Thermodynamics in the Spekkens Toy Theory

Jamie Y. Parkinson

Supervised by Dan E. Browne and Ian J. Ford
Submitted towards the MSci in Theoretical Physics at UCL

I declare that all material presented in this report is entirely my own work except where explicitly and individually indicated, and that all sources used in its preparation and all quotations are clearly cited.

Abstract

Quantum theory is venerated by physicists and has been experimentally confirmed countless times. Despite this, we have no physical justification for the mathematical postulates upon which it is built. At the opposite end of the spectrum, we have thermodynamics, an old and highly physical branch of physics which has also had a great deal of experimental verification. This report looks at the unification of thermodynamics and the pursuit of a justification for quantum theory.

We examine a toy theory of quantum states developed by Rob Spekkens [1], which reproduces a great deal of quantum theory by considering states as "states of knowledge" and applying a simple knowledge-based axiom to them, and we attempt to introduce some theory of thermodynamics to it. In doing so, we build upon existing work on representations of this theory, and using results from a broad variety of fields such as large deviations theory and the theory of open quantum systems we successfully introduce a notion of energy to the theory and construct a simple engine within its context. We explore the thermodynamic implications of the theory's axiom and, finally, suggest avenues for further work.

Contents

1 The Context for Our Work **1**

2 The Spekkens Toy Theory **3**
- 2.1 Terminology . . . 3
- 2.2 The Basics of the Theory . . . 4
- 2.3 Representational Connections to Quantum Theory . . . 4
  - 2.3.1 Elementary Systems Are Like Qubits . . . 4
  - 2.3.2 The Stabilizer Formalism . . . 5
  - 2.3.3 Toy Stabilizers . . . 6
    - 2.3.3.1 (Mathematical) Operations on Toy Stabilizers . . . 6
    - 2.3.3.2 A Warning . . . 7
  - 2.3.4 Composite System Toy Stabilizers . . . 7
- 2.4 Geometric Representation of States . . . 8
  - 2.4.1 A Mapping To 3D Space for Single Systems . . . 8
  - 2.4.2 Generalising the Mapping to Composite Systems . . . 10
    - 2.4.2.1 Mapping the 2-bit States to 15D Space . . . 11
    - 2.4.2.2 Mappings for Arbitrary Dimensions . . . 11
  - 2.4.3 Shortcomings of This Approach . . . 11

3 Introducing Thermodynamics to The Spekkens Toy Theory **13**
- 3.1 Defining Energy . . . 13
  - 3.1.1 Energy as a Property of the Ontic States . . . 13
    - 3.1.1.1 Closeness to Validity . . . 14
  - 3.1.2 Energy in Composite Systems . . . 16
    - 3.1.2.1 Counting the Symmetries . . . 17
    - 3.1.2.2 Validity After Measurement for Composite Systems . . . 17
  - 3.1.3 Numerical Verification of the Results . . . 18
- 3.2 The Szilard Engine . . . 19
  - 3.2.1 Background . . . 19
  - 3.2.2 Constructing the Szilard Engine . . . 20
    - 3.2.2.1 Framework and Notation . . . 20
    - 3.2.2.2 Operation of the Engine . . . 21
  - 3.2.3 A Circuit Representation of the Szilard Engine . . . 22
    - 3.2.3.1 Quantum Circuits . . . 22
    - 3.2.3.2 The Szilard Engine Circuit . . . 23
  - 3.2.4 The Expansion Step . . . 24
    - 3.2.4.1 Quantum Operations . . . 24
    - 3.2.4.2 Constructing the Expansion Gate . . . 25

3.2.4.3 Work Extraction in the Expansion Step 27
3.2.5 The Erasure Step 28
3.2.6 The Szilard Engine in the Spekkens Theory 28
3.2.6.1 Translating the Steps to Toy Stabilizer Notation 29
3.2.6.2 A Diversion: Irreversible Transformations on Spekkens States 30
3.2.6.3 Violation of the Knowledge Balance Principle 31

4 Conclusions 33

A Map of 2-bit States and Toy Stabilizers 34

B Code Listings 37

C Derivation of the Upper Bound on Expected Divergence 39

D Exorcising Maxwell's Demon (Again) 41

Chapter 1

The Context for Our Work

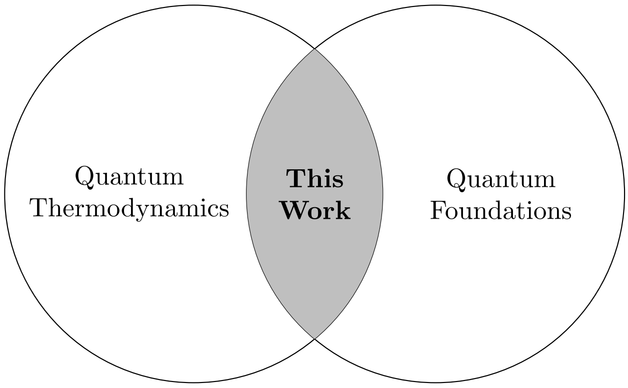
Before I begin the business of reporting on the work completed in the undertaking of this project, I will offer some context for the work and some thoughts on the importance of such work. The work lies somewhere around the intersection of the fields of **quantum thermodynamics** (QT) and **quantum foundations** (QF).

Quantum thermodynamics is the unification of the old and the new: it is the study of thermodynamics in small (quantum scale) systems, and at some level it hopes to unify the seemingly disjoint fields of thermodynamics and quantum theory by seeing that one may emerge from the other. The origins of classical thermodynamics were in the wholly practical application of the study of engines (most notably performed by Sadi Carnot [2]), but then the development of a statistical theory of thermodynamics in the late 19^{th} century offered us an explanation of the behaviour of macroscopic matter through simple - yet fallible - models of the microscopic. In contrast, the early 20^{th} century offered a theory of the microscopic behaviour of matter that appears to this day to be infallible: quantum mechanics. Whilst advances in statistical thermodynamics have drawn heavily on results from quantum mechanics, it remains a largely classical theory based on assumptions on the nature of microscopic behaviour that - while elegant and indeed effective - are not really rigourously justifiable. That said, the classical laws of thermodynamics are widely seen as absolute truths in the macroscopic limit, despite there being no truly watertight justification of, in particular, the famous second law.

Hopefully this account makes it clear that there is a clear appeal in studying QT and attempting to understand thermodynamics and quantum mechanics together. It may be worth noting that, whilst it seems that the consensus is that we should wish thermodynamics to emerge from quantum mechanics, the converse is also appealing to some.

Furthermore, it should become readily apparent that, if we are to suggest a possibility of the emergence of one set of foundational principles from another, we should understand the motivation for the statement of the fundamental set of principles. It's no secret that many are dissatisfied with the fact that there is no universally accepted or provable physical interpretation of our entirely mathematical postulates of quantum mechanics, and while we've no doubt as to the quantum nature of reality, there is seemingly very little rationale for this reality. Attempts at a resolution of this uncomfortable truth form the field of quantum foundations. Explicitly, I feel that - particularly when taking the point of view that quantum theory may emerge from thermodynamics - ideas from QF are at the very least a useful tool for attacking problems in QT, and at most may be an intrinsic part of the latter programme of study.

The Spekkens toy theory is an extremely conceptually simple classical theory of local hidden variables that replicates a large number of "quantum" phenomena. It lies comfortably within the domain of QF. The work documented herein takes the Spekkens toy theory, and attempts to develop some kind of framework for thermodynamics in it, thus offering a small number of results within the remit of the programme outlined above.



Chapter 2

The Spekkens Toy Theory

Here we will present a brief overview of the aforementioned Spekkens toy theory. Introduced by Robert Spekkens in a 2007 Physical Review A paper [1], the theory offers a conceptually simple model of states based the principle that an observer's knowledge about the exact specification of the state is restricted. Spekkens uses a novel representation of these states of knowledge, one that is essentially colouring in squares subject to some simple rules.

Despite this seeming highly arbitary and lacking in complexity, it turns out that the Spekkens theory can reproduce large amounts of quantum theory: entanglement, teleportation, action at a distance and more. It is perhaps worth noting that the theory was later reformulated more generally by the original author [3] but here we will deal only with the original formulation and with another formulation developed by Matthew Pusey [4].

2.1 Terminology

There are a few non-standard terms and concepts used in the context of the Spekkens theory that are key to all of the work contained in this report; a glossary of sorts follows:

Ontic states These are wholly real states of reality; sharp points in phase space. From the Greek *ontos*, "to be".

Epistemic states These are states of knowledge; probability distributions over phase space and hence over ontic states. From the Greek *episteme*, "to know".

Ontic base The ontic states that a given epistemic state spans.

The knowledge balance principle The principal axiom of the Spekkens theory, elaborated on below. Broadly speaking, a statement that the observer may only have at most half of the available knowledge about a system.

Elementary systems The building blocks of the theory, consisting of ontic states and considered as being "in" an epistemic state, and subject to the knowledge balance principle.

2.2 The Basics of the Theory

The knowledge balance principle relies on having a good definition of knowledge, an end that is achieved by considering questions that can be asked about a system. Take a 4-element set, say $\{1, 2, 3, 4\}$, and choose any 1 element of it. see that we can find out what this element is by asking 2 questions that respectively divide the set into 2 upon asking them: for example "is the choice either 1 or 2?" and "is the choice either 2 or 3?".

Knowledge of the answers to these questions is what we mean by knowledge in this original formulation of the Spekkens theory. Imposing that the most knowledge an observer is allowed is half of that available makes it clear that in this 4-element case we could ask one question of the 2, and hence the 4-element set is in some sense the smallest that we can deal with. In the language of the toy theory, this means our elementary systems are composed of 4 ontic states, and allowed epistemic states of maximal (allowed) knowledge have an ontic base consisting of 2 ontic states. Trivially there is also the case of no knowledge: an epistemic state with an ontic base covering the whole elementary system.

We represent systems by a matrix of squares representing the ontic states, and represent the epistemic states by shading the squares corresponding to their ontic bases. The 7 allowed epistemic states for the elementary system are reproduced in full for clarity in Fig. 2.1.

Composite systems can be constructed by combining any number elementary systems. There are various subtleties to the rules and behaviours regarding such systems, which are investigated in quite some detail in Spekkens' original paper.

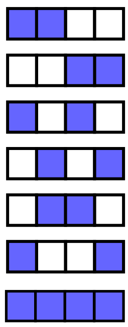


Figure 2.1: All 7 allowed epistemic states for the elementary system of the Spekkens toy theory

2.3 Representational Connections to Quantum Theory

As noted previously, the Spekkens theory manages to reproduce a variety of predictions of quantum mechanics. Here we talk in some detail about the connections between quantum states (and their various representations), and Spekkens' epistemic states.

2.3.1 Elementary Systems Are Like Qubits

Following the original work, we draw a direct analogy between qubit states and the 7 allowed epistemic states for an elementary system, illustrated in Fig. 2.2. Note that

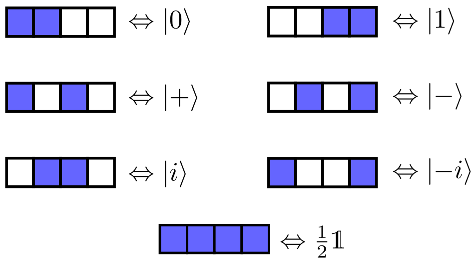


Figure 2.2: The analogs between certain qubit states and the allowed epistemic states of the Spekkens theory. Note that we have represented zero-knowledge state (at the bottom of the figure) as a density operator rather than a state vector: explicitly, recall the notation that $\frac{1}{2}\mathbb{1} = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|)$, the fully mixed state.

epistemic states with disjoint ontic bases correspond to states that are antipodal on the Bloch sphere. Considering this appealing analogue, we will often refer to the elementary systems as bits, and to composite systems composed of n elementary systems as n-bit systems.

2.3.2 The Stabilizer Formalism

Before continuing, we take a brief diversion into the stabilizer formalism for qubits. Developed by Daniel Gottesman in 1996 in the context of quantum error correction [5], the stabilizer formalism is an elegant representation of quantum states and of operations upon them. At its core are the observables represented by the Pauli matrices, here denoted as X, Y and Z but otherwise in their usual form. We also include the 2 by 2 identity matrix I:

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}; \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \tag{2.1}$$
$$Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}; \quad I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

As is well-known, these matrices all have eigenvalues of ± 1, and in the stabilizer formalism we use this fact to represent states: a single-qubit state is represented by the Pauli observable which has an expectation value of 1, eg. $|0\rangle \Leftrightarrow Z$, $|-\rangle \Leftrightarrow -X$ and so on.

Multi-qubit states require some more notation: we use subscripts on the $\{X, Y, Z, I\}$ to indicate which qubit they are operating upon and combine them to form n-qubit observables (eg. $X_1Z_2I_3$ would be the 3-qubit observable corresponding to X on qubit 1, Z on qubit 2, and the identity on qubit 3), and specify states by *groups* of these observables, where each member of the group has an expectation value of 1. Furthermore, for notational compactness, we usually write the generator of the group, denoted by the use of angle brackets, rather than its full form, and where I_j appears we do not write it as its existence is implicit. For example, the Bell state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ has expectation values of 1 for the group of observables $\{I_1I_2, X_1X_2, Y_1Y_2, Z_1Z_2\}$, but if we recall the cyclic property of the Pauli matrices we can see that these observables are not all independent and we need only write 2 of them to fully specify the group in terms of its generators: $\langle X_1X_2, Z_1Z_2\rangle$.

The stabilizer formalism is treated in detail by various authors (eg Refs [5, 6, 7]) but the above summary will suffice for our purposes, which are to be elucidated shortly.

2.3.3 Toy Stabilizers

Hopefully the juxtaposition of the previous 2 sections will have shown the appeal in trying to develop some kind of stabilizer formalism for the Spekkens bits. Here we reproduce the results of Matthew Pusey [4, 8] although we do introduce some slightly different - but entirely equivalent - definitions. Refs [4, 8] contain a much more detailed and rigorous analysis of the toy stabilizer formalism.

As stated above, we are interested in using some kind of stabilizer formalism in the context of the Spekkens theory, and we begin by defining 4 single-system observables:

$$\begin{aligned} \mathcal{X} &:= (1,-1,1,-1) \\ \mathcal{Y} &:= (1,-1,-1,1) \\ \mathcal{Z} &:= (1,1,-1,-1) \\ \mathcal{I} &:= (1,1,1,1) \end{aligned} \tag{2.2}$$

In this notation, it is useful to imagine that the $+1$s correspond to the ontic states in the ontic base of an epistemic state (shaded), and the -1s to those not in the ontic base (unshaded). It follows that, in addition to the direct qubit analogy in Section 2.3.1/Fig. 2.2 we now have another representation of the single-system epistemic states, which is illustrated in Fig. 2.3.

Whilst the assignments of these vectors to observable names is arbitrary, the choice made here emphasises and simplifies the connection between these 3 equivalent representations we have shown here: the toy stabilizer observables for the Spekkens states are named the same as the qubit stabilizer observables for the qubit states in Fig. 2.2.

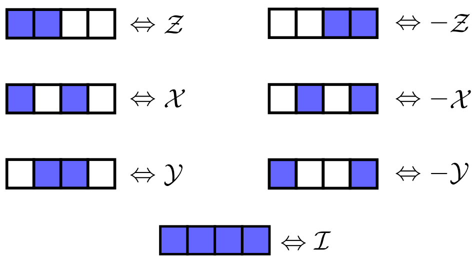


Figure 2.3: Following the definitions in Eq. 2.2, it is easy represent the epistemic states by these observables. Note that, once we have arbitrarily defined the observables, the assignment of epistemic states to them is *not* arbitrary, as it is in Fig. 2.2, as the $+/-1$s in the definitions of the observables directly correspond to known/unknown ontic states.

2.3.3.1 (Mathematical) Operations on Toy Stabilizers

We introduce some definitions/notation for convenience. We call the core group of single-system observables $\mathcal{P} = \{\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{I}\}$. The product of 2 toy stabilizer observables *on the same bit* is the pointwise/Hadamard product:

$$(\mathcal{P}_i \mathcal{P}_j)_m := (\mathcal{P}_i \odot \mathcal{P}_j)_m = (\mathcal{P}_i)_m \cdot (\mathcal{P}_j)_m$$

For multiple-system observables, we use the same notation as with the qubit stabilizer formalism in which subscripts indicate which system the single-system observable is acting on, and use an (almost) outer product to compute the 2-system observable like so, where $\mathcal{A}, \mathcal{B} \in \mathcal{P}$:

$$[\mathcal{A}_1\mathcal{B}_2]_i := [diag(\mathcal{A}) \otimes diag(\mathcal{B})]_{ii} = (\mathcal{A})_{[i \mod 4]}(\mathcal{B})_i \tag{2.3}$$

For composite systems with more than 2 subsystems the natural extension of this outer product-like operation is used. Finally, we define $\mathcal{O}_n$ as the group of all valid toy stabilizer observables (excluding the identity $\mathcal{I}^{\otimes n}$) on n bits. All other notation used is the same as for the qubit stabilizer formalism.

2.3.3.2 A Warning

Despite the appealing link between the stabilizer formalism and the Spekkens model, it is important to note that the toy stabilizers and the qubit stabilizers are **not exactly equivalent**. Specifically, note that for $P_i = (X, Y, Z)_i$ we have the well-known cyclic property that $P_iP_j = i\varepsilon_{ijk}P_k + \delta_{ij}I$, whereas in the toy stabilizer formalism the analogous relation would be $\mathcal{P}_i\mathcal{P}_j = (1-\delta_{ij})\mathcal{P}_k + \delta_{ij}\mathcal{I}$. As such, when working in the toy stabilizer formalism we must be take care to avoid thinking of it as being exactly the same as the qubit stabilizer formalism.

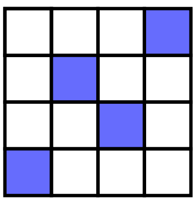
2.3.4 Composite System Toy Stabilizers

As was briefly mentioned in Section 2.3.2, the representation of single-bit systems by observables with expectation values of +1 is generalised to multiple-bit systems by the use of groups of such observables, and indeed by the generators of such groups. A brief example for the 2-bit case is given here. We can write down the 15-tuple[1] of all valid observables on 2 bits, $\mathcal{O}_2$:

$$\begin{aligned}\mathcal{O}_2 = (&\mathcal{I}_1\mathcal{X}_2, \mathcal{I}_1\mathcal{Y}_2, \mathcal{I}_1\mathcal{Z}_2, \mathcal{X}_1\mathcal{I}_2, \mathcal{Y}_1\mathcal{I}_2,\\ &\mathcal{Z}_1\mathcal{I}_2, \mathcal{X}_1\mathcal{Y}_2, \mathcal{Y}_1\mathcal{X}_2, \mathcal{X}_1\mathcal{Z}_2, \mathcal{Z}_1\mathcal{X}_2,\\ &\mathcal{Y}_1\mathcal{Z}_2, \mathcal{Z}_1\mathcal{Y}_2, \mathcal{X}_1\mathcal{X}_2, \mathcal{Y}_1\mathcal{Y}_2, \mathcal{Z}_1\mathcal{Z}_2)\end{aligned} \tag{2.4}$$

where the vector representations of these observables (should we wish to find them) can be calculated using Eq. 2.3. We can then represent 2-bit states by groups of these observables, and in turn by generators of groups, just like for the stabilizers on qubits. An example for a specific 2-bit composite system is given in Fig. 2.4

It follows that we can also, by generalising $\mathcal{O}_2$ and applying a similar procedure, represent n-bit states by similar group generators.



$$\Leftrightarrow \{\mathcal{I}_1\mathcal{I}_2, \mathcal{Z}_1\mathcal{X}_2, \mathcal{X}_1\mathcal{Z}_2, \mathcal{Y}_1\mathcal{Y}_2\} = \langle \mathcal{Z}_1\mathcal{X}_2, \mathcal{X}_1\mathcal{Z}_2\rangle$$

Figure 2.4: An example mapping of a 2-bit system to a group of toy stabilizer observables, and hence to their generator (in the angle brackets).

[1] The slightly pedantic use of a tuple rather than a set here becomes clear later on: its ordering is arbitrary but it will be necessary for an ordering to be present.

2.4 Geometric Representation of States

It is well known that all pure qubit states (2D vectors in Hilbert space) lie on the Bloch sphere, a 2-sphere on which antipodal points correspond to orthogonal vectors in Hilbert space, and that all mixed qubit states lie within this sphere [7].

Specifically, for a general pure state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, the consideration that the state must be normalised and the knowledge that we are interested only in the relative phase between the basis vectors means it is natural to introduce a transformation where $\alpha = \cos(\theta/2)$ and $\beta = e^{i\varphi}\sin(\theta/2)$. The usual use of $\{\theta, \varphi\}$ in spherical coordinates leads to the Bloch sphere. The general relation (see, for example, Ref. [7]) between a density matrix ρ and a vector $\boldsymbol{r}$ lying on/within the Bloch sphere is:

$$\rho = \frac{1}{2}(\mathbb{1} + \boldsymbol{r}\cdot\boldsymbol{\sigma}) \tag{2.5}$$

where $\mathbb{1}$ is the 2D identity matrix, $\boldsymbol{\sigma} \coloneqq (\sigma_x, \sigma_y, \sigma_z)$ and the σ_i are the usual Pauli matrices.

Given this elegant mapping of qubit states into an intuitively accessible geometric representation, and the various connections between the Spekkens theory and qubit states, it is appealing to try to find a similar representation of the epistemic states of elementary systems. Work towards this aim follows.

2.4.1 A Mapping To 3D Space for Single Systems

Firstly we introduce yet another piece of notation, albeit a simple one. We represent Spekkens states as vectors of 0s and 1s for the unshaded and shaded squares, respectively. Figure 2.5 gives an example of this mapping.

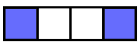
 $\Leftrightarrow (1,0,0,1)$

Figure 2.5: A demonstration of the vector representation of a Spekkens state (for a single system).

With this simple notation it turns out that it is very easy to map these 4D "Spekkens vectors" to points in 3D space, analogously to Eq. 2.5.

Result 2.4.1. *We can define a transformation $M : \mathbb{R}^4 \to \mathbb{R}^3$ that takes a Spekkens vector v_s to 3D Euclidean space, analogously to the transformation that takes 2D Hilbert space vectors to the complex projective line (Bloch sphere):*

$$M(v_s) \coloneqq \frac{1}{2}\begin{pmatrix} \mathcal{X} \\ \mathcal{Y} \\ \mathcal{Z} \end{pmatrix} \cdot v_s = \frac{1}{2}\begin{pmatrix} 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \end{pmatrix} \cdot v_s^T \tag{2.6}$$

This transformation bounds all valid epistemic states within an octahedron, analogous to the Bloch sphere. Recalling the definition of the d-dimensional cross polytope C_d^Δ (of which the octahedron is the 3D case) in terms of the 1-norm as [9]:

$$C_d^\Delta \coloneqq \left\{ \mathbf{x} \in \mathbb{R}^d \;\middle|\; \|\mathbf{x}\|_1 \coloneqq \sum_i |x_i| \leq 1 \right\}$$

we can define a measure of state validity $\sigma : \mathbb{R}^4 \to \mathbb{R}^1$ that is ≤ 1 for all valid epistemic states and their convex combinations, with the inequality saturated in the former case and the measure equal to zero for the zero-knowledge state.

$$\sigma(v_s) = \|M(v_s)\|_1$$

We can prove the statement that convex combinations of allowed states lie within the octahedron as follows:

Proof. Consider that allowed states lie at the points given by permutations of $(\pm 1, 0, 0)$; these are $\pm$ the set of unit vectors $\mathbf{e}_i$. Now form a general convex combination of all of these:

$$\mathbf{x} = \sum_{i=1}^{3} \lambda_i^+ \mathbf{e}_i - \sum_{i=1}^{3} \lambda_i^- \mathbf{e}_i$$

where $\sum_{i=1}^{3} (\lambda_i^+ - \lambda_i^-) = 1$ and $\lambda_i^+ - \lambda_i^- \geq 0 \quad \forall i$ by the definition of convex combination. Now:

$$\begin{aligned}
\mathbf{x} &= \sum_{i=1}^{3} (\lambda_i^+ - \lambda_i^-)\, \mathbf{e}_i \\
\|\mathbf{x}\|_1 &= \left\| \sum_{i=1}^{3} (\lambda_i^+ - \lambda_i^-)\, \mathbf{e}_i \right\|_1 \\
&\leq \sum_{i=1}^{3} \|(\lambda_i^+ - \lambda_i^-)\, \mathbf{e}_i\|_1 \\
&= \sum_{i=1}^{3} \|\mathbf{e}_i\|_1 \cdot \|(\lambda_i^+ - \lambda_i^-)\|_1 \\
&= 1 \cdot \sum_{i=1}^{3} (\lambda_i^+ - \lambda_i^-) = 1
\end{aligned}$$

Where the inequality is due to the triangle inequality, $\|\mathbf{e}_i\|_1 = 1$ was taken outside in the last line because that equality holds $\forall i$, and the 1-norm was removed in the last line because of the previously stated requirement that the $(\lambda_i^+ - \lambda_i^-)$ are all positive.

Now recall the definition of the octahedron by the 1-norm and see that all these convex combinations of allowed states $\mathbf{x}$ must lie within the octahedron. $\square$

Note that invalid Spekkens states may still be in some sense valid quantum states - they may lie between the octahedron and the Bloch sphere. An example of such a state is shown in Fig. 2.6.

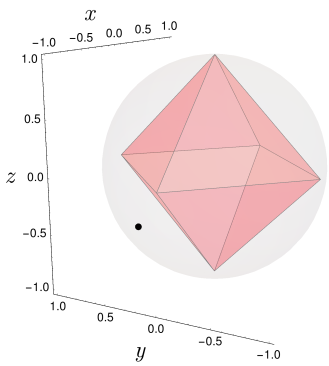
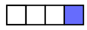


Figure 2.6: An illustration of how a disallowed Spekkens state may still lie within the Bloch sphere. The sphere is illustrated in pale grey and the Spekkens octahedron is in red, and the point $\frac{1}{2}(-1, 1, -1)$ corresponding to the disallowed state  is shown as a black dot.

2.4.2 Generalising the Mapping to Composite Systems

It turns out that the mapping of states to a lower-dimensional space (although not the state validity function) given in Result 2.4.1 can be fairly easily generalised to composite systems.

First we must generalise the vector notation introduced in 2.4.1 to the composite systems. The operation in question is similar to the usual operation of matrix vectorisation *but with non-standard indices.* To see this, consider that in Spekkens' paper (Ref. [1]) the convention, as here, for 2-bit systems is to index the ontic states as if they were 2D Cartesian coordinates rather than like matrix elements: the bottom-left ontic state is $(1, 1)$ and the top-right is $(4, 4)$. It follows that the indexing for n-bit case corresponds to n-dimensional Cartesian coordinates.

The vector notation is now developed. We convert shaded/unshaded squares in the diagrams to 1/0s as before, then vectorise with the last dimension as the slowest-changing index. A diagram, Fig. 2.7, makes this slightly awkward description clear for the 2-bit case.

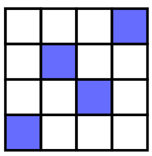
 $\Leftrightarrow (1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1)$

Figure 2.7: An example of the vector v_s for a 2-bit composite system. Note that the first element in the vector corresponds to the bottom-left $(1, 1)$ square, the 2nd to the square at $(2, 1)$, the third at $(3, 1)$ and so on up until the last element, corresponding to the top-right $(4, 4)$ square.

2.4.2.1 Mapping the 2-bit States to 15D Space

We take the tuple of observables on 2 systems given in Eq. 2.4 and construct a transformation function M_2 similar to M in Result 2.4.1:

$$M_2(v_s) := \frac{1}{4}\begin{pmatrix} \mathcal{I}_1\mathcal{X}_2 \\ \mathcal{I}_1\mathcal{Y}_2 \\ \vdots \\ \mathcal{Z}_1\mathcal{Z}_2 \end{pmatrix} \cdot v_s^T = \frac{1}{4}\mathsf{mat}(\mathcal{O}_2) \cdot v_s^T \tag{2.7}$$

where $\mathsf{mat}(\mathcal{O}_2)$ is the matrix representation of the tuple $\mathcal{O}_2$, with the matrix's rows being the vectors given by the relation in Eq. 2.3.[2]

2.4.2.2 Mappings for Arbitrary Dimensions

Whilst it turns out that this approach has its shortcomings - to be discussed in Section 2.4.3 - it does have another use in addition to the geometric representation. Pusey offers a few examples of 2-bit Spekkens states and the corresponding generators of toy stabilizer groups [4, 8] but does not give us any deterministic mapping from a Spekkens state to the toy stabilizer groups. We present such a mapping below.

Result 2.4.2. *We can determine the complete group of toy stabilizer observables of any n-bit Spekkens system as follows. In an extension of our previous results we can write down the tuple of all valid observables on n bits, $\mathcal{O}_n$:*

$$\begin{aligned} \mathcal{O}_n &= (\text{all permutations of } n \text{ single-bit observables [excl. } \mathcal{I}^{\otimes n}]) \\ &= (\mathcal{I}_1\mathcal{I}_2\cdots\mathcal{X}_n, \mathcal{I}_1\mathcal{I}_2\cdots\mathcal{Y}_n, \cdots, \cdots, \mathcal{Z}_1\mathcal{Z}_2\cdots\mathcal{Z}_n) \end{aligned}$$

Note that there are $4^n - 1$ observables on n bits. Using the definition of the Spekkens vector v_s as well as that of $\mathsf{mat}(\mathcal{O}_n)$ from above, the corresponding toy stabilizer group on n bits, $\mathcal{G}_n$, is given by:

$$\mathcal{G}_n = \left\{ o_i \cdot [\mathcal{O}_n]_i \;\middle|\; |o_i| = 1, o_i \in \frac{1}{2^n}\mathsf{mat}(\mathcal{O}_n) \cdot v_s^T \right\}$$

or, to put this into words, we can determine the group of toy stabilizer observables for a Spekkens state by performing a transformation of the Spekkens vector to a $(4^n - 1)$-d vector in which the ± 1s indicate the signs of the corresponding observables in $\mathcal{O}_n$, with all other observables not present in the group.

Appendix A gives a complete map of all 2-bit states to their corresponding toy stabilizers, determined using the method described above.

2.4.3 Shortcomings of This Approach

Although these matrices of observables are clearly useful, they do have some shortcomings which we have not resolved.

We know that the number of pure stabilizer states on n qubits is given by $n_s = 2^n \prod_{i=1}^{n}(2^i + 1)$ (see, for example, Ref. [10]) and hence there are this many maximal-knowledge Spekkens states for n bits also. This fact alone means that our 1-norm

[2] The ordering present in a tuple is now being used; otherwise we would not know what order to put the rows of the matrix in.

measure of state validity from Result 2.4.1 cannot work for $n > 1$, as the cross-polytope in d dimensions that the 1-norm describes is known to have $2d = 2(4^n - 1) \neq n_s$ vertices [9], so cannot form the convex hull of valid states.

Furthermore, we can argue that there are (far) more degrees of freedom than are in some sense "necessary" for a geometric mapping. For $n = 2$, we can see that $n_s = 60$, but then consider that the 2-bit mapping to 15-d space given in Eq. 2.7 gives us 3 ± 1s distributed in the 15 dimensions of the resulting vector (all other elements are 0) and there are $2^3 \cdot \binom{15}{3} = 3640$ ways to distribute these ± 1s - but only 60 of these are valid states! This leads us to suggest that there may be some polytope in lower-dimensional space which *does* have vertices corresponding to the maximal-knowledge states.

More simply than the above argument - it seems unlikely that the best representation of something (toy theory states) that are similar a subset of 4D Hilbert space (stabilizer states) requires a whole 15 dimensions. Previous work has been completed that tries to generalise the Bloch sphere to 2 qubits, and manages to do so in the context of a mere 3-sphere (4 dimensions) [11].

Considering all of the above, we suggest that more work could be performed on the geometric representation of Spekkens states, with a goal of being able to determine polytopes that are the convex hull of all valid maximal-knowledge states.

Chapter 3

Introducing Thermodynamics to The Spekkens Toy Theory

Now that we have developed a better understanding of the Spekkens toy theory and its behaviour, and have learned of and developed some tools that make the theory more mathematically tractable, we turn to our main aim of introducing thermodynamics into the theory. We will first look at possible definitions of energy in the theory, and will then attempt to apply the results of this to building some kind of heat engine in the context of the theory.

3.1 Defining Energy

Given all of the various similarities of the Spekkens theory to (subsets of) quantum theory that we described in Section 2.3.1, we could perhaps be led to think that a definition of energy would not be too difficult to find. After all, for some quantum state $|\psi_n\rangle$ we know that energy measurements upon it will always give E_n, the eigenvalues of the Hamiltonian operator.

Given our previous assertion that the valid epistemic states for elementary systems correspond to a certain set of qubit states, we could be led to simply define energies as being properties of the epistemic states, but if we make this claim we reach an uncomfortable conclusion: we would be defining energies as properties of probability distributions, of states of knowledge, rather than as properties of states of reality.

We argue that to require energy to be an "objective" property of real physical states is not an unreasonable desideratum. Indeed, this is reminiscent of the usual (and certainly non-controversial) setting of statistical mechanics in which the microstates have defined energies but we cannot "access" these - we instead have an expectation value of energy for the macrostate, given by $\langle E\rangle = -\frac{\partial \ln Z}{\partial \beta}$. We explore whether we can reasonably think of the energy of an epistemic state as an expectation value over the ontic states below.

3.1.1 Energy as a Property of the Ontic States

We first try to assign energies (completely generally, at this stage) to the ontic states that compose an elementary system, such that the ith ontic state has some energy E_i - this is illustrated in Fig. 3.1a.

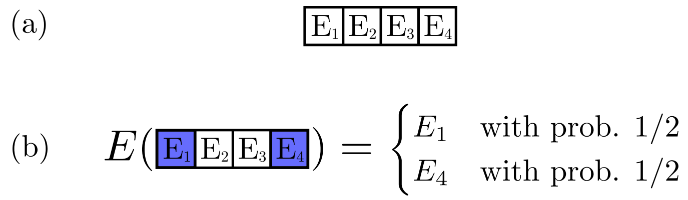


Figure 3.1: (a) Illustrates how we assign the energies E_i to the ith ontic states
(b) Illustrates how, if we were to measure the energy of a single elementary system in some epistemic state, the returned value would tell us which ontic state the system was in, thus violating the knowledge balance principle.

However, we immediately run into a problem: suppose that we can perform an energy measurement on some given epistemic state (recall that for epistemic states of elementary systems, we will know that the system is in either 1 of 2 states - or of 4 states, for the no-knowledge case). Then we get back the "actual" energy of the system; that of the ontic state which it is in. But this means that we know which ontic state the system is in and hence the knowledge balance principle is violated! This issue is illustrated in Fig. 3.1b.

We offer a simple resolution of this problem: we say that a given epistemic state of an elementary system is in fact a representation of some kind of Platonic ensemble of a very large number (N) of systems in the same epistemic state, and that an energy measurement on the state is a performance of the measurement of the type illustrated in Fig 3.1b on all members of the ensemble.

As such the law of large numbers tells us that in the limit of very large N we recover the energy given by the average of the energies of the ontic states almost surely, and hence have no extra knowledge about the state of the system. For the illustrated example this means that we could say that this system has an expectation value for energy equal to $\frac{1}{2}(E_1+E_4)$. Furthermore we can see by Hoeffding's inequality [12] that the probability that the average energy after N measurements $\bar{E}_N$ deviates from the average of the ontic state energies $\langle E\rangle$ by more than a small amount δ is:

$$\Pr\left(\left|\bar{E}_N - \langle E\rangle\right| \geq \delta\right) \leq 2\exp\left(-\frac{2N\delta^2}{\left[\max\{E_i\} - \min\{E_i\}\right]^2}\right) \tag{3.1}$$

Which clearly tends to 0 as $N \to \infty$, as expected.

3.1.1.1 Closeness to Validity

It is interesting to see how quickly we approach a no-extra-knowledge state. We will refer to the *probability distribution implied by the energy measurements*: for example, if 5 energy measurements on the zero-knowledge state were to return $(E_1, E_2, E_2, E_4, E_3)$ we would have a corresponding distribution of $(\frac{1}{5}, \frac{2}{5}, \frac{1}{5}, \frac{1}{5})$, where the ith bin corresponds to the ith ontic state. In the interests of absolute clarity we note that these distributions have no dependence whatsoever on the actual values of the E_i. We will also refer to the *canonical distribution*, that which corresponds to the probability distribution represented by the epistemic state in the original setting (or, completely equivalently, to the expected distribution implied by energy measurements).

Consider that, while the canonical distribution certainly gives us no extra knowledge, there are likely other distributions that give no extra knowledge. For example, if

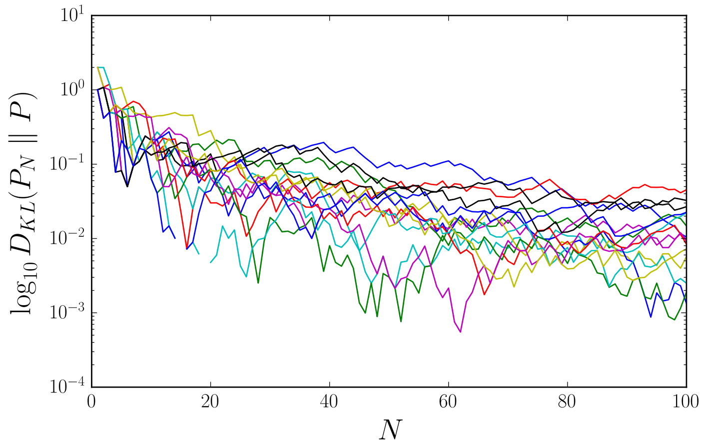


Figure 3.2: A plot of 15 paths taken by $D_{KL}(P_N \parallel P)$ as states are added to the Platonic ensemble (meaning that a new ensemble is not created for each value of N, rather that a new energy measurement is "added" to P_N at each step. The y-axis is logarithmically scaled and as such zero values have been clipped to make the plot clearer.

the probabilities of being in the 4 ontic states were $(\frac{1}{6}, \frac{1}{3}, \frac{1}{6}, \frac{1}{3})$ then this is clearly just corresponding to a convex combination of 2 allowed and disjoint epistemic states.

As such, we choose to define the "distance" that a state is from the no-extra-knowledge state to be the relative entropy (also known as Kullback-Leibler divergence) between the 2 probability distributions. This is defined, for 2 distributions P and Q, as [7]:

$$D_{KL}(P \parallel Q) := \sum_i P_i \log_2 \frac{P_i}{Q_i} \tag{3.2}$$

Note that D_{KL} is not symmetric in its arguments. We take the second argument to be the canonical distribution and call it P, and the first argument to be the distribution implied by the energy measurements and call it P_N. Given that in our case P is uniform then it is easy to show that this relative entropy is in fact the same as the entropy difference between the distributions, which in turn is some constant minus the Shannon entropy H of P_N, but given the appeal in the meaning of D_{KL} we will not use that form.

In the spirit of work on smooth entropies [13], we now introduce a notion of being in an ϵ-valid state, where the relative entropy between the state implied by the energy measurements and the allowed state is some small amount ϵ.

Result 3.1.1. *Energy measurements on a system in some epistemic state, in our setting of the state representing a Platonic ensemble of N systems in the same state, give us knowledge about the state. We define P_N as the probability distribution implied by N energy measurements, and P as the (uniform) canonical distribution corresponding to the epistemic state. We say that the state given by P_N is ϵ-valid if:*

$$D_{KL}(P_N \parallel P) \leq \epsilon$$

*Then, remembering that **(i)** D_{KL} only differs from $H(P_N)$ by a constant and **(ii)** $H(P)$ is the max-entropy state, we can use a result of Antos and Kontoyiannis [14] and state* *that:*

$$\Pr\left(D_{KL}(P_N \parallel P) > \epsilon\right) \leq \exp\left(-\frac{N\epsilon^2}{2\left[\log_2 N\right]^2}\right)$$

which, as expected, goes to 0 as $N \to \infty$. Note that this has no dependence on the number of ontic states in the base of the epistemic state, but is quite a loose bound. Thus, we can set arbitrary confidence integrals that we are in an ϵ-valid state.

Fig. 3.2 shows 15 sample paths taken by $D_{KL}(P_N \parallel P)$ as states are added to the ensemble, up to $N = 100$. The asymptotic behaviour is clearly visible (note that the y axis is log-scaled). The code used to generate this plot is available in Appendix B.

3.1.2 Energy in Composite Systems

The above approach of thinking of energy as expectation values can be extended to composite systems. We treat only the 2-bit case but all of this could be quite easily extended to the n-bit case. First, we must consider what form the energies of the $4^n = 16$ ontic states can take (remember also that maximal-knowledge 2-bit states have 4 states in their ontic base, and the only allowed less-than-maximal-knowledge states have 8). Recalling the Cartesian-like indexing of the ontic states of a 2-bit system from Section 2.4.2, we make the reasonable assumption that the energies combine like $E_{ij} = E_i + E_j$ (implicitly we are imposing here that the 2 elementary systems that compose the composite system have the same E_i). This labelling is illustrated in Fig. 3.3.

$E_4 + E_1$	$E_4 + E_2$	$E_4 + E_3$	$2E_4$
$E_3 + E_1$	$E_3 + E_2$	$2E_3$	$E_3 + E_4$
$E_2 + E_1$	$2E_2$	$E_2 + E_3$	$E_2 + E_4$
$2E_1$	$E_1 + E_2$	$E_1 + E_3$	$E_1 + E_4$

Figure 3.3: Our convention for assigning energies to the ontic states of a 2-bit system, following $E_{ij} = E_i + E_j$ with the indices (i, j) being defined as in Section 2.4.2.

Looking at this labelling it is apparent that there are some interesting degeneracies and symmetries in the 2-bit energies:

- The energies are always symmetric across the diagonal: as such, if the epistemic state on which an energy measurement is being performed has ≥ 2 diagonally opposite ontic states within its ontic base, any single energy measurement returning an energy corresponding to one of these states does *not* in fact specify the exact ontic state of the system exactly, as we found for the elementary system.

- The energies on the diagonal are unique: a measurement returning one of these energies does specify the exact ontic state of the system. That is, unless:

- If any 2 of the E_i are equal, or in fact (more strongly) if the sum of any 2 the E_i are equal to twice one of the others.[1]

[1]Note an interesting connection here: this is much like the situation of degenerate energy gaps of a Hamiltonian, meaning that there can exist some partitioning of the system and the bath such that they are non-interacting and is a case that causes problems when examining thermalisation of quantum systems [15]. It would be interesting to look into the meaning of this degeneracy in our setting.

3.1.2.1 Counting the Symmetries

We developed a few mathematical tools to count the symmetries of the Spekkens states given only their toy stabilizer representation and hence make it easier to identify the 3 bullet points listed above.

Result 3.1.2. *Given a toy stabilizer group generator, $\langle G \rangle$, and recalling that any toy stabilizer observable can be represented as a vector (by Eq. 2.3, and with* $\mathsf{vec}(g)$ *denoting the vector representation of an observable g), then define a generator vector $\mathbf{x}$:*

$$\mathbf{x}\left(\langle G \rangle\right) := \frac{1}{|\langle G \rangle|} \sum_{g \in \langle G \rangle} \mathsf{vec}(g)$$

Note that the vector $\mathbf{x}$ is similar (but not identical) to the Spekkens vector v_s. Now introduce 2 functions of an index i, $\Delta(i)$ and $\phi(i)$, which identify whether an element i is on the diagonal and map i across the diagonal [2]*, respectively.*

$$\begin{aligned} \Delta(i) &:= \begin{cases} 1 & \text{if } \left\lfloor \frac{i-1}{4} \right\rfloor = (i-1) \bmod 4 \\ 0 & \text{otherwise} \end{cases} \\ \phi(i) &:= 4\left[(i-1) \bmod 4\right] + \left\lfloor \frac{i-1}{4} \right\rfloor + 1 \end{aligned}$$

Now we can introduce our counting functions, $n_{diag}(\mathbf{x})$ and $n_{opp}(\mathbf{x})$, that count the number of on-diagonal elements and the number of diagonally opposite pairs, respectively.

$$\begin{aligned} n_{diag}(\mathbf{x}) &:= \#\left\{\Delta(i) \cdot x_i = 1 \mid x_i \in \mathbf{x}\right\} \\ n_{opp}(\mathbf{x}) &:= \frac{1}{2}\#\left\{\frac{1}{2}\left(x_i + x_{\phi(i)}\right)\left(1 - \Delta(i)\right) = 1 \,\middle|\, x_i \in \mathbf{x}\right\} \end{aligned}$$

Where $\#\{\cdots\}$ indicates the number of elements in the set.

3.1.2.2 Validity After Measurement for Composite Systems

We can immediately apply the results of Section 3.1.1.1 to the special case of 2-bit epistemic states that satisfy any of the following:

- $n_{diag} = n_{opp} = 0$
- $n_{diag} = 4$
- $n_{opp} = 2$ for a maximal-knowledge state
- $n_{opp} = 4$ for a submaximal-knowledge state

As these states have canonical distributions that are uniform. For the states other than this - which contain additional degeneracies and symmetries and hence have non-uniform canonical distributions - Result 3.1.1 does not follow immediately, but it is fairly obvious to see that, by the law of large numbers, in the limit of large N the

[2] The choice of notation ϕ indicates that this function, given the previously discussed Cartesian-like indices for Spekkens vectors, is somewhat analogous to a rotation (by π) in Cartesian space.

distribution implied by the energy measurements will be almost surely equal to the canonical distribution and hence D_{KL} will tend to 0.

We can in fact now introduce another bound, that is far tighter and more general than that in Section 3.1.1.1, but introduces a dependence on the size of the ontic base.

Result 3.1.3. *Given any epistemic state, the probability distribution implied by energy measurements will be ϵ-close to the canonical distribution. We quote Theorem 11.2.1 from Cover and Thomas [16]:*

$$\Pr\left(D(P_N \parallel P) > \epsilon\right) \leq 2^{-N\epsilon} \cdot 2^{-|P|\log_2(N+1)}$$

And identify that $|P|$ is the size of the ontic base of the state in question. See that, as predicted, we will be in an ϵ-valid state with some probability tending to 1 for $N \to \infty$.

3.1.3 Numerical Verification of the Results

We present a numerical verification of the above results claiming that the large-N Platonic ensemble of epistemic states does not violate the knowledge balance principal after energy measurement. To do so we introduce yet another bound, following Roy [17]:

Result 3.1.4. *We can upper-bound the expectation value of $D_{KL}(P_N \parallel P)$ in the special case of 2-bit maximal-knowledge epistemic states with either $n_{diag} = n_{opp} = 0$ or $n_{diag} = 4$, as in Section. 3.1.2.2 but restricted further to the case of the maximal-knowledge states, which have an ontic base of size 4. This bound takes the form:*

$$\langle D_{KL}(P_N \parallel P)\rangle \leq \log_2\left(\frac{3+N}{N}\right) \tag{3.3}$$

Which, once again, tends to 0 as $N \to \infty$.

A derivation of Eq. 3.3 can be found in Appendix C.

The advantage of this kind of bound (ie. on an expectation value) is that it can be tested numerically in a very obvious way: Fig. 3.4 illustrates this. The code used to generate the figure can be found in Appendix B.

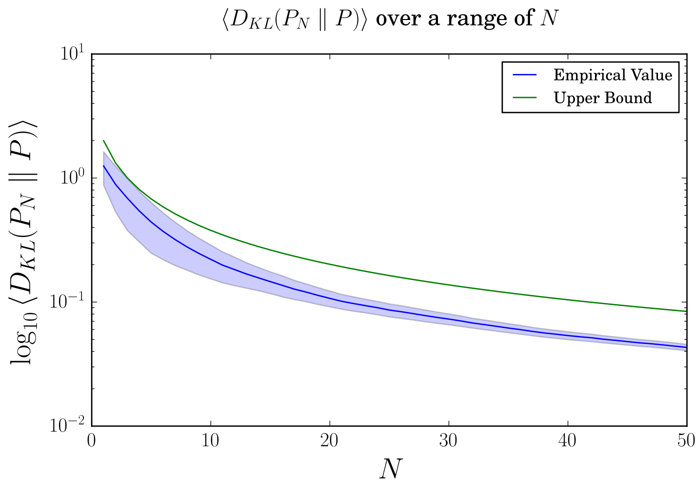


Figure 3.4: A plot of the empirically calculated expectation value of $D_{KL}(P_N \parallel P)$ (in blue), where the average was taken over 5000 paths (cf. Fig 3.2). The upper bound of Eq. 3.3 is shown in green, and the pale blue shaded area indicates $2\sigma^2$ - twice the variance - about the expectation value at each point. See that the upper bound holds and that both the empirical value and the upper bound are tending to 0.

3.2 The Szilard Engine

Now that we have found a consistent way to define energy in the Spekkens toy theory, we try to put this new knowledge to use. The origins of thermodynamics are in the study of heat engines [2], and fittingly we here try to implement a very simple engine in the context of the toy theory.

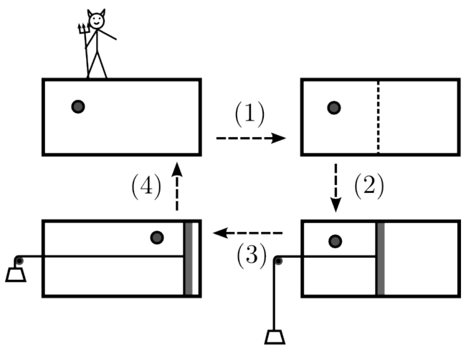


Figure 3.5: A simple illustration of the operation of the Szilard engine. The steps (1)-(4) are explained in the text. Note that this example illustrates the case where the demon's measurement shows that the particle is on the left - the alternative process would have the particle on the right.

3.2.1 Background

The Szilard engine [18] can be thought of as a variant on the infamous "Maxwell's Demon" *gedankenexperiment*, in which the ability of a "demon" to observe the positions and velocities of gas particles allows it to separate the hot and cold particles and hence decrease the entropy of the gas and (apparently) violate the second law. For now we will not concern ourselves with the second law issues, although we do delve into that debate in Appendix D.

The setting of the Szilard engine is as follows. We begin with a box in a heat bath containing a single ideal gas particle, and a demon who can sit atop the box and, should he wish, look inside and observe where the particle is. The demon also has a piston attached to a weight which he can insert into either side of the box. The engine cycle consists of the following steps:

(1) The demon measures the position of the particle to determine whether it is on the left or the right hand side of the box.

(2) The demon inserts the piston/weight on the opposite side to the particle (using the result of the measurement).

(3) The single-particle gas expands isothermally, moving the piston out and hence lifting the weight and extracting work.

(4) The demon removes the piston and the system is once again in the initial state. Note that the demon no longer knows where the particle is.

A simple illustration of this cycle is illustrated in Fig. 3.5.

3.2.2 Constructing the Szilard Engine

Whilst the Szilard engine is conceptually fairly simple, it is far from clear how we could implement it in the Spekkens theory or in any of the analogous representations that we have previously discussed. Concepts like isothermal expansion and insertion of a piston seem at first glance to be highly physical, and despite our work on energy definition in the toy theory we are still mostly beholden to Spekkens' accurate observation that "the toy theory contains almost no *physics*".

In the next few sections we shall construct a model of the Szilard engine with no reference to the toy theory, and will then demonstrate that the model we have constructed can easily be represented in the toy theory.

3.2.2.1 Framework and Notation

We claim that one can construct a Szilard engine from 3 binary (ie 2-state) systems:

- A demon D with a left state (L) and a right state (R)
- A system S (the box and the particle) with states L and R
- A weight W with a ground state 0 and a lifted/excited state 1

Whilst there is nothing going on in the Szilard engine that is in any sense quantum, it is very convenient to think of these states as being kets in the computational basis of qubits. That is to say that the states $\{L, R\}$ of the demon or the system, and the states $\{0, 1\}$ of the weight, will be written like $\{|0\rangle, |1\rangle\}$ respectively. We will also use the usual notational convention that subscripts on the kets indicate which system they describe, eg. $|0\rangle_S$ would indicate the L state of the system. In the interests of clarity we shall always list the 3 systems in the order DSW, following Eames [19].

This notation allows us to describe statistical mixtures of states. For example, if we wished to describe that the location of the gas particle was unknown and hence that the gas in some sense "filled" the box, we could describe the system with a density matrix equal to $\frac{1}{2}\left(|0\rangle\langle 0|_S + |1\rangle\langle 1|_S\right)$.

We can also describe the energy difference between the ground and lifted weight states rather succinctly by giving it a Hamiltonian, $H_W := E_l\,|1\rangle\langle 1|_W$. The Hamiltonians of the demon and the system are degenerate due to the symmetry of the demon/system states, and hence are considered to be 0.

3.2.2.2 Operation of the Engine

We can now "translate" the steps of the engine cycle from Section 3.2.1 into our bra-ket notation for the binary representation.

Working in density matrix notation, we say that the demon starts in the $|0\rangle\langle 0|_D$ state, the weight starts in its ground state $|0\rangle\langle 0|_W$, and the system starts in the aforementioned mixed state which will shall now write as $\frac{1}{2}\mathbb{1}_S$. We call the state of the whole engine at some point in its cycle $\rho(t)$, where t indexes the stage of the cycle. The initial state $\rho(0)$ of the whole engine is:

$$\rho(0) = |0\rangle\langle 0|_D \otimes \frac{1}{2}\mathbb{1}_S \otimes |0\rangle\langle 0|_W$$

Now we can run the engine, and we can write down the state of the engine after each step of the cycle:

(1) The demon makes the measurement, meaning that it takes on the state of the system, and thus the engine state becomes:

$$\rho(1) = \frac{1}{2}(|00\rangle\langle 00|_{DS} + |11\rangle\langle 11|_{DS}) \otimes |0\rangle\langle 0|_W$$

(2) The demon inserts the piston and thus, because of its informed choice as to which side to insert the piston on, puts the system into a specific state with probability 1. Without loss of generality - but at the cost of diverging slightly from the literal physical interpretation - we say that this specific state is $|1\rangle_S$. It follows that:

$$\rho(2) = \frac{1}{2}\mathbb{1}_D \otimes |1\rangle\langle 1|_S \otimes |0\rangle\langle 0|_W$$

(3) Now the gas expands and the weight is lifted to its higher-energy state:

$$\rho(3) = \frac{1}{2}\mathbb{1}_D \otimes \frac{1}{2}\mathbb{1}_S \otimes |1\rangle\langle 1|_W$$

(4) Everything must be reset to its original state. We can interpret the resetting of the weight as representing the replacement of the excited weight by a new ground-state weight, such that the removed weight(s) act as an energy store.

$$\rho(4) = \rho(0) = |0\rangle\langle 0|_D \otimes \frac{1}{2}\mathbb{1}_S \otimes |0\rangle\langle 0|_W$$

Whilst this presentation may be vaguely appealing in its apparent simplicity, it is lacking and unconvincing in several areas. We have simply stated the changes in the system state with no description of the operations that caused these changes, the justification of step (2) is arguably unconvincing and hard to understand, and perhaps most importantly we don't seem to have learned anything new about the thermodynamics - there doesn't seem to be any thermodynamics involved in the maths at all.

We now proceed to construct the above in terms of operations on the systems, and the thermodynamic results follow.

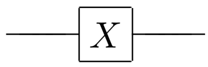
3.2.3 A Circuit Representation of the Szilard Engine

We will represent our engine using the circuit model of quantum computation, and as such a brief summary of this follows.

3.2.3.1 Quantum Circuits

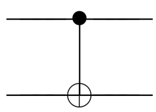
The circuit model of quantum computation represents computations by assemblies of "quantum gates", which represent reversible operations on the qubits [7]. Furthermore, these quantum circuits can be conveniently illustrated as circuit diagrams much like classical logic circuit diagrams and electric circuit diagrams. Quantum circuits have given rise to various well-known quantum computational results, such as Shor's celebrated prime factoring algorithm [20].

In the quantum circuit setting, qubits are represented by wires, and the gates by symbols on the wires. Time is implied to run from left to right. An example of a gate acting on 1 qubit is the X gate, which is represented as



and performs a "bit-flip" that takes $|0\rangle \rightarrow |1\rangle$ and vice versa in the computational basis, much like the classical NOT gate. The matrix representation of the operation is the Pauli matrix σ_x.

We can also have gates that act on more than 1 qubit. One of the most common 2-qubit gates is the controlled-NOT or CNOT gate, which is represented as



and performs the X operation (as above) on the target qubit denoted by the crossed circle (here on the bottom) if and only if the control qubit denoted by the dot (here on the top) is in the state $|1\rangle$. The matrix representation of the CNOT is, for an input state of (control $\otimes$ target):

$$CNOT = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{3.4}$$

As an aside to show the "quantumness" of these gates, note that the CNOT can entangle qubits: taking the control to be the equal superposition of $|0\rangle$ and $|1\rangle$, and the target to be in $|0\rangle$, the result of the CNOT will be a maximally entangled (Bell) state.

Despite this, when in the computational basis without linear superpositions (and hence in our treatment) there is no real difference between these gates and their classical equivalents, but we stick with the quantum notation due to the toolbox of notation and techniques that it brings along with it.

3.2.3.2 The Szilard Engine Circuit

We first present our circuit representation of the Szilard engine, and then attempt to explain it.

Result 3.2.1. *We can represent a Szilard engine as the following quantum circuit:*

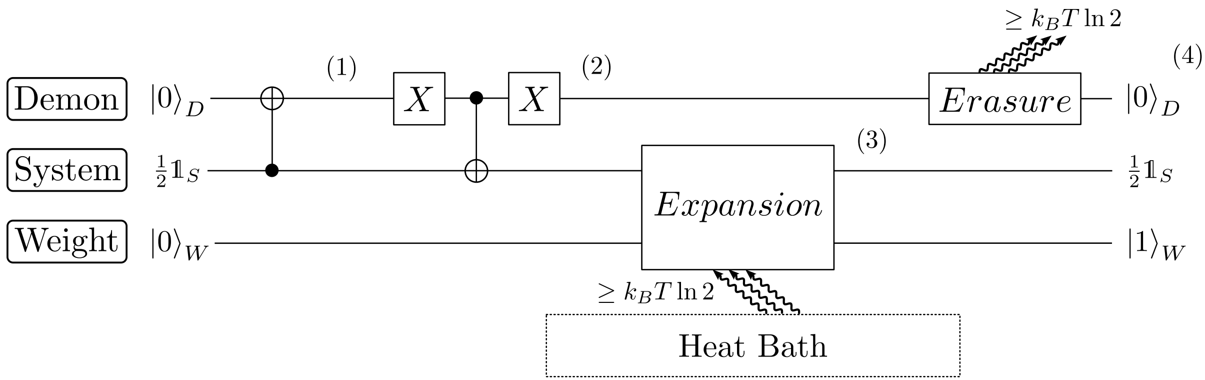


Figure 3.6: The quantum circuit comprising Result 3.2.1. Refer to the text for explanations.

The bracketed numbers correspond to the items in the list in Section 3.2.2.2. The X and CNOT gates operate as described above, and we elaborate on the "Expansion" and "Erasure" gates in the text below. The wiggly arrows on these gates represent heat transfer to/from the heat bath, and the inequalities are what we claim the work extraction/cost to be for the gates. This is also justified in the text.

The introduction of the CNOT for the first step very simply justifies the transformation of the engine state to $\rho(1)$, using the notation from Section 3.2.2.2. The Demon starts in the left-hand state and only "changes his mind" if the control qubit - the system - is in the right-hand state.

The second step - the transformation to $\rho(2)$ - was previously difficult to justify but now becomes quite simple. The demon's 2-state mind is too simple to be able to work out how to put the piston on the opposite side on its own, so he "flips" his mind using the X gate, so he is now thinking of the side opposite to the particle, and he then controls a CNOT on the system (inserting the piston) - thus putting the system in state $|1\rangle_S$ with certainty. Recall from the definitions in Section 3.2.2.1 that this is completely analogous to confining the gas particle in one half of the box.

The demon then flips his mind back to its previous state using another X gate (recall that the Pauli matrices square to the identity, and that in these computational basis states the CNOT does not modify the control qubit). This second flip is not strictly necessary - it doesn't really matter what state the demon is in - but ensures complete consistency with our initial model of the engine.

The third step and fourth steps, illustrated as an "Expansion" gate and an "Erasure" gate, respectively, require rather more detail to be explained.

3.2.4 The Expansion Step

Firstly we reiterate that the expansion step takes the engine from its state after the piston insertion, $\rho(2) = \frac{1}{2}\mathbb{1}_D \otimes |1\rangle\langle 1|_S \otimes |0\rangle\langle 0|_W$, to the state where the weight has been lifted and the gas has expanded, $\rho(3) = \frac{1}{2}\mathbb{1}_D \otimes \frac{1}{2}\mathbb{1}_S \otimes |1\rangle\langle 1|_W$.

Consider that there are in some sense 2 processes going on here - the lifting of the weight, and the expansion of the gas in the system. Whilst it is surely unphysical to do so, at this stage we develop the expansion step by at first considering these two processes to happen separately.

As we know that the system begins in $|1\rangle_S$ and the weight in $|0\rangle_W$, it is apparent that the lifting of the weight can be represented by a CNOT controlled on the system - this is our first substep of expansion, which we do not claim to be physical.

Now we need an operation to expand the gas; that is perform the transformation $|1\rangle_S \xrightarrow{expansion} \frac{1}{2}\mathbb{1}_S$. But we've run into a problem here! Quantum gates represent unitary operations on the qubits (at least in the textbook setting), which correspond to rotations of the Bloch Sphere, but the mixed state $\frac{1}{2}\mathbb{1}_S$ lies within the Bloch sphere. In order to perform the required transformation we need a non-unitary transformation, and for this we shall invoke the quantum operations formalism [7].

3.2.4.1 Quantum Operations

Consider that a quantum system coupled to an environment can be totally described by the density matrix:

$$\rho_{world} = \rho_{sys} \otimes \rho_{env}$$

We assume that the total system (the world) evolves unitarily under some transformation U, so we can say that that the world evolves to:

$$\rho'_{world} = U(\rho_{sys} \otimes \rho_{env})U^\dagger$$

Now we can trace out the environment and see that the system has been transformed non-unitarily to:

$$\rho'_{sys} = \mathrm{Tr}_{env}\left[\rho'_{world}\right] = \mathrm{Tr}_{env}\left[U(\rho_{sys} \otimes \rho_{env})U^\dagger\right]$$

It can be shown [7] that we can represent the transformation on the system by a sum of operators acting on its initial state. That is to say:

$$\rho'_{sys} = \sum_k E_k \rho_{sys} E_k^\dagger$$

where the $\{E_k\}$ are a set of operators known as the "operation elements" which satisfy the constraint that $\sum_k E_k E_k^\dagger \leq \mathbb{1}$. Indeed, this representation goes further than the tracing-out technique and can (when the inequality is not saturated) represent non-trace-preserving transformations.

We now give an example of a specific quantum operation that is of interest to us, known as the Generalised Amplitude Damping (GAD) channel, which represent "dissipation to an environment at finite temperature" - information on all of this can once again be found in Ref. [7], with a derivation of the channel available in (for example) Ref. [21]. The $\{E_k\}$ for the GAD channel are given by:

$$\begin{aligned} E_0 &= \sqrt{p}\begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix} \\ E_1 &= \sqrt{p}\begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix} \\ E_2 &= \sqrt{1-p}\begin{pmatrix} \sqrt{1-\gamma} & 0 \\ 0 & 1 \end{pmatrix} \\ E_3 &= \sqrt{1-p}\begin{pmatrix} 0 & 0 \\ \sqrt{\gamma} & 0 \end{pmatrix} \end{aligned} \tag{3.5}$$

With the stationary state of the system given by:

$$\rho(t \to \infty) = \begin{pmatrix} p & 0 \\ 0 & 1-p \end{pmatrix} \tag{3.6}$$

Where $p \in [0,1]$, and γ is a parameter that can be generally written (again, see Ref. [21] for a derivation) as a function of time, $\gamma(t) = 1 - e^{-\alpha t}$ with α being some function of bath temperature and the strength of the interaction between the bath and the system.

Hopefully the reason that we have presented this channel is clear: isothermal expansion of a gas in a box that is in a heat bath is the same thing as dissipation to an environment at finite temperature. We can now construct this component of the "Expansion" gate.

3.2.4.2 Constructing the Expansion Gate

Since we want to describe the thermalisation of the system S, it is appropriate to take its final state to be the state at thermal equilibrium, which for a general system with energy eigenstates $|\psi_n\rangle$ in a bath at temperature $T = (k\beta)^{-1}$ is described by (see for example Ref. [15]):

$$\rho_{eq} = \frac{\sum_n e^{-\beta E_n} |\psi_n\rangle \langle\psi_n|}{\sum_n e^{-\beta E_n}}$$

It is easy to see that for a degenerate Hamiltonian, meaning that all of the E_n are equal, ρ_{eq} is the maximally mixed state $\frac{1}{2}\mathbb{1}$. Then Eq. 3.6 immediately tells us that our Generalised Amplitude Damping channel has $p = \frac{1}{2}$. The matrix representation of the initial state $|1\rangle\langle 1|_S$ is:

$$\rho_S = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

Using the operators in Eq. 3.5, we find that:

$$\rho_s' = \sum_k E_k \rho_s E_k^\dagger = \begin{pmatrix} p\gamma & 0 \\ 0 & 1-p\gamma \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\gamma & 0 \\ 0 & 1-\frac{1}{2}\gamma \end{pmatrix}$$

Recalling that $\gamma(t) = 1 - e^{-\alpha t}$, see that in the limit of $t \to \infty$, we do indeed recover the equilibrium state, as expected:

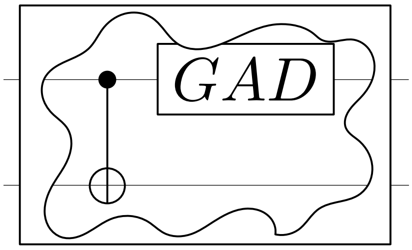


Figure 3.7: An illustration of how we have constructed our “Expansion” gate. The “GAD” gate represents the Generalised Amplitude Damping channel. Once again we emphasise that this should not be considered to be physical: the 2 steps shown within the gate happen simultaneously.

$$\rho'_s(t \to \infty) = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \frac{1}{2}\mathbb{1}$$

Now, combining this with our previous description of the weight-lifting CNOT gate, we can construct the full “Expansion” gate. Fig. 3.7 illustrates this construction, but once again we must reiterate that it is unphysical to consider the 2 constituent steps separately. To resolve this, we can just use the matrix forms of the 2 steps given in Eq. 3.4 for the CNOT and Eq. 3.5 for the GAD channel. Remembering the correct ordering of the matrix multiplication - AB for operation B performed first - and noting that the GAD channel acts on 1 qubit only, we find that the set of operators $\{E_k^{exp}\}$ for the full expansion channel are given by the following identity, where the outer subscripts denote the system that the operator acts upon:

$$(E_k^{exp})_{S\otimes W} = \left[(E_k^{GAD})_S \otimes \mathbb{1}_W\right] \cdot CNOT_{S\otimes W}$$

We can now explicitly write the operators for the channel.

Result 3.2.2. *The “Expansion” channel, which models the expansion of a gas lifting a weight, can be represented by the (non-unique) operation elements:*

$$\begin{aligned}
E_0 &= \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \sqrt{1-\gamma} \\ 0 & 0 & \sqrt{1-\gamma} & 0 \end{pmatrix} \\
E_1 &= \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 0 & 0 & \sqrt{\gamma} \\ 0 & 0 & \sqrt{\gamma} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\
E_2 &= \frac{1}{\sqrt{2}}\begin{pmatrix} \sqrt{1-\gamma} & 0 & 0 & 0 \\ 0 & \sqrt{1-\gamma} & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \\
E_3 &= \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \sqrt{\gamma} & 0 & 0 & 0 \\ 0 & \sqrt{\gamma} & 0 & 0 \end{pmatrix}
\end{aligned}$$

Where we have inserted the previously-derived value of $p = \frac{1}{2}$ for the GAD channel, and γ takes the form from before, $\gamma(t) = 1 - e^{-\alpha t}$.

3.2.4.3 Work Extraction in the Expansion Step

In Fig. 3.6 in Result 3.2.1, we illustrated that the expansion step in some sense involved the interaction of the engine with a heat bath. We also wrote an inequality next to the arrows representing this interaction, and claimed that this related to the work extraction. We present 2 justifications:

- **Justification by entropy change**[3]
  Consider the (Gibbs) entropy change from step (2) to step (3): both before and after we trivially have $S_G(W) = 0$, but $S_G(S)$ increases from 0 (pre-expansion; definitely in state $|1\rangle_S$) to:

  $$S_G(S') = -2(\frac{1}{2}k_B \ln \frac{1}{2}) = k_B \ln 2$$

  where k_B is Boltzmann's constant. Recalling that $\mathrm{d}S = \mathrm{d}Q/T$ we argue that this entropy change is associated with a work extraction of at least $k_B T \ln 2$ (where T is the bath temperature), as in the non-irreversible case we would actually find that $\mathrm{d}S = (\mathrm{d}Q/T) + \mathrm{d}S_i$. [22]

- **Operational Justification**
  Alternatively, we can use a new result of Faist et al (2015) [23] to calculate the minimum work cost of any operation. The authors give a precursor to their main result that we shall use:

  $$W^{\epsilon=0} \geq k_B T \ln 2 \cdot \log_2 \left\| \mathcal{E}\left(\Pi_X\right) \right\|_\infty$$

  where the superscript $\epsilon = 0$ indicated that there is no entropy smoothing involved, and $\mathcal{E}(\Pi_X)$ indicates some quantum operation $\mathcal{E}$ acting on Π_X, the projector onto the support of the input state[4] to the operation, X. For our input state of $|1\rangle\langle 1|_S \otimes |0\rangle\langle 0|_W$, the projector Π_X is identical to the input state, and we find that:

  $$\begin{aligned} W^{\epsilon=0} &\geq k_B T \ln 2 \cdot \log_2 \left\| \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \gamma & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2-\gamma \end{pmatrix} \right\|_\infty \\ &= k_B T \ln 2 \cdot \log_2 \left[\frac{1}{2}(2-\gamma) \right] \text{ for } \gamma \in [0,1] \\ &= k_B T \ln 2 \cdot \log_2 \frac{1}{2} = -k_B T \ln 2 \text{ as } t \to \infty \end{aligned}$$

  Seeing that a negative work cost corresponds to a work extraction, we recover

[3] This sort of thinking is indebted to Renato Renner's brief but excellent lecture notes on quantum thermodynamics. However, these notes have recently disappeared from the ETH website and so cannot be referenced, and the method is recalled from memory.

[4] For any state with a (general) density matrix $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$, the projector onto its support is given by $\Pi = \sum_i |\psi_i\rangle\langle\psi_i|$

the result from the entropy change justification. Note that an advantage of this approach is that we can calculate entropy changes for any input states, including truly quantum states - that is to say, off-diagonal density matrix elements.

This means that the Hamiltonian that we wrote down for the weight qubit in Section 3.2.2.1, $E_l \left|1\right\rangle \left\langle 1\right|_W$, should have $E_l = k_B T \ln 2$ or else we may find that the weight is "too heavy" for the engine to lift.

3.2.5 The Erasure Step

The erasure step takes the mixed state (and in principle, any state) to some predetermined state which we require to be $\left|0\right\rangle_D$. It is clearly logically irreversible in that there is no way to tell what the initial state was given the post-erasure state. The stated work cost of (at least) $k_B T \ln 2$ is given by the Landauer principle [24], which claims that the entropy decrease of the "memory" (ie the demon qubit) must be accompanied by an equal or greater entropy increase in the environment and can be partially justified by the same kind of entropic arguments as above.

The Landauer principle, whilst broadly accepted, is somewhat contentious among thermodynamicists and indeed philosophers [25, 26]. We will not involve ourselves in this debate here but Appendix D discusses the thermodynamic issues in the Szilard engine and argues resolutions using our model, and hence does wade into the debate somewhat. For now we consider the Landauer principle to be acceptable, and can refer the reader to the literature for justifications and derivations of the stated work cost:

- del Rio et al present an excellent and highly readable account of the erasure of some pure state in Ref. [27]
- Weilenmann et al argue that they can justify the thermodynamic meanings of information theoretic entropies without recourse to the Landauer principle in Ref. [28]
- Hilt et al argue that the Landauer principle can indeed be derived from considering some harmonic oscillator coupled to a bath in Ref. [29]
- Bérut et al claim to have experimentally verified the Landauer principle in Ref. [30] (perhaps this is the most convincing argument in its favour!)

Finally we note that the equality of the work cost of erasure and the work extraction due to expansion would appear to have saved the second law - this is effectively the Bennett resolution of the Szilard engine paradox [31].

3.2.6 The Szilard Engine in the Spekkens Theory

Finally we are in a position to give the reader what we promised long ago at the start of Section 3.2.2: a construction of the engine in the context of the Spekkens toy theory. Our use of computational basis qubit states means that we can directly use the *toy* stabilizer notation to "operate" our quantum circuit (of Result 3.2.1) as the differences between the toy stabilizers and the stabilizers on qubits do not come in to play for these simple states.

3.2.6.1 Translating the Steps to Toy Stabilizer Notation

Expressing the qubit states given in Section 3.2.2.2 in the stabilizer notation is very simple. We only ever consider operations on 2 of the qubits at any time and so for clarity will talk only about 2-qubit states, where the third is implicitly in the state it was last described to be in. Following Section 3.2.2.2, we are initially in:

$$\rho(0) = |0\rangle\langle 0|_D \otimes \frac{1}{2}\mathbb{1}_S \otimes |0\rangle\langle 0|_W$$

so we can write the DS and W stabilizers as:

$$\begin{aligned} \langle G\rangle_{DS}(0) &= \langle \mathcal{Z_D}\rangle \\ \langle G\rangle_W(0) &= \mathcal{Z}_W \end{aligned}$$

Now we can operate the engine:

(1) Noting that the CNOT gate controlled on the first bit is equivalent to the following transformation of toy stabilizers [4]:

$$\mathcal{Z}_1 \to \mathcal{Z}_1,\ \mathcal{Z}_2 \to \mathcal{Z}_1\mathcal{Z}_2$$

$$\mathcal{X}_1 \to \mathcal{X}_1\mathcal{X}_2,\ \mathcal{X}_2 \to \mathcal{X}_2$$

We can immediately write the state of $\langle G\rangle_{DS}$ after measurement:

$$\begin{aligned} \langle G\rangle_{DS}(1) &= \langle \mathcal{Z}_D\rangle \xrightarrow{CNOT} \langle \mathcal{Z}_D\mathcal{Z}_S\rangle \\ &= \langle \mathcal{Z}_D\mathcal{Z}_S\rangle \end{aligned}$$

(2) Next we recall that the X gate transforms the stabilizers like:

$$\mathcal{Z} \to -\mathcal{Z}$$

And so we write the piston insertion step (consisting of 3 sequential operations: X gate on the demon, CNOT controlled by the demon, X gate on the demon) as:

$$\begin{aligned} \langle G\rangle_{DS}(2) &= \langle \mathcal{Z}_D\mathcal{Z}_S\rangle \xrightarrow{X} \langle -\mathcal{Z}_D\mathcal{Z}_S\rangle \xrightarrow{CNOT} \langle -\mathcal{Z}_S\rangle \xrightarrow{X} \langle -\mathcal{Z}_S\rangle \\ &= \langle -\mathcal{Z}_S\rangle \end{aligned}$$

where the fact that the final transformation on the 1st line does nothing to the generator is due to the fact that a bit flip acting on a fully mixed state still gives a fully mixed state. Our previous statement that this 2nd X gate was unnecessary is now clearly justified.

(3) Now we look at the SW stabilizers in order to perform the expansion step: we are at first in a pure state where the system is in $-\mathcal{Z}_S$ and the weight is in $\mathcal{Z}_W$:

$$\langle G\rangle_{SW}(2) = \langle -\mathcal{Z}_S, \mathcal{Z}_W\rangle$$

We return to the unphysical representation of this stage as consisting of a CNOT and a GAD channel; see that the GAD channel takes any operator to the identity, and recall the CNOT transformations from item (1). The CNOT stage (we arbitrarily call the time index for this 2.5) is:

$$\begin{aligned} \langle G\rangle_{SW}(2.5) &= \langle -\mathcal{Z}_S, \mathcal{Z}_W\rangle \xrightarrow{CNOT} \langle -\mathcal{Z}_S, \mathcal{Z}_S\mathcal{Z}_W\rangle \\ &= \{\mathcal{I}_S\mathcal{I}_W, -\mathcal{Z}_S\mathcal{I}_W, \mathcal{Z}_S\mathcal{Z}_W, -\mathcal{I}_S\mathcal{Z}_W\} \\ &= \langle -\mathcal{Z}_S, -\mathcal{Z}_W\rangle \end{aligned}$$

and the GAD stage then transforms this to:

$$\langle G \rangle_{SW}(3) = \langle -\mathcal{Z}_S, -\mathcal{Z}_W \rangle \xrightarrow{GAD} \langle -\mathcal{Z}_W \rangle$$

(4) Finally we must perform the erasure step and take the identity state of the demon, $\mathcal{I}_D$, back to the demon's initial state $\mathcal{Z}_D$. The generator for the identity is empty and so this is slightly awkwardly notated as:

$$\langle G \rangle_D(4) = \langle \rangle \xrightarrow{erasure} \langle \mathcal{Z}_D \rangle$$

and for clarity we rewrite the SW state:

$$\langle G \rangle_{SW}(4) = \langle G \rangle_{SW}(3) = \langle -\mathcal{Z}_W \rangle$$

Which clearly corresponds exactly to the state $\rho(4)$ given in Section 3.2.2.2.

While we do not do this here, it is clear to see that all of the above 2-bit states can be represented in the original diagrammatic notation by using the mappings given in Appendix A.

3.2.6.2 A Diversion: Irreversible Transformations on Spekkens States

In the seemingly innocuous act of writing down the action of the GAD channel and the erasure on the toy stabilizers we have inadvertently introduced a notion of irreversible transformations into the theory, something that has not previously been studied. By "irreversible" we mean logically irreversible; there is no way to tell the initial state from looking at the final state.

In the original notation this is quite appealingly illustrated, in that in addition to moving around the ontic states, several states are mapped into the same place or one state is mapped to several places: the size of the ontic base is not conserved. Fig. 3.8 is an illustration of the transformation performed in the expansion stage (the third step) of the engine cycle.

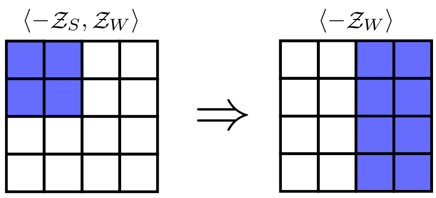


Figure 3.8: An illustration of the irreversible transformation of ontic states brought about by the expansion stage of the engine's cycle.

Furthermore, we note that there is an appealing geometric interpretation of the irreversible transformations. It is common to illustrate quantum operations as deformations of the Bloch sphere [7], and so we suggest that irreversible transformations in the Spekkens theory can be illustrated as deformations of the octahedron of Section 2.4.1. In fact, considering that this octahedron which is defined by the 1-norm is entirely bounded by the sphere of radius 1, it follows immediately that all valid quantum operations (that is, those that correspond to affine transformations on the Bloch sphere) can potentially be applied to the Spekkens theory.

We give an example of such a deformation for the GAD case. We know that the GAD operators (Eq. 3.5) transform the Bloch sphere to [7]:

$$\left(\frac{x}{\sqrt{1-\gamma}}\right)^2 + \left(\frac{y}{\sqrt{1-\gamma}}\right)^2 + \left(\frac{z-\gamma(2p-1)}{1-\gamma}\right)^2 = 1$$

And so it turns out to be very easy to write down the transformed Spekkens state octahedron:

$$\left|\frac{x}{\sqrt{1-\gamma}}\right| + \left|\frac{y}{\sqrt{1-\gamma}}\right| + \left|\frac{z-\gamma(2p-1)}{1-\gamma}\right| = 1$$

This shape is illustrated in Fig. 3.9 for specific values of p and γ.

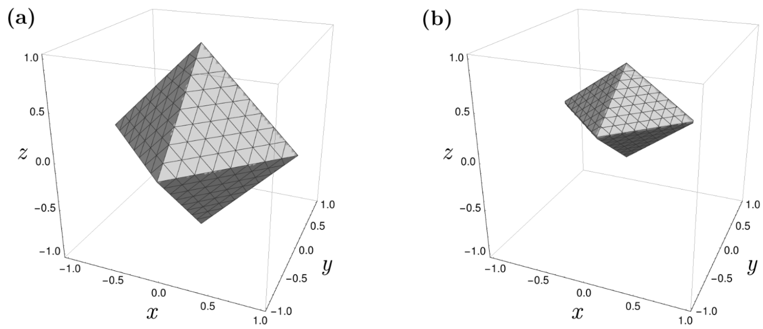


Figure 3.9: An illustration of the deformation of the Spekkens state octahedron that the Generalised Amplitude Damping (GAD) channel causes. **(a)** shows the original octahedron (defined by the 1-norm), and **(b)** shows it after applying a GAD channel with $p = 0.8$ and $\gamma = 0.5$.

3.2.6.3 Violation of the Knowledge Balance Principle

We now briefly and in a rather hand-waving manner consider the behaviour of the engine when we allow the knowledge balance principle to be violated. We consider the case where the weight is lifted to a disallowed state, illustrated in Fig. 3.10

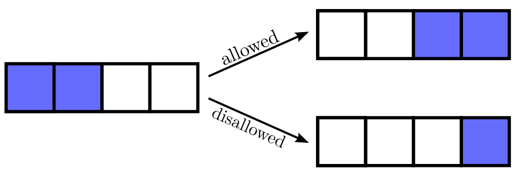


Figure 3.10: On the left is the weight state before it is "lifted", and on the right we have illustrated the allowed state that it has been lifted to in previous sections as well as the disallowed state which we are now considering.

Our first indication that something is awry (in addition to the knowledge balance principle violation) is that the Shannon entropy of the weight changes when it is lifted. For this single state with a uniform canonical probability distribution the entropy is dependent only on the size of the ontic base[5] $|P|$:

$$S = \log_2 |P|$$

[5]It may be of historical interest to the reader that this entropy function is also known as Hartley entropy, and predates Shannon entropy [32].

So the entropy change ΔS is equal to $\log_2 1 - \log_2 2 = -1$. Again assuming the Landauer principle to be valid, we have 2 choices when trying to interpret this thermodynamically: if the weight is not considered to be in thermal contact with the bath, it implies that the expansion step of the gas has provided the work that we presume to have effected this entropy decrease, and hence the engine appears to be extracting more work than the second law should allow. Alternatively, if we allow the weight to exchange energy with the bath, this entropy change (at least partially) cancels out that of the gas expansion and so the engine extracts less work from the bath.

Using the geometric mapping from Section 2.4.1, this final disallowed state corresponds to the point $\frac{1}{2}(-1, 1, -1)$ within the Bloch sphere (this is in fact the state illustrated in Fig. 2.6). Using Eq. 2.5 - the relationship between the density matrix and this point - we can see that in some sense it corresponds to the quantum state:

$$\begin{aligned}
\rho &= \frac{1}{2}(\mathbb{1} + \boldsymbol{r} \cdot \boldsymbol{\sigma}) \\
&= \frac{1}{2}\left(\mathbb{1} + \frac{1}{2}(-\sigma_x + \sigma_y - \sigma_z)\right) \\
&= \frac{1}{4}\begin{pmatrix} 1 & -1-i \\ -1+i & 3 \end{pmatrix}
\end{aligned}$$

And so, assuming that nothing has changed with the expansion of the gas (ie the system being irreversibly transformed to the mixed state), the final state of SW is given by:

$$\begin{aligned}
\rho'_{SW} &= \frac{1}{2}\mathbb{1}_S \otimes \frac{1}{4}\begin{pmatrix} 1 & -1-i \\ -1+i & 3 \end{pmatrix}_W \\
&= \frac{1}{8}\begin{pmatrix} 1 & -1-i & 0 & 0 \\ -1+i & 3 & 0 & 0 \\ 0 & 0 & 1 & -1-i \\ 0 & 0 & -1+i & 3 \end{pmatrix}
\end{aligned}$$

Not only does this look rather different to the final state density matrix found in Section 3.2.4.3, but as the input state is the same as before we can once again apply the result of Faist et al:

$$\begin{aligned}
W^{\epsilon=0} &\geq k_B T \ln 2 \cdot \log_2 \left\| \rho'_{SW} \right\|_\infty \\
&= k_B T \ln 2 \cdot \log_2 \left[\frac{1}{8}\left(3 + \sqrt{2}\right)\right] \\
&\approx -0.86 k_B T \ln 2
\end{aligned}$$

Noting that using this assumes that the weight is in thermal contact with the bath. As we claimed, the engine is now less efficient. We are aware that this final section of our investigations is by no means rigorous, but we do hope that the reader is convinced that the knowledge balance principle - and adherence to it - does have thermodynamic implications.

Chapter 4

Conclusions

The Spekkens toy theory is an elegant and simple model that, by simply colouring in squares according to one fundamental rule, presents an appealing argument that almost all of quantum theory can be reproduced by taking the view that quantum states are states of knowledge rather than of reality. However, by the author's own admission the theory "contains almost no physics". In this report, we have attempted to introduce something truly and tangibly physical to the theory: thermodynamics.

We have shown that, when considering the states of knowledge of the theory to be representative of huge ensembles of similar states, we can extend the theory to include energies of states in a way that violates neither the fundamental rule of the theory (the knowledge balance principle) nor some reasonable physical assumptions about how energy should behave. Building upon this definition, we have implemented a simple heat engine in the toy theory and shown that it is completely operable within the theory's framework. Finally, we have argued that the knowledge balance principle, and our choice of whether to adhere to it, has thermodynamic implications.

In our rather circuitous route to the above results, we have made progress towards a geometric representation of states and have used this representation to help us with the results. We have used results from large deviations theory to quantify the way in which the aforementioned definition of energy does or does not adhere to the knowledge balance principle and we have combined group theoretic and geometric techniques in order to examine properties related to symmetries and degeneracies of states. We have formulated the infamous Szilard engine as a series of quantum operations upon qubits, which so far as we are aware is a new result.

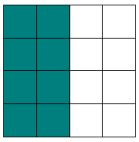
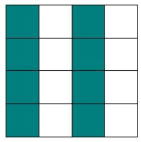
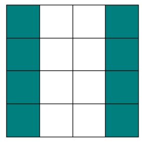
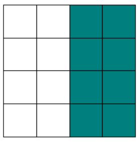
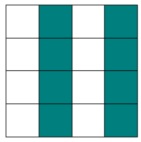
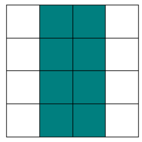
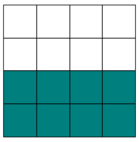
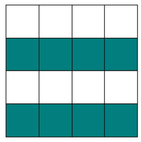
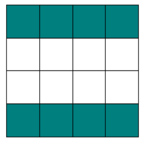
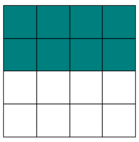
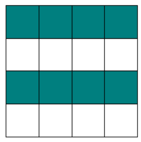
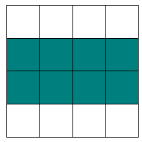
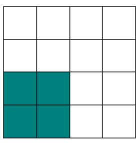
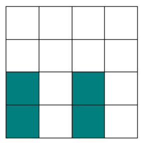
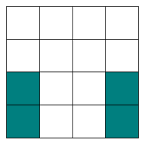
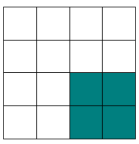
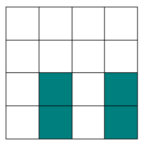
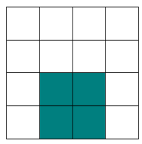
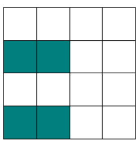
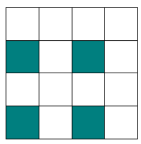
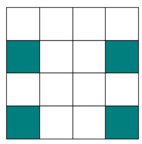
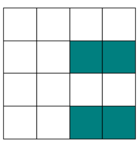
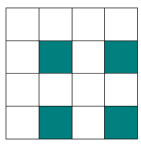
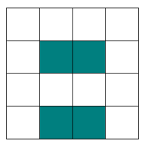
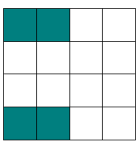
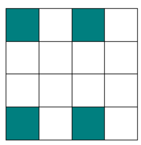
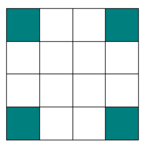
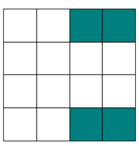
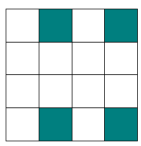
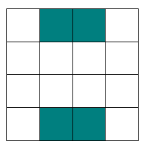
We suggest that the work on the geometric representation of Spekkens theory states could be extended and solidified, and that in turn this framework could potentially be used to find new results in the context of the theory. In particular, some kind of deterministic "plug-in" measure of knowledge balance principle violation that quantified how much the principle was violated for completely general states would be very appealing.

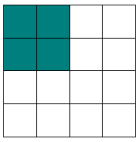
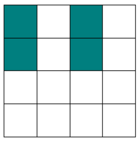
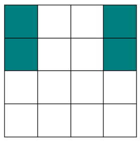
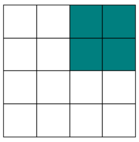
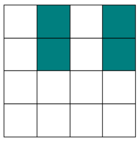
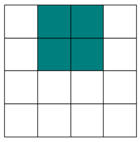
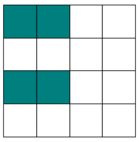
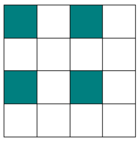
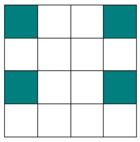
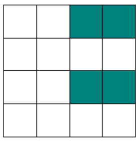
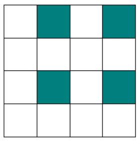
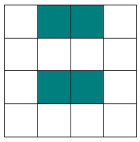
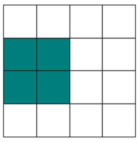
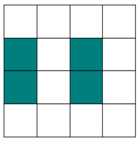
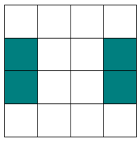
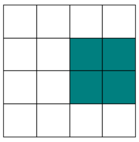
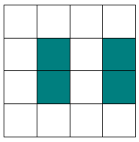
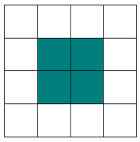
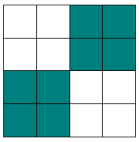
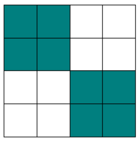
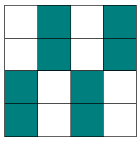
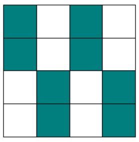
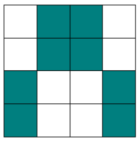
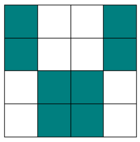
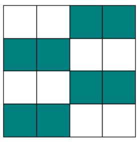
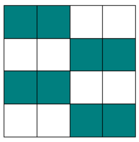
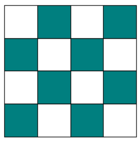
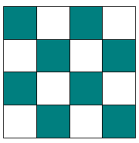
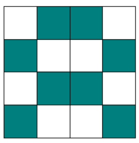
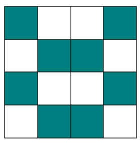
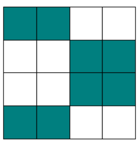
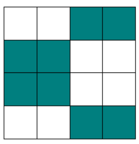
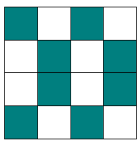
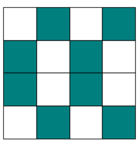
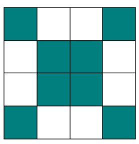
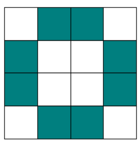
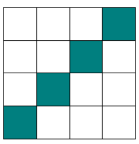
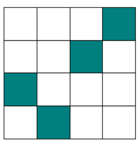
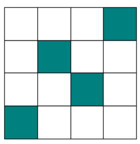
In addition, it would be interesting to try to implement more engines/refrigerators in the toy theory and see how these behave, particularly with regard to knowledge balance principle adherence. Indeed, any further work trying to link principle adherence to thermodynamic behaviour would be extremely interesting to see.

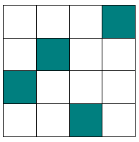
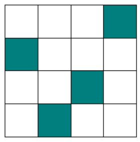
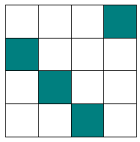
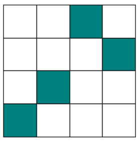
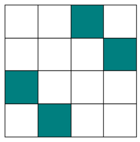
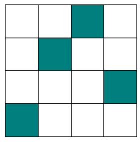
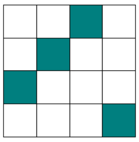
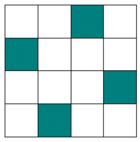
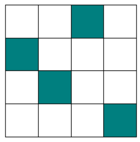
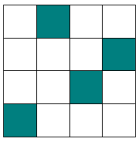
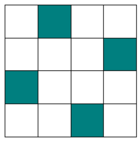
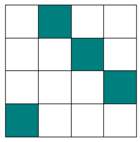
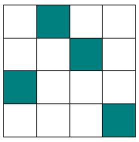
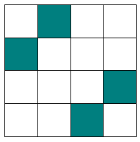
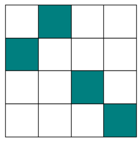
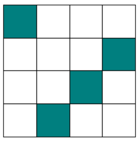
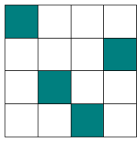
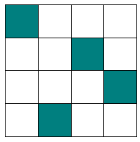
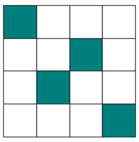
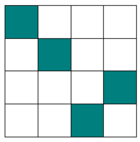
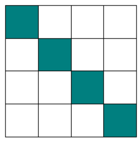
Quantum mechanics and quantum information theory are no doubt highly successful theories with great explanatory and quantitative power, but it is still unsatisfactory to see that we are yet to fully justify our present formulation of quantum mechanics from physical principles. This report is in some vague sense an expression of the hope that such a justification, if it exists, may emerge from thermodynamic principles.

Appendix A

Map of 2-bit States and Toy Stabilizers

$\langle \mathcal{Z}_2 \rangle$		$\langle \mathcal{X}_2 \rangle$		$\langle \mathcal{Y}_2 \rangle$	
$\langle -\mathcal{Z}_2 \rangle$		$\langle -\mathcal{X}_2 \rangle$		$\langle -\mathcal{Y}_2 \rangle$	
$\langle \mathcal{Z}_1 \rangle$		$\langle \mathcal{X}_1 \rangle$		$\langle \mathcal{Y}_1 \rangle$	
$\langle -\mathcal{Z}_1 \rangle$		$\langle -\mathcal{X}_1 \rangle$		$\langle -\mathcal{Y}_1 \rangle$	
$\langle \mathcal{Z}_1, \mathcal{Z}_2 \rangle$		$\langle \mathcal{Z}_1, \mathcal{X}_2 \rangle$		$\langle \mathcal{Z}_1, \mathcal{Y}_2 \rangle$	
$\langle \mathcal{Z}_1, -\mathcal{Z}_2 \rangle$		$\langle \mathcal{Z}_1, -\mathcal{X}_2 \rangle$		$\langle \mathcal{Z}_1, -\mathcal{Y}_2 \rangle$	
$\langle \mathcal{X}_1, \mathcal{Z}_2 \rangle$		$\langle \mathcal{X}_1, \mathcal{X}_2 \rangle$		$\langle \mathcal{X}_1, \mathcal{Y}_2 \rangle$	
$\langle \mathcal{X}_1, -\mathcal{Z}_2 \rangle$		$\langle \mathcal{X}_1, -\mathcal{X}_2 \rangle$		$\langle \mathcal{X}_1, -\mathcal{Y}_2 \rangle$	
$\langle \mathcal{Y}_1, \mathcal{Z}_2 \rangle$		$\langle \mathcal{Y}_1, \mathcal{X}_2 \rangle$		$\langle \mathcal{Y}_1, \mathcal{Y}_2 \rangle$	
$\langle \mathcal{Y}_1, -\mathcal{Z}_2 \rangle$		$\langle \mathcal{Y}_1, -\mathcal{X}_2 \rangle$		$\langle \mathcal{Y}_1, -\mathcal{Y}_2 \rangle$	

$\langle -\mathcal{Z}_1, \mathcal{Z}_2 \rangle$		$\langle -\mathcal{Z}_1, \mathcal{X}_2 \rangle$		$\langle -\mathcal{Z}_1, \mathcal{Y}_2 \rangle$	
$\langle -\mathcal{Z}_1, -\mathcal{Z}_2 \rangle$		$\langle -\mathcal{Z}_1, -\mathcal{X}_2 \rangle$		$\langle -\mathcal{Z}_1, -\mathcal{Y}_2 \rangle$	
$\langle -\mathcal{X}_1, \mathcal{Z}_2 \rangle$		$\langle -\mathcal{X}_1, \mathcal{X}_2 \rangle$		$\langle -\mathcal{X}_1, \mathcal{Y}_2 \rangle$	
$\langle -\mathcal{X}_1, -\mathcal{Z}_2 \rangle$		$\langle -\mathcal{X}_1, -\mathcal{X}_2 \rangle$		$\langle -\mathcal{X}_1, -\mathcal{Y}_2 \rangle$	
$\langle -\mathcal{Y}_1, \mathcal{Z}_2 \rangle$		$\langle -\mathcal{Y}_1, \mathcal{X}_2 \rangle$		$\langle -\mathcal{Y}_1, \mathcal{Y}_2 \rangle$	
$\langle -\mathcal{Y}_1, -\mathcal{Z}_2 \rangle$		$\langle -\mathcal{Y}_1, -\mathcal{X}_2 \rangle$		$\langle -\mathcal{Y}_1, -\mathcal{Y}_2 \rangle$	
$\langle \mathcal{Z}_1 \mathcal{Z}_2 \rangle$		$\langle -\mathcal{Z}_1 \mathcal{Z}_2 \rangle$		$\langle \mathcal{Z}_1 \mathcal{X}_2 \rangle$	
$\langle -\mathcal{Z}_1 \mathcal{X}_2 \rangle$		$\langle \mathcal{Z}_1 \mathcal{Y}_2 \rangle$		$\langle -\mathcal{Z}_1 \mathcal{Y}_2 \rangle$	
$\langle \mathcal{X}_1 \mathcal{Z}_2 \rangle$		$\langle -\mathcal{X}_1 \mathcal{Z}_2 \rangle$		$\langle \mathcal{X}_1 \mathcal{X}_2 \rangle$	
$\langle -\mathcal{X}_1 \mathcal{X}_2 \rangle$		$\langle \mathcal{X}_1 \mathcal{Y}_2 \rangle$		$\langle -\mathcal{X}_1 \mathcal{Y}_2 \rangle$	
$\langle \mathcal{Y}_1 \mathcal{Z}_2 \rangle$		$\langle -\mathcal{Y}_1 \mathcal{Z}_2 \rangle$		$\langle \mathcal{Y}_1 \mathcal{X}_2 \rangle$	
$\langle -\mathcal{Y}_1 \mathcal{X}_2 \rangle$		$\langle \mathcal{Y}_1 \mathcal{Y}_2 \rangle$		$\langle -\mathcal{Y}_1 \mathcal{Y}_2 \rangle$	
$\langle \mathcal{Z}_1 \mathcal{Z}_2, \mathcal{X}_1 \mathcal{X}_2 \rangle$		$\langle \mathcal{Z}_1 \mathcal{Z}_2, -\mathcal{X}_1 \mathcal{Y}_2 \rangle$		$\langle \mathcal{Z}_1 \mathcal{X}_2, \mathcal{X}_1 \mathcal{Z}_2 \rangle$	

$\langle \mathcal{Z}_1\mathcal{X}_2, -\mathcal{X}_1\mathcal{Y}_2\rangle$		$\langle \mathcal{X}_1\mathcal{Z}_2, -\mathcal{Z}_1\mathcal{Y}_2\rangle$		$\langle \mathcal{X}_1\mathcal{X}_2, -\mathcal{Z}_1\mathcal{Y}_2\rangle$	
$\langle \mathcal{Z}_1\mathcal{Z}_2, \mathcal{X}_1\mathcal{Y}_2\rangle$		$\langle \mathcal{Z}_1\mathcal{Z}_2, -\mathcal{X}_1\mathcal{X}_2\rangle$		$\langle \mathcal{Z}_1\mathcal{Y}_2, \mathcal{X}_1\mathcal{Z}_2\rangle$	
$\langle \mathcal{Z}_1\mathcal{Y}_2, -\mathcal{X}_1\mathcal{X}_2\rangle$		$\langle \mathcal{X}_1\mathcal{Z}_2, -\mathcal{Z}_1\mathcal{X}_2\rangle$		$\langle \mathcal{X}_1\mathcal{Y}_2, -\mathcal{Z}_1\mathcal{X}_2\rangle$	
$\langle \mathcal{Z}_1\mathcal{X}_2, \mathcal{X}_1\mathcal{Y}_2\rangle$		$\langle \mathcal{Z}_1\mathcal{X}_2, -\mathcal{X}_1\mathcal{Z}_2\rangle$		$\langle \mathcal{Z}_1\mathcal{Y}_2, \mathcal{X}_1\mathcal{X}_2\rangle$	
$\langle \mathcal{Z}_1\mathcal{Y}_2, -\mathcal{X}_1\mathcal{Z}_2\rangle$		$\langle \mathcal{X}_1\mathcal{X}_2, -\mathcal{Z}_1\mathcal{Z}_2\rangle$		$\langle \mathcal{X}_1\mathcal{Y}_2, -\mathcal{Z}_1\mathcal{Z}_2\rangle$	
$\langle \mathcal{Y}_1\mathcal{Z}_2, -\mathcal{Z}_1\mathcal{Y}_2\rangle$		$\langle \mathcal{Y}_1\mathcal{X}_2, -\mathcal{Z}_1\mathcal{Y}_2\rangle$		$\langle \mathcal{Y}_1\mathcal{Z}_2, -\mathcal{Z}_1\mathcal{X}_2\rangle$	
$\langle \mathcal{Y}_1\mathcal{Y}_2, -\mathcal{Z}_1\mathcal{X}_2\rangle$		$\langle \mathcal{Y}_1\mathcal{X}_2, -\mathcal{Z}_1\mathcal{Z}_2\rangle$		$\langle \mathcal{Y}_1\mathcal{Y}_2, -\mathcal{Z}_1\mathcal{Z}_2\rangle$	

Appendix B

Code Listings

All code was written in a Jupyter 4.1.0 iPython notebook, with the environment:

```
Python 3.5.1 |Anaconda 2.5.0 (64-bit)| (default, Dec 7 2015, 11:16:01)
[GCC 4.4.7 20120313 (Red Hat 4.4.7-1)]
```

And with SciPy 0.17.0 and NumPy 1.10.4.

The code used to generate both Fig. 3.2 and Fig. 3.4 was:

```python
import numpy as np
import scipy.stats

def newMeasurement():
        return np.random.random_integers(low = 1, high = 4)

def updatePDF(pdf, new , n):
        pdf *= n
        pdf[new-1] += 1.0
        pdf /= (n + 1.0)
        return pdf

def H_D1(pdf):
        return np.log2(4) - scipy.stats.entropy(pk = pdf, base = 2)

def H_D1_bound(n):
        return np.log2((3. + n)/n)
H_D1_bound = np.vectorize(H_D1_bound)

def genChain(length):
        pdf = np.array([1., 0., 0., 0.])
        np.random.shuffle(pdf)
    nrange = np.arange(1, length + 1)
        d1s = [H_D1(updatePDF(pdf, newMeasurement(), n)) for n in nrange]
        return np.array(d1s)

AVG_SIZE = 5000
F_SIZE = 50
chains = [genChain(F_SIZE) for i in range(AVG_SIZE)]
x = np.arange(1, F_SIZE + 1)
d1b_avg = np.mean(chains, axis = 0)
d1b_var = np.var(chains, axis = 0)
```

```
N_SAMPLES = 15
SAMPLE_RANGE = 100
samples = np.array([genChain(SAMPLE_RANGE) for i in range(N_SAMPLES)])
samplesx = np.arange(1, SAMPLE_RANGE + 1)
```

Fig. 3.2 plotted the rows of `samples` against `samplesx`, and Fig. 3.4 plotted `d1b_avg` against x, with the variance lines being given by `d1b_avg` ± `2*d1b_var` and the upper bound given by `H_D1_bound(x)`, all also against x.

Appendix C

Derivation of the Upper Bound on Expected Divergence

This derivation almost entirely follows Ref. [17].

Firstly we consider the expectation value of the Shannon entropy, here denoted as H:

$$\begin{aligned}
\langle H(P_N\rangle &= \left\langle -\sum_i P_{N,i}\log_2 P_{N,i}\right\rangle \\
&= -\sum_i \langle P_{N,i}\log_2 P_{N,i}\rangle \qquad \text{(C.1)}
\end{aligned}$$

Given we are sampling from N bins, see that $P_{N,i} = \frac{x_i}{N}$ where x_i is the number of samples returning bin i.

$$\begin{aligned}
\langle P_{N,i}\log_2 P_{N,i}\rangle &= \left\langle \frac{x_i}{N}\log_2\frac{x_i}{N}\right\rangle \\
&= \sum_{k=0}^{N}\Pr(x_i = k)\frac{k}{N}\log_2\frac{k}{N} \\
&= \sum_{k=0}^{k}\binom{N}{k}p_i^k(1-p_i)^{N-k}\frac{k}{N}\log_2\frac{k}{N} \\
&= \frac{1}{n}\sum_{k=0}^{N}\frac{N!}{(N-k)!k!}p_i^k(1-p_i)^{N-k}\frac{k}{N}\log_2\frac{k}{N} \\
&= \frac{1}{n}\sum_{k=1}^{N}\frac{N!}{(N-k)!k!}p_i^k(1-p_i)^{N-k}\frac{k}{N}\log_2\frac{k}{N} \\
&= p_i\sum_{k=1}^{N}\frac{(N-1)!}{(N-k)!(k-1)!}p_i^{k-1}(1-p_i)^{N-k}\log_2\frac{k}{N}
\end{aligned}$$

Where the p_i are the probabilities corresponding to the expected distribution of P_N, which is P. Now let $j = k - 1$ and $m = N - 1$:

$$\begin{aligned}
&= p_i \sum_{j=0}^{m} \frac{m!}{(m-j)!j!} p_i^j (1-p_i)^{m-j} \log_2 \frac{j+1}{m+1} \\
&= p_i \sum_{j=0}^{m} \Pr(x_i = j) \log_2 \frac{j+1}{m+1}
\end{aligned}$$

Now recall Jensen's inequality, which states that for a convex function f : $f(\langle X \rangle) \leq \langle f(X) \rangle$.

$$\begin{aligned}
&\leq p_i \log_2 \left[\sum_{j=0}^{m} \Pr(x_i = j) \frac{j+1}{m+1} \right] \\
&= p_i \log_2 \frac{m \cdot p_i + 1}{m+1} \\
&= p_i \log_2 \left(p_i + \frac{1-p_i}{N} \right)
\end{aligned}$$

So we can plug this back into C.1 to get:

$$\langle H(P_N) \rangle \geq - \sum_i p_i \log_2 \left(p_i + \frac{1-p_i}{N} \right)$$

Next, we move to calculating the expectation value of D_{KL}. Recall the equality:

$$\begin{aligned}
D_{KL}(P_N \quad &\| \quad P) = H(P_N, P) - H(P_N) \\
\langle D_{KL}(P_N \| P) \rangle &= \langle H(P_N, P) \rangle - \langle H(P_N) \rangle \\
&= H(P) - \langle H(P_N) \rangle \\
&= H(P) + \sum_i p_i \log_2 \left(p_i + \frac{1-p_i}{N} \right)
\end{aligned}$$

Now, recall that the sum over i is the sum over the number of ontic states, and for the case we are considering this is 4. Furthermore, we are considering the case when the canonical distribution P is uniform, and so $H(P) = \log_2 4$ and $p_i = \frac{1}{4}$.

$$\begin{aligned}
\langle D_{KL}(P_N \| P) \rangle &\leq \log_2 4 + \sum_{i=1}^{4} \frac{1}{4} \log_2 \left(\frac{1}{4} + \frac{1 - \frac{1}{4}}{N} \right) \\
&= \log_2 \left(\frac{4 + N - 1}{N} \right) \\
&= \log_2 \left(\frac{3 + N}{N} \right) \qquad \square
\end{aligned}$$

Appendix D

Exorcising Maxwell's Demon (Again)

Here[1] I will argue a resolution of the apparent second law violation that the Szilard engine presents, that differs slightly from the "textbook" Bennett resolution and is informed by the representation of Result 3.2.1.

One might expect that the repeated exorcisms of Maxwell's demon would have at some point over the past 150 years finally banished him, but he persists to this day. I will not even attempt to review the literature on this but will instead concentrate on the standard resolution due to Bennett [31].

The standard presentation of this is to say that the demon starts in some "ready" state - where he knows nothing about the state of the system - and then, having been put into the "left" or "right" state by his measurement, must be returned to the ready state in order to restart the cycle and make his next measurement. Landauer's principle is invoked, and we say that the apparent destruction of information performed by this resetting of the demon has a work cost of at least $k_B T \ln 2$. I do not claim that this resolution is wrong per se, but I do claim that its apparent weaknesses may be more effectively argued against by presenting the problem slightly differently.

Ford [25] asks why the demon must start out in the "ready" state, rather than "left" or "right", and the model of the Szilard engine in the main text immediately gives an answer to this question: there is indeed no requirement that the demon starts in the "ready" state. Instead, I argue, the requirement is that the demon starts in a *known* state. That is to say it doesn't matter whether he starts in the "left" or "right" state so long as we know which of those it is. The erasure step is therefore necessary to have certainty about the state of the demon.

Consider the case that we do not perform the erasure and so we do not know whether we start in "left" or "right". Using our circuit representation where the measurement is represented by a CNOT, 50% of the time we will start in the "left" state and so the cycle will work just fine. However, in the other 50% of cases where we start in the "right" state, the CNOT will mean that the demon measures the particle to be on the opposite side to where it is, and so will extract no work. This 50/50 probability of work extraction is of course the same as guessing, and so there is no second law violation as the engine doesn't work!

[1] I could not justify shoehorning this slightly self-indulgent section into the main narrative of this report, but did think it was worth including somewhere.

The obvious argument against this resolution is that the CNOT is an overly simplistic representation of the measurement process - why should it matter what state we start in? Why not just measure where the particle is and then set the demon to the result of the measurement, regardless of his initial state? Of course, there is no reason that we could not have such a measurement, but (and now we once again return to the helpful language offered by quantum formalism, but still really are dealing with classical states) now *this* would be an operation that takes a mixed state to a pure state and so - and we are invoking Landauer here - arguably has a work cost of $k_B T \ln 2$. Indeed, this is in my opinion quite well justified by both del Rio et al [27] and Faist et al [23].

I have offered absolutely nothing to support the Landauer principle here, so can't claim to have fixed this interpretational issue. However, I do think that this setting, as well as quite concretely showing that no "ready" state of the demon is required, makes it clear that one must either choose a model of the Szilard engine where there is an erasure (with a work cost) or where there is a measurement with a work cost: not because these are 2 different methods of exorcism, but because the engine demands that if we do not have one, we must have the other. Whichever choice is made, the model admits no entropy decrease at any time: uncertainty is "moved" from the system to the demon, and then out to the heat bath by whichever work-expending step was chosen.

References

[1] Robert W. Spekkens. In defense of the epistemic view of quantum states: a toy theory. *Physical Review A*, 75(3), March 2007. arXiv: quant-ph/0401052.

[2] Sadi Carnot. *Reflections on the Motive Power of Fire: And Other Papers on the Second Law of Thermodynamics.* Courier Corporation, May 2012.

[3] Robert W. Spekkens. Quasi-quantization: classical statistical theories with an epistemic restriction. *arXiv:1409.5041 [quant-ph]*, September 2014. arXiv: 1409.5041.

[4] Matthew F. Pusey. Stabilizer notation for Spekkens' toy theory. *Foundations of Physics*, 42(5):688–708, May 2012. arXiv: 1103.5037.

[5] Daniel Gottesman. Class of quantum error-correcting codes saturating the quantum Hamming bound. *Phys. Rev. A*, 54(3):1862–1868, September 1996.

[6] Scott Aaronson and Daniel Gottesman. Improved Simulation of Stabilizer Circuits. *arXiv:quant-ph/0406196*, June 2004. arXiv: quant-ph/0406196.

[7] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, Cambridge ; New York, 10th anniversary edition edition, December 2010.

[8] Matthew F. Pusey. *A few connections between Quantum Computation and Quantum Non-Locality.* MRes, Imperial College London, December 2010.

[9] Günter M. Ziegler. *Lectures on Polytopes.* Springer Science & Business Media, 1995.

[10] D. Gross. Hudson's theorem for finite-dimensional quantum systems. *Journal of Mathematical Physics*, 47(12):122107, December 2006.

[11] Rémy Mosseri and Rossen Dandoloff. Geometry of entangled states, Bloch spheres and Hopf fibrations. *J. Phys. A: Math. Gen.*, 34(47):10243, 2001.

[12] Wassily Hoeffding. Probability Inequalities for Sums of Bounded Random Variables. *Journal of the American Statistical Association*, 58(301):13–30, March 1963.

[13] R. Renner and S. Wolf. Smooth Renyi entropy and applications. In *International Symposium on Information Theory, 2004. ISIT 2004. Proceedings*, pages 233–, June 2004.

[14] András Antos and Ioannis Kontoyiannis. Convergence properties of functional estimates for discrete distributions. *Random Struct. Alg.*, 19(3-4):163–193, October 2001.

[15] Noah Linden, Sandu Popescu, Anthony J. Short, and Andreas Winter. Quantum mechanical evolution towards thermal equilibrium. *Phys. Rev. E*, 79(6):061103, June 2009.

[16] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, November 2012.

[17] Brandon C Roy. Bounds on the expected entropy and KL-divergence of sampled multinomial distributions (Draft). June 2011. Available at: http://alumni.media.mit.edu/~bcroy/papers/bcroy_expected-entropy.pdf.

[18] L. Szilard. über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen. *Z. Physik*, 53(11-12):840–856, November 1929. Translation available online at: http://www.weizmann.ac.il/complex/tlusty/courses/InfoInBio/Papers/Szilard1929.pdf.

[19] Gloria Koenig. *Charles & Ray Eames, 1907-1978, 1912-1988: Pioneers of Mid-Century Modernism*. TASCHEN Gmbh, Köln ; Los Angeles, 1st edition edition, November 2005.

[20] P. Shor. Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. *SIAM Rev.*, 41(2):303–332, January 1999.

[21] Carlo Cafaro and Peter van Loock. Approximate quantum error correction for generalized amplitude-damping errors. *Phys. Rev. A*, 89(2):022316, February 2014.

[22] Ian Ford. *Statistical Physics: An Entropic Approach*. John Wiley & Sons, March 2013.

[23] Philippe Faist, Frédéric Dupuis, Jonathan Oppenheim, and Renato Renner. The Minimal Work Cost of Information Processing. *Nature Communications*, 6:7669, July 2015. arXiv: 1211.1037.

[24] R. Landauer. Irreversibility and Heat Generation in the Computing Process. *IBM Journal of Research and Development*, 5(3):183–191, July 1961.

[25] Ian J. Ford. Maxwell's demon and the management of ignorance in stochastic thermodynamics. *arXiv:1510.03222 [cond-mat]*, October 2015. arXiv: 1510.03222.

[26] John D. Norton. Waiting for Landauer. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 42(3):184–198, August 2011.

[27] Lídia del Rio, Johan Åberg, Renato Renner, Oscar Dahlsten, and Vlatko Vedral. The thermodynamic meaning of negative entropy. *Nature*, 474(7349):61–63, June 2011.

[28] Mirjam Weilenmann, Lea Krämer, Philippe Faist, and Renato Renner. Axiomatic relation between thermodynamic and information-theoretic entropies. *arXiv:1501.06920 [cond-mat, physics:quant-ph]*, January 2015. arXiv: 1501.06920.

[29] Stefanie Hilt, Saroosh Shabbir, Janet Anders, and Eric Lutz. Landauer's principle in the quantum regime. *Phys. Rev. E*, 83(3):030102, March 2011.

[30] Antoine Bérut, Artak Arakelyan, Artyom Petrosyan, Sergio Ciliberto, Raoul Dillenschneider, and Eric Lutz. Experimental verification of Landauer/'s principle linking information and thermodynamics. *Nature*, 483(7388):187–189, March 2012.

[31] Charles H. Bennett. Notes on Landauer's principle, reversible computation, and Maxwell's Demon. *Stud. Hist. Phil. Mod. Phys.*, 34(3):501–510, September 2003.

[32] R. V. L. Hartley. Transmission of Information. *Bell System Technical Journal*, 7(3):535–563, July 1928.