcal{S}$ is now defined as a sum of two contributions

$$\mathcal{S}(\rho) = H(\rho) + K(\rho) \quad (1.3)$$

where $H(\rho)$ is given by Eq. (1.1) and $K(\rho)$ is the algorithmic information content (also known as algorithmic randomness) of the state of the system. Algorithmic information content $K(\rho)$ is equal to the minimal size of the description (i.e. the shortest computer program) which allows one to reconstruct the state described by ρ with the requisite accuracy.

Measurements decrease statistical entropy $H(\rho)$ but increase the amount of information, and, consequently, the algorithmic information content $K(\rho)$. In a sense, a measurement converts the *potential information* given by the pre-measurement density matrix $\rho_{\text{eff}, \text{pre}}$, $H(\rho_{\text{eff}, \text{pre}})$, into the *actual information* $K(\rho_{\text{eff}, \text{post}})$, where $\rho_{\text{eff}, \text{post}}$ stand-

for the final density matrix. Furthermore, in an equilibrium ensemble (i.e. whenever $-Tr \rho_{before} \log \rho_{before}$ is at a maximum consistent with the already available data) this conversion

$$H(\rho_{before}) \rightarrow K(\rho_{after})$$

conserves the sum of these two quantities. That is,

$$\langle H(\rho_{before}) - H(\rho_{after}) \rangle = \langle K(\rho_{after}) - K(\rho_{before}) \rangle \quad (1.4)$$

In this sense H and K can be regarded as analogous of potential and kinetic energy E_K and E_P . The process which converts E_K into E_P is the dynamical evolution. The process which converts H into K is the measurement. The physical entropy S -- the sum of H and K -- is conserved under the optimal (dissipationless) measurement and thus deserves treatment similar to the *total* energy, which is a constant of motion.

This paper is organized into two key sections: The next section introduces the concept of algorithmic information content and discusses its role in the physical context. Section III describes in some detail how the use of the physical entropy $S(\rho)$ eliminates the threat Maxwell's demons seem to pose to the second law of thermodynamics. The material discussed in these two sections is in part covered in the two more detailed papers by this author which have been already published^{1,2}. Some of the ideas employed below have been anticipated by the seminal papers of Bennett^{3,4}. Furthermore, a rigorous and careful discussion of the algorithmic information theory in the physical context has been recently carried out by Caves⁵.

II. Definitions of Algorithmic Randomness

Algorithmic randomness -- also known as *algorithmic complexity*, *algorithmic information content*, or *algorithmic entropy* -- provides a measure of the information content of the known (that is, defined without a recourse to ensembles and probabilities) characteristics of a physical object -- in other words, its description to which we shall refer as

"image." It is usually denoted by $K(s)$ and is given by the length — in the number of bits — of the shortest "message" which, when sent to the addressee, suffices to uniquely reproduce the image of this object. To partially eliminate the dependence of $K(s)$ on the "addressee" it is usually assumed that only a universal Turing machine U is an acceptable recipient of the message. I shall apply the concept of algorithmic randomness in the physical context, where an information gathering and using system — an "IGUS" or an "observer" is a natural recipient of the message. The applicability of the algorithmic concepts will, therefore, be based on a conjecture that information processing of all intelligent beings capable of observations can be simulated by universal computers. This assumption is known as Church-Turing thesis⁶. With all of the above caveats in mind one can write down algorithmic randomness of the image s with respect to the universal computer U as

$$K_U(s) = |s_U^*| \quad (2.1)$$

where the vertical "absolute value" lines denote the length, in bits, of the binary string inbetween them and s_U^* stands for the minimal (shortest) program which, upon execution on U , yields s as the output⁷.

In what follows we shall suppress the cumbersome subscript " U ": One needs, however, to keep in mind that the programs are intended for a given computer. Therefore, even if the "designated" computers are always universal — so that the programs for one of them (e.g. U) can be executed on any other universal computer U' provided that a finite sized prefix $r_{UU'}$ "translating" instructions from U to U' is supplied — the length of the minimal program will differ. This difference will be, however, bounded by the size of the prefix:

$$|r_{UU'}| \leq K_U(s) - K_{U'}(s) \leq |r_{UU'}| \quad (2.2)$$

Joint algorithmic randomness of two strings, s, t , is given by the size of the smallest program which can output them both on the output tape. It satisfies the "commuting"

property:

$$K(s, t) = K(t, s) + O(1) \quad (2.3)$$

Error term $O(1)$ which appears above is related to the "computer dependence" mentioned before. Below, I shall omit $O(1)$ and use approximate versions of the equality and inequality signs to simplify notation.

Conditional algorithmic information can be also defined. By analogy with the Shannon's information theory⁸ it is given by the size of the minimal program, which must be supplied in addition to certain image t to reproduce another image s . It satisfies the expected equality:

$$K(s) = K(t) + K(s|t) + O(K(K(t))) \quad (2.4)$$

only with the logarithmic error terms ($K(K(t)) \simeq \log_2 K(t)$ for a "typical string").

Mutual algorithmic information is defined

$$K(s : t) = K(s) + K(t) - K(s, t) \quad (2.5)$$

in the usual manner. Furthermore one may define algorithmic distance either as:

$$\Delta(s, t) = K(s|t) + K(t|s) \quad (2.6a)$$

or as:

$$\Delta(s, t) = K(s, t) - K(s : t). \quad (2.6b)$$

Because of the logarithmic errors in Eq. (2.4) these two definitions are not exactly identical. However, both of them are positive, symmetric, and satisfy triangle inequality¹. Hence, they induce a metric on the space of binary sequences or "images" they correspond to.

Programs which can be used to define algorithmic randomness must satisfy certain requirements. In particular, they must halt after a finite number of steps. Moreover, it is convenient to demand that they should be "self delimiting" — that is, the program itself

must determine when to stop reading its input. Consequently, minimal program must contain the information about its size.

This last “self-delimiting” requirement is reflected in the estimates of the typical sizes of the minimal programs required to represent objects such as natural numbers. Without it typical algorithmic randomness of a natural number n will be given (for self-explanatory reasons) by:

$$K(n) \simeq \log_2 n \quad (2.7a)$$

Self-delimiting requirement alters this by adding (to the leading order) an extra term

$$K(n) \simeq \log_2 n + \log_2(\log_2 n) \quad (2.7b)$$

which is needed to encode the length of the minimal program.

The main advantage of using self-delimiting codes comes from the fact that they are closely related to the so-called prefix-free codes in the context of the coding theory⁸. One can therefore directly apply probabilistic arguments of coding theory to the algorithmic information theory. In particular, one can show that the so-called Kraft inequality⁸:

$$\sum_s 2^{-K(s)} \leq 1 \quad (2.8)$$

holds for the self-delimiting definition of the algorithmic information content. This inequality is crucial in exploring the relationship between the statistical (ensemble) entropy and the algorithmic definition of the information content, which is central in the physical applications of $K(s)$.

Another requirement which is worth spelling out explicitly is the demand that, after the minimal program stops, the output tape should contain nothing but the desired image: In other words, it is not acceptable to output a whole collection of images one of which is the “right one” and leave the task of locating it to someone else. This requirement

avoids paradoxes which would arise because it is often possible to specify more concisely an ensemble — a whole set of images — than to point out precisely the “right” one.

The format of the reproduction of the image is somewhat arbitrary: A plot which, with the desired accuracy, reproduces the locations of gas particles in the phase space would be an acceptable “proof” that the message was really “understood” by the recipient. In order to facilitate this method of reproduction one may consider a “polycephalic” universal Turing machine with several heads which can access several multidimensional tapes (Fig. 1). Such machines resemble more closely modern-day computers. In spite of their improved “hardware” algorithmic complexity defined for them would differ only slightly (by a constant) from the randomness defined by the more “traditional” designs of the universal Turing machines with just a single infinite one-dimensional tape.

Algorithmic randomness can be *defined* but it cannot be *computed* in a systematic fashion^{7,9,10}. That is, there can be no algorithm κ which would, when supplied with the string s as an input, compute and print out $K(s)$ on the output tape and halt in a finite time. This uncomputability of algorithmic randomness is necessary for the consistency of its definition. For, if $K(s)$ was computable, one could devise a program which would output an algorithmically random binary strings ($K(s) \simeq |s|$) by means of a code with the size smaller than $K(s)$, which would be a contradiction. This feat could be accomplished by writing a program α which generates, in a sequence, all binary strings, computes their algorithmic randomness $K(s)$ and outputs this s for which $K(s)$ exceeds for the first time some predetermined number larger than the size of the program $|\alpha|$ (e.g. $K(s) \geq 10|\alpha|$). This would give one a way to generate s by means of a program shorter than the “minimal” program, which is a contradiction.

In spite of this uncomputability of the algorithmic information content of specific strings or images, one can usually reliably estimate algorithmic information content of the whole classes (ensembles) of states of physical systems. For example, by estimating

the size of the program needed to plot typical microstate s_{GAS} of the ideal gas in a container of volume V at temperature T one can derive Sackur-Tetrode equation²:

$$K(s_{GAS}) \simeq N \left\{ \log_2 \frac{V}{N \Delta V} + \frac{D}{2} \log_2 \frac{m k_B T}{(\Delta p)^2} \right\} \quad (2.9)$$

Above, N is the number of particles, ΔV and $(\Delta p)^D$ are volumes of the regular (cubic) cells in the phase space which define the resolution of the grid, D is the dimension of the system and m is the mass of the single particle. This close correspondence between algorithmic randomness and physical entropy (Eq. (2.9)) could be made even more reminiscent of its textbook version by recognizing that quantum mechanics demands that the phase space cells have volumes $\Delta V (\Delta p)^D = h^D$, where h is Planck's constant. Bennett³ has pointed out that the inequality:

$$H(\mathcal{E}) \leq \sum_i p_i K(s_i) < H(\mathcal{E}) + K(\mathcal{E}) + O(1) \quad (2.10)$$

established by mathematicians in the seventies^{9,10}, implies that the ensemble averaged value of the algorithmic entropy of the microstate which occurs with the certain probability p_i in the ensemble $\mathcal{E} \equiv \{p_i, s_i\}$ is almost identical to the statistical entropy

$$H(\mathcal{E}) = -\sum_i p_i \log_2 p_i \quad (2.11)$$

for the ensembles which are usually encountered in thermodynamic arguments. For, algorithmic information content of the description of a typical thermodynamic ensemble $K(\mathcal{E}) = K(p_i, s_i)$ is minute compared with its statistical entropy $H(\mathcal{E})$: Specifying temperature, number of particles, and the size/shape of the container suffices to write a concise algorithm which will continue listing microstates of $\mathcal{E}$ in some convenient order (e.g. of decreasing probability p_i). By contrast, specifying a single microstate on this list will require its relatively lengthy "address". (One encounters here the counter-intuitive property of K mentioned before: it is often easier to specify an ensemble than a single microstate!)

The close relation between typical $K(s)$ and $H(\mathcal{E})$ has led Bennett to suggest the $K(s)$ could be regarded as a true entropy of the physical system³. I shall argue — in the next

section — that from the point of view of the observer it is the “hybrid” quantity, *physical entropy*:

$$S = H + K , \quad (2.12)$$

given by the sum of the observer's ignorance (H) and of the disorder in the data about the system already available to the observer (K) — that determines the maximum amount of the useful work which can be extracted.

III. Thermodynamic Engines, the Second Law, and Shannon's Noiseless Coding Theorem

The aim of this section is to analyze the second law in the context of thermodynamic engines and Maxwell's demons. I shall carry out this analysis *from the point of view of the “demon”* — that is, of the observer which makes measurements on a physical system and takes advantage of their outcomes by adjusting its actions to maximize “profits” — which, in the context of thermodynamics correspond to the amount of the extracted useful energy.

Measurement has been always regarded as a threat to the validity of the second law. After all, it decreases ignorance, which, in effect, lowers the statistical entropy

$$H(\rho) = -\text{Tr } \rho \log_2 \rho \quad (3.1)$$

of the measured system. Maxwell¹¹ pointed out this threat. Smoluchowski¹² managed to “contain” it by demonstrating that only “intelligent beings” could be possibly “clever enough” to take advantage of the information obtained, thus ruling out “trap doors,” “ratchet and pawl” arrangements and other “unintelligent” alternatives (e.g. diodes) as potential second law violators. Szilard has further sharpened the problem by inventing his famous “Szilard's engine” (Fig. 2) which directly converts information gained into work¹³. To save the second law he was forced to postulate that the act of measurement must be thermodynamically expensive, and that the price must be no less than $k_B T$ per bit of

information gain. Szilard suspected that the expensive part of the measuring process is the erasure of the useless information, which is required if the demon is to operate in a cyclic fashion, and which is in turn necessary to truly endanger the second law. Landauer's¹⁴ discussion of thermodynamics of computation confirmed Szilard's suspicion that indeed "information disposal" carries the price of $k_B T$ per bit. Bennett^{3,4} took Landauer's ideas to their logical conclusion by demonstrating that if the cost of erasure is accounted for, threat to the second law in Szilard's version of the Maxwell's demon disappears. He has also essentially clarified the question of the thermodynamic cost of information processing by demonstrating that every computation can be always restated so that it is logically reversible. That is, every computation can be reformulated so that the computer continues to "remember" the logical "route" of the computation and, therefore, the computation can be in principle performed in a thermodynamically reversible fashion³.

These developments laid a foundation for a new formulation of statistical physics^{1,2,5}: Its aim is to express the laws of thermodynamics from the point of view -- in the language of -- the observer who is directly involved in the process of energy extraction, the "IGUS", which plays the role of the Maxwell's demon.

In the traditional formulation of the process of measurement the roles of the observer and the "legislator" or "narrator" -- that is, the intelligent being which analyzes from the outside the actions of the observers directly involved in the process of measuring (for instance, the author of the textbook on statistical mechanics of measurements) -- are clearly divided. The observer performs the measurement on a system which is always (explicitly or implicitly) regarded as a "member of an ensemble." In the language of the "narrator," the observer "becomes correlated" with the "measured system," which leads to an increase of the mutual information between the two. The "narrator" can now describe the observers alternative "states of mind" in terms of "an ensemble of outcomes" and switch from the simple "narrator" to the "legislator" by pointing out that the "ensemble

of correlated observer - system pairs" must satisfy such - and - such inequalities (total entropy cannot decrease) and, therefore, that "the laws of thermodynamics (in particular, the second law) must be necessarily obeyed."

These two levels of description — the level of "observer-demon" and the level of "narrator-legislator" — lead to paradoxes in statistical mechanics and in quantum physics. One classic example is the well-known quantum mechanical "Wigner's Friend" paradox¹⁶. The central issue is invariably the necessity to reconcile two descriptions of the same event, one according to the demon (friend) directly involved in the process of measurement and aware of its outcome, the other due to a "bystander" (narrator) who knows that the measurement has occurred, but does not know what is its result. Within the context of the ensemble description, it is impossible to convince the observer-demon that it must fail in its attempts to violate the second law: After all, the laws are formulated in terms of ensembles, while the observer is dealing, after each specific measurement, with a definite outcome and the corresponding definite record. Why should he be concerned?

To explain to the demon its own limitations in its own language one must recognize in addition to the Bennett-Landauer-Szilard point about the cost of erasure^{3,4,13,14} several additional ideas which have originated in the related contexts of information transmission (Shannon's coding theorem⁸), and information processing (Bennett's conclusion that computation can be made reversibly³) and the algorithmic information theory discussed in the preceding section. I shall explain their role while considering an example — a "gedanken experiment" — inspired by Szilard's discussion of the "measurement-powered engine".

Consider a thermodynamic system described by the density matrix ρ . Its standard entropy $H(\rho)$ is given by:

$$H(\rho) = -\text{Tr } \rho \log_2 \rho \quad (3.2)$$

The “ensemble” is characterised by the density matrix which is associated with the record (description) r and the corresponding algorithmic information content $K(\rho) \equiv K(r) \equiv |r^*|$. In accord with the discussion of the preceding section, $K(r)$ plays the role of $K(\mathcal{E})$: It is the size of the smallest description of the density matrix, and, therefore, may be small compared with $H(\rho)$, which has the value comparable with the algorithmic randomness of a typical microstate (see Eq. (2.10)).

Let us now suppose that the Maxwell’s demon makes a dissipationless measurement of the observable $\hat{\Pi} = \sum_i \pi_i \hat{P}_i$. (Note: for this measurement to be dissipationless $\hat{\Pi}$ and ρ must commute. I assume that this is indeed the case.)

From the point of view of the demon the outcome is some definite $\hat{P}_i$, which can be recorded by a description of the outcome, record r_i . From the point of view of the “narrator” dissipationless measurement has established correlation between the memory of the demon and the state of the system. In other words, when a measurement (e.g. by a “narrator”) of the demon’s memory will reveal a record r_i , further measurements of the system will always find it in the state described by the density matrix:

$$\rho_i = - \frac{\hat{P}_i \rho \hat{P}_i}{\text{Tr}(\hat{P}_i \rho)} \quad (3.3)$$

with the corresponding statistical entropy:

$$H(\rho_i) = -\text{Tr} \rho_i \log_2 \rho_i \quad (3.4)$$

Given the ensemble of records r_i , each encountered with the probability:

$$p_i = \text{Tr} \hat{P}_i \rho \quad (3.5)$$

one computes that — on the narrator (i.e. ensemble averaged) level of description — the decrease of entropy of the system S with respect to the demon:

$$\Delta H_S = H(\rho) - \sum_i p_i H(\rho_i) \quad (3.6)$$

corresponds to the increase of the uncertainty about the demons "state of mind":

$$\Delta H_D = -\sum p_i \log_2 p_i \quad (3.7)$$

Obviously, since the measurement was assumed to be dissipationless, the total entropy has to be conserved. Hence,

$$\Delta H_S = \Delta H_D \quad (3.8)$$

For measurements in which the measured observable $\hat{\Pi}$ and the pre-measurement density matrix ρ do not commute

$$\Delta H_S < \Delta H_D, \quad (3.9)$$

as the gain of information is in part offset by the dissipation caused by the "redefinition" of the basis system. (However, even then $\Delta H_S \geq 0$, as it was conjectured by Groenewold and demonstrated by Lindblad¹⁷).

Following the "footsteps" of the Szilard's demon, the observer can now attempt to extract useful work from the system by confining it in a dissipationless fashion to the phase space volume corresponding to the projection operator $\hat{P}_i$. The system can be then relaxed in a slow, isothermal expansion to the original ρ . The gain of useful work is:

$$\Delta W_i^+ = k_B T (H(\rho) - H(\rho_i)) \quad (3.10)$$

As we have already noted, the second law — from the observers viewpoint — is valid only on the ensemble averaged level: Thus, the the narrator may have difficulty in convincing the demon of its ultimate validity since the ensemble language is not immediately relevant for the observer — especially if the demon got lucky in the first few trials.

There is, however, an argument which can be formulated directly in the demon's language: It is based on two facts: (i) cost of erasure is at least $k_B T$ per bit. (ii) The sizes of minimal descriptions are limited in a manner which can be traced back to the coding theory⁸. The first fact, suspected by Szilard¹¹, justified by Landauer¹⁴, and put to use by

Bennett in Szilard's version of Maxwell's demon³, has already been discussed. The second fact is known as Shannon's noiseless channel coding theorem⁴. It states that the signals emitted by the source of information with the "premeasurement surprise" given by the source entropy H can be always encoded in the description of average length equal to H per message:

$$H \leq \sum_i p_i |r_i^*| < H + \epsilon \quad (3.11)$$

Here ϵ is small compared to H ($\epsilon \leq 1$). Above, I have used the notation of the past section to indicate that this inequality refers to the smallest, most concise records.

The importance of keeping records concise follows from the first fact: The cost of erasure of a record is computed on the "per bi." basis. To close the engine cycle the demon must get rid of the no longer useful records of the past measurements. For a single record r_i , the price of erasure cannot be made lower than:^{1,2}

$$\Delta W_i^{(-)} = k_B T (K(r_i) - K(r)) \quad (3.12)$$

Algorithmic complexities $K(r_i)$ and $K(r) \equiv K(\rho)$ refer to the information content of demon's mind before and after the most economical (in the thermodynamic sense) erasure of the no longer useful record of the measurement outcome " r_i ." The proof of Eq. (3.12) and discussion of its thermodynamic implications is given in Ref. 1.

In the language of the observer, the second law will be satisfied on the case by case average, since the least cost of restoring the memory to the "ready to measure" state is given by Eq. (3.12) which, because of the Shannon's theorem, mandates that the net gain of work is:

$$\begin{aligned} \Delta W^{(+)} + \Delta W^{(0)} - \Delta W^{(-)} &\sim k_B T (\Delta H - \Delta K) \\ &= k_B T [H(\rho) + K(r) - (H(\rho_i) + K(r_i))]. \end{aligned} \quad (3.13)$$

Given the structure and physical implications of this equation, it is useful to think of

the hybrid quantity

$$S(\rho) = H(\rho) + K(\rho) \quad (3.14)$$

as the only entropy observer directly involved in measurement will truly need to formulate thermodynamics on its “own” terms.

From the point of view of the “narrator” the validity of the second law is secured by the fact that, on the average, measurement of an equilibrium ensemble will increase the number of bits in demon’s memory by no less than ΔH (see Fig. 3). From the demon’s point of view this efficiency corresponding to ΔH can be achieved only when the minimal descriptions for the measurement outcomes are discovered. The actual efficiency will therefore be tied to the demon’s ability to “compress” descriptions. Some of the difficulties associated with this process are related to the Gödel’s undecidability via Turing’s halting problem.^{1,2}

It is important to emphasize that the algorithmic component to the efficiency is somewhat subjective not only because it depends on the available data as well as on the demon’s ability to discover concise records, but also because of more fundamental reasons: Minimal descriptions are always conditional — they may depend on the other information, already present in the demon’s “mind”. This is because the algorithmic randomness is defined with respect to a certain universal machine U . Hence, by the same token

$$S_U \equiv H(\rho) + K_U(\rho) \quad (3.14a)$$

is the “correct” way of stating the definition of physical entropy.

The role of this subjectivity in the formulation of thermodynamics and — perhaps more importantly, in the discussion of adaptive complex systems which can play roles of a demon and which can be modeled by a universal computer U — is only beginning.

I would like to thank Charles H. Bennett, Carlton M. Caves, Stirling A. Colgate, Murray Gell-Mann, James B. Hartle, William G. Unruh and John A. Wheeler for useful discussions, and the Santa Fe Institute for facilitating research on the subject.

References

1. W. H. Zurek: **Phys. Rev. A**, in the press.
2. W. H. Zurek: **Nature** **341** (1989) 119.
3. C. H. Bennett: **Int. J. Theor. Phys.** **21** (1983) 305; also **IBM J. Res. Dev.** **17** (1973) 525 and *ibid* **32** (1988) 16.
4. C. H. Bennett: **Scient. American** **255** (11) (1987) 108.
5. C. M. Caves: *Complexity, Entropy, and Physics of Information*, W. H. Zurek, ed. (Addison-Wesley, in press).
6. See, e.g. D. Hofstadter: *Gödel, Escher, Bach* (Random House, New York, 1979).
7. R. J. Solomonoff: **Inf Control** **7** (1964) 1; A. N. Kolmogorov, **Inf. Transmission** **1** (1965) 3; G. J. Chaitin, **Journal of the Assoc. Comp. Mach.** **13** (1966) 547.
8. C. E. Shannon and W. Weaver: *The Mathematical Theory of Communication* (University of Illinois Press 1949); R. W. Hamming: *Coding and Information Theory*, (Prentice-Hall, Englewood Cliffs, 1987).
9. A. K. Zvonkin and L. A. Levin: **Uspekhi Mat. Nauk** **25** (6) (1970) 602.
10. G. J. Chaitin: **Journal of the ACM** **22** (1975) 323
11. J. C. Maxwell: *Theory of Heat*, 4th edition, (Longmans, Green & Co., London 1875), pp. 328-329; See H. Lef: **Am. J. Phys.**, in press, for references discussing "Maxwell's demon".
12. M. Smoluchowski: *Vorträge über die Kinetische Theorie der Materie und der Elektrizität* (Teubner, Leipzig, 1914).
13. L. Szilard: **Z. Phys.** **53** (1929) 840. English translation in **Behavioral Science** **9** (1964) 301, reprinted in *Quantum Theory and Measurement*, J. A. Wheeler and W. H. Zurek, eds. (Princeton University Press, Princeton, 1983).
14. R. Landauer: **IBM J. Res. Dev.** **3** (1961) 113.
15. W. H. Zurek: *Frontiers of Nonequilibrium Statistical Mechanics* eds. Moore, G. F.

- and Scully, M. O. (Plenum, New York, 1986) 151.
16. E. P. Wigner: *The Scientist Speculates*, I. J. Good, ed. (Heinemann, London, 1961 and Basic Books, New York, 1962). (Heinemann, London, 1961 and Basic Books, New York, 1962).
17. H. J. Groenewold: **Int. J. Theor. Phys.** **4** (1971) 327; G. Lindblad: **Commun. Math. Phys.** **28** (1972) 245.

Figure Captions

Fig. 1: A *polycephalic* Turing machine can access several multidimensional tapes. It will be convenient to imagine that the minimal program containing the description of the state of a physical system (e.g. ideal gas) is supplied on the one-dimensional input tape, but the corresponding “image” is plotted on a 2D-dimensional “phase space tape.” (See text and Section III, Ref. 2 for further discussion).

Fig. 2: Classical and quantum¹⁵ versions of the Szilard engine¹³. The gas consists of a single molecule. Measurement determines whether – after the partition is inserted – all of the “gas” is on the left or right of the partition. The partition is then used as a piston. As the system is assumed to be in contact with the thermal reservoir at temperature T , $k_B T \ln 2$ of useful work can be extracted in each cycle of the engine. This would result in a violation of the second law unless – as assumed by Szilard¹³ – the measurement carries a cost of $k_B T$ per bit. Bennett^{3,4}, following an earlier analysis of costs of erasure in the process of computation by Landauer¹⁴, has demonstrated that the original suspicions of Szilard were correct: *Erasure* of one bit of information acquired by the demon about the location of the molecule in the environment at temperature T can be accomplished only at the cost of $k_B T$, which exactly cancels the gain of useful work and reaffirms the validity of the second law. This conclusion was also reached in the quantum case¹⁵. Szilard’s analysis also hints at the profound, physical (and not just purely formal) connection between the thermodynamic entropy and information.

Fig. 3: Measurements on an equilibrium ensemble decrease ignorance H_I by supplying the data d , which, in the end, could be detailed enough to leave a pure state with $H_I = 0$. Algorithmic information content of the available data $K(d)$ increases so that the sum $S = H_I + K(d)$ remains approximately constant in the typical instances, and is exactly constant on the ensemble averaged level.