entropy

MDPI

Article

Entropic Probability and Context States

Benjamin Schumacher [1,*] and Michael D. Westmoreland [2]

[1] Department of Physics, Kenyon College, Gambier, OH 43022, USA
[2] Department of Mathematics, Denison University, Granville, OH 43023, USA; westmoreland@denison.edu
* Correspondence: schumacherb@kenyon.edu; Tel.: +1-740-427-5832

Abstract: In a previous paper, we introduced an axiomatic system for information thermodynamics, deriving an entropy function that includes both thermodynamic and information components. From this function, we derived an entropic probability distribution for certain uniform collections of states. Here, we extend the concept of entropic probability to more general collections, augmenting the states by reservoir and context states. This leads to an abstract concept of free energy and establishes a relation between free energy, information erasure, and generalized work.

Keywords: axiomatic thermodynamics; information; entropy; Maxwell's demon; probability

check for updates

Academic Editor: Miguel Rubi

Received: 8 January 2025
Revised: 4 February 2025
Accepted: 7 February 2025
Published: 12 February 2025

Citation: Schumacher, B.; Westmoreland, M.D. Entropic Probability and Context States. *Entropy* **2025**, *27*, 187. https://doi.org/10.3390/e27020187

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In [1], we developed an axiomatic system for thermodynamics that incorporated information as a fundamental concept. This system was inspired by previous axiomatic approaches [2,3] and discussions of Maxwell's demon [4,5]. The basic concept of our system is the *eidostate*, which is a collection of possible states from the point of view of some agent. The axioms imply the existence of additive conserved quantities called *components of content* and an entropy function $\mathbb{S}$ that identifies reversible and irreversible processes. The entropy includes both thermodynamic and information components.

One of the surprising things about this axiomatic system is that despite the absence of probabilistic ideas in the axioms, a concept of probability emerges from the entropy $\mathbb{S}$. If state e is an element of a uniform eidostate E, then we can define

$$P(e|E) = \frac{2^{\mathbb{S}(e)}}{2^{\mathbb{S}(E)}}. \tag{1}$$

States in E with higher entropy are assigned higher probability. As we will review below, this distribution has a uniquely simple relationship to the entropies of the individual states and the overall eidostate E.

The emergence of an entropic probability distribution motivates us to ask several questions. Can this idea be extended beyond uniform eidostates? Can we interpret an arbitrary probability distribution over a set of states as an entropic distribution within a wider context? What does the entropic probability tell us about probabilistic processes affecting the states within our axiomatic system? In this paper, we will address these questions.

2. Axiomatic Information Thermodynamics

In physics, a formal axiomatic system can serve several purposes. It can serve to ensure consistent reasoning with unfamiliar or counter-intuitive mathematical objects. Furthermore, it may clarify the underlying concepts of a subject and their relations. This in turn can help to understand how widely, and under what circumstances, those concepts

may apply to the physical world. Here, we review the system in [1] and describe the ideas represented by the definitions and axioms.

Our system aims to capture how an agent with finite knowledge might describe states and processes in the external world. The set of *states* is denoted $\mathscr{S}$, but the agent's knowledge is represented by an *eidostate*, which is a finite collection of *possible* states in $\mathscr{S}$. (The term *eidostate* comes from the Greek *eidos*, meaning *to see*.) Thus, the agent's knowledge of the state may be incomplete. No probability distribution is assumed a priori for the eidostate; it is a simple enumeration of possibilities. The set of eidostates is called $\mathscr{E}$. (Without too much confusion, we may identify a state $a \in \mathscr{S}$ with the singleton eidostate $\{a\} \in \mathscr{E}$ so that $\mathscr{S}$ can be regarded as a subset of $\mathscr{E}$.)

Eidostates may be combined by the operation $+$, which simply represents the Cartesian product of the sets. Thus, if A and B are distinct eidostates, the combination $A + B$ is not the same as $B + A$. Two eidostates A and A' are similar ($A \sim A'$) if they are formed by the same Cartesian factors perhaps put together in a different way. Our first axiom is

Axiom 1 (Eidostates)**.** $\mathscr{E}$ is a collection of sets called *eidostates* such that

(a) Every $A \in \mathscr{E}$ is a finite non-empty set with a finite prime Cartesian factorization.

(b) $A + B \in \mathscr{E}$ if and only if $A, B \in \mathscr{E}$.

(c) Every non-empty subset of an eidostate is also an eidostate.

The agent may control state transformations, including those that involve the acquisition, use, or deletion of information. This produces a relation $\to$ on $\mathscr{E}$. We interpret $A \to B$ to mean that there exists a dynamical evolution that transforms A into B, perhaps in the presence of some other "apparatus" states that undergo no net change. We assume that any simple rearrangement of an eidostate's components is always possible, that transformations can be composed in succession, and so on:

Axiom 2 (Processes)**.** Let eidostates $A, B, C \in \mathscr{E}$, and $s \in \mathscr{S}$.

(a) If $A \sim B$, then $A \to B$.

(b) If $A \to B$ and $B \to C$, then $A \to C$.

(c) If $A \to B$, then $A + C \to B + C$.

(d) If $A + s \to B + s$, then $A \to B$.

Formally, a *process* is a pair $\langle A, B \rangle$ of eidostates, and it is said to be possible if either $A \to B$ or $B \to A$. An eidostate A is *uniform* if, for all $a, b \in A$, $\langle a, b \rangle$ is possible. Uniform eidostates, as we will see, have particularly tractable properties. The existence of *non*-uniform eidostates is neither forbidden nor guaranteed by the axioms, and there are models both with them and without them.

Consider a non-deterministic process in which initial state a might evolve into either b_1 or b_2. We describe this as a deterministic process on eidostates (that is , $\{a\} \to \{b_1, b_2\}$). The principle underlying the "Second Law" in our axiomatic system is simple to state. Given a collection A of possible states, no process can deterministically eliminate one or more of the elements of A without any other effect. Possibilities cannot simply be "deleted" by the agent. That is,

Axiom 3. If $A, B \in \mathscr{E}$ and B is a proper subset of A, then $A \nrightarrow B$.

To borrow an example from the next section, suppose the agent does not know whether a coin is showing "heads" (h) or "tails" (t). This might be represented by the eidostate $\{h, t\}$. The agent might examine the coin to discover its actual state, but this is not a *deterministic*

process, and it might end up in either h or t. To *guarantee* that the coin is h, the agent will need to intervene in the coin's state, which in principle may involve changes to other states.

On the other hand, it is reasonable to assume the existence of certain "conditional" processes:

Axiom 4 (Conditional processes)**.**

(a) Suppose $A, A' \in \mathscr{E}$ and $b \in \mathscr{S}$. If $A \to b$ and $A' \subseteq A$ then $A' \to b$.

(b) Suppose A and B are uniform eidostates that are each disjoint unions of eidostates: $A = A_1 \cup A_2$ and $B = B_1 \cup B_2$. If $A_1 \to B_1$ and $A_2 \to B_2$ then $A \to B$.

We now introduce information concepts into the system. This is based on the idea of a *record state*. State r is a record state if there exists another state such that $a + r \leftrightarrow a$. That is, a *particular* record state may be freely created or deleted. An information state is an eidostate containing only record states; the set of these is called $\mathscr{I}$. A bit state I_{b} is an information state with exactly two elements, and a bit process is a process of the form $\langle r, I_b \rangle$. Then, we have

Axiom 5 (Information)**.** There exist a bit state and a possible bit process.

Devices such as Maxwell's demon accomplish state transformations by acquiring information. We accommodate these into our theory via a new axiom:

Axiom 6 (Demons)**.** Suppose $a, b \in \mathscr{S}$ and $J \in \mathscr{I}$ such that $a \to b + J$.

(a) There exists $I \in \mathscr{I}$ such that $b \to a + I$.

(b) For any $I \in \mathscr{I}$, either $a \to b + I$ or $b + I \to a$.

The first clause asserts that what one demon can do (transforming a into b by acquiring the information in J), another demon can undo (transforming b into a by acquiring information in I). The second clause envisions nearly reversible demons.

Two more axioms serve more technical purposes. The first regularizes the $\to$ relation by its asymptotic behavior:

Axiom 7 (Stability)**.** Suppose $A, B \in \mathscr{E}$ and $J \in \mathscr{I}$. If $nA \to nB + J$ for arbitrarily large values of n, then $A \to B$.

The second posits the existence of states in which conserved quantities may be stored—e.g., the height of a weight used for storing energy. We have

Axiom 8 (Mechanical states)**.** There exists a subset $\mathscr{M} \subseteq \mathscr{S}$ of *mechanical states* such that

(a) If $l, m \in \mathscr{M}$, then $l + m \in \mathscr{M}$.

(b) For $l, m \in \mathscr{M}$, if $l \to m$ then $m \to l$.

(This axiom is identical to one posited by Giles [2], whose system was a model for our own.) Note that the axiom does not assert that $\mathscr{M}$ is non-empty; it is perfectly consistent to have a model for the axioms with no mechanical states.

We introduce one additional axiom that allows us to relate uniform eidostates to singleton states. Consider, for instance, a situation in which a one-particle "Szilard" gas is confined to a volume. This situation corresponds to a single state s_0 (as in the first part of Figure 2 below). Now, we introduce a partition so that the gas particle might be either in one side or the other (that is, state s_a or s_b). If we remove the partition, the original state s_0 is restored. Thus, $a \leftrightarrow \{a_1, a_2\}$. Considering many such examples, we arrive at our final axiom:

Axiom 9 (State equivalence)**.** If E is a uniform eidostate, then there exist states $e, x, y \in \mathscr{S}$ such that $x \to y$ and $E + x \leftrightarrow e + y$.

The system yields some striking results. For instance, we can establish the existence of *components of content*, which are numerical quantities that are conserved in every possible process. A component of content Q is a real-valued additive function on the set of states $\mathscr{S}$. (Additive in this context means that $Q(a + b) = Q(a) + Q(b)$.) In a uniform eidostate E, every element has the same values of all components of content, so we can without ambiguity refer to the value $Q(E)$. The set of uniform eidostates is denoted $\mathcal{U}$. This set includes all singleton states in $\mathscr{S}$, all information states in $\mathscr{I}$ and so forth, and it is closed under the $+$ operation.

Perhaps the most far-reaching theorem established by the axioms is the following (Theorem 8 in [1]):

Theorem 1. *There exist an entropy function $\mathbb{S}$ and a set of components of content Q on $\mathcal{U}$ with the following properties:*

(a) For any $E, F \in \mathcal{U}$, $\mathbb{S}(E + F) = \mathbb{S}(E) + \mathbb{S}(F)$.

(b) For any $E, F \in \mathcal{U}$ and component of content Q, $Q(E + F) = Q(E) + Q(F)$.

(c) For any $E, F \in \mathcal{U}$, $E \to F$ if and only if $\mathbb{S}(E) \leq \mathbb{S}(F)$ and $Q(E) = Q(F)$ for every component of content Q.

(d) $\mathbb{S}(m) = 0$ for all $m \in \mathscr{M}$.

On uniform eidostates, the entropy function $\mathbb{S}$ is completely determined (up to a non-mechanical component of content) by the $\to$ relations among the eidostates.

3. Coin-and-Box Model

Now that we have described our system in some detail, we will introduce a very simple model of the axioms. None of our later results depend on this model, but a definite example will be convenient for explanatory purposes. Our theory deals with configurations of coins and boxes; the states are arrangements of coins, memory records, and closed boxes containing coins. We have the following:

- *Coin states*, which can be either h (heads) or t (tails) or combinations of these. It is also convenient to define a *stack state* s_n to be a particular combination of n coin states h: $s_n = h + (h + (h + \cdots))$. The *coin value* Q of a compound of coin states is just the total number of coins involved. A finite set K of coin states is said to be Q-uniform if every element has the same Q-value.
- Record states r. As the name suggests, these should be interpreted as specific values in some available memory register. The combination of two record states is another record state. Thus, r, $r + r$, $r + (r + r)$, etc., are all distinct record states. Record states are not coins, so $Q(r) = 0$.
- Box states. For any Q-uniform set of coin states C, there is a sequence of *box states* b_n^K. Intuitively, this represents a kind of closed box containing $nQ(K)$ coins so that $Q(b_n^K) = nQ(K)$. If $K = \{h, t\}$, then we denote the corresponding "basic" box states by b_n.

An *eidostate* is any finite, non-empty, Q-uniform set of states.

We now must define the relation $\to$ among eidostates.

- We assume that similar eidostates can be transformed into each other in accordance with our axioms. As far as the $\to$ relation is concerned, we can freely rearrange the "pieces" in a compound eidostate.

- For coin states, $h \leftrightarrow t$.
- If r is a record state, $a + r \leftrightarrow a$ for any a. In a similar way, for an *empty* box state, $a + b_0^K \leftrightarrow a$.
- If K is a Q-uniform eidostate of coin states, $b_n^K + K \leftrightarrow b_{n+1}^K$.

Now, we add some rules that allow us to extend these to more complex situations and satisfy the axioms. In what follows, A, A', B, etc., are eidostates, and s is a state.

Transitivity. If $A \to B$ and $B \to C$, then $A \to C$.

Augmentation. If $A \to B$, then $A + C \to B + C$.

Cancelation. If $A + s \to B + s$, then $A \to B$.

Subset. If $A \to s$ and $A' \subseteq A$, then $A' \to s$.

Disjoint union. If A and B are both disjoint unions $A = A_1 \cup A_2$ and $B_1 \cup B_2$, and both $A_1 \to B_1$ and $A_2 \to B_2$, then $A \to B$.

Using these rules, we can prove a lot of $\to$ relations. For example, for a basic box state, we have $b_n + \{h, t\} \leftrightarrow b_{n+1}$. From the subset rule, we have $b_n + h \to b_{n+1}$ (but not the reverse). Then, we can say

$$b_n + h \to b_{n+1} \to b_n + \{h, t\}, \tag{2}$$

from which we can conclude (via transitivity and cancelation) that $h \to \{h, t\}$. The use of a basic box allows us to "randomize" the state of one coin.

Or consider two coin states and distinct record states r_0 and r_1. Then

$$h \to h + r_0 \qquad \text{and} \qquad t \to t + r_1 \to h + r_1, \tag{3}$$

from which can show that $\{h, t\} \to h + \{r_0, r_1\}$. That is, we can set an unknown coin state to h if we also make a record of which state it is. A pretty similar argument establishes the following:

$$\begin{aligned} (h + \{r_0, r_1\}) + b_n &\to \{h + r_0, t + r_1\} + b_n \\ &\to (\{h, t\} + r_0) + b_n \to \{h, t\} + b_n \to b_{n+1}. \end{aligned} \tag{4}$$

The bit state $\{r_0, r_1\}$ can be deleted at the cost of a coin absorbed by the basic box. The basic box is a coin-operated deletion device, and since each step above is reversible, we can also use it to dispense a coin together with a bit state (that is, an unknown bit in a memory register).

These examples help us to clarify an important distinction. What is the difference between the box state b_1 and the eidostate $\{h, t\}$? Could we simply replace all box states b_n^K with a simple combination $K + (K + \ldots)$ of possible coin eidostates? We cannot, because such a replacement would preclude us from using the subset rule to obtain Equation (2). The whole point of the box state is that the detailed state of its contents is *entirely inaccessible* for determining possible processes. From the point of view of the agent, putting a coin in a box effectively randomizes it.

It is not difficult to show that our model satisfies all of the axioms presented in the the last section with the mechanical states in $\mathscr{M}$ identified as coin states. The key idea in the proof is that we can reversibly reduce any eidostate to one with a special form:

$$A \leftrightarrow s_q + I_k, \tag{5}$$

where s_q is a stack state of $q = Q(A)$ coins and I_k is an *information state* containing k possible record states. (This is true because all box states can be reversibly "emptied", and any

Q-uniform eidostate of coin states can be reversibly transformed into a stack state and an information state, as in Equation (5)). Relations between eidostates are thus reduced to relations between states of this form. We note that the coin value q is conserved in every $\to$ relation, and no relation allows us to decrease the value of k. In our model, there is just one independent component of content (Q itself), and the entropy function is $\mathbb{S}(A) = \log k$. (We use base-2 logarithms throughout).

4. The Entropy Formula and Entropic Probability

Now, let us return to the general axiomatic system. One general result of [1] is a formula for computing the entropy of a uniform eidostate E in terms of the entropies of its elements e. This is

$$\mathbb{S}(E) = \log\left(\sum_{e\in E} 2^{\mathbb{S}(e)}\right). \tag{6}$$

It is this equation that motivates our definition of the entropic probability of e within the eidostate E:

$$P(e|E) = \frac{2^{\mathbb{S}(e)}}{2^{\mathbb{S}(E)}}. \tag{7}$$

Then, $P(e|E) \geq 0$ and the probabilities sum over E to 1. As we have mentioned, the entropy function $\mathbb{S}$ may not be quite unique; nevertheless, two different admissible entropy functions lead to the *same* entropic probability distribution. Even better, our definition gives us a very suggestive formula for the entropy of E:

$$\mathbb{S}(E) = \sum_{e\in E} P(e|E)\mathbb{S}(e) - \sum_{e\in E} P(e|E)\log P(e|E) \tag{8}$$

$$= \left\langle \mathbb{S}(a) \right\rangle + H(\vec{P}), \tag{9}$$

where the mean $\langle\cdots\rangle$ is taken with respect to the entropic probability, and $H(\vec{P})$ is the Shannon entropy of the distribution P [6,7].

Equation (9) is very special. If we choose an arbitrary distribution (say $P'(e|E)$), then with respect to this probability, we find

$$\mathbb{S}(E) \geq \left\langle \mathbb{S}(a) \right\rangle_{P'} + H(\vec{P}'), \tag{10}$$

with equality if and only if P' is the entropic distribution [7]. Therefore, we might *define* the entropic probability to be the distribution that maximizes the sum of average state entropy and Shannon entropy—a kind of "maximum entropy" characterization.

How is the entropic probability related to other familiar probability concepts? To quote [1],

> Every formal basis for probability emphasizes a distinct idea about it. In Kolmogorov's axioms [8], probability is simply a measure on a sample space. High-measure subsets are more probable. In the Bayesian approach of Cox [9], probability is a rational measure of confidence in a proposition. Propositions in which a rational agent is more confident are also more probable. Laplace's early discussion [10] is based on symmetry. Symmetrically equivalent events—two different orderings of a shuffled deck of cards, for instance—are equally probable. (Zurek [11] has used a similar principle of "envariance" to discuss the origin of quantum probabilities.) In algorithmic information theory [7], the algorithmic probability of a bit string is related to its complexity. Simpler bit strings are more probable.

We may remark further that axiom systems like those of Kolmogorov or Cox do not assign actual probabilities in any situation; they simply enforce rules that any such assignment must satisfy. Entropic probability is an actual assignment of a particular distribution, which is determined by objective facts about state transformations.

5. Uniformization

A unique entropic probability rule arises from our $\rightarrow$ relations among eidostates, which in the real world might summarize empirical data about possible state transformations. But so far, this entropic probability distribution $P(e|E)$ is only defined within a uniform eidostate E.

In part, this makes sense. An eidostate represents represents the knowledge of an agent—i.e., that the state must be one of those included in the set. This is the knowledge upon which the agent will assign probabilities, which is why we have indicated the eidostate E as the *condition* for the distribution. Furthermore, these might be the only eidostates, since the axioms themselves do not guarantee that any non-uniform eidostates exist. (Some models of the axioms have them, and some do not.) But can we generalize the probabilities to distributions over non-uniform collections of states?

Suppose $A = \{a, a', \ldots\}$ is a finite set of states that is possibly not uniform. Then, we say that A is *uniformizable* if there exists a uniform eidostate $\hat{A} = \{a + m_a, a' + m_{a'}, \ldots\}$, where the states m_a are mechanical states in $\mathscr{M}$. The idea is that the states in A, which vary in their components of content, can be extended by mechanical states that "even out" these variations. Since $\hat{A}$ is uniform, then $a + m_a \rightleftharpoons a' + m_{a'}$ for any $a, a' \in A$. The abstract process $\langle a, a' \rangle$ is said to be *adiabatically possible* [2]. Mechanical states have $\mathbb{S}(m_a) = 0$, so the entropy of the extended $\hat{A}$ is just

$$\mathbb{S}(\hat{A}) = \log\left(\sum_{a \in A} 2^{\mathbb{S}(a)}\right), \tag{11}$$

which is independent of our choice of the uniformizing mechanical states.

What is the significance of this entropy? Suppose A and B are not themselves uniform, but their union $A \cup B$ is uniformizable. Then, we may construct uniform eidostates $\hat{A} = \{a + m_a, \ldots\}$ and $\hat{B} = \{b + m_b, \ldots\}$ such that either $\hat{A} \rightarrow \hat{B}$ or $\hat{B} \rightarrow \hat{A}$, depending on whether $\mathbb{S}(\hat{A}) \geq \mathbb{S}(\hat{B})$ or the reverse. In short, the entropies of the extended eidostates determine whether the set of states A can be turned into the set B *if* we imagine that these states can be augmented by mechanical states, embedding them in a larger, uniform context.

Given the entropy of the extended state, we can define

$$P(a|A) = P(a + m_a|\hat{A}) = \frac{2^{\mathbb{S}(a)}}{2^{\mathbb{S}(\hat{A})}}. \tag{12}$$

This extends the entropic probability to the uniformizable set A.

Let us consider an example from our coin-and-box model. We start out with the non-uniform set $B = \{b_n, b_{n+1}\}$. These two basic box states have different numbers of coins. But we can uniformize this set by adding stack states, so that $\hat{B} = \{b_n + s_{m+1}, b_{n+1} + s_m\}$ is a uniform eidostate. The entropy of a basic box state is $\mathbb{S}(b_n) = n$, so we have

$$\mathbb{S}(\hat{B}) = \log\left(2^n + 2^{n+1}\right) = \log(3 \cdot 2^n) = n + \log 3. \tag{13}$$

The entropic probabilities are thus

$$P(b_n|B) = \frac{1}{3} \qquad \text{and} \qquad P(b_{n+1}|B) = \frac{2}{3}. \tag{14}$$

6. Reservoir States

So far, we have uniformized a non-uniform set A by augmenting its elements with mechanical states, which act as a sort of "reservoir" of components of content. These mechanical states have no entropy of their own. But we can also consider a procedure in which the augmenting states act more like the states of a thermal reservoir in conventional thermodynamics.

We begin with a mechanical state μ, and posit a sequence of *reservoir states* θ_n, which have the following properties.

- For any n, $\theta_n + \mu \to \theta_{n+1}$.
- $\theta_k + \theta_l \leftrightarrow \theta_m + \theta_n$ if and only if $k + l = m + n$.

The reservoir states θ_n form a ladder. We can ascend one rung in the ladder by "dissolving" the mechanical state μ into the reservoir. If we have more than one reservoir state, we can ascend one ladder provided we descend another by the same number of rungs.

For any n and m, we have that $\mathbb{S}(\theta_n) + \mathbb{S}(\theta_{m+1}) = \mathbb{S}(\theta_{n+1}) + \mathbb{S}(\theta_m)$, so that

$$\sigma = \mathbb{S}(\theta_{n+1}) - \mathbb{S}(\theta_n) = \mathbb{S}(\theta_{m+1}) - \mathbb{S}(\theta_m) \tag{15}$$

is a non-negative constant for the particular sequence of reservoir states. This sequence $\{\theta_n\}$ is characterized by the state μ and the entropy increment σ. Note that we can write $\mathbb{S}(\theta_n) = n\sigma + S_0$, where $S_0 = \mathbb{S}(\theta_0)$.

For example, in our coin-and-box model, the basic box states b_n act as a sequence of reservoir states with a mechanical (coin) state $\mu = h$ and an entropy increment $\sigma = \log 2 = 1$. The more general box states b_n^K form a reservoir state sequence with $\mu = s_q$ and $\sigma = \log k$, where $q = Q(K)$ and k is the number of states in K. For each of these box-state reservoir sequences, $S_0 = \mathbb{S}(\theta_0) = 0$.

One particular type of reservoir is a *mechanical* reservoir consisting of the states μ, $\mu + \mu$, $\mu + (\mu + \mu)$, etc. We denote the nth such state by μ_n. For the μ_n reservoir states, $\sigma = 0$. If we have a finite set of states $A = \{a, a', \ldots\}$ that can be uniformized by the addition of the μ_n states, they can also be uniformized by a corresponding set of non-mechanical reservoir states θ_n:

$$\hat{A} = \{a + \theta_{n_a}, a' + \theta_{n_{a'}}, \ldots\}. \tag{16}$$

As before, we can find the entropy of this uniform eidostate and define entropic probabilities. But the θ reservoir states now contribute to the entropy and affect the probabilities.

First, the entropy:

$$\mathbb{S}(\hat{A}) = \log\left(\sum_{a\in A} 2^{\mathbb{S}(a)+\mathbb{S}(\theta_{n_a})}\right) = S_0 + \log\left(\sum_{a\in A} 2^{\mathbb{S}(a)} 2^{n_a\sigma}\right). \tag{17}$$

The entropic probability—which now depends on the choice of reservoir states—is

$$P_\theta(a|A) = P(a + \theta_{n_a}|\hat{A}) = \frac{2^{\mathbb{S}(a)+n_a\sigma}}{\left(\sum_{a\in A} 2^{\mathbb{S}(a)+n_a\sigma}\right)}. \tag{18}$$

The reservoir states affect the relative probabilities of the states. For example, suppose $\mathbb{S}(a) = \mathbb{S}(a')$ for a pair of states in A. We might naively think that these states would end up with the same entropic probability, as they would if we uniformized A by mechanical

states. But since we are uniformizing using the θ reservoir states, it may be that θ_{n_a} and $\theta_{n_{a'}}$ have different entropies. Then, the ratio of the probabilities is

$$\frac{P_\theta(a|A)}{P_\theta(a'|A)} = \frac{2^{n_a\sigma}}{2^{n_{a'}\sigma}} = 2^{(n_a - n_{a'})\sigma}, \tag{19}$$

which may be very different from 1.

Again, let us consider our coin-and-box model. We begin with the non-uniform set $A = \{h, t, h+t\}$. Each of these states has the same entropy $\mathbb{S}$, that is, zero. We choose to uniformize using basic box states b_n. For instance, we might have

$$\hat{A} = \{h + b_1, t + b_1, (h+t) + b_0\}. \tag{20}$$

Recalling that $\sigma = 1$, the entropy is

$$\mathbb{S}(\hat{A}) = \log\left(2^1 + 2^1 + 2^0\right) = \log 5. \tag{21}$$

This yields probabilities

$$P_b(h|A) = \frac{2}{5} \qquad P_b(t|A) = \frac{2}{5} \qquad P_b(h+t|A) = \frac{1}{5}. \tag{22}$$

As an illustration of these ideas, consider the version of Maxwell's demon shown in Figure 1.

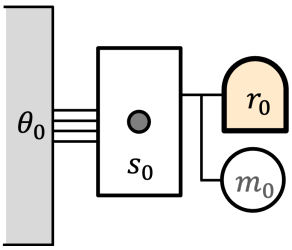


Figure 1. A simple Maxwell's demon.

The demon is a reversible computer with an initial memory state r_0. It is equipped with a reversible battery for storing energy, initially in mechanical state m_0. The demon interacts with a one-particle "Szilard" gas, in which the single particle can move freely within its volume (state s_0). The gas is maintained in thermal equilibrium with a heat reservoir whose initial state is θ_0. We might denote the overall initial state by $((r_0 + m_0) + s_0) + \theta_0$.

Now, the demon introduces a partition into the gas, separating the enclosure into unequal subvolumes, as in Figure 2. The two resulting states are s_a and s_b, which are not equally probable. The probabilities here are entropic probabilities due to the difference in entropy of s_a and s_b.

Now, the demon records the location of the particle in its memory and uses this to control the isothermal expansion of the one-particle gas. The work is stored in the battery. At the end of this process, the demon retains its memory record, and the battery is in one of two mechanical states m_a and m_b. The gas is again in state s_0. But different amounts of heat have been extracted from the reservoir during the expansion, so the reservoir has two different states θ_a and θ_b.

The overall final eidostate might be represented as

$$F = \{((r_a + m_a) + s_0) + \theta_a, ((r_b + m_b) + s_0) + \theta_b\}. \tag{23}$$

The states of the demon and the gas, $(r_a + m_a) + s_0$ and $(r_b + m_b) + s_0$, have different energies and the same entropy. It is the reservoir states θ_a and θ_b that (1) make F uniform (constant energy) and (2) introduce the entropy differences leading to different entropic probabilities for the two states.

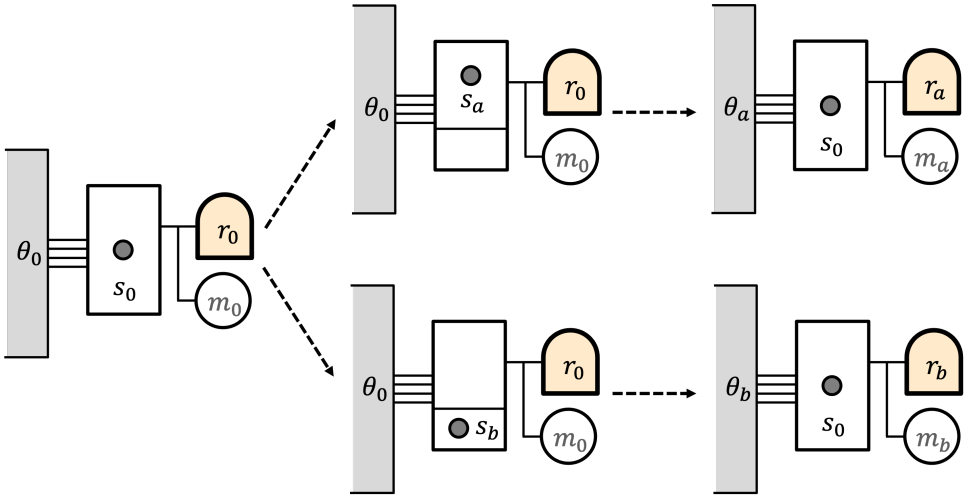


Figure 2. Extraction of work by dividing gas enclosure into unequal volumes.

A conventional view would suppose that the unequal probabilities for the two final demon states comes from their history—that is, that the probabilities are inherited from the unequal partition of the gas. In the entropic view, the unequal probabilities are due to differences in the *environment* of the demon, which is represented by the different reservoir states θ_a and θ_b. The environment, in effect, serves as the "memory" of the history of the process.

7. Context States

When we uniformize a non-uniform A by means of a sequence of reservoir states, the reservoir states affect the entropic probabilities. We can use this idea more generally.

For example, in our coin-and-box model, suppose we flip a coin but do not know how it lands. This might be represented by the eidostate $F = \{h, t\}$. Without further information, we would assign the coin states equal probability $1/2$, which is the simple entropic probability. But suppose we have additional information about the situation that would lead us to assign probabilities $1/3$ and $2/3$ to the coin states. This additional information—this *context*—must be reflected in the eidostate. The example in Equation (14) tells us that this does the job:

$$\hat{F} = \{h + (b_n + s_{m+1}), t + (b_{n+1} + s_m)\}. \tag{24}$$

The extended coin-flip state $\hat{F}$ includes extra context so that the entropic probability reflects our additional information.

In general, we can adjust our entropic probabilities by incorporating *context states*. Suppose we have a uniform eidostate $E = \{e_1, e_2, \ldots\}$, but we wish to *specify* a particular non-entropic distribution p_k over these states. Then, for each e_k, we introduce eidostates C_k, leading to an extended eidostate

$$\hat{E} = \bigcup_k (e_k + C_k), \tag{25}$$

which we assume is uniform. The C_k values are the context states. Our challenge is to find a set of context states so that the entropic probability in $\hat{E}$ equals the desired distribution p_k.

We cannot always do this exactly, but we can always approximate it as closely as we like. First, we note that we can always choose our context eidostates to be information states. The information state I_n containing n record states has entropy $\log n$. Now, for each k, we closely approximate the ratio $p_k/2^{\mathbb{S}(e_k)}$ by a rational number; and since there are finitely many of these numbers, we can represent them using a common denominator. In our approximation,

$$\frac{p_k}{2^{\mathbb{S}(e_k)}} = \frac{n_k}{N}. \tag{26}$$

Now, choose $C_k = I_{n_k}$ for each k. The entropy of $\hat{E}$ becomes

$$\begin{aligned} \mathbb{S}(\hat{E}) &= \log\left(\sum_k 2^{\mathbb{S}(e_k+\log n_k)}\right) \\ &= \log\left(\sum_k n_k 2^{\mathbb{S}(e_k)}\right) = \log\left(\sum_k p_k N\right) = \log N. \end{aligned} \tag{27}$$

From this, we find that the entropic probability is

$$P(e_k + I_k|\hat{E}) = \frac{n_k 2^{\mathbb{S}(e_k)}}{N} = p_k, \tag{28}$$

as desired.

We find, therefore, that the introduction of context states C_k allows us to "tune" the entropic probability to approximate any distribution p_k that we like. This is more than a trick. The distribution p_k represents additional implicit information (beyond the mere list of states $E = \{e_k\}$), and such additional information must have a physical representation. The context states are that representation.

8. Free Energy

The tools we have developed can lead to some interesting places. Suppose we have two sets of states, $A = \{a_i\}$ and $B = \{b_j\}$, endowed with a priori probability distributions p_i and q_j, respectively. We wish to know when the states in A can be turned into the states in B, which are perhaps augmented by reservoir states. That is, we wish to know when $\hat{A} \to \hat{B}$.

We suppose we have a mechanical state μ, leading to a ladder of mechanical reservoir states $\mu_n = \mu + (\mu + \ldots)$. The mechanical state μ is non-trivial in the the sense that $s + \mu \not\to s$ for any s. This means that there is a component of content Q such that $Q(\mu) \neq 0$. The set $A \cup B$ can be uniformized by augmenting the a_i and b_j states by μ_n mechanical reservoir states.

However, we still need to realize the p_i and q_j probabilities. We do this by introducing as context states a corresponding ladder of reservoir states θ_n such that $\sigma = \mathbb{S}(\theta_{n+1}) - \mathbb{S}(\theta_n)$ is very small. Essentially, we assume that the reservoir states are "fine-grained" enough that we can approximate any positive number by $2^{n\sigma}$ for some positive or negative integer n. Then, if we augment the a_i and b_j states by combinations of μ_n and θ_n states, we can uniformize $A \cup B$ and also tune the entropic probabilities to match the a priori p_i and q_j. The final overall uniform eidostate is

$$\{a_i + (\mu_{l_i} + \theta_{k_i}), b_j + (\mu_{h_j} + \theta_{n_j})\}, \tag{29}$$

for integers l_i, k_i, h_j and n_j. The uniformized $\hat{A}$ and $\hat{B}$ eidostates are subsets of this and thus are themselves uniform eidostates. The entropic probabilities have been adjusted so that

$$p_i = \frac{2^{\mathbb{S}(a_i)+k_i\sigma+\mathbb{S}(\theta_0)}}{2^{\mathbb{S}(\hat{A})}} \quad \text{and} \quad q_j = \frac{2^{\mathbb{S}(b_j)+n_j\sigma+\mathbb{S}(\theta_0)}}{2^{\mathbb{S}(\hat{B})}}. \tag{30}$$

We now choose a component of content Q such that $Q(\mu) = \varepsilon > 0$. Since the overall state is is uniform, it must be true that

$$Q(a_i) + l_i\varepsilon + k_i\varepsilon = Q(b_j) + h_j\varepsilon + n_j\varepsilon = \text{constant} \tag{31}$$

for all choices of i, j. Of course, if all of these values are the same, we can average them together and obtain

$$\langle Q(a_i)\rangle_p + \langle l_i\rangle_p\varepsilon + \langle k_i\rangle_p\varepsilon = \langle Q(b_j)\rangle_q + \langle h_j\rangle_q\varepsilon + \langle n_j\rangle_q\varepsilon. \tag{32}$$

We can write the average change in the Q-value of the mechanical state as

$$\Big(\langle h_j\rangle_q - \langle l_i\rangle_p\Big)\varepsilon = \Big(\langle k_i\rangle_p - \langle n_j\rangle_q\Big)\varepsilon + \langle Q(a_i)\rangle_p - \langle Q(b_j)\rangle_q. \tag{33}$$

Since all of the states lie within the same uniform eidostate, $\hat{A} \to \hat{B}$ if and only if $\mathbb{S}(\hat{A}) \leq \mathbb{S}(\hat{B})$—that is,

$$H(\vec{p}) + \langle \mathbb{S}(a_i)\rangle_p + \langle k_i\rangle_p\sigma \leq H(\vec{q}) + \langle \mathbb{S}(b_j)\rangle_q + \langle n_j\rangle_q\sigma. \tag{34}$$

From this, it follows that

$$\Big(\langle k_i\rangle_p - \langle n_j\rangle_q\Big) \leq \frac{1}{\sigma}\Big(H(\vec{q}) - H(\vec{p}) + \langle \mathbb{S}(b_j)\rangle_q - \langle \mathbb{S}(a_i)\rangle_p\Big). \tag{35}$$

If we substitute this inequality into Equation (33), we obtain

$$\begin{aligned}\Big(\langle h_j\rangle_q - \langle l_i\rangle_p\Big)\varepsilon - \frac{\varepsilon}{\sigma}(H(\vec{q}) - H(\vec{p})) \quad \leq \quad & +\frac{\varepsilon}{\sigma}\Big(\langle \mathbb{S}(b_j)\rangle_q - \langle \mathbb{S}(a_i)\rangle_p\Big) \\ & -\Big(\langle Q(b_j)\rangle_q - \langle Q(a_i)\rangle_p\Big).\end{aligned} \tag{36}$$

We can obtain insight into this expression as follows. Given the process $\hat{A} \to \hat{B}$,

- $\Big(\langle h_j\rangle_q - \langle l_i\rangle_p\Big)\varepsilon$ is the average increase in Q-value of the mechanical state, which we can call $\langle \Delta Q_\mu \rangle$. Intuitively, this might be regarded as the "work" stored in the $\hat{A} \to \hat{B}$ process.
- We can denote the change in the Shannon entropy of the probabilities by $\Delta H = H(\vec{q}) - H(\vec{p})$. Since each a_i or b_j state could be augmented by a corresponding record state, this is the change in the information entropy of the stored record.
- For each state a, we can define the *free energy* $F(a) = Q(a) - \frac{\varepsilon}{\sigma}\mathbb{S}(a)$. We call this free "energy", even though Q does not necessarily represent energy, because of the analogy with the familiar expression $F = E - TS$ for the Helmholtz free energy in conventional thermodynamics. The average change in the free energy F is

$$\langle \Delta F\rangle = \Big(\langle Q(b_j)\rangle_q - \langle Q(a_i)\rangle_p\Big) - \frac{\varepsilon}{\sigma}\Big(\langle \mathbb{S}(b_j)\rangle_q - \langle \mathbb{S}(a_i)\rangle_p\Big). \tag{37}$$

The free energy F depends on the particular reservoir states θ_n only via the ratio ε/σ. Given this value, $\langle \Delta F\rangle$ depends only on the a_i and b_j states, together with their a priori probabilities.

- To return to our coin-and-box example, suppose we use the basic box states b_n as reservoir states θ_n, and we choose the coin number Q as our component of content. Then, $\varepsilon = 1$ and $\sigma = 1$, so that the free energy function $F(a) = Q(a) - \mathbb{S}(a)$. (If we use different box states b_n^K as reservoir states, the ratio ε/σ is different.)

With these definitions, Equation (36) becomes

$$\langle \Delta Q_\mu \rangle - \frac{\varepsilon}{\sigma} \Delta H \leq -\langle \Delta F \rangle. \tag{38}$$

Increases in the average stored mechanical work, and decreases in the stored information, must be paid for by a corresponding decrease in the average free energy.

Many useful inferences can be drawn from this. For example, the erasure Q-cost of one bit of information in the presence of the θ-reservoir is ε/σ. This cost can be paid from either the mechanical Q-reservoir state, the average free energy, or from a combination of these. This amounts to a very general version of Landauer's principle [12]: one that involves any type of mechanical component of content.

Author Contributions: Conceptualization, B.S. and M.D.W.; Investigation, B.S. and M.D.W.; Writing—original draft, B.S. and M.D.W.; Writing—review & editing, B.S. and M.D.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors gratefully acknowledge the comments of the referees, which have substantially improved the clarity and self-sufficiency of this paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Hulse, A.; Schumacher, B.; Westmoreland, M.D. Axiomatic Information Thermodynamics. *Entropy* **2018**, *20*, 237. [CrossRef] [PubMed]
2. Giles, R. *Mathematical Foundations of Thermodynamics*; Pergamon Press Ltd.: Oxford, UK, 1964.
3. Lieb, E.H.; Yngvason, J. A Guide to Entropy and the Second Law of Thermodynamics. *Not. Am. Math. Soc.* **1998**, *45*, 571–581.
4. Szilard, L. On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings. *Z. Phys.* **1929**, *53*, 840–856; English translation in *Behav. Sci.* **1964**, *9*, 301–310. [CrossRef]
5. Bennett, C.H. The Thermodynamics of Computation—A Review. *Int. J. Theor. Phys.* **1982**, *21*, 905–940. [CrossRef]
6. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. 623–656. [CrossRef]
7. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; John Wiley and Sons: Hoboken, NJ, USA, 2006.
8. Kolmogorov, A.N. *Foundations of the Theory of Probability*, 2nd ed.; Chelsea: New York, NY, USA, 1956.
9. Cox, R.T. *The Algebra of Probable Inference*; Johns Hopkins Press: Baltimore, MD, USA, 1961.
10. de Laplace, P.S. *A Philosophical Essay on Probabilities*; Truscott, F.W., Emory, F.L., Translators; Dover: New York, NY, USA, 1951.
11. Zurek, W.H. Probabilities from entanglement, Born's rule $p_k = |\psi_k|^2$ from envariance. *Phys. Rev. A* **2005**, *71*, 052105. [CrossRef]
12. Landauer, R. Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.* **1961**, *5*, 183–191. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.