2013

Conditioning, Correlation and Entropy Generation in Maxwell's Demon

Neal G. Anderson
University of Massachusetts Amherst, anderson@ecs.umass.edu

Entropy **2013**, *15*, 4243-4265; doi:10.3390/e15104243

OPEN ACCESS

entropy

ISSN 1099-4300

www.mdpi.com/journal/entropy

Article

Conditioning, Correlation and Entropy Generation in Maxwell's Demon

Neal G. Anderson

Department of Electrical and Computer Engineering, University of Massachusetts Amherst, Amherst, MA 01003-9292, USA; E-Mail: anderson@ecs.umass.edu; Tel.: +1-413-545-0765; Fax: +1-413-545-4611

Received: 29 July 2013; in revised form: 22 September 2013 / Accepted: 24 September 2013 / Published: 9 October 2013

Abstract: Maxwell's Demon conspires to use information about the state of a confined molecule in a Szilard engine (randomly frozen into a state subspace by his own actions) to derive work from a single-temperature heat bath. It is widely accepted that, if the Demon can achieve this at all, he can do so without violating the Second Law only because of a counterbalancing price that must be paid to erase information when the Demon's memory is reset at the end of his operating cycle. In this paper, Maxwell's Demon is analyzed within a "referential" approach to physical information that defines and quantifies the Demon's information via correlations between the joint physical state of the confined molecule and that of the Demon's memory. On this view, which received early emphasis in Fahn's 1996 classical analysis of Maxwell's Demon, information is erased *not* during the memory reset step of the Demon's cycle, but rather during the expansion step, when these correlations are destroyed. Dissipation and work extraction are analyzed here for a Demon that operates a generalized quantum mechanical Szilard engine embedded in a globally closed composite, which also includes a work reservoir, a heat bath and the remainder of the Demon's environment. Memory-engine correlations lost during the expansion step, which enable extraction of work from the Demon via operations conditioned on the memory contents, are shown to be dissipative when this decorrelation is achieved unconditionally so no work can be extracted. Fahn's essential conclusions are upheld in generalized form, and his quantitative results supported via appropriate specialization to the Demon of his classical analysis, all without external appeal to classical thermodynamics, the Second Law, phase space conservation arguments or Landauer's Principle.

Keywords: Maxwell's Demon; Szilard engine; correlation entropy; mutual information; thermodynamics of computation

---

1. Introduction and Background

Maxwell's fictitious Demon [1] has been analyzed from many points of view since the Demon's "birth" in 1871 ([2–4] and references therein). The Demon's notoriety derives from his alleged capacity to convert heat from a single-temperature source into work (in apparent violation of the Second Law) by acquiring and using information about the randomly evolving microstate of an engine's working substance. Szilard's 1929 analysis [5] and other influential studies kept attention focused on measurement costs as the likely bulwark preventing the Demon from violating the Second Law, and measurement was widely regarded as the culprit for most of the Demon's life. Understanding of the Demon took a crucial turn, however, with the work of Bennett [6], who argued that the Demon could *reversibly* acquire information about the state of a molecule in a Szilard engine, but that the act of resetting the Demon's memory—required for *sequences* of reversible measurements—is necessarily dissipative. Applying Landauer's Principle [7] to the reset step of a Szilard-engine-based Demon utilizing reversible measurement, Bennett showed that the dissipative cost of this step was sufficient to counterbalance the work that the Demon is able to produce. Information erasure has since been widely (if not universally) credited with the salvation of the Second Law within the context of Maxwell's Demon.

In a 1996 paper [8], Paul Fahn put a twist on Bennett's analysis that has implications for this diagnosis of the Demon's inability to beat the Second Law. Fahn regarded information erasure and memory reset (widely regarded as synonymous) as distinct processes that can occur in separate steps of the Demon's cycle. To introduce this distinction and its consequences, both of which are central concerns of this work, we provide a brief sketch of Fahn's analysis.

Fahn, like Bennett, used reversible classical thermodynamics to analyze a Demon, $\mathcal{D}$, consisting of a Szilard engine and a one-bit memory. The Szilard engine consists of a single molecule in a closed cylinder that is equipped with a removable partition and frictionless pistons that can be selectively inserted into the cylinder from the left and right. The memory has three (equal entropy) states: two "data" states (labeled "L" and "R") that can be used to register outcomes of a coarse-grained measurement of the molecule position, and a reset state labeled "0." The Demon, in perpetual contact with its environment (a heat bath, $\mathcal{B}$, at temperature T), goes through a cycle that converts heat from the bath into work. The Demon's cycle, which is comprised of a sequence of steps that can generally change the entropy of the Demon, the entropy of the bath and the total (Demon+bath) entropy, proceeds as follows:

Initial: The cycle begins with the molecule moving freely in the unpartitioned cylinder and the memory in the "0" state.

A. Insertion: The partition is inserted, trapping the molecule in either the left or right half of the cylinder. This process is considered to be both energetically and entropically benign; the partition is considered to be frictionless, and the entropy reduction that results from halving of the molecule's accessible volume is exactly balanced by the increase that results from uncertainty about *which* volume (left or right) holds the molecule. (This assignment of entropy has been controversial, as will be discussed in Section 4.1.). Both the memory and environment are assumed to be unaffected.

B. Measurement: A reversible position measurement senses the location of the molecule and registers the outcome in the memory ("L" for left half and "R" for right half, respectively) bringing the states of the memory and molecule into correlation. The molecule and environment are assumed to be unaffected. The local entropy of the memory is increased, since it reflects the randomness of the molecule position, but the entropy of the Demon as a whole is unchanged, because of the correlation created between the molecule and memory states.

C. Expansion: One of the two pistons is selected for insertion based on the memory state (right piston for "L" outcome and left piston for "R" outcome). The piston is slid up to the partition, the partition is removed, and the molecule is allowed to push the piston back to the end of the cylinder and do reversible work in an isothermal expansion process. The molecule is returned to its initial state in this process, and its local entropy is unchanged (*cf.* Step A above), while the loss of correlation between the molecule and the memory increases the entropy of the Demon as a whole by $k_B \ln(2)$. The environment entropy is, by classical thermodynamics, reduced by precisely this same amount as a result of the heat delivered to the molecule during the volume-doubling isothermal expansion, so there is no total entropy change. The memory is unaffected in this step. We should note that although Fahn took the piston insertion, partition removal, and gas expansion to three distinct steps in his work, we have combined this sequence of three processes into a single "Expansion" step for consistency with our treatment of a generalized Demon.

D. Memory Reset: The Demon's memory is unconditionally reset, reducing its local entropy—and the entropy of the Demon as a whole—by $k_B \ln(2)$ and returning the Demon to its initial state. Fahn attributes an equal and opposite entropy increase to the environment, citing conservation of phase-space volume and its consistency with the environmental cost of "Landauer erasure" ($k_B \ln(2)$ Joule per Kelvin-bit). This completes the Demon's cycle.

Table 1 summarizes step-by-step entropy changes in the Demon ($\Delta S^{\mathcal{D}}$), the bath ($\Delta S^{\mathcal{B}}$) and their sum ($\Delta S_{tot} = \Delta S^{\mathcal{D}} + \Delta S^{\mathcal{B}}$) according to Fahn's analysis. Following Fahn, all entropies in Table 1 have been expressed in "information theoretic" units. (One bit of entropy in information-theoretic units corresponds to $k_B \ln(2)$ Joules per Kelvin in "thermodynamic" units.) Work has been produced, but the Second Law has not been violated, since ΔS_{tot} is nondecreasing for the cycle. This is entirely consistent with Bennett's final result and with the bulk of current "conventional wisdom" regarding the Demon's capabilities and resource requirements.

What differs in Fahn's analysis is the increase in the Demon's entropy that occurs in the expansion step from the loss of the molecule-memory correlations, and his association of this entropy increase with information erasure. This association is perhaps counterintuitive, since the memory remains in one of the "data states" (L or R) until it is reset in the subsequent *memory reset* step, which might then seem obviously to be the information erasure step. Fahn, however, adopts the point of view that information is bound up in the molecule-memory correlations, since the memory holds information *about* the molecule position only when these correlations exist. These information-bearing correlations are lost during the expansion step, and once they are "gone", the memory state can no longer refer to anything having to do with the molecule's state.

Table 1. Entropy changes throughout the Demon's cycle, according to Fahn's analysis (adapted from [8]).

Step	Demon $\Delta S^{\mathcal{D}}$	Bath $\Delta S^{\mathcal{B}}$	Total ΔS_{tot}
A. Insertion	0	0	0
B. Measurement	0	0	0
C. Expansion	$+1$	-1	0
D. Memory Reset	-1	$+1$	0
Cycle Total	0	0	0

In this work, we offer a new analysis, based on our "referential approach" to physical information theory [9–12], to a Demon that operates a generalized quantum-mechanical Szilard engine in a cycle analogous to Fahn's. The referential approach is based on an overtly relational conception of information, similar in spirit to that adopted by Fahn in his analysis of the Demon, which views physical information *exclusively* in terms of correlation (*i.e.*, as *mutual* information) and is well suited for fundamental analyses of dissipation in information processing scenarios. The Szilard engine used by our Demon is much more general than that used by the Demon of Fahn's analysis, but includes the engine of Fahn's analysis as a special case. Our analysis arrives at Fahn's broad conclusions via a very different route and lends support to several of his quantitative results through appropriate specialization. We note that although we view the Demon through a generalization of Fahn's analysis in this work, and although Fahn's explicit association of information with system-memory correlations is by far the exception in Maxwell's Demon studies, an emphasis on mutual information can be found in other studies of Demons and Szilard engines. These include a few studies that predate Fahn (e.g., [13,14]) and some very recent work (e.g., [15–19]).

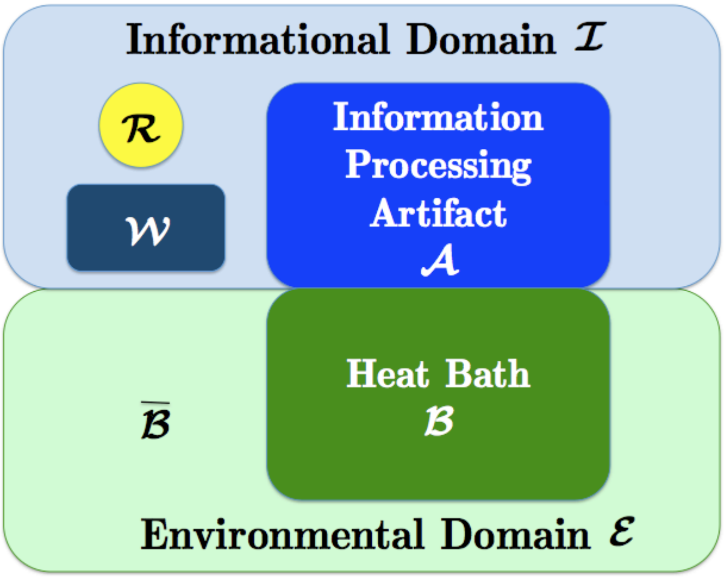
The remainder of this paper is organized as follows: In Section 2, we sketch the features of the referential approach germane to this work and show how it is used to describe elementary operations like those performed during the Demon's cycle. A Demon utilizing a generalized Szilard engine is then described within the referential approach in Section 3, where the full operating cycle is described, step-by-step entropy changes are tabulated and extraction of work from the Demon is analyzed. Several aspects of our analysis and its connection to Fahn's work, some of which are of broader significance for studies of Maxwell Demons, are discussed in Section 4. The generality of our analysis is highlighted, as is its resistance to criticisms leveled against some classical analyses (Section 4.1). Results of Fahn's analysis are recovered from our generalized Demon through appropriate specialization (Section 4.2), and a conjecture of Fahn's regarding the entropy cost of information is considered within our approach (Section 4.3). Finally, the implications of correlations and interactions that lie beyond the Demon-bath composite, but that are naturally captured in our referential analysis, are discussed (Section 4.4). The paper concludes in Section 5.

2. Referential Information and Entropy Generation in Physical Processes

2.1. Referential Physical Information

We begin by sketching the "referential" approach to physical information theory [9–12] that we use to analyze Maxwell's Demon in this work and by introducing key definitions and notation. The referential approach aims to provide a fundamental physical description of classical information processing and its physical costs in globally closed, multicomponent quantum systems. The generic multicomponent system configuration relevant to the present work is depicted in Figure 1.

Figure 1. Generic multicomponent physical system, built around an information-processing artifact, $\mathcal{A}$, and heat bath $\mathcal{B}$, used here to obtain bounds on the dissipative costs of processing information about $\mathcal{R}$ in $\mathcal{A}$. The composite $\mathcal{IE}$ is globally closed.



A physical system of interest for informational analysis—called an *information processing artifact* $\mathcal{A}$—belongs (by definition) to part of the *informational domain*, $\mathcal{I}$, of the multicomponent composite, as does a *referent system*, $\mathcal{R}$, about which $\mathcal{A}$ may bear information. Additional systems that can interact with $\mathcal{R}$ and $\mathcal{A}$ may also belong to $\mathcal{I}$ (e.g., $\mathcal{W}$ in Figure 1). The complementary *environmental domain*, $\mathcal{E}$, consists of the part, $\mathcal{B}$ (typically a heat bath), of $\mathcal{E}$ that can interact directly with $\mathcal{A}$. Furthermore, included in $\mathcal{E}$ is the "remote environment", $\bar{\mathcal{B}}$, which can interact with $\mathcal{B}$ (but not $\mathcal{A}$) and can "rethermalize" $\mathcal{B}$ whenever it is driven from thermal equilibrium. $\bar{\mathcal{B}}$ includes everything required to ensure that the full composite, $\mathcal{IE}$, is globally closed.

Two signature features characterize the referential approach. First, information is regarded as a relational property of two systems. Any attribution of "information content" to $\mathcal{A}$ requires specification of a "referent" system, $\mathcal{R}$, about which $\mathcal{A}$ is taken to bear information, and the amount of information *about* $\mathcal{R}$ that is in $\mathcal{A}$—hereafter, the "$\mathcal{R}$-information" in $\mathcal{A}$—is exclusively associated with correlations between the statistical physical states of $\mathcal{R}$ and $\mathcal{A}$. This is to say that, in the referential approach, "information" can only mean *mutual* information. Second, the $\mathcal{R}$–$\mathcal{A}$ correlations are regarded as "information" *if and only if* $\mathcal{R}$ is, like $\mathcal{A}$, a part of $\mathcal{I}$; correlations between $\mathcal{A}$ and subsystems of the "environmental domain", $\mathcal{E}$, are not granted this status. This distinction is largely terminological—the mutual information measure is used to quantify correlations between subsystems both within and across the two domains—but it gives special significance to correlations that do not cross the boundary between $\mathcal{I}$ and $\mathcal{E}$. This sort of distinction is routine in familiar contexts: Components of an electromagnetic field at a receiver that was excited by a transmitter are, for example, granted the status of "signal," while field components correlated to other (environmental) sources are regarded as "noise." Here, we choose the $\mathcal{I}$-$\mathcal{E}$ boundary so $\mathcal{I}$ includes only systems over which the Demon has direct control in at least part of his operating cycle.

The referential approach is formalized within non-relativistic quantum theory as follows: First, as per the fundamental quantum postulates, state description and dynamics are as follows:

- **States**: The quantum state of the global system ($\mathcal{IE}$) at any time, t, is described by a density operator, $\hat{\rho}^{\mathcal{IE}}(t) \equiv \hat{\rho}(t)$, defined on the universal Hilbert space $\mathcal{H} = \mathcal{H}^{\mathcal{I}} \otimes \mathcal{H}^{\mathcal{E}}$. Local states of subsystems are described by reduced density operators obtained via the partial trace rule, e.g., so the local state of $\mathcal{RA}$ is:
$$\hat{\rho}^{\mathcal{RA}}(t) = Tr_{\mathcal{WE}}[\hat{\rho}(t)] \tag{1}$$

- **Dynamics**: Consistent with Schrodinger dynamics, the global state, $\hat{\rho}(t)$, is assumed to evolve unitarily as:
$$\hat{\rho}(t_f) = \hat{U}(t_i, t_f)\hat{\rho}(t_i)\hat{U}(t_i, t_f)^\dagger \tag{2}$$
over any time interval, $[t_i, t_f]$, where:
$$\hat{U}(t_i, t_f) = \mathcal{T}\left[\exp\left(-\frac{i}{\hbar}\int_{t_i}^{t_f} \hat{H}(t')dt'\right)\right] \tag{3}$$
$\hat{H}(t)$ is the global time-dependent Hamiltonian and $\mathcal{T}$ is the time-ordering symbol. Note that the time-dependent *local* states of subsystems (e.g., $\hat{\rho}^{\mathcal{RA}}(t)$) need not, and generally do not, evolve unitarily.

Next, information and entropy are defined and quantified as follows:

- **Information**: The amount of information about $\mathcal{R}$ in $\mathcal{A}$ at time t is quantified by the quantum mutual information (or correlation entropy):
$$I^{\mathcal{RA}}(t) = S(\hat{\rho}^{\mathcal{R}}(t)) + S(\hat{\rho}^{\mathcal{A}}(t)) - S(\hat{\rho}^{\mathcal{RA}}(t)) \tag{4}$$
where $S(\cdot)$ is the von Neumann entropy, defined for any density operator, $\hat{\rho}$, as $S(\hat{\rho}) \equiv -Tr[\hat{\rho}\log_2\hat{\rho}]$, and the local states of $\mathcal{R}$ and $\mathcal{A}$ are given by $\hat{\rho}^{\mathcal{R}}(t) = Tr_{\mathcal{A}}[\hat{\rho}^{\mathcal{RA}}(t)]$ and $\hat{\rho}^{\mathcal{A}}(t) = Tr_{\mathcal{R}}[\hat{\rho}^{\mathcal{RA}}(t)]$, respectively. Note that, for any process in which $\mathcal{R}$ plays the role of a stable referent system, $\hat{\rho}^{\mathcal{R}}(t)$ is assumed to evolve only trivially over the time interval associated with the process.

- **Total Entropy**: The *total* entropy for a pair of subsystems is defined at time t as the sum of local entropies. Three such quantities will, in the context of the Demon, be of interest in this work:
$$S_{tot}^{\mathcal{RA}}(t) = S\left(\hat{\rho}^{\mathcal{R}}(t)\right) + S\left(\hat{\rho}^{\mathcal{A}}(t)\right) \tag{5}$$

$$S_{tot}^{\mathcal{AB}}(t) = S\left(\hat{\rho}^{\mathcal{A}}(t)\right) + S\left(\hat{\rho}^{\mathcal{B}}(t)\right) \tag{6}$$

$$S_{tot}(t) = S\left(\hat{\rho}^{\mathcal{RA}}(t)\right) + S\left(\hat{\rho}^{\mathcal{E}}(t)\right) \tag{7}$$

The total entropy defined for a composite system is not to be confused with the joint entropy, $S(\hat{\rho})$, of the composite taken as a whole. The difference between the two composite system entropies is, in fact, the mutual information reflecting correlations between the states of the two systems (cf. (4)). Joint entropy is conserved under unitary evolution, whereas total entropy is not.

The highly constrained definition of information used in the referential approach, i.e., the association of information exclusively with correlations between the physical states of systems, allows us to define and quantify information in an unambiguous and thoroughly physical manner. The approach was designed for, and is particularly well suited to, problems like Maxwell's Demon that involve both information processing and thermodynamics, since standard thermodynamic process steps are accommodated in a manner that allows information, entropy and system-environment correlations to be tracked through sequences of processes and physical costs to be quantified without external appeal to the Second Law. This is enabled by global closure of the composite $\mathcal{IE}$ together with the embedding of $\mathcal{A}$'s immediate surroundings ($\mathcal{B}$) within a "remote environment" ($\bar{\mathcal{B}}$) that acts to rethermalize $\mathcal{B}$ after it is driven from equilibrium. The resulting ability to describe familiar thermodynamic process steps without losing track of correlations created in previous steps enables the comprehensive analysis of the Demon's cycle presented in Sections 3 and 4. To set the stage for this analysis, we now describe the treatment of process steps within the referential approach.

2.2. Physical Processes in the Referential Approach

Consider a physical process involving a system, $\mathcal{A}$, that (*i*) transforms the state of $\mathcal{A}$ in a specified, generally irreversible, manner and (*ii*) begins and ends with $\mathcal{A}$'s immediate surroundings (here, the bath $\mathcal{B}$) in a thermal state at some temperature T. Such processes are ubiquitous in thermodynamics and arise in a wide variety of contexts, including the operating cycle of Maxwell's Demon. They are, in fact, also incompatible with the global closure of $\mathcal{AB}$: dynamical constraints imposed by Schrodinger evolution require that nonunitary local state transformation of $\mathcal{A}$ achieved through global unitary evolution of $\mathcal{AB}$ drive $\mathcal{B}$ away from its initial thermal state. If an irreversible transformation of $\mathcal{A}$ is to leave $\mathcal{B}$ in the same (thermal) state it was in at the beginning of the process, then $\mathcal{B}$ must be "rethermalized" by a greater environment $\bar{\mathcal{B}}$ within which it is embedded.

This is handled within the referential approach by dividing each such process step into two consecutive phases: a "control" phase and a "restoration" phase. In the control phase, the state of $\mathcal{A}$ is transformed, and $\mathcal{B}$ is generally driven away from its initial thermal state. In the subsequent restoration phase, $\mathcal{B}$ is rethermalized through interaction with the greater environment, $\bar{\mathcal{B}}$. The global state transformation for a process step occurring over the time interval $[t_i, t_f]$ is thus of the form:

$$\hat{\rho}(t_f) = \hat{U}(t_i, t_f)\hat{\rho}(t_i)\hat{U}(t_i, t_f)^{\dagger} = \hat{U}(t_c, t_f)\hat{U}(t_i, t_c)\hat{\rho}(t_i)\hat{U}(t_i, t_c)^{\dagger}\hat{U}(t_c, t_f)^{\dagger} \tag{8}$$

or:

$$\hat{\rho}(t_f) = \hat{U}(t_c, t_f)\tilde{\rho}(t_c)\hat{U}(t_c, t_f)^{\dagger} \tag{9}$$

where:

$$\tilde{\rho}(t_c) = \hat{U}(t_i, t_c)\hat{\rho}(t_i)\hat{U}(t_i, t_c)^{\dagger} \tag{10}$$

with:

$$\hat{U}(t_i, t_c) = \hat{U}^{\mathcal{IB}}(t) \otimes \hat{U}^{\bar{\mathcal{B}}} \tag{11}$$

and:

$$\hat{U}(t_c, t_f) = \hat{U}^{\mathcal{I}} \otimes \hat{U}^{\mathcal{B}\bar{\mathcal{B}}} \tag{12}$$

Here $\hat{U}(t_i, t_c)$ and $\hat{U}(t_c, t_f)$ are the unitary evolution operators for the control and restoration phases of the process, respectively. The unitary $\hat{U}^{\mathcal{IB}}(t)$ embodies all time-dependent forces acting on and within $\mathcal{I}$ to transform the local state of $\mathcal{A}$ as required during the control phase, as well as all interactions coupling $\mathcal{A}$ and $\mathcal{B}$ that allow for system-bath energy and entropy exchange. The unitary $\hat{U}^{\mathcal{B}\bar{\mathcal{B}}}$ reflects all interactions between $\mathcal{B}$ and $\bar{\mathcal{B}}$ that, together with the state of $\bar{\mathcal{B}}$, ensure that $\mathcal{B}$ is unconditionally rethermalized in the restoration phase, i.e., that this phase always ends with the system left in a global state of the form $\hat{\rho}(t_f) = \hat{\rho}^{\mathcal{I}\bar{\mathcal{B}}}(t_f) \otimes \hat{\rho}^{\mathcal{B}}_{th}$. ($\hat{U}^{\bar{\mathcal{B}}}$ and $\hat{U}^{\mathcal{I}}$ are benign self-evolution operators.) Note that all energy and entropy flows between $\mathcal{I}$ and $\mathcal{E}$ occur (via $\mathcal{B}$) during (and only during) the control phase of a process, and that all energy and entropy flows between $\mathcal{B}$ and $\bar{\mathcal{B}}$ occur during (and only during) the restoration phase. Note also that any $\mathcal{I}$-$\mathcal{E}$ correlations that are created during the control phase are preserved during the restoration phase, but are necessarily transformed from $\mathcal{I}$-$\mathcal{B}$ correlations to $\mathcal{I}$-$\bar{\mathcal{B}}$ correlations upon rethermalization of $\mathcal{B}$ by $\bar{\mathcal{B}}$. Finally, note that in Equation (10) and hereafter, a tilde will be used to denote states and quantities relevant to the conclusion of the control phase of a process.

The heterogeneous environment model described above enables preservation and tracking of $\mathcal{I}$-$\mathcal{E}$ correlations that influence the total entropy change ΔS_{tot} for processes. The plausibility of the underlying environmental decomposition ($\mathcal{E} = \mathcal{B}\bar{\mathcal{B}}$) and dynamical decomposition ($\hat{U}(t_i, t_f) = \hat{U}(t_c, t_f)\hat{U}(t_i, t_c)$) hinges on the plausibility of a few key assumptions: $\mathcal{B}$ is assumed to have vastly many more degrees of freedom than $\mathcal{A}$; $\bar{\mathcal{B}}$ is assumed to have vastly many more degrees of freedom than $\mathcal{B}$ (as well as the resources to repeatedly rethermalize $\mathcal{B}$); and it must be possible to select the "boundary" between $\mathcal{B}$ and $\bar{\mathcal{B}}$, such that any disturbance of $\mathcal{B}$ by $\mathcal{A}$ does not "reach" $\bar{\mathcal{B}}$ until transformation of the state of $\mathcal{A}$ is complete. We claim that these assumptions are implicit and widely used in analyses of systems coupled to heat baths and that they can be met to an extremely good approximation in realistic system-bath composites with proper selection of the $\mathcal{B}$-$\bar{\mathcal{B}}$ boundary. We also claim that these assumptions (or something very much like them) are *required* if one is to properly describe sequences of generally irreversible processes in a closed system-environment composite that each begin and end with the system's immediate surroundings in the same thermal state (as widely assumed in thermodynamic analyses). The heterogeneous environment model allows the description of processes involving $\mathcal{AB}$ that transform the state of $\mathcal{A}$ in a generally irreversible manner, while beginning *and* ending with $\mathcal{B}$ in a thermal state, with resulting changes in system-environment correlations fully preserved. (See [11] for further discussion.)

Bounds on the physical costs of processes like those described above can be determined from key aspects of the initial and final states, specification of those subsystems of $\mathcal{I}$ that can interact during the two phases of the process and fundamental physical constraints imposed by the unitary nature of time evolution in globally closed composite systems. This is the basis for our analysis of the Demon.

3. Maxwell's Demon with a Generalized Szilard Engine

3.1. Description of the Demon in the Referential Approach

The Demon, $\mathcal{D}$, of this work is a composite quantum system consisting of a closed quantum subsystem, $\mathcal{S}$, that functions as the "working substance" of a generalized Szilard engine and a memory subsystem, $\mathcal{M}$. The Demon performs sequences of operations corresponding to Steps A–D of the cycle described in Section 1. The role of the artifact, $\mathcal{A}$, in Figure 1 is played by $\mathcal{S}$ in Step A, by $\mathcal{MS}$ in Step B, by $\mathcal{S}$ in Step C and by $\mathcal{M}$ in Step D, and the role of the referent, $\mathcal{R}$, is played by $\mathcal{M}$ in Step C. Each operation involves application of a time-dependent potential to the subsystem playing the role of $\mathcal{A}$. In Step C, this potential is conditioned on the state of the memory, $\mathcal{M}$, which plays the role of the referent, $\mathcal{R}$, in this step, while a pre-defined potential is unconditionally applied in all other steps.

In the expansion step of the cycle, the Demon delivers work to a "work reservoir" $\mathcal{W}$ while interacting with the heat bath $\mathcal{B}$, as when an expanding gas ($\mathcal{S}$) drawing heat from a bath ($\mathcal{B}$) displaces a piston that lifts a suspended weight ($\mathcal{W}$). We allow interactions with $\mathcal{W}$ to change its energy, but not its entropy (a defining characteristic of work reservoirs (cf. [20])) and include an additional condition that $\mathcal{W}$ cannot hold or gain information about the Demon's state. An example of a system that would be *disqualified* as a work reservoir by these criteria would be a pair of identical weights, each of which can be lifted by one of the two pistons in Fahn's Demon. Although symmetry dictates that the amount of work delivered to the two-weight system would not depend on which of the two weights were lifted (i.e., on which side of the cylinder the molecule is trapped in Step A), the final state of the two-weight system would belie the pre-expansion state of the molecule. This amounts to a transfer of information about $\mathcal{S}$ to the two-weight system, and an agent wishing to tap this work would have to possess this information to know how to extract it. This requirement is incompatible with the notion of a "pure" work reservoir, so we do not consider it as such. A slightly less restrictive definition of a work reservoir could allow for some entropy transfer between $\mathcal{S}$ and $\mathcal{W}$, provided that what the entropy transfer does is not accompanied by transfer of $\mathcal{M}$-information to $\mathcal{W}$, but we adopt the more restrictive definition in this work.

With this, we describe the Demon's full cycle and analyze the associated physical costs via the referential approach. For comparison with Fahn's analysis, which explicitly concerns the Demon-bath composite, our initial focus is on the Demon-bath subsystem. The full composite will be considered in Section 4.4.

3.2. The Demon's Cycle

For each step of the Demon's cycle, we now consider the states of the Demon $\mathcal{D} = \mathcal{MS}$, the bath, $\mathcal{B}$, and the Demon-bath composite, $\mathcal{DB}$, during the control phase of each process step, during which $\mathcal{DB}$ evolves unitarily. As mentioned previously, all energy exchange between the Demon and bath and associated entropy changes occur during this phase of each process. Bath rethermalization ensures that the bath is always in a thermal state at the beginning of each process step, so the initial state of $\mathcal{DB}$ for the η-th step is always of the form $\tilde{\rho}^{\mathcal{D}}_{\eta-1} \otimes \hat{\rho}^{\mathcal{B}}_{th}$. States associated with completion of the control phase of the η-th step are denoted with tildes (e.g., as $\tilde{\rho}^{\mathcal{D}}_{\eta}$).

Initial: The cycle begins at time $t = 0$ with the memory and engine system in standard states, $\hat{\rho}_0^{\mathcal{M}}$ and $\hat{\rho}_0^{\mathcal{S}}$, respectively. The bath is in a thermal state, $\hat{\rho}_{th}^{\mathcal{B}}$, and all three subsystems are uncorrelated. The initial states of $\mathcal{D}$, $\mathcal{B}$ and $\mathcal{DB}$ are thus:

$$\hat{\rho}_0^{\mathcal{D}} = \hat{\rho}_0^{\mathcal{M}} \otimes \hat{\rho}_0^{\mathcal{S}} \tag{13}$$

$$\hat{\rho}_0^{\mathcal{B}} = \hat{\rho}_{th}^{\mathcal{B}} \tag{14}$$

$$\hat{\rho}_0^{\mathcal{DB}} = \hat{\rho}_0^{\mathcal{M}} \otimes \hat{\rho}_0^{\mathcal{S}} \otimes \hat{\rho}_{th}^{\mathcal{B}} \tag{15}$$

Since the states of $\mathcal{M}$ and $\mathcal{S}$ are uncorrelated, there is initially no "$\mathcal{S}$-information", i.e., information about the system state, in the memory ($I_0^{\mathcal{MS}} = 0$).

A. Insertion: In the "insertion" step ($0 < t \leq t_A$), the self-Hamiltonian of the Demon is unconditionally varied from $\hat{H}_0^{\mathcal{D}}$ to $\hat{H}_A^{\mathcal{D}}$. We do not make any *a priori* assumptions about the form of either Hamiltonian, but require that any potential barriers included in $\hat{H}_A^{\mathcal{D}}$ are finite if arbitrarily high. (This is important for reasons to be discussed in Section 4.1.) We do not assume that this potential "insertion" is benign; we allow that the insertion potential generally changes both the energy and entropy of $\mathcal{S}$ and, because $\mathcal{S}$ and $\mathcal{B}$ remain in contact throughout the process, the energy and entropy of the environment as well.

The unitary evolution operator for the control phase of the insertion step is of the form:

$$\hat{U}_A^{\mathcal{IB}}(t) = \hat{I}^{\mathcal{M}} \otimes \hat{U}_A^{\mathcal{SB}}(t) \otimes \hat{I}^{\mathcal{W}} \tag{16}$$

The states of $\mathcal{D}$ and $\mathcal{DB}$ after the control phase of Step A are thus:

$$\tilde{\rho}_A^{\mathcal{D}} = \hat{\rho}_0^{\mathcal{M}} \otimes \tilde{\rho}_A^{\mathcal{S}} \tag{17}$$

$$\tilde{\rho}_A^{\mathcal{DB}} = \hat{\rho}_0^{\mathcal{M}} \otimes \tilde{\rho}_A^{\mathcal{SB}} \tag{18}$$

with $Tr_{\mathcal{B}}[\tilde{\rho}_A^{\mathcal{SB}}] = \hat{\rho}_A^{\mathcal{S}}$.

B. Measurement: In the "measurement" step ($t_A < t \leq t_B$), $\mathcal{MS}$ unconditionally undergoes a transformation that is equivalent to a quantum measurement with discrete outcomes, with $\mathcal{S}$ and $\mathcal{M}$ playing the roles of "measured system" and "apparatus," respectively. The i-th of these possible outcomes obtains with a probability p_i, which depends on the initial state, $\hat{\rho}_A^{\mathcal{S}}$, and the nature of the measurement interaction and is registered in the memory subsystem, $\mathcal{M}$, by leaving this system in state $\hat{\rho}_i^{\mathcal{M}}$. When the i-th outcome *is* realized, the subsystem, $\mathcal{S}$, is left in a corresponding post-measurement state, $\hat{\rho}_i^{\mathcal{S}}$. We assume that the $\hat{\rho}_i^{\mathcal{M}}$ have support on disjoint subspaces of $\mathcal{M}$'s state space, ensuring that they are mutually distinguishable and, thus, able to instantiate distinct measurement outcomes, but make no assumptions about the self entropies of the various memory states, $\hat{\rho}_i^{\mathcal{M}}$, or their distinguishability from the standard state, $\hat{\rho}_0^{\mathcal{M}}$. Nor do we make any assumptions about the self entropies of the various $\hat{\rho}_i^{\mathcal{S}}$, their distinguishability from the standard state $\hat{\rho}_0^{\mathcal{S}}$ *or from one another*. We also allow that the environment may play a role in the measurement process, e.g., to accommodate the energy and entropy changes of $\mathcal{MS}$, environment-induced decoherence, *etc*., and may thus end up correlated to the measurement outcome. These liberal conditions allow for a very broad class of ideal and imperfect quantum measurements.

The unitary evolution operator for the control phase of the measurement step is of the form:

$$\hat{U}_B^{\mathcal{IB}}(t) = \hat{U}_B^{\mathcal{MSB}}(t) \otimes \hat{I}^{\mathcal{W}} \tag{19}$$

The states of $\mathcal{D}$, $\mathcal{B}$ and $\mathcal{DB}$ at the conclusion of Step B are thus:

$$\tilde{\rho}_B^{\mathcal{D}} = \sum_i p_i \left(\hat{\rho}_i^{\mathcal{M}} \otimes \hat{\rho}_i^{\mathcal{S}}\right) \tag{20}$$

$$\tilde{\rho}_B^{\mathcal{B}} = \sum_i p_i \tilde{\rho}_{Bi}^{\mathcal{B}} \tag{21}$$

$$\tilde{\rho}_B^{\mathcal{DB}} = \sum_i p_i \left(\hat{\rho}_i^{\mathcal{M}} \otimes \tilde{\rho}_{Bi}^{\mathcal{SB}}\right) \tag{22}$$

with $Tr_{\mathcal{B}}[\tilde{\rho}_{Bi}^{\mathcal{SB}}] = \hat{\rho}_i^{\mathcal{S}}$.

Note that the memory and engine are correlated at the conclusion of Step B, so there is $\mathcal{S}$-information in $\mathcal{M}$ and *vice versa*.

C. Expansion: In the crucial "expansion" step ($t_B < t \leq t_C$), the potential that was imposed on the engine $\mathcal{S}$ in the insertion step is evolved back to its initial configuration while delivering work to $\mathcal{W}$. Two aspects of this operation are critical. First, if the Demon is to use information to deliver work to $\mathcal{W}$ during this step—his signature maneuver—then the inserted potential must be removed in a manner that is conditioned on the state of the memory. The same is true of the $\mathcal{S}$-$\mathcal{W}$ coupling, *i.e.*, the interactions that play the role of the linkages required in the classical case to ensure that $\mathcal{S}$ does the same work on $\mathcal{W}$ regardless of the engine state. This is to say that when the memory is in state $\hat{\rho}_i^{\mathcal{M}}$, then the system Hamiltonian must be restored from $\hat{H}_A^{\mathcal{S}}$ (at time t_B) to $\hat{H}_0^{\mathcal{S}}$ (at time t_C) by a time-dependent Hamiltonian $\hat{H}_{Ci}^{\mathcal{S}}(t)$ *that is dictated by the measurement outcome*, and an appropriate time-dependent Hamiltonian $\hat{H}_i^{\mathcal{SW}}$ must be "activated" during this same interval. Second, unless the decision-making capability required for conditional application of the $\hat{H}_{Ci}^{\mathcal{S}}(t)$ is to be "outsourced" beyond the boundaries of the composite $\mathcal{DE}$ (and any associated decision costs potentially hidden from view), then this capability must be "built in" to the global evolution operator in Step C.

The unitary evolution operator:

$$\hat{U}_C^{\mathcal{IB}}(t) = \hat{U}_C^{\mathcal{MSWB}}(t) = \sum_i \left(\hat{\Pi}_i^{\mathcal{M}} \otimes \hat{U}_{Ci}^{\mathcal{WSB}}(t)\right) \tag{23}$$

has the structure required to enforce this conditioning, where $\hat{\Pi}_i^{\mathcal{M}}$ is the projector onto the support subspace of the i-th data state of the memory. The unitary $\hat{U}_{Ci}^{\mathcal{WSB}}(t)$ applied to $\mathcal{WSB}$ is conditioned on the state of $\mathcal{M}$, but leaves $\mathcal{M}$ intact. The states of $\mathcal{D}$, $\mathcal{B}$ and $\mathcal{DB}$ at the conclusion of Step C are:

$$\tilde{\rho}_C^{\mathcal{D}} = \left(\sum_i p_i \hat{\rho}_i^{\mathcal{M}}\right) \otimes \hat{\rho}_0^{\mathcal{S}} \tag{24}$$

$$\tilde{\rho}_C^{\mathcal{B}} = \sum_i p_i \hat{\rho}_{Ci}^{\mathcal{B}} \tag{25}$$

$$\tilde{\rho}_C^{\mathcal{DB}} = \sum_i p_i \left(\hat{\rho}_i^{\mathcal{M}} \otimes \tilde{\rho}_{Ci}^{\mathcal{SB}} \right) \tag{26}$$

with $Tr_{\mathcal{B}}[\tilde{\rho}_{Ci}^{\mathcal{SB}}] = \hat{\rho}_0^{\mathcal{S}}$ for all i. Note that the evolution operator, $\hat{U}_C^{\mathcal{IB}}(t)$, describes unitary evolution of the composite $\mathcal{WMSB}$ in this (and only this) step, which leaves open the possibility that the evolution of $\mathcal{MSB}$ is nonunitary. However, because the work reservoir, $\mathcal{W}$, must be transformed in the same manner for all outcomes stored in $\mathcal{M}$ (so only "pure work" is delivered to $\mathcal{W}$), $\mathcal{MSB}$ evolves unitarily in this step as well.

D. Memory Reset: In the "memory reset" step, the Demon's memory is returned to its initial state. This is achieved via an *unconditional* local operation on $\mathcal{S}$; there are no subsystems in the accessible $\mathcal{I}$-domain that still hold $\mathcal{S}$-information and could thus be used in Step D to perform the reset conditionally.

The unitary evolution operator for the control phase of the reset step is of the form:

$$\hat{U}_D^{\mathcal{IB}}(t) = \hat{I}^{\mathcal{S}} \otimes \hat{U}_D^{\mathcal{MB}}(t) \otimes \hat{I}^{\mathcal{W}} \tag{27}$$

The states of $\mathcal{D}$, $\mathcal{B}$ and $\mathcal{DB}$ after the control phase of Step D are thus:

$$\tilde{\rho}_D^{\mathcal{D}} = \hat{\rho}_0^{\mathcal{M}} \otimes \hat{\rho}_0^{\mathcal{S}} \tag{28}$$

$$\tilde{\rho}_D^{\mathcal{B}} = \sum_i p_i \tilde{\rho}_{Di}^{\mathcal{B}} \tag{29}$$

$$\tilde{\rho}_D^{\mathcal{DB}} = \hat{\rho}_0^{\mathcal{M}} \otimes \tilde{\rho}_0^{\mathcal{S}} \otimes \tilde{\rho}_D^{\mathcal{B}} \tag{30}$$

The control phase of Step D returns the Demon, $\mathcal{D}$, to its initial state, after which the restoration phase returns the bath, $\mathcal{B}$, to its initial (thermal) state to complete the Demon's cycle. Note that, unlike the composite, $\mathcal{DB}$, the global system has been altered by the cycle in two ways: work has been delivered to the work reservoir, $\mathcal{W}$, and the greater environment has been altered in a manner that depends upon what transpired throughout the cycle (see Section 4.4).

3.3. Entropy and Information

We now consider changes in the self and total entropies of the Demon-bath composite during the control phase of each process step, as well as the memory-system mutual information. General expressions for these changes, as they will be denoted for the η-th process step, are:

$$\Delta \tilde{S}_\eta^{\mathcal{D}} = \Delta \tilde{S}_\eta^{\mathcal{M}} + \Delta \tilde{S}_\eta^{\mathcal{S}} - \Delta \tilde{I}_\eta^{\mathcal{MS}} \tag{31}$$

$$\Delta \tilde{S}_\eta^{\mathcal{B}} = \Delta \tilde{I}_\eta^{\mathcal{DB}} - \Delta \tilde{S}_\eta^{\mathcal{M}} - \Delta \tilde{S}_\eta^{\mathcal{S}} + \Delta \tilde{I}_\eta^{\mathcal{MS}} \tag{32}$$

$$\Delta \tilde{S}_{tot}^{\mathcal{DB}}|_\eta = \Delta \tilde{I}_\eta^{\mathcal{DB}} \tag{33}$$

where tildes again denote quantities relevant to the conclusion of the control phase. Equation (31) follows from Equation (4) for $\mathcal{D} = \mathcal{MS}$. Equation (33) follows from Equation (4) for $\mathcal{DB}$ and recognition that

$\Delta\tilde{S}_{\eta}^{\mathcal{DB}} = 0$, since $\mathcal{DB}$ evolves unitarily during the control phase of each process step and von Neumann entropy is invariant under unitary evolution. Equation (32) follows from Equations (31) and (33) and the defining relation $\tilde{S}_{tot}^{\mathcal{DB}}|_{\eta} = \tilde{S}_{\eta}^{\mathcal{D}} + \tilde{S}_{\eta}^{\mathcal{B}}$.

The entropy changes on each step are summarized along with the cycle totals in Table 2. These entropy changes are obtained from straightforward application of Equations (31)–(33) to the states of $\mathcal{DB}$ tabulated in Section 3.2, together with the definitions $\Delta S_A^{\mathcal{S}} = S(\tilde{\rho}_A^{\mathcal{S}}) - S(\tilde{\rho}_0^{\mathcal{S}})$, $\Delta S_B^{\mathcal{S}} = S(\tilde{\rho}_B^{\mathcal{S}}) - S(\tilde{\rho}_A^{\mathcal{S}})$, $\Delta S_B^{\mathcal{M}} = S(\tilde{\rho}_B^{\mathcal{M}}) - S(\tilde{\rho}_0^{\mathcal{M}})$, and $I_B^{\mathcal{MS}} = \Delta S_B^{\mathcal{M}} + \Delta S_B^{\mathcal{S}}$.

Table 2. Demon-bath entropy changes for demon operating generalized szilard engine.

Step	**Demon** $\Delta\tilde{S}^{\mathcal{D}}$	**Bath** $\Delta\tilde{S}^{\mathcal{B}}$	**Total** $\Delta\tilde{S}_{tot}^{\mathcal{DB}}$
A. Insertion	$\Delta S_A^{\mathcal{S}}$	$\tilde{I}_A^{\mathcal{SB}} - \Delta S_A^{\mathcal{S}}$	$\tilde{I}_A^{\mathcal{SB}}$
B. Measurement	$\Delta S_B^{\mathcal{M}} + \Delta S_B^{\mathcal{S}} - I_B^{\mathcal{MS}}$	$\tilde{I}_B^{\mathcal{DB}} - \Delta S_B^{\mathcal{M}} - \Delta S_B^{\mathcal{S}} + I_B^{\mathcal{MS}}$	$\tilde{I}_B^{\mathcal{DB}}$
C. Expansion	$-\Delta S_A^{\mathcal{S}} - \Delta S_B^{\mathcal{S}} + I_B^{\mathcal{MS}}$	$\tilde{I}_C^{\mathcal{DB}} + \Delta S_A^{\mathcal{S}} + \Delta S_B^{\mathcal{S}} - I_B^{\mathcal{MS}}$	$\tilde{I}_C^{\mathcal{DB}}$
D. Memory Reset	$-\Delta S_B^{\mathcal{M}}$	$\tilde{I}_D^{\mathcal{MB}} + \Delta S_B^{\mathcal{M}}$	$\tilde{I}_D^{\mathcal{MB}}$
Cycle Total	0	$\tilde{I}_A^{\mathcal{SB}} + \tilde{I}_B^{\mathcal{DB}} + \tilde{I}_C^{\mathcal{DB}} + \tilde{I}_D^{\mathcal{MB}}$	$\tilde{I}_A^{\mathcal{SB}} + \tilde{I}_B^{\mathcal{DB}} + \tilde{I}_C^{\mathcal{DB}} + \tilde{I}_D^{\mathcal{MB}}$

Note that, in expressing the entropy change of the Demon in the expansion step, we have used the fact that cumulative change in the self entropy of $\mathcal{S}$ in the insertion and measurement steps is undone in the expansion step, since $\mathcal{S}$ is reset to its initial state in this step ($\Delta\tilde{S}_C^{\mathcal{S}} = -\Delta S_A^{\mathcal{S}} - \Delta S_B^{\mathcal{S}}$). Note also that all memory-engine mutual information $\Delta\tilde{I}_B^{\mathcal{MS}} = I_B^{\mathcal{MS}}$ created in the measurement step is lost in the expansion step (so, $\Delta\tilde{I}_C^{\mathcal{MS}} = -I_B^{\mathcal{MS}}$). (The memory and engine states are completely uncorrelated throughout steps A and D.) Finally, note that although the amounts $\Delta\tilde{I}_{\eta}^{\mathcal{DB}} = \tilde{I}_{\eta}^{\mathcal{DB}}$ of Demon-bath correlation created in each of the steps is dependent upon the details of the associated interactions and processes, they are always nonnegative (and zero for reversible processes). It follows that the cumulative total entropy for the Demon-bath composite is nondecreasing for the cycle:

$$\sum_{\eta} \Delta\tilde{S}_{tot}^{\mathcal{DB}}|_{\eta} = \sum_{\eta} \tilde{I}_{\eta}^{\mathcal{DB}} \geq 0 \tag{34}$$

3.4. Work Extraction

We now consider extraction of work from the Demon during the expansion Step C. Within the present framework, this amounts to a global evolution of $\mathcal{WMSB}$ that increases the expected energy of the work reservoir, $\mathcal{W}$, (*i*) by the same amount for all measurement outcomes held in $\mathcal{M}$ and (*ii*) without changing the entropy of $\mathcal{W}$. For the i-th measurement outcome, $\mathcal{WSB}$ evolves unitarily (since the state of $\mathcal{M}$ remains unchanged). Conservation of energy on $\mathcal{WSB}$ for this i-th process yields:

$$\Delta\langle E^{\mathcal{W}}\rangle = -\Delta\langle E_i^{\mathcal{S}}\rangle - \Delta\langle E_i^{\mathcal{B}}\rangle \tag{35}$$

We can lower bound $\Delta\langle E_i^{\mathcal{B}}\rangle$ with the help of Partovi's inequality [21]: For unitary evolution $\hat{\rho}^{\mathcal{SB}'} = \hat{U}(\rho^{\mathcal{S}} \otimes \hat{\rho}_{th}^{\mathcal{B}})\hat{U}^{\dagger}$ of a system initially in *any* state, $\hat{\rho}^{\mathcal{S}}$, and an environment initially in a thermal state, $\hat{\rho}_{th}^{\mathcal{B}}$, at temperature T,

$$\Delta\langle E^{\mathcal{B}}\rangle \geq k_B T \ln(2)\Delta S^{\mathcal{B}} \tag{36}$$

where $\Delta\langle E^{\mathcal{B}}\rangle = \langle E^{\mathcal{B}'}\rangle - \langle E^{\mathcal{B}}\rangle = Tr[\hat{\rho}^{\mathcal{B}'}\hat{H}^{\mathcal{B}}] - Tr[\hat{\rho}^{\mathcal{B}}\hat{H}^{\mathcal{B}}]$. Here, $\hat{H}^{\mathcal{B}}$ is the self-Hamiltonian of the environment. We thus have for the i-th process:

$$\Delta\langle E_i^{\mathcal{B}}\rangle \geq -k_B T\ln(2)\Delta S_i^{\mathcal{WS}} = -k_B T\ln(2)\Delta S_i^{\mathcal{S}} \tag{37}$$

which is negative for processes (like the isothermal expansion of a gas) that increase the entropy of the engine's working substance. With (35), this yields:

$$\Delta\langle E^{\mathcal{W}}\rangle \leq -\left[\Delta\langle E_i^{\mathcal{S}}\rangle - k_B T\ln(2)\Delta S_i^{\mathcal{S}}\right]. \tag{38}$$

Interpreting the bracketed term as a quantum mechanical form for the free energy of the engine system, this inequality says that, for the i-th process, the work extracted from the engine during the expansion step of any cycle can never exceed the average reduction in the free energy of the engine system.

Averaging $\Delta\langle E^{\mathcal{W}}\rangle$ over all processes, with proper weighting for the outcome probabilities p_i, we obtain:

$$\Delta\langle E^{\mathcal{W}}\rangle = -\sum_i p_i\Delta\langle E_i^{\mathcal{S}}\rangle - \sum_i p_i\Delta\langle E_i^{\mathcal{B}}\rangle \equiv -\langle\Delta\langle E_i^{\mathcal{S}}\rangle\rangle - \langle\Delta\langle E_i^{\mathcal{B}}\rangle\rangle \tag{39}$$

With Equation (37):

$$\langle\Delta\langle E_i^{\mathcal{B}}\rangle\rangle \equiv \sum_i p_i\Delta\langle E_i^{\mathcal{B}}\rangle \geq -k_B T\ln(2)\langle\Delta S_i^{\mathcal{S}}\rangle \tag{40}$$

and we finally have:

$$\Delta\langle E^{\mathcal{W}}\rangle \leq -\left[\langle\Delta\langle E_i^{\mathcal{S}}\rangle\rangle - k_B T\ln(2)\langle\Delta S_i^{\mathcal{S}}\rangle\right] \tag{41}$$

Two aspects of this work bound deserve elaboration: First, note that the requirement that the Demon deliver the same amount of work to the work reservoir on every cycle, regardless of the outcome of the Demon's measurement, limits work extraction to that associated with the measurement outcome that is "worst" in this regard. This ties the work-producing capacity and dissipation of the engine directly to the post-insertion states and the measurement performed by the Demon and may be expected to be maximized for measurements that are characterized by highly symmetric post-measurement states. Any excess work producing capacity not utilized in the expansion step must be accommodated as dissipation elsewhere, since the engine must always be returned to its initial state.

Second, note that because the $\hat{\rho}_i^{\mathcal{M}}$ are mutually orthogonal, and because $-\Delta\tilde{I}_C^{\mathcal{MB}} = I_B^{\mathcal{MS}}$, we can write

$$\langle\Delta S_i^{\mathcal{S}}\rangle = \Delta S^{\mathcal{S}} - \Delta\tilde{I}_C^{\mathcal{MB}} = \Delta S^{\mathcal{S}} + I_B^{\mathcal{MS}} \tag{42}$$

for the expansion step, where $S^{\mathcal{S}} = S(\hat{\rho}^{\mathcal{S}}) = S\left(\sum_i p_i\hat{\rho}_i^{\mathcal{S}}\right)$ denote the entropy of the ensemble density operator for the system. With this, Equation (41) can be rewritten as:

$$\Delta\langle E^{\mathcal{W}}\rangle \leq -\left[\Delta\langle E^{\mathcal{S}}\rangle - k_B T\ln(2)\Delta S^{\mathcal{S}}\right] + k_B T\ln(2) I_B^{\mathcal{MS}} \tag{43}$$

where $\langle E^{\mathcal{S}}\rangle = Tr[\hat{\rho}^{\mathcal{S}}\hat{H}^{\mathcal{S}}]$. This form of Equation (41), which upper bounds the work the Demon can deliver in terms of statistical properties of the ensemble and the amount of information gained in the measurement step, is similar in form to a work bound obtained for a demon via the stochastic thermodynamic approach of Sagawa and Ueda (Equation (8) of [18]). Detailed comparison of the quantities defined in these inequalities and in the underlying Demon models, physical processes and analyses is, however, beyond the scope of this work.

3.5. Analysis Summary

Above, we analyzed a Demon operating a general quantum mechanical Szilard engine, focusing specifically on the Demon-bath composite and those intervals within steps of the operating cycle where Demon-bath energy and entropy exchanges take place. The conclusions of this analysis can be summarized as follows:

> A Demon who acquires information about the state of a Szilard engine's working substance and uses it to conditionally reinitialize the engine state in an expansion process can deliver an amount of work:
>
> $$\Delta\langle E^{\mathcal{W}}\rangle \leq -\left[\langle\Delta\langle E_i^{\mathcal{S}}\rangle\rangle - k_B T\ln(2)\Delta\langle S_i^{\mathcal{S}}\rangle\right] \tag{44}$$
>
> to a work reservoir every engine cycle, increasing (on average) the entropy of the engine's surroundings, $\mathcal{B}$, and the total entropy of the Demon-bath composite by respective amounts
>
> $$\langle\Delta\langle E_i^{\mathcal{B}}\rangle\rangle \geq -k_B T\ln(2)\langle\Delta S_i^{\mathcal{S}}\rangle \tag{45}$$
>
> $$\Delta\tilde{S}_{tot}^{\mathcal{SD}} \geq 0 \tag{46}$$
>
> Here, $\Delta\langle E_i^{\mathcal{S}}\rangle$ and $\Delta S_i^{\mathcal{S}}$ are the changes in the expected energy and in the entropy of the working substance during the expansion step of a cycle in which the measurement yielded the i-th outcome.

4. Discussions

In this section, we consider several aspects of the generalized Demon of this work and its connection to the classical Demon of Fahn's analysis. We first discuss the generality of our analysis and point out ways in which it sidesteps objections that have been raised in the classical context (Section 4.1). We then show how the results of Fahn's analysis are supported by our generalized analysis when appropriately specialized (Section 4.2). Next, within the context of our generalized analysis, we consider an additional theme explored by Fahn: the entropy cost of information that is suggested by comparative analyses of the work-producing Demon and a Demon who opts *not* to use the information he has acquired to do work (Section 4.3). Finally, we discuss Demon-environment correlations, which are neglected in Fahn's analysis, but are naturally preserved in the referential approach, and other considerations that arise in a "closed-universe" treatment of the Demon (Section 4.4).

4.1. Generality

The Demon of this work is, to our knowledge, among the most general studied to date. Studies of quantum heat engines and demons are often specific to particular implementation, e.g., partitioned potential wells with controllable barriers [13,22,23], magnetically coupled spin systems [24] and coupled qubits in superconducting circuits [25], and to those measurement observables defined by these implementations. The Demon studied here uses a general quantum system as a working substance and can make very general quantum measurements on this system. The few restrictions *do* apply to our

Demon and his analysis: One is that the environment with which the Demon directly interacts is a large *but finite* reservoir that is in a thermal state at the beginning of each step in the Demon's cycle. (See [27] for consideration of engines coupled to nonthermal reservoirs.) Another is that work done by the Demon must be delivered to a work reservoir that is explicitly included within the Demon's universe. The general nature of our Demon and his operating cycle frees him of many assumptions that are commonly encountered in the literature on Maxwell's Demon, some of which have raised objections in the classical context (see [26] and the references therein). We emphasize a few important aspects of this generality below.

First, we note that any inserted potential structure that would provide the functionality of a classical partition, i.e., that would confine the working substance to a state subspace until work can be extracted, is assumed here to be of finite (if arbitrarily large) height, so a single state space applies to the working substance of the engine at all times throughout the full cycle. This is significant in light of Norton's objection to the construction and thermodynamic interpretation of ensembles of partitioned classical Szilard engines [26]. The members of such ensembles confine molecules to opposite sides of impenetrable partitions, and are thus characterized by disjoint state spaces that are dynamically inaccessible to one another. The assumption of perfect confinement is implicitly replaced here by the weaker requirement that the statistical state of the engine arrived at in the control phases of the insertion and measurement steps remain stable on the time scale of bath rethermalization (i.e., throughout the restoration phase), which indeed suggests the use of *high* potential barriers. As long as these barriers are finite, however, the engine system has access to its entire state space after insertion and measurement, and the concerns that Norton has raised about the construction of ensembles in classical demons with impenetrable partitions are sidestepped.

Second, we note that measurements made by the Demon in Step B can be arbitrarily noisy and can generally affect the state of the working substance. Post-measurement states associated with various measurement outcomes need not be confined to the disjoint subspaces defined by confining potentials, or to any set of disjoint subspaces for that matter. The only requirement on measurement is that it yields definite, distinguishable outcomes and that the statistical post-measurement states of the working substance remaining at the conclusion of the control phase of the measurement process are specified for each outcome. Note again that we have also allowed that measurements may have outcome-dependent effects on the environment.

Third, we stress that the results obtained above and in the following section are not obtained using relations or phase-space volume conservation arguments from classical thermodynamics, by invoking the Second Law, or through "off the shelf" application of Landauer's Principle. All results follow from the specified forms of the global system states at specific times throughout the cycle, unitary Schrodinger dynamics, established entropic inequalities, energy conservation and the assumption that the bath, $\mathcal{B}$, is in a thermal state characterized by temperature T at the *beginning* of each process step. The invariance of von Neumann entropy under unitary evolution—a purely dynamical fact—has been used liberally, but the nondecreasing total entropy for the full-cycle is a *result* that is obtained with no external appeal to the Second Law. Although several quantities appearing in the results of this work (particularly, Section 4.2) are reminiscent of Landauer's Principle (LP) [7], they emerge from our analysis independent of LP or classical phase-space conservation arguments upon which classical derivations are based.

The above considerations notwithstanding, our results are inescapably tied to the general forms of the physical states assumed for the global composite at specified times throughout the cycle and the assumption that they can be connected through globally unitary evolution. We have not specified example physical processes that correspond to those in the operating cycles or proposed concrete realizations of the Demon, nor have we demonstrated that realization of the required state transformations is feasible. We acknowledge the same, but note that our objective is not to demonstrate realization of a Demon or to argue for his ability to achieve his objective. Rather, we aim to *weaken* the restrictions the Demon must obey by freeing him of many common idealizations and assumptions—thus accommodating a wider range of realistic scenarios—and to establish fundamental bounds on what a Demon that could meet these more liberal requirements within the constraints of physical law could achieve and at what cost. Additionally, as emphasized throughout, we aim to do this in a manner that generalizes and clarifies relationships between entropy generation and information that were discussed by Fahn.

The above analysis showed that the total Demon-bath entropy is non-increasing for the Demon's cycle, even in this very general context, and that this is ensured in part by dissipation from erasure of information about the memory contents from the working substance *during the expansion step*. We further explore this relationship and the connection to Fahn's conclusions in the remainder of this section.

4.2. Recovering Fahn

We now tailor our analysis for the classical Demon of Fahn, specifying the required features of the generalized Demon's states on each step:

A. Insertion: In the insertion step, the state of the engine system is transformed from the initial state, $\hat{\rho}_0^{\mathcal{S}}$ (with entropy $S_0^{\mathcal{S}}$), to a uniform mixture:

$$\tilde{\rho}_A^{\mathcal{S}} = \frac{1}{2}\tilde{\rho}_L^{\mathcal{S}} + \frac{1}{2}\tilde{\rho}_R^{\mathcal{S}} \tag{47}$$

of *orthogonal* states, $\tilde{\rho}_L^{\mathcal{S}}$ and $\tilde{\rho}_R^{\mathcal{S}}$, with respective entropies $S_L^{\mathcal{S}} = S_R^{\mathcal{S}} = S_0^{\mathcal{S}} - 1$. The post-insertion entropy is:

$$\tilde{S}_A^{\mathcal{S}} = H_2\left(\frac{1}{2}\right) + \frac{1}{2}S_L^{\mathcal{S}} + \frac{1}{2}S_R^{\mathcal{S}} = S_0^{\mathcal{S}} \tag{48}$$

ensuring that the partition insertion is entropically benign ($\Delta\tilde{S}_A^{\mathcal{S}} = 0$) as in Fahn. (Here $H_2(\cdot)$ is the classical binary entropy function.) $\tilde{I}_A^{\mathcal{DB}} = 0$, since the partition insertion is assumed to be thermodynamically reversible with equal and opposite system-bath entropy changes. The form (48) follows directly from the "grouping property" of von Neumann entropy for mixtures of orthogonal states: For a set $\hat{\rho}_i$ of density operators with support on orthogonal subspaces, the von Neumann entropy of the convex combination $\hat{\rho} = \sum_i p_i\hat{\rho}_i$ is:

$$S(\hat{\rho}) = H(\{p_i\}) + \sum_i p_i S(\hat{\rho}_i) \tag{49}$$

where $H(\{p_i\}) = -\sum_i p_i \log_2 p_i$ is the Shannon entropy of the probability mass function $\{p_i\}$.

B. Measurement: The measurement step corresponds to a projective measurement onto the orthogonal subspaces supporting $\tilde{\rho}_L^{\mathcal{S}}$ and $\tilde{\rho}_R^{\mathcal{S}}$, which does not affect the system state but does bring the memory into correlation with the system:

$$\tilde{\rho}_B^{\mathcal{MS}} = \frac{1}{2}\tilde{\rho}_L^{\mathcal{M}} \otimes \tilde{\rho}_L^{\mathcal{S}} + \frac{1}{2}\tilde{\rho}_R^{\mathcal{M}} \otimes \tilde{\rho}_R^{\mathcal{S}} \tag{50}$$

The memory is transformed from the null state, $\hat{\rho}_0^{\mathcal{M}}$, to a uniform mixture of orthogonal states, $\hat{\rho}_L^{\mathcal{M}}$ and $\hat{\rho}_R^{\mathcal{M}}$, that, consistent with Fahn, have the same entropy as the null state ($S_0^{\mathcal{M}} = S_L^{\mathcal{M}} = S_R^{\mathcal{M}}$). This yields $\Delta\tilde{S}_B^{\mathcal{M}} = 1$, $\Delta\tilde{S}_B^{\mathcal{S}} = 0$ (since $\mathcal{S}$ is unaffected by the measurement), and $\Delta\tilde{I}_B^{\mathcal{MS}} = 1$. $\Delta\tilde{I}_B^{\mathcal{DB}} = 0$, since it is assumed that the measurement is performed reversibly with no entropic consequences for the bath.

C. Expansion: $\Delta\tilde{I}_C^{\mathcal{DB}} = 0$, since, for each measurement outcome from Step B, Step C requires a *conditional* isothermal expansion process, each of which is thermodynamically reversible with equal and opposite system-bath entropy changes ($\Delta\tilde{I}_{Ci}^{\mathcal{SB}} = 0$).

D. Memory Reset: The memory reset reverses the one-bit entropy increase of $\mathcal{M}$ that occurred in Step B. $\tilde{I}_D^{\mathcal{MB}} = 0$, since the reset process is assumed to be thermodynamically reversible with equal and opposite system-bath entropy changes.

With substitution of these values, Table 2 simplifies (with the bounds saturated) to Table 1, supporting the results of Fahn's analysis.

4.3. The No-Work Demon and the Entropy Cost of Information

To explore the entropic consequences of using information to extract work, Fahn revised his analysis for a Demon who takes no action on the information acquired about the state of the molecule in the engine. Instead of conditionally inserting a piston and extracting work via isothermal expansion in Step C, the "no work" Demon simply removes the partition unconditionally and allows the gas to expand freely. This transforms the state of the molecule and decorrelates the molecule from the memory exactly as it does for the Demon who *does* extract work, and leaves the molecule in the same final state, but without drawing the heat from the environment that would be required to sustain the molecule's ability to do work throughout the stroke. This eliminates the one-bit environmental entropy decrease that occurs during the expansion step in Fahn's analysis of the Demon that *does* do work (Table 1), which revises the entropy costs for the cycle as shown in Table 3. The entropy of the expansion step is increased by one bit, as is the cumulative total entropy for the full cycle.

Description of the "no-work" Demon within our generalized analysis requires both that the work reservoir, $\mathcal{W}$, is decoupled from the engine system, $\mathcal{S}$, throughout the cycle (since no attempt is made to extract work) and that the memory, $\mathcal{M}$, is decoupled from $\mathcal{S}$ during expansion (since the expansion/decorrelation is achieved unconditionally by a local operation on $\mathcal{S}$). The engine system, $\mathcal{S}$, can couple to the bath, $\mathcal{B}$, alone during the expansion step, with which it interacts in an uncontrollable fashion that cannot be conditioned on the measurement outcome held in the memory. The expansion step that does not produce work is thus equivalent to an erasure operation on $\mathcal{S}$, which results in unconditional and irreversible loss of all $\mathcal{M}$-information from $\mathcal{S}$ to the bath.

Table 3. Entropy Changes for a Demon who acquires information, but does no work (adapted from [8]).

Step	**Demon** $\Delta S^{\mathcal{D}}$	**Bath** $\Delta S^{\mathcal{B}}$	**Total** ΔS_{tot}
A. Insertion	0	0	0
B. Measurement	0	0	0
C. Expansion	$+1$	0	$+1$
D. Memory Reset	-1	$+1$	0
Cycle Total	0	$+1$	$+1$

The consequences of this for analysis of our generalized Demon are as follows. The evolution operator for Step C in the no-work Demon is:

$$\hat{U}_C^{\mathcal{IB}}(t) = \hat{I}^{\mathcal{M}} \otimes \hat{U}_C^{\mathcal{MS}}(t) \otimes \hat{I}^{\mathcal{W}} \tag{51}$$

This operation corresponds precisely to the unconditional erasure operation treated within the referential approach in [9]. Because $\mathcal{S}$ is transformed in precisely the same way as it is for the Demon who delivers work, in the sense that the initial and final states are the same in both cases, the general forms of the Demon-bath entropy changes of Table 2 are unchanged. The unconditional nature of the erasure operation does, however, have an important consequence. Conservation of mutual information in the unitarily evolving tripartite system, $\mathcal{MSB}$, *and* the bipartite subsystem, $\mathcal{SB}$, together with the complete loss of $\mathcal{MS}$ correlations during expansion, imply that:

$$\tilde{I}_C^{\mathcal{DB}} = I_B^{\mathcal{MS}} \tag{52}$$

Specializing the results of Table 2 for Fahn's classical Demon (as in Section 4.2) and making appropriate adjustments for the no-work Demon ($\tilde{I}_C^{\mathcal{DB}} = \tilde{I}_C^{\mathcal{MB}} = I_B^{\mathcal{MS}} = 1$ bit), we recover Fahn's results for this Demon (Table 3).

More generally, translating the entropy and energy bounds obtained in Section 3 to the case at hand, we obtain the following general result for the no-work Demon:

> A Demon who acquires information about the state of a Szilard engine's working substance and who resets the engine state via local operations on the engine that are *not* conditioned on the outcomes can deliver no work:
>
> $$\Delta\langle E^{\mathcal{W}}\rangle = 0 \tag{53}$$
>
> The unconditional expansion operation increases the entropy of the engine's surroundings $\mathcal{B}$ (on average) and the total entropy of the Demon-bath composite by respective amounts:
>
> $$\langle\Delta\langle E_i^{\mathcal{B}}\rangle\rangle \geq -k_BT\ln(2)\langle\Delta S_i^{\mathcal{S}}\rangle + k_BT\ln(2)I_B^{\mathcal{MS}} \tag{54}$$
>
> $$\Delta\tilde{S}_{tot}^{\mathcal{SD}} \geq k_BT\ln(2)I_B^{\mathcal{MS}} \tag{55}$$
>
> Here, $\langle\Delta S_i^{\mathcal{S}}\rangle$ is the average change in the entropy of the working substance during the expansion step and $I_B^{\mathcal{MS}}$ is the memory-engine mutual information gained during the measurement step (and irreversibly lost during expansion).

Modification of bath entropy change during expansion, which is derived from classical thermodynamic arguments in Fahn's work, has a different origin here: the increased bath entropy is a consequence of Demon-bath correlations that are necessarily created as the memory-system correlations are destroyed.

Comparing the results he obtained for the work-producing Demon and the no-work Demon, Fahn conjectured the following general principle regarding the entropy cost of information:

> "The entropy cost of information is the degree to which the information is *not* used to obtain work from the system. This cost is assessed when the correlation between information and the system is lost." (from [8])

Comparison of our conclusions for the generalized Demon that produces work (Section 3.5) and the Demon that does not do work (this subsection) lend support to Fahn's conjecture, and in a broader context. When the Demon does not use the memory-engine correlations to condition the expansion operations and extract work, the lower bound on the bath entropy change increases precisely by the amount of information that was embodied within these correlations prior to expansion ($I_B^{\mathcal{MS}}$). This entropic cost, which is exclusively associated with the expansion step, survives in the lower bound on the increase in total Demon-bath entropy for the full cycle, while contributions from the subsystem self-entropy changes distributed throughout the Demon's cycle ultimately sum to zero.

4.4. Demon-Environment Correlations and Global Closure

Demon-environment correlations have already played an important role in this work, but only as Demon-bath correlations created during the control phases of the operations performed in the Demon's cycle. This focus is appropriate for comparing Demon-bath entropy generation obtained here with results from Fahn and with others who limit their attention to the Demon and bath, since all such changes take place during the control phases. It does not, however, tell the whole story of entropy generation in a closed universe that includes the Demon-bath composite as a subsystem. One may wonder about the cumulative total entropy change of the entire, globally closed Demon-*environment* composite. It is arguably *this* total entropy, obtained for the full "universe" containing the Demon, that one might expect to be nondecreasing by analogy with the Second Law.

The total (Demon + environment) entropy for the full "universe", with $\bar{\mathcal{B}}$ included in the environment $\mathcal{E}$, is indeed nondecreasing. The stepwise and full-cycle entropy changes, $\Delta S^{\mathcal{D}}$, $\Delta S^{\mathcal{E}}$ and $\Delta S_{tot}^{\mathcal{DE}} = \Delta S^{\mathcal{D}} + \Delta S^{\mathcal{E}}$, are as in Table 2, but with the Demon-bath correlation entropies, $\tilde{I}_A^{\mathcal{SB}}$, $\tilde{I}_B^{\mathcal{DB}}$, $\tilde{I}_C^{\mathcal{CB}}$ and $\tilde{I}_D^{\mathcal{MB}}$, replaced by respective correlation entropy changes, $\Delta I^{\mathcal{D}\bar{\mathcal{B}}}$ ($\eta = A., B, C, D$). Note that, even if the Demon and bath states are uncorrelated at the beginning of each step, the same need not be true for correlations between the Demon and the *greater environment*, $\bar{\mathcal{B}}$. Any $\mathcal{D}$-$\mathcal{B}$ correlations created during the control phase of any step are transformed into $\mathcal{D}$-$\bar{\mathcal{B}}$ correlations during the restoration phase, where they persist as initial Demon-environment correlations for the next step. While the presence of these initial $\mathcal{D}$-$\bar{\mathcal{B}}$ correlations has no effect on entropy exchange or energy flow in the Demon-bath evolution associated with the following step, they contribute to the total "universe" entropy ($\Delta I^{\mathcal{D}\bar{\mathcal{B}}} \geq 0$) whenever they are present. Analysis of the Demon using the referential approach enables tracking of these "remote" correlations throughout the Demon's cycle.

A final point is in order regarding the globally closed nature of the full composite in which the Demon resides. If no external work is done on the global system, the cycles described in this work cannot proceed forever. Although the Demon and his local surroundings return to their initial states at the end of each cycle, the work-reservoir is left richer (if the Demon has achieved his objective) and the greater environment has expended resources to rethermalize the bath. Eventually, the Demon's universe will suffer its own kind of heat death. This picture would be modified somewhat if one allows an external "operator" that supplies work to help the Demon achieve his tasks on each step and/or lowers the entropy of the greater environment, but all energy, entropy and information flows associated with this delivery of resources would have to be included in the balance sheet when comparing the Demon's work delivery capacity to his overall operating costs. We have, perhaps, implicitly allowed for the presence of such an external operator in our analysis, since we have not required energy conservation in Step A, B or D of the cycle. In our view, there is nothing inherently objectionable about the presence of such an operator, provided that this operator just takes the Demon through the steps of the cycle without having to make decisions or acquire information about the Demon's state in the process of delivering work to the Demon. Such an external agent could even provide work for Step C without affecting our general conclusions about the delivery of work to the reservoir, as long as dissipation not associated with coupling of $\mathcal{WS}$ involves $\bar{\mathcal{B}}$ and not $\mathcal{B}$. It is, however, hard to see how an external operator tasked with staving off the heat death of the Demon's universe by lowering the entropy of the greater environment could be informationally benign, since this process would ultimately require unloading of records that the greater environment holds about the Demon's history.

5. Conclusions

In this work, we have considered a Maxwell Demon residing in a globally closed universe who uses measurements on a quantum mechanical Szilard engine to convert heat into work. The generalized Demon was defined so that several idealizations and requirements common in analyses of Demons are relaxed, and the Demon's cycle was analyzed using a physical-information-theoretic approach for determination of the fundamental physical costs of operations. Our study is, like Fahn's 1996 classical analysis [8], rooted in a relational view of physical information, and we have viewed the generalized Demon through the prism of Fahn's analysis. Our results support Fahn's view that information erasure and resetting of the Demon's memory—widely regarded as being synonymous and credited with saving the Second Law—are distinct phenomena that contribute to a balancing of the entropic books in separate steps of the Demon's cycle.

We suspect that more insight into Maxwell's Demon is likely available through further generalization and application of closed-system approaches that, like the referential approach used to analyze the Demon in this work, define and quantify information in explicitly physical terms. Global closure allows explicit consideration of the larger universe in which the Demon is situated, helping to ensure that analyses of his actions are comprehensive and fully consistent with physical law and, more generally, that the Demon has nowhere left to hide.

Acknowledgments

The author thanks the National Science Foundation for partial support of this work (Grant CCF-0916156), İlke Ercan for comments on the manuscript and two anonymous reviewers for constructive suggestions.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Maxwell, J.C. *Theory of Heat*; Longmans, Green, and Co.: London, UK, 1871.
2. Leff, H.S.; Rex, A.F. *Maxwell's Demon 2: Entropy, Information Computing*; Institute of Physics Publishing: Bristol, UK, 2003.
3. Maruyama, K.; Nori, F.; Vedral, V. Colloquium: The physics of Maxwell's demon and information. *Rev. Modern Phys.* **2009**, *81*, 1–23.
4. Sagawa, T. Review of Maxwell's Demon. In *Thermodynamics of Information Processing in Small Systems*; Springer: Tokyo, Japan, 2013; pp. 9–16.
5. Szilard, L. On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings. *Z. Phys.* **1929**, *53*, 840–856.
6. Bennett, C.H. The thermodynamics of computation—A review. *Int. J. Theor. Phys.* **1982**, *21*, 905–940.
7. Landauer, R. Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.* **1961**, *5*, 183–191.
8. Fahn, P. Maxwell's demon and the entropy cost of information. *Found. Phys.* **1996**, *26*, 71–93.
9. Anderson, N.G. Information erasure in quantum systems. *Phys. Lett. A* **2008**, *372*, 5552–5555.
10. Anderson, N.G. On the physical implementation of logical transformations: Generalized L-machines. *Theor. Comp. Sci.* **2010**, *411*, 4179–4199.
11. Anderson, N.G. Overwriting information: Correlations, physical costs, and environment models. *Phys. Lett. A* **2008**, *376*, 1426–1433.
12. Ganesh, N.; Anderson, N.G. Irreversibility and dissipation in finite-state automata. *Phys. Lett. A*, in press.
13. Zurek, W.H. Maxwell's Demon, Szilard's Engine, and Quantum Measurements. In *Frontiers of Nonequilibrium Statistical Physics*; Moore, G.T., Scully, M.O., Eds.; Plenum Press: New York, NY, USA, 1984; pp. 151–161.
14. Lloyd, S. Use of mutual information to decrease entropy: Implications for the second law of thermodynamics. *Phys. Rev. A* **1989**, *39*, 5378–5386.
15. Quan, H.T.; Liu, Y.-X.; Sun, C.P.; Nori, F. Quantum thermodynamic cycles and quantum heat engines. *Phys. Rev. E* **2007**, *76*, 031105.

16. Toyabe, S.; Sagawa, T.; Ueda, M.; Muneyuki, E.; Sano, M. Experimental demonstration of information-to-energy conversion and validation of the generalized Jarzynski equality. *Nat. Phys.* **2010**, *6*, 988–992.
17. Vaikuntanathan, S.; Jarzynski, C. Modeling Maxwell's demon with a microcanonical Szilard engine. *Phys. Rev. E* **2011**, *83*, 061120.
18. Sagawa, T.; Udea, M. Fluctuation theorem with information exchange: Role of correlations in stochastic thermodynamics. *Phys. Rev. Lett.* **2012**, *109*, 180602.
19. Park, J.J.; Kim, K-H.; Sagawa, T.; Kim, S.W. Heat Engine Driven by Purely Quantum Information. 2013. arxiv.org/abs/1302.3011.
20. Anacleto, J. Work reservoirs in thermodynamics. *Eur. J. Phys.* **2010**, *31*, 617–624.
21. Partovi, M.H. Quantum thermodynamics. *Phys. Lett. A* **1989**, *137*, 440–444.
22. Kim, S.W.; Sagawa, T.; de Liberato, S.; Ueda, M. Quantum Szilard engine. *Phys. Rev. Lett.* **2011**, *106*, 070401.
23. Li, H.; Zou, J.; Li, J.-G.; Shao, B.; Wu, L.-A.; Revisiting the quantum Szilard engine with fully quantum considerations. *Ann. Phys.* **2012**, *327*, 2955–2971.
24. Lloyd, S. Quantum-mechanical Maxwell's demon. *Phys. Rev. A* **1997**, *56*, 3374–3382.
25. Quan, H.T.; Wang, Y.D.; Liu, Y-x.; Sun, C.P.; Nori, F. Maxwell's demon assisted thermodynamic cycle in superconducting quantum circuits. *Phys. Rev. Lett.* **2006**, *97*, 180402.
26. Norton, J.D. Waiting for Landauer. *Stud. Hist. Philos. Modern Phys.* **2011**, *137*, 184–198.
27. De Liberato, S.; Ueda, M. Carnot's theorem for nonthermal stationary reservoirs. *Phys. Rev. E* **2011**, *84*, 051122.