Landauer Defended: Reply to Norton

James Ladyman, Katie Robertson

Department of Philosophy, University of Bristol

March 7, 2013

Abstract

Ladyman et al (2007) proposed a model of the implementation of logical operations by physical processes in order to clarify the exact statement of Landauer's Principle, and then offered a new proof of the latter based on the construction of a thermodynamic cycle, arguing that if Landauer's Principle were false it would be possible to harness a machine that violated it to produce a violation of the second law of thermodynamics. In a recent paper in this journal, John Norton (2011) directly challenges the consistency of that proof. In the present paper we defend the proof given by Ladyman et al against his critique. In particular, contrary to what Norton claims, we argue that the processes used in the proof cannot be used to construct a cycle that enacts erasure in a thermodynamically reversible way, and that he does not show that the processes used in the proof violate the Second Law of Thermodynamics.

1   Introduction

There is a consensus among physicists that there is a connection between information processing and thermodynamics (see, for example, Blundell and

Blundell (2010), and this is foundational to work in work in other areas of physics, including, for example, cosmology (Lloyd 2002). In particular, it is standardly thought that logically irreversible computations can only be implemented by thermodynamically irreversible processes. This is known as Landauer's Principle, hereafter LP, the quantitative form of which says that there is an entropy increase of $k \ln 2$ associated with the resetting of a single bit of data in a computational device. John Norton (sometimes together with John Earman) has repeatedly criticised this consensus and argued that the purported proofs of both the qualitative and quantitative forms of LP are not sound (see Earman and Norton 1998, 1999, and Norton 2005). Norton's critique led Charles Bennett (2003) to concede that LP cannot be used to show that there cannot be a Maxwell Demon. It also exposed deficiencies in the extant discussions of LP in the wider literature, demonstrated that LP could not be established by reasoning about particular cases, and showed that a general proof of LP had not been given. Ladyman et al (2007), hereafter LPSG, proposed a model of the implementation of logical operations by physical processes in order to clarify the exact statement of LP, and then offered a new proof of the latter based on the construction of a thermodynamic cycle, arguing that if LP were false it would be possible to harness a machine that violated it to produce a violation of the second law of thermodynamics. In a recent paper in this journal (2011), John Norton directly challenges the consistency of that proof. He also gives a number of arguments that he takes to undermine the foundations of the thermodynamics of computation (against which he also argues in his (forthcoming)). In the present paper we reply to Norton and defend the proof given by LPSG against his critique.

2 The Connection between Logical and Thermodynamic Irreversibility

LPSG proposed an analysis of the implementation of computation by physical systems in order to clarify exactly what LP says and to provide a basis for its evaluation and proof. They argue that much of the confusion in the literature results from a failure to clearly distinguish between logical transformations and the physical processes that implement them, for example, by loosely talking about logically irreversible processes. To be clear, we emphasise that a logical transformation is a mathematical entity not a physical one. Physical processes are the direct subject of thermodynamics which applies to computations only derivatively via applying to the processes used to implement them. Once this distinction is made it is obvious that it makes no sense to talk of the implementation of a logical transformation by a physical process, rather, in so far as logical transformations are implemented using physical systems, they are implemented by families of physical processes. For example, it is not correct to talk of '*the* erasure process' because, depending upon what the input state is, a different member of the family of physical processes will occur. In general, if a logical transformation is a single-valued map L from a finite set X of input states, into a finite set Y of output states, then it will be implemented by a family of physical processes equinumerous with the number of logical input states.[1] (For example, in the case of RESET there are two input states usually labelled 0 and 1.) We say that a logical transformation, L, is *logically reversible* if and only if $L : X \to Y$ is a one-to-one (injective) mapping. Hence, with a reversible logical transformation, we can uniquely reconstruct the input state from the output state. If L is not a one-to-one mapping, we say that it is *logically irreversible*. For the physical system to implement the logical transformation reliably, the family

---

[1]Like LPSG, we are only concerned with deterministic computation and not with probabilistic operations such as Maroney's RAND (2005).

$$x \quad \xrightarrow{\;L\;} \quad y$$
$$\| \qquad\qquad \|$$
$$D_{in}(x) \quad \xrightarrow[\Lambda_L]{} \quad D_{out}(y)$$

Figure 1: An illustration of the relationship between the logical states x and y and their representative physical states $D_{in}(x)$ and $D_{out}(y)$, showing the logical transformation L and the physical time evolution operator Λ_L. (reproduced from Ladyman et al 2007)

of processes must take each of the physical states that represent the logical input states to the appropriate physical state, that is the one that represents the right logical output state.

This is all summarised in the definition of L-machine given by LPSG. The key features of their analysis are briefly reviewed below.

To physically implement a logical transformation, we require: A physical device, D, a specification of which physical states of that device correspond to the possible logical states (we call the former *representative states*), and a time evolution operator of that device. We refer to this combined system as an *L-machine.*

The time evolution operator, Λ_L must generate the relevant family of processes, and the reliability of the implementation consists in the time evolution operator being such as to ensure that *whichever* of the representative physical states the device is prepared in, it ends up in the appropriate representative state.

LPSG emphasised that everything about the behaviour of the device must be incorporated into the time evolution operator and external agents may not intervene during its operation. In particular this prohibits any such external agent affecting the time evolution of the system by making use of information about its state while it is running.

4

An L-machine $\{D, \{D_{in}(x)|x \in X\}, \{D_{out}(y)|y \in Y\}, \Lambda_L\}$ physically implements L in the following sense. If D is prepared in the input state $D_{in}(x)$ corresponding to the logical input state $x \in X$, and is then evolved using Λ_L, it will be left in the output state $D_{out}(y)$ corresponding to the logical output state $y = L(x) \in Y$. We will denote this physical process by p_x.

A process p is *thermodynamically reversible* if and only if $\Delta S_{tot}(p) = 0$, where S_tot is the total entropy of the whole system.

If $\Delta S_{tot}(p) > 0$, the physical process p cannot be run in reverse, as the reverse process p' would have $\Delta S_{tot}(p') < 0$, and hence violate the second law. We therefore refer to any process p for which $\Delta S_{tot}(p) > 0$ as *thermodynamically irreversible*. As is well known, there are a number of formulations of the second law that are provably equivalent to this, modulo certain assumptions. The other one to which we will refer is the Kelvin statement of the second law according to which there is no cyclic process whose sole effect is the conversion of heat into work (see Uffink 2001, p. 328).

A family of physical processes is thermodynamically irreversible if and only if at least one of its members is.

LPSG's proof uses these definitions to establish the connection between logical irreversibility and thermodynamic irreversibility as follows. LP can be stated precisely as the claim that if L is logically irreversible then any L-machine will use a thermodynamically irreversible family of physical processes to implement L.

3 Norton's Critique

In Norton (2011), John Norton argues that LPSG's proof is not sound for two reasons. The first is that, according to him it is possible to use processes from the above thermodynamic cycle to violate the Second Law of thermodynamics

(hereafter SL). Hence, in effect, where LPSG argue that LP must be true by reductio, because assuming its negation plus the admissibility of a certain set of processes entails the violation of SL, Norton argues that inferring the truth of LP is fallacious because the processes entail the violation of the second law on their own. Hence, his first charge is that LPSG are inconsistent in affirming both that their processes are admissible and SL. He argues that the processes in question violate SL in two principle ways. Firstly, he claims that a different cycle can be constructed with them that violates SL. Secondly, he claims that one process in particular violates the second law. In sum, then where LPSG claim that $p_1 \wedge ...p_4 \wedge \neg LP \vdash \neg SL$ (where $p_1..p_4$ are premises asserting that the respective processes $P_1..P_4$ are admissible) and so infer that LP is true, Norton argues that $\{P\} \vdash \neg SL$ making LPSG's inference to LP gratuitous.

Norton's second objection to the soundness of LPSG's proof is that according to him, their cycle can be adapted to provide an example of 'dissipationless erasure' which is his term for an implementation of RESET that is thermo-dynamically reversible in direction contradiction of LP. In what follows we first explain the cycle the Norton constructs as the basis of his discussion and how it relates to the cycle used by LPSG. We then list the processes used, before arguing that Norton uses a particular kind of 'controlled operation' that differs from any of those used by LPSG, and which is inadmissible. Hence, we argue that Norton's counterexample fails, and that if it is reconstructed using admissible processes used by LPSG then it reduces to reset of known data which is logically reversible and so not a counterexample to LP. We also argue that the processes that LPSG regard as admissible do not violate the second law, and that Norton's cycle that violates SL cannot be constructed from admissible processes.

3.1 Norton's cycle and the LPSG cycle compared

Norton constructs a modified version of the cycle that LPSG use in their proof which he claims is a particular case of their cycle, where the device is also a one-molecule gas in a box. As a consequence of this, the two states of the device in step 2 (M_L and M_R), have the same entropy. From this characterisation he then constructs a list of processes with the aim of:

1) Showing these processes are inconsistent with asserting the 2nd law.

2) Using these processes to construct 2 counterexamples; 'dissipationless erasure' and reset by the repeated removal of the partition.

Initial conditions: A one-molecule gas occupies the whole box, B. The device is initially in M_L.

Step 1 (same as LPSG): A partition is inserted into the box.

Step 2: A controlled operation is performed on the device depending on the position of the molecule in B, a) If the molecule is found on the LHS, the device is set to the LHS (this means doing nothing as an action is only triggered if the box molecule is found on the right).

b) If the molecule in the box is found on the RHS, the device is set to right hand." The shift is performed by a reversible thermodynamic process. Since the thermodynamic entropies of M_L and M_R are the same, no heat passes to or from the surroundings." (Norton 2011, p. 188) (This step is uncontroversial and the same as in LPSG's cycle.)

Step 3 (same as LPSG): A controlled operation is performed on the box depending on the state of the device, a) If the device is in the state M_L, the piston inserted on the RHS.

b) If the device is in state M_R, the piston is inserted on the LHS.

Step 4: "The erasure process is performed. It transforms the memory device from the probabilistically mixed state of M_L or M_R with equal probability, to the initial state of M_L" (ibid)

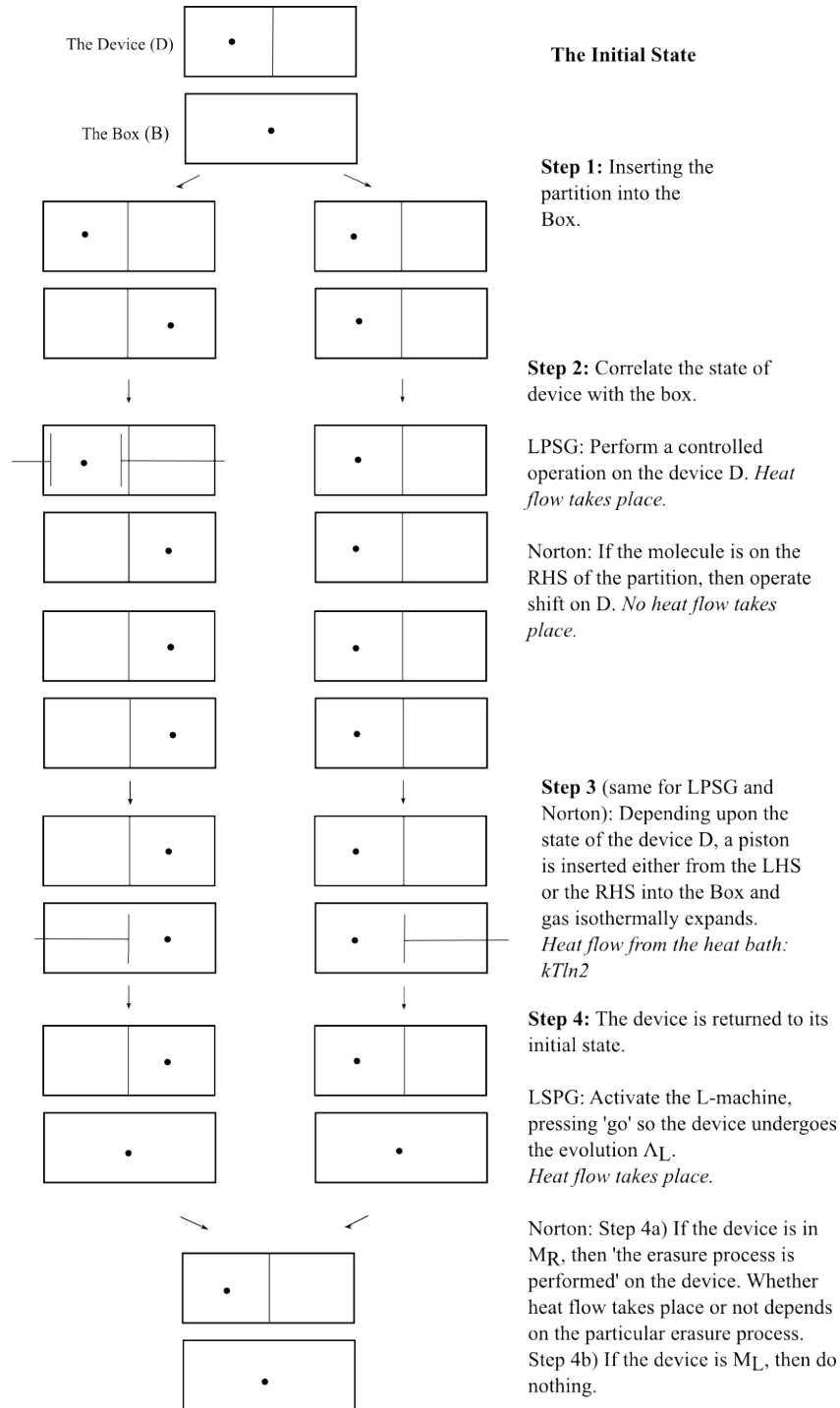Norton's cycle is intended to use exactly the same set of processes as those

The Device (D)

The Box (B)

The Initial State

Step 1: Inserting the
partition into the
Box.

Step 2: Correlate the state of
device with the box.

LPSG: Perform a controlled
operation on the device D. *Heat
flow takes place.*

Norton: If the molecule is on the
RHS of the partition, then operate
shift on D. *No heat flow takes
place.*

Step 3 (same for LPSG and
Norton): Depending upon the
state of the device D, a piston
is inserted either from the LHS
or the RHS into the Box and
gas isothermally expands.
*Heat flow from the heat bath:
kTln2*

Step 4: The device is returned to its
initial state.

LSPG: Activate the L-machine,
pressing 'go' so the device undergoes
the evolution Λ_L.
Heat flow takes place.

Norton: Step 4a) If the device is in
M_R, then 'the erasure process is
performed' on the device. Whether
heat flow takes place or not depends
on the particular erasure process.
Step 4b) If the device is M_L, then do
nothing.

Figure 2: *A diagram of Norton's cycle*

that LPSG use in their's. In the next section we list them and discuss how his usage of them differs from LPSG.

3.2   List of Admissible Processes

Process 1a: Isothermal expansion. This a reversible process that takes place at constant temperature; a piston is inserted and the gas does work on it, increasing the volume of the gas. The amount of work done by the gas is equal to the heat flow *from* the reservoir to the gas. In an isothermal expansion to twice its initial volume, the work done by the gas is $-kT \ln 2$.

Process 1b: Isothermal compression. A reversible process at a constant temperature. The piston reduces the volume of the gas. The work done on the gas by the piston is equal to the heat delivered *to* the heat bath. The work done by the piston is $-kT \ln 2$.

Process 2a: Removal of the partition. The partition (which traps the gas on one side of a box) is removed. (There is no heat flow.)

Process 2b: Insertion of the partition. The partition is inserted trapping the gas in a smaller volume.[2] We will assume the partition is inserted halfway along the box. (There is no heat flow.)

(These two processes are used by LPSG just as Norton says they are.)

Process 3a: Detection. The location of the molecule can be determined without the detection incurring any thermodynamic cost.

Process 3b: Detect and Trigger. "According to the whether the outcome of a detection is L or R, $process_L$ or $process_R$, respectively, may be initiated,

---

[2]Clearly, a distinction is needed here. Our gas consists of only one molecule so inserting a partition in the middle of the box will compress the gas to half its volume with certainty. However, in a many molecule gas the probability of the gas being compressed to half the volume is very small.

without the initiation passing heat to the heat bath, where these are any two admissible processes." (ibid, p. 189)

The key difference and infidelity to LSPG is Norton's use of process 3b. Process 3b is a controlled operation: if the control state is detected to be X, then perform process p_X on Z. If the control state is detected to be Y, then perform process p_Y on Z. This process is used in step 2 and 3 of both LPSG's cycle and Norton's particular case. However, in step 4 of Norton's cycle he performs a controlled operation on the *same* state.

Further, a controlled operation on the same state is a crucial step in his counterexample of 'dissipationless erasure.' This counterexample proceeds by using the next process.

Process 4: Shift. "If a system has states M_1 or M_2 of equal thermodynamic entropy, then a shift process moves the system from one state to the other without passing heat to the heat bath." (ibid)

In LPSG, the shift process is used in step 2. However, in step 2, which process happened to the device (representing the target bit) depended on the state of the gas (representing the control bit). On the other hand, Norton's cycle uses 'shift' in his fourth step, in such a way that whether the location of the molecule is shifted (the target bit) depends upon the location of the molecule (the control bit), because the state of the device is used to represent both bits. Hence, in Norton's cycle shift is used as a 'controlled operation' by part of the system on itself.

So Norton replaces the above step 4 with step 4*: "If the memory device is measured to be in state M_R, a shift process is initiated that moves the molecule from the state M_R to state M_L with equal thermodynamic entropy by a process that passes no heat to the heat bath." (ibid)

The second counterexample resets a memory device initially in either M_L or M_R to M_L. Norton says "the process detects whether the molecule is

trapped in the right side of the chamber. If it is, the partition is removed and replaced" (ibid, p. 190). This process is repeated until the molecule is detected on the left (the probability become overwhelmingly high (≈ 0.999) after 10 repetitions). However, this process might not be successful in resetting and as such it does not correspond to a logical transformation, rather it is a probabilistic operation taking M_R to M_L or M_R each with a probability of $\frac{1}{2}$. As such, this is not a counterexample against LPSG who state in their premises that they do not consider probabilistic transformations. This is also a 'controlled operation' on the same state: which action (remove partition or nothing) is performed on the device, depends on the state of the same device (molecule on LH or RH side).

Whether you can perform a 'controlled operation' on the same state is pivotal to the success or failure of Norton's reply. In the next section, we argue that this is not possible.

4   Responding to Norton's Critique

In the first subsection below we consider controlled operations and the particular operation that Norton uses in his critique, arguing that it is not among the operations that LPSG use, and that it is in any case inoperable and not rightly called a 'controlled operation' in the sense in which the term is usually meant hence the scare quotes above. In the second subsection we consider Norton's arguments that the processes LPSG use are inconsistent with the second law of thermodynamics and argue that each of them fails.

4.1   Controlled operations

Controlled operations are commonly discussed in the literature. A controlled operation is a logical transformation that maps an input state of at least 2

bits to an output state of at least two bits in such a way that how one of the bits, the target bit, is transformed depends on the value of the other bit, the control bit. The most commonly discussed example of such an operation is CNOT which has the following truth table, where bit_1 is the target bit and bit_2 is the control bit.

$Input_1$	$Input_2$	$Output_1$	$Output_2$
0	0	0	0
1	0	1	0
0	1	1	1
1	1	0	1

The heart of the Norton's objections lies in the fact that his cycle requires that a controlled operation can be implemented by a physical system that has only one degree of freedom which must therefore be used to represent both the target bit and the control bit. On the contrary, LPSG are explicit that "the same bit cannot be both the control and the target of a controlled operation" (Ladyman et al 2007, p. 23 note 7).

In particular, considering Norton's alleged counterexample of a cycle using LPSG's processes that violates SL, step 4* requires that a controlled operation be performed from a physical degree of freedom to itself. That is instead of:

$Target_{in}$	$Control_{in}$	$Target_{out}$	$Control_{out}$
Z	0	$p_0(Z)$	0
Z	1	$p_1(Z)$	1

we have:

$Target_{in}$	$Control_{in}$	$Target_{out}$	$Control_{out}$
0	0	$p_0(0)$	0
1	1	$p_1(1)$	1

In the first table which process p_x occurs depends only the value of the control bit. This table implicitly contains another two rows as the physical system Z

which represents the target bit can be in two different states. However, the second table cannot have an additional two rows, as the same physical state represents both the target and the control bit. Further, unless both p_1 and p_0 are 'do nothing' then the physical state that previously represented the control bit does not exist after the transformation.

However, it makes no sense to allow which bit a physical degree of freedom represents to change during the processes that implement a computation. Furthermore, for physical degrees of freedom to represent different bits they must be independent of each other. This is because, for a physical system to represent a set of n bits it must have sufficient, i.e. 2^n, different configurations of its degrees of freedom. For example, if the only two alternative states for a one molecule gas are B_L molecule trapped on the LHS, or B_R (trapped on RHS) then it does not make sense to say this represents 001010 as this string represents 6 bits of information and the system would need at least 2^6 different configurations. If this were not so the different bits could not vary independently of each and that they do so is what makes them different bits in the first place. For example, there is a difference between having four different logical states a, b, c, and d, and having four different configurations of two bits 11, 10, 01 and 00. While in each case there are four possible states, they are computationally different because they involve different alphabets. The former four states cannot be used as inputs to AND or $CNOT$ whereas the latter four can. It is crucial to the identity of logical transformations whether or not they map the former set of input states to output states or the latter. Accordingly, when we consider physical states that represent logical states in the context of implementation, it matters whether the former are composed of independent degrees of freedom because otherwise measuring one bit cannot be done without measuring all of them. Equivalently, it would not be possible to make a process that transforms a bit according to some rule the same regardless of the value of other bits. In the particular case of CNOT, it is required that the physical process that transforms the target bit

depend only on the value of the control bit and not on the value of the target bit itself (Z).

For example, the four different input rows of CNOT could be physically represented by a pin on four different places on a chess board (here we do not have independent degrees of freedom). Since there are two different positions corresponding to the single value of the control bit, which physical process acts on the system depends on the value of both bits, not just the control bit.

There is another reason why different bits cannot be represented by the same physical state at different times, namely that to allow them to do completely trivialises the physical implementation of logical transformations. For example, consider COPY. If relabelling is allowed no physical change in a system is required for it to implement COPY because we simply stipulate that whatever state the physical degree of freedom happens to be in now represents the state being copied.

Norton's alleged counterexample relies crucially on the use of a process that consists in a degree of freedom performing a 'controlled operation' on itself. Norton claims to show that allowing P_1 to P_4 plus P_N entails that LP is false. However, P_N is not among the processes that LPSG use. Their controlled operations are standard because their inputs consist of a control bit and a target bit represented by two distinct physical degrees of freedom, and their outputs also consist of two distinct physical degrees of freedom that also represent the control bit and the target bit where the former is in the same state as before. Norton's 'controlled operation' is not of this kind because the control bit is not represented at all at the end of the operation. Furthermore P_N entails that LP is false without the rest of the cycle. This is because an instance of P_N would be RESET in a thermodynamically reversible way as follows. Consider a memory device with a single degree of freedom. If it can perform a controlled operation on itself then let that operation be

14

to do nothing if the state is 0 and to perform SHIFT if it is in the state 1. Since, doing nothing and shift are both thermodynamically reversible then P_N entails not-LP and hence the rest of Norton's cycle using LPSG's processes is redundant. Hence, where Norton claims $p_1 \wedge ... \wedge p_4 \wedge p_N \vdash \neg SL$ actually $p_N \vdash \neg SL$. It is clearly question-begging in the dialectical context to assume the use of a process that on its own entails not-LP. In any case, such controlled operations are not controlled operations in the sense in which CNOT and COPY, since there is no degree of freedom representing the controlled bit at the end of the operation.

Norton claims that the idea of 'controlled operation' from a degree of freedom to itself is vindicated since a robot may be posited to enact it. However, contrary what Norton supposes, any such robot would have internal degrees of freedom that would store the state of the control bit (the state of the device), thereby effectively remembering the state of the target bit, and hence requiring resetting for a cycle to be completed. It is also arguable that such 'controlled operations' of a degree of freedom on itself are not viable because any operation would require an auxiliary system whose internal state would determine which operation was performed. For example, a piece of paper that says 'destroy me' cannot read and destroy itself but would have to be destroyed by a system that read it, that is copied it, first.

In order to reconstruct P_N from processes that are used by LSPG, a copy of the state of D (which is acting as the control and target bit) must be taken. This ensures that there is a distinct physical state representing the control bit that continues to be represented after the physical process has acted upon the target bit. At the end of the cycle this state must then be reset. However, if LP were false, this reset would be unproblematic as the 'dissipationless erasure' procedure could then be implemented to perform it.[3] However, this amounts to the reset of *known* data, and LSPG made it clear (p. 22) that it is agreed on all sides that reset of known data can be done without increasing

---

[3]This does then just amount to asserting LP is false.

entropy because it is logically reversible. As Feynman (1996, p. 144) says "If we know the atoms position, then we expend no energy in resetting, irrespective of where the atom starts out." Thus, he says it is only *unknown* data to which has a cost associated with resetting: "Only if we do not know which side of the compartment the atom is in do we expend free energy." (Ibid.) As shown in LPSG (p. 23), the reset of known data is actually the logical transformation 'UNCOPY' which is logically reversible. As shown by Bennett (1973), logically reversible transformations can be physically realised in a thermodynamically reversible way and are irrelevant to LP which is only concerned with logically irreversible transformations.

Elsewhere Norton (2012) uses a distinction between known and 'random data' that is commonly found in the literature (Leff and Rex (200. For 'random data' the molecule may be found on the LHS or the RHS, where as an example 'known' data would be a string of devices with the molecule on the LHS (reset to zero). This is a bizarre definition. Why would whether the state of one device is known or not depend on the state of other devices?[4] In contrast, LSPG make the distinction between known and unknown data in the following way for a device, D, that contains both a register R and a memory M (which can be read). If the physical states of R and M represent the same logical state, the data contained in the device is *known*. If the states of R and M are not correlated, then the data is *unknown*.

We can operationally define whether a physical state represents known or unknown data. For physical states representing known data, process 1a can happen as the physical state representing the memory bit acts as the control bit to determine which side the piston should enter from. For unknown data, before process 1a can occur the location of the molecule (whether it is on the LHS or RHS) must be detected and stored in the memory (which then acts

---

[4]Note it is misleading to talk of 'thermalised data' being represented by the molecule occupying the whole chamber as such a state does not really represent data at all rather it is more like an empty register.

as the control bit).

By taking a copy of the state in order to perform step 4, there is a memory (the copy) which is correlated with the register to be acted upon (the detected state). Thus, so reconstructed, Norton's erasure process acts on *known* data: a process established to be thermodynamically reversible (as it is logically reversible).

In the light of the above, we conclude that Norton's alleged counterexample fails to exhibit a violation of LP. If his cycle does violate LP it is simply because he assumes the admissibility of a process (P_N), that LPSG do not use, and which is sufficient on its own to violate LP. If we reconstruct Norton's process using only LPSG's processes then it is not RESET of unknown data and hence not a counterexample to LP.

4.2 Does the 'standard inventory' of processes allow for violation the second law?

As mentioned above, even if this alleged counterexample fails, Norton has a separate argument against LPSG, namely that the processes that they use, which as he notes, are standardly employed in the literature (Bennett 2003), allow for violation of SL. If this is true then it is no surprise that these processes conjoined with the assumption of the falsity of LP entail violation of SL, and of course one could not conclude from an argument to that effect that LP must be true after all. In particular, LPSG argue that there is a missing entropy that must come from the RESET of the device in their cycle, but if the processes they use violate SL then that missing entropy is accounted for by them alone and is an artefact resulting from accounting for the difference between the entropy reducing processes and the assertion of the SL. Norton therefore charges LSPG with inconsistency since he alleges they assume SL, while also using a set of processes that allow for violations

of it. It is important to note that this charge of inconsistency has a wider scope; if these processes (which are widely used in the literature do indeed allow for violation of SL then they cannot be used for any proof of LP that follows the sound horn of Earman's and Norton's dilemma.

4.2.1   A second law violating cycle?

Norton constructs a cycle as follows:

Initial state: The molecule occupies the whole box.

1) The partition is inserted, trapping the one-molecule gas on one side of the partition. *Process 2b*

2) The piston is inserted, and the gas does work in an isothermal expansion. *Process 1a*

The gas has returned to its initial state so a cycle has been constructed whose sole outcome is the conversion of heat to work: a violation of SL.

However, the above cycle is inoperable. In order to know on which side to insert the piston, one needs to know which side the molecule is on. But Norton faces a dilemma. Either there is an agent outside the system that determines on which side to insert the piston, or not. If the former, then this cycle is irrelevant since everything that determines the time evolution of the device must be contained within it as argued at length by LPSG and reiterated above. On the other hand, if there is no external agent then the cycle uses a 'controlled operation' of a degree of freedom on itself which we argued above is not possible, and in any case, the assumption that it is possible entails that LP is false and is therefore question-begging in this context.

Again, in order to make this cycle operable using admissible processes, there must be a distinct physical state which controls from which side the piston enters. This cannot be just some arbitrary state; it needs to be different

depending on which side of the partition the molecule is, so that the piston entering will varying accordingly. In order for the process to be cyclic, this 'control state' must be reset to its initial state and by LP, there is a thermodynamic cost associated with this. Hence, this revised cycle doesn't violate SL.

4.2.2 The partition problem: process 2b

Even if all his other arguments fail, Norton also objects to LPSG's proof on the grounds that one of the processes they admit, namely 2b above, itself violates SL. If the processes involved are thermodynamically reversible, the entropy difference between two states A and B, $S(B) - S(A)$, should be independent of the path between them. If state A is the gas occupying the entire box, and state B the gas occupying half the box then we expect the entropy difference to be the same regardless of the reversible process taken to reach one from the other. As such, prima facie, we would expect the entropy difference between A and B to be the same regardless whether the volume was halved via a isothermal compression or by inserting the partition. However, in step 1 of the cyclic process, inserting the partition is assumed not to change the entropy of the gas.[5] In contrast, isothermally compressing the gas reduces its entropy by $kT \ln 2$ (along with a corresponding increase of the entropy of the heat bath).

So does inserting the partition decrease the entropy of the gas? As SL is statistical, it holds that *on average* the entropy does not decrease. For a many-molecule gas composed of n molecules, it is very unlikely ($Pr = 1/2^n$) that all the molecules will be on one side when you insert the partition. Therefore, in this case inserting the partition is not a violation of SL. However, for a one-molecule gas the volume it can occupy will be halved with

---

[5]The temperature of the gas doesn't change either, as whilst the volume has halved the pressure has doubled: $pV = nRT$.

certainty.

Premise 1: Inserting the partition is a reversible process that halved the volume of the one molecule gas with certainty.

Premise 2: The entropy difference between two states A and B is the same, regardless of path taken (provided thermodynamically reversible processes are used).

Premise 3: Isothermally compressing a gas *reduces* the entropy of the gas by $kT \ln 2$ and *increases* the entropy of the heat bath by $kT \ln 2$.

Premise 4: The state $B_{compression}$ (the gas isothermally compressed to half the box) *is the same* as state $B_{partition}$ (the gas reduced to half its original volume by inserting the partition).

Conclusion: Inserting the partition reduces the entropy of the gas by $kT \ln 2$ with certainty: a violation of SL.

However, LPSG can reject premise 4. There is an important distinction between $B_{partition}$ and $B_{compression}$; in the former case you do not know which side of the box the particle is confined to. By considering the statistical mechanical definition of entropy: $S = k \sum_i p_i \ln p_i$, where p_i is the probability for the system to be state i.

$$S = k \sum_i p_i \ln p_i \tag{1}$$

The entropy associated with the box in state A can be written:[6]

$$-S_{before}/k = p_L \ln p_L + p_R \ln p_R \tag{2}$$

When you compress the gas isothermally with the piston entering from the right, $p_L \to 1$ and $p_R \to 0$. Thus, (as $\ln 1 = 0$), $S(B_{compression}) \to 0$. As

---

[6]There is no fundamental reason to split the box in 2, rather 3 or any other number, but 2 allows for an easy comparison with state B.

expected, the entropy of the gas has decreased (and that of the heat bath has increased).

However, the case of inserting the partition is different, since one does not know which side the molecule is on. Because of this, $S(B_{partition}) = k(p_L \ln p_L + p_R \ln p_R)$. Hence, contrary to what Norton's says, inserting the partition does not change the entropy of the gas. Hence, premise 4 is false and the partition does not violate SL.

It may be objected that the only difference is that in the final state $B_{compression}$ one does know which side the gas is trapped on, whereas in the partition method the final state $B_{partition}$ is equally likely to be on the right or left, and that this difference is solely in our epistemic position, rather than a difference in the state properties of the gas. Further, it seems that any difference in the information content of the system (if information is a some way a property of the system) only applies *on average*. For the 'average' state $B_{partition}$ there is a difference in the information (compared to the state $B_{compression}$). But is hard to see what 'information' about the state $B_{partition}$ on the LHS is different from the 'information content' of the state $B_{compression}$ just because the former *could* have been on the RHS. Unlike a quantum superposition, there *is* a matter of fact whether the gas is on the LHS or RHS.

However, by returning to some familiar phenomenological concepts such as work and heat the difference between state $B_{compression}$ and state $B_{partition}$ can be elucidated. The key difference is how you can extract work from the two states. In order to do an isothermal expansion, you need to know which side of the box the molecule is on to know whether to insert the piston on the left or right hand side. Given 100 systems in the state $B_{compression}$, you can extract work from all of them. However, given 100 systems in the state $B_{partition}$ half of the time you will have inserted the piston on the wrong side, doing work (as you are compressing the gas) rather than extracting work. This is why the fact that $B_{partition}$ *could* have been on the other side is relevant: there

is no one procedure to extract work from $B_{partition}$ with certainty.[7] Norton questions why we should think that whether we *know* about which state a device is in affect matters. But it does matter: thermodynamics is about the properties of matter and how we can exploit or use these properties to do work. Knowing what state a device is in changes which operations you are able to perform.

Norton does not acknowledge this difference and so allows (via process 3b) that the position of the particle in $B_{partition}$ can be detected and the piston inserted accordingly. Thus, on Norton's view, work can be extracted from the state $B_{partition}$. However, this view is mistaken as it would involve performing 'controlled operations' on the same state. As shown earlier, this is not possible and therefore the insertion of the partition does not violate the Kelvin version of the second law, in addition to not violating the information-theoretic version (because the information-theoretic entropy does not decrease).

There is a further way to consider whether the insertion of the partition decreases the entropy and therefore violates SL, namely by reference to Clausius' equation. The Clausius equation directly relates entropy to heat flow, as discussed in Norton section 2.2, and we could use it to argue that there is no entropy change through the insertion of the partition because, unlike an isothermal compression, there is no associated heat flow. Hence, if we assume that the entropy of a system only changes if there is a heat flow, then the insertion of the partition does not change the entropy. However, he argues that the Clausius formula is not valid in this context because the entropy change between the initial and final states in question is not path-independent, since if the volume of the gas is halved by inserting the partition then a different entropy change (0) is associated with the state than if the path is isothermal compression ($kln2$). However, if we regard the final state as only being the

[7]The above example also helps to further illuminate the statistical nature of SL: it could happen that the all 100 are on the LHS, and you extract work from all of them; you were just lucky. However *on average*, you cannot extract work.

same as the state reached by isothermal compression if it is known where the particle is (in contrast with what is assumed in premise 4 above), then the entropy is path-independent.

We now compare the two paths quantitatively using the method defended by Ladyman et al 2008, hereafter LPS, who show how to calculate the entropy of a statistical mixture of definite thermodynamic states. They argue that if a system is in one of two states with equal probability then the entropy that should be assigned to the system is half the sum of the entropies associated with those two definite states, plus a term of $kln2$ associated with the uncertainty about in which state it is.

Initially the gas in the box has entropy X. The path taken by isothermal compression is as follows: the piston is inserted then compression, entropy of the gas reduced by kln2 (corresponding increase in the heat bath). The gas is known to be on the RHS (for example) and can be used to do work on the piston (by isothermal expansion).

The other path is as follows. The partition is inserted and the molecule is either on the LHS or RHS. Using the formula above for the entropy of a statistically mixed state the entropy of the gas is $\frac{1}{2}(S_L + S_r) + kln2$, where S_L and S_R are the well-defined thermodynamic entropies associated with a gas trapped on one side (via isothermal compression). S_L and S_R are $X - kln2$. After detecting which side of the partition of molecule is on, we can then extract work from the system. Furthermore, according to the above formula the entropy of the gas has reduced by $kln2$ (as there is no longer any uncertainty and no entropy associated with this). This state is now the same as the final state reached via isothermal compression. Therefore, again it can be seen that inserting the partition is thermodynamically reversible and, provided the true final states are considered, the entropy associated is path-independent.[8]

---

[8]Note that this deals with Norton's objections to the result of LPS (2008) which are

5   Issues with the one-molecule gas

Norton argues that LP has not been proven, and thus that no link has been
established between logical and thermodynamical irreversibility. This link is
the foundation of the thermodynamics of computation which Norton claims is
a 'spurious science' (Norton 2012). According to him, the Landauer-Bennett
orthodoxy which is at the heart of the latter is unlikely ever to be estab-
lished because of the need to consider the effect of fluctuations in microscale
implementations of computations. Further, he claims that the treatment of
fluctuations including by LPSG is inconsistent. In the following section, we
consider Norton's characterisation of processes at the the molecular level and
see that the treatment of fluctuations implicit in LPSG is not inconsistent
after all.[9]

5.1   Norton, the Second Law and the molecular scale

For Norton, the fact that molecules make up the macroscopic entities that
thermodynamics talks about ushers in violations of SL because he says "a
kinetic gas could, with very low probability, spontaneously recompress to a
much smaller volume, in contradiction with the second law of thermodynam-
ics" (2011, p. 185). Furthermore, he claims that Brownian motion is an
example of SL being violated on the molecular scale because he regards the
dancing of the pollen grains as a result of heat energy from the water being
converted into work (the motion of the pollen). This he claims even violates
the Kelvin formulation of SL (Norton 2012).

This perspective on processes at the molecular scale explains his discontent

_____

precisely that the thermodynamic entropy is not well-defined for the states involved since
it he argues it is path-dependent, and that the processes used are inadmissible.

[9]Note there is no inherent reason that the thermodynamics of computation *has* to be
about the microscopic level. Rather, the motivation for considering small systems such as
one-molecule gases is to consider the lower limit on what is possible.

with the partition process. Fluctuations cause the molecule to move from one side of the box to another: inserting the partition is 'locking' in one of these fluctuations. Why does Norton object to 'locking in' a fluctuation? Accumulating and amplifying these microscopic violations is precisely how a Maxwell's demon would operate. Thus, Norton objects to processes that 'lock in' fluctuations to prove LP, given that LP is supposed (for example, Blundell 2006) to help us escape from Maxwell's demon.

According to Norton, the reason why we do not have to abandon SL, is because 'the microscopic violations' of SL never accumulate to a macroscopic difference. For example, we cannot harness the work done by the pollen grains to drive a macroscopic change as they are moving in random directions. So the molecular constitution of matter does not stop SL being true as an emergent fact about macroscale processes.

However, Norton is mistaken. Firstly, a kinetic gas spontaneously compressing to a smaller volume is *not* a violation of SL. It is overwhelmingly unlikely to happen, as is throwing a fair coin and getting 1000 heads consecutively. However, we do not think that 1000 heads violates the law of 'head or tails with 50/50 probability'. Unlikely events *can* happen. Further, it is a mistake to characterise brownian motion as a 'thermodynamic fluctuation phenomenon' that violates SL. It is not that the 'thermal energy' of the water that is being translated into 'work done' by the pollen grains jiggling. Rather, some of the kinetic energy of the water molecules is transferred to the pollen grains by collisions; this is entirely unmysterious. Heat and work are properties that belong to macroscopic entities, such as gases and heat baths, rather than molecules or pollen grains.

But if Norton can't talk about heat and work at the scale of pollen grains, how can LSPG (and other proponents of LP) talk about one molecule gas in a box doing work and exchanging heat with the heat bath? A molecule in a box can be treated as an ideal gas. To treat 1000 molecules in a box as

a gas (with constant pressure) you have to average over a short time. The same reasoning applies to one molecule in a box, except for the behaviour to count as a 'gas' you must average over a long(er) time. Thus, as it can be treated as a gas, it is legitimate to talk about work and heat.

Further, the same mischaracterisation is the basis of the claim that inserting a partition is 'locking in' a fluctuation. The molecule moving from one side to the other is not a density fluctuation of the gas. This is a category mistake: the molecule and the gas are not the same entity. The gas has a volume equal to that of the box, whereas the molecule's volume is minute (of the order $10^{-33}/m^3$). Moreover, the molecule moving from one side of the box, colliding with the walls are *part* of it behaving like an ideal gas. If the molecule remained in the same location for an extended period of time, it would no longer be acting as an ideal gas. As such, the molecule moving from one side to the other is *not* a fluctuation in the sense of a deviation from equilibrium.

In the picture below are three different snapshots of the configuration of the molecules in a gas. It is an error to draw the line around the molecules (as shown) and claim that the 'gas' has a density fluctuation between the different snapshots.
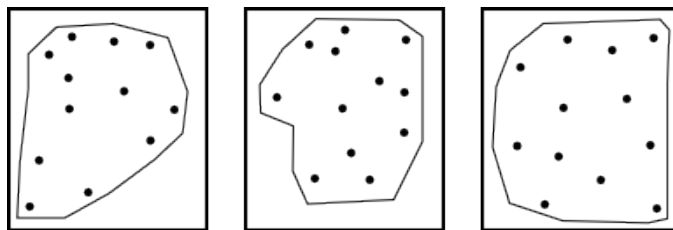

Figure 3: *The different positions of molecules at three consecutive moments of time*

It is this mischaracterisation of thermal fluctuations at the molecular level that seems to motivate Norton's discontent with the processes employed by LSPG: he thinks the processes *proceed* by fluctuations. As the movement of

the molecule is seen as a density fluctuation, he thinks isothermal expansion also proceeds by fluctuations. This is the basis of his claim that the thermodynamics of computation selectively ignores some fluctuations (such as the position of the piston) but not others (density fluctuations of the gas). Further, he thinks "that [it] is especially awkward since the one occasion on which processes connected to fluctuations are not ignored is when the theory treats the thermodynamics of erasure" (Norton 2012). However, we have seen that this view stems from a mischaracterisation of the one-molecule gas. *All* fluctuations are ignored, so the charge of inconsistency fails. Despite this, whether the assumption that fluctuations can be ignored remains an open question. In Norton (forthcoming) he claims that fluctuations fatally disrupt processes at the molecular level. The Smoluchowski trapdoor fails because it has to be light enough that a collision with a faster moving molecule would open it but not a slower moving molecule, so it is in fact flapping about wildly. Norton claims that the piston in an isothermal expansion would face the same problem; it would be as likely to go from being in the middle of the box to the end of the box via an isothermal expansion as via a fluctuation. However, the piston could just be really heavy and the time taken for the isothermal expansion could be very long.

6   Conclusion

We have argued that Norton's critique of the proof of LP given by LPSG fails. In particular, the processes they use do not violate the second law of thermodynamics, and the cycle he constructs to perform 'dissipationless erasure' requires a process that is not among those LPSG use and is inadmissible, and when revised to be admissible the cycle reduces to reset of known data which is not logically irreversible. Finally, the treatment of fluctuations by LPSG is consistent and the response to Norton's critique of LPSG also counters his criticisms of the result of LPS. We have not addressed Norton's

most recent paper that criticises the whole science of the thermodynamics of computation and that remains a subject for future work.

Acknowledgements

References

S. Blundell and K. Blundell, *Concepts in Thermal physics*, (2010)

Bennett, C. H. (2003). Notes on Landauer's principle, reversible computation, and Maxwell's demon. *Studies in the History and Philosophy of Modern Physics, 34*, 501-510.

Earman, J., and Norton, J. D. (1998). Exorcist XIV: The wrath of Maxwell's demon. Part I: From Maxwell to Szilard. *Studies in the History and Philosophy of Modern Physics, 29*, 435-471.

Earman, J., and Norton, J. D. (1999). Exorcist XIV: The wrath of Maxwell's demon. Part II: From Szilard to Landauer and beyond. *Studies in the History and Philosophy of Modern Physics, 30*, 1-40.

Feynman, R. P. (1996). *Feynman Lectures on Computation*, edited by J. G. Hey and W. Allen. Reading, MA: Addison-Wesley

Ladyman J., Presnell, S., Short, A. and Groisman, B. *The connection between logical and thermodynamic irreversibility* in Studies in History and Philosophy of Modern Physics 38, (2007)

Ladyman J., Presnell, S., Short, A., *The Use of the Information-Theoretic Entropy in Thermodynamics* in Studies in History and Philosophy of Science Part B, 39 (2): 315-324 (2008)

Leff, H. S., and Rex, A. F. (eds.) *Maxwells demon: Entropy, information, computing* Adam Hilger, Bristol. (1990)

Lloyd, S. *Computational capacity of the universe* Vol. 88, 273901-4

Norton J.D., *Waiting for Landauer*, Studies in the history and philosophy of modern physics, (2011)

Norton J.D., *When a good theory meets a bad idealisation: The failure of the thermodynamics of computation* http://www.pitt.edu/ jdnorton/Goodies/Idealization/index.html (Accessed: March 2012)

Norton J. D., *The End of the Thermodynamics of Computation: A No Go Result* Philosophy of Science (forthcoming)

Maroney, O. J. E. (2005). The (absence of a) relationship between thermodynamic and logical reversibility. *Studies in History and Philosophy of Modern Physics, 36*, 355-374.

Norton, J. D. (2005). Eaters of the lotus: Landauer's principle and the return of Maxwell's demon. *Studies in the History and Philosophy of Modern Physics, 36*, 375-411.

Uffink, J. (2001). Bluff Your Way in the Second Law of Thermodynamics. *Studies In History and Philosophy of Modern Physics, 32*, 305-394.