Studies in History and Philosophy of Modern Physics 68 (2019) 90–105

Contents lists available at ScienceDirect

Studies in History and Philosophy of Modern Physics

journal homepage: www.elsevier.com/locate/shpsb

The physics of implementing logic: Landauer's principle and the multiple-computations theorem

Meir Hemmo [a, *], Orly Shenker [b]

[a] Philosophy Department, University of Haifa, Haifa 31905, Israel
[b] Program in the History and Philosophy of Science, The Hebrew University of Jerusalem, Jerusalem 91905, Israel

ARTICLE INFO

Article history:
Received 4 January 2019
Received in revised form
22 March 2019
Accepted 9 July 2019
Available online 26 July 2019

Keywords:
Entropy
Individuation of computation
Landauer's principle
Logical (ir)reversibility
Second law of thermodynamics

ABSTRACT

This paper makes a novel linkage between the *multiple-computations theorem* in philosophy of mind and *Landauer's principle* in physics. The multiple-computations theorem implies that certain physical systems implement simultaneously more than one computation. Landauer's principle implies that the physical implementation of “logically irreversible” functions is accompanied by minimal entropy increase. We show that the multiple-computations theorem is incompatible with, or at least challenges, the universal validity of Landauer's principle. To this end we provide accounts of both ideas in terms of low-level fundamental concepts in statistical mechanics, thus providing a deeper understanding of these ideas than their standard formulations given in the high-level terms of thermodynamics and cognitive science. Since Landauer's principle is pivotal in the attempts to derive the universal validity of the second law of thermodynamics in statistical mechanics, our result entails that the multiple-computations theorem has crucial implications with respect to the second law. Finally, our analysis contributes to the understanding of notions, such as “logical irreversibility,” “entropy increase,” “implementing a computation,” in terms of fundamental physics, and to resolving open questions in the literature of both fields, such as: what could it possibly mean that a certain physical process implements a certain computation.

1. Introduction

In this paper we make a direct and hitherto unnoticed linkage between the multiple computations theorem in philosophy of mind and cognitive science and Landauer's principle in physics. Both are central in their respective fields. We show that this linkage has very surprising consequences in physics and in philosophy of mind and cognitive science (see more details in this introductory section).

The *multiple-computations* theorem (also called *pan-computationalism* or the *indeterminacy of computations* theorem[1]) says that certain physical systems that implement computations, implement simultaneously more than one computation.[2] One implication of this theorem is that if the mind is understood as a computation implemented in the brain, it may be that a single brain implements two entire minds simultaneously, thus challenging the computational theory of mind.[3]

Landauer's principle says that there is a systematic connection between abstract logical properties of computations and the physical properties of the computers on which they are implemented: it claims that the physical implementation of “logically irreversible” functions is accompanied by some minimal dissipation of energy (increase of thermodynamic entropy). This principle is pivotal in contemporary defenses of the universality of the second law of thermodynamics, since it is taken to be crucial in showing that Maxwell's Demon, which is a counter example for the second law, is physically impossible.

The tasks undertaken in this paper are these:

I. Show that the multiple-computations theorem challenges the universal validity of *Landauer's principle*. Combining the insights concerning “*physical (implementation of) computation*” from both physics and cognitive science, our surprising result is that the

* Corresponding author.
E-mail addresses: meir@research.haifa.ac.il (M. Hemmo), orly.shenker@mail.huji.ac.il (O. Shenker).

[1] See for example (Piccinini, 2017).

[2] The distinction between implementing “computation” and implementing “formal structures” is addressed later, in Sections 3 and 8. In the context of Landauer's principle, which is our focus here, the terminology in the literature does not make these subtle distinctions. The “formal structures” in question here are single-valued maps from input states to output states, typically logical transformations. See discussion of the relevant notion of computation in (Ladyman, 2009, Section 3).

[3] On the computational theory of mind, see (Rescorla, 2017).

https://doi.org/10.1016/j.shpsb.2019.07.001
1355-2198/© 2019 Elsevier Ltd. All rights reserved.

multiple-computations theorem in cognitive science has important implications with respect to the second law of thermodynamics. We open this paper with a presentation of the two theses, in Sections 2 and 3. The incompatibility between them is first demonstrated in Section 4.

II. Provide the multiple-computations theorem with a general physical underpinning, by formulating it in terms of *fundamental physics*, based on recent findings in the philosophy of physics. This task involves clarifying what the notion of "*implementing a computation*" could mean in terms of fundamental physics, given the insights on this topic from recent literature in the philosophy of computation and cognitive science. This physical underpinning sheds light on some aspects of this theorem that are (still) under debate in contemporary literature, e.g., under what conditions can we justifiably say that a given physical system uniquely implements a given computation. This task is undertaken in Sections 5 and 8.

III. Formulate Landauer's principle in terms of *fundamental physics* (rather than in terms of thermodynamics or statistical mechanics), thus clarifying where (according to fundamental physics) this principle holds and where it doesn't. This task involves clarifying the notions of "*implementing a computation*", "*logical irreversibility*" and "*entropy increase*" in terms of fundamental physics, given insights from recent literature on the philosophy of physics and in particular of statistical mechanics. This task is undertaken in Sections 5, 6, and 7. The results of tasks II and III support the results of task I.

2. Introducing Landauer's principle

In 1961, in a paper published in the *IBM Journal for Research and Development*, Rolf Landauer suggested an idea that, if true, is quite astonishing. Technically, the idea – that since then came to be known as *Landauer's principle* – is this: The physical implementation of logically irreversible operations (namely those in which from the output one cannot infer the input) is necessarily accompanied by dissipation of energy (i.e. increase of entropy) in the amount of (at least) $k$log2 per loss of one bit of information (see also Landauer, 1992, 1996).[4] This typically means that implementing logically irreversible properties – but not logically reversible ones – is responsible for the generation of *heat* in the minimum amount of kTlog2 per lost bit. The simplest example of a logically irreversible operation is "*logical erasure*" (hereafter for short: "erasure"[5]), which is the mapping 1 → 1 and 0 → 1 (or equivalently 1 → 0 and 0 → 0). In this operation one bit of information – e.g. an answer to one yes/no question – is "lost", since the input cannot be recovered from the output.[6] According to Landauer's principle the physical implementation of this logical operation must be accompanied by entropy increase of at least $k$log2.

Why is this idea astonishing? Landauer's principle posits a substantive but non-reductive connection between the physical level and the logical level that seem to be straightforwardly in tension with the thesis of physicalism. The first issue is this: Landauer's principle does not satisfy the condition of *supervenience*; see Fig. 1.

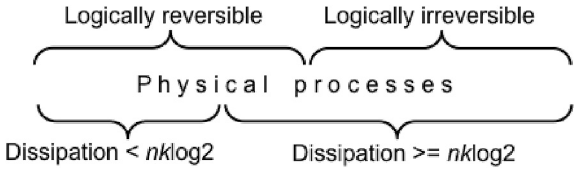

Fig. 1. Entropy increase and logical (ir)reversibility; failure of supervenience.

This is because Landauer's principle entails that physical processes that result in entropy increase *above* the minimum may implement both logically reversible and logically irreversible computations, so that Landauer's minimum bound on entropy increase is only partly correlated with logical irreversibility. Since supervenience is usually taken to be the (minimal) hallmark of physicalism (e.g., Kim, 1990, 2012), it turns out that the notion of logical (ir)reversibility in Landauer's principle is (in this sense) a *non-physical* one. While the idea that logical properties are not physical properties may, in itself, be acceptable for many (according to main stream views, at least[7]), the astonishing idea is that facts, that (as we now see) are non-physical, are claimed to put strong measurable constraints on possible physical facts (concerning entropy increase), and moreover that this idea is (claimed to be) based on considerations within physics (see Sections 2, 5, 6, 7 on these physical considerations).

The second implication of Landauer's principle that is in tension with (reductive) physicalism is that Landauer's principle entails that the property of the logical reversibility of computations is *multiply realizable* in the following sense: *Reversible* computations may be implemented by physical processes that result in entropy increase above the minimum as well as by physical processes that result in entropy increase below the minimum. So, according to Landauer's principle, also the property of logical reversibility of computations is not fully correlated with the minimum entropy bound set by the principle.[8]

When reading Landauer's (1961) and later literature on the subject, it is not easy to see this problem: the connection between the logical (ir)reversibility of the implemented function and the nature of the physical implementation is treated as if it is straightforward, almost as if the logical properties are inherent in the physical properties of the implementing system, or that (conversely) the physical properties are inherently implementing the logical properties.

The fact that in Landauer's principle logical properties do not supervene on physical properties may be a hint that problems lurk in it, waiting to be uncovered (as we do in this paper). (This is the first, but not the last, insight to be taken from the philosophy of mind and cognitive science and applied in the foundations of physics. We shall encounter more as we proceed.)

Why is Landauer's principle significant? Landauer's principle is key to the contemporary establishment of the universal validity of the second law of thermodynamics, since it is pivotal in the attempts to solve the riddle of *Maxwell's Demon*. Maxwell's Demon is a *thought experiment* proposed in 1867 by J.C. Maxwell as a *perpetuum mobile* of the second kind, that is, as a counter example for the second law of thermodynamics. Since then physicists and philosophers have tried, in a variety of ways, to ensure the universal validity of the second law by showing that Maxwellian Demons are incompatible with the principles of fundamental physics. There is, of course, overwhelming empirical evidence supporting the second

[4] Entropy has no physical units; it is a number. This holds for all expressions of entropy, in all the different approaches to statistical mechanics; see e.g., (Frigg, 2008).

[5] We take "erasure" to be a logical function; erasure by, for example, blowing up a computer is not what we are after.

[6] See discussion of this term in (Hoefer, 2016, Section 2.3) Note that accounting for irreversible logical functions in terms of *bi-directionally-deterministic* physical theories (e.g., classical mechanics) requires a *macroscopic* description, since in a bi-directionally deterministic world, there is *no* microscopic erasure; see Section 5. For more on the notion of 'determinism' and its connection with causation, see (Hoefer, 2016; Ben-Menahem, 2018).

[7] See (Szabo, 2012) on physicalism concerning mathematics.

[8] The thesis of multiple realizability cannot be reconciled with physicalism but we will not argue for this point here; see (Hemmo & Shenker, 2019). Our arguments below do not depend on this issue.

law of thermodynamics, and yet, such a thought experiment is taken by many to be problematic. The most recent attempts to solve this problem are based on Landauer's principle: The currently prevalent view is that if Landauer's principle is true then Maxwell's Demon is "exorcised",[9] and the universality of the (probabilistic counterpart of the) second law is preserved.[10] Hence the great importance ascribed to Landauer's principle.[11]

Most of the physicists as well as philosophers take Landauer's principle to be established and validly supported by physics (see e.g., Bennett, 1982, 2003; Bub, 2001; Feynman, 1996; Ladyman, 2009; Ladyman & Robertson, 2014), although some arguments have been put forward that criticize it (including e.g., Earman & Norton, 1998, 1999; Maroney, 2005; Norton, 2005, 2011; Hemmo & Shenker, 2010, 2012, 2013).

How should the acceptability of Landauer's principle be decided?

Empirical testing: First of all, Landauer's principle entails empirical predictions, and attempts are constantly made to establish either its truth or its falsity on the basis of empirical findings.[12] At this point, however, we do not take these findings to be conclusive, and we shall not go into their details.

A-priori considerations: As is usual, we do not take Landauer's principle to be assumed a-priori, or to be a tautological definition, that identifies loss of n bits with energy dissipation of $nk\log2$.[13,14]

Compatibility with the theories of physics: In this paper, and as is usual, we take Landauer's principle to concern the physics of computing systems, that needs to be supported by the theories of physics that are relevant for such systems. Accordingly, the criterion that we employ for deciding on the acceptability of this principle is its support by these theories of physics. We undertake this task in two ways:

(1) In Section 4 we provide a counter example for Landauer's principle, that is based on the multiple-computations theorem (presented in Section 3), which we take (in turn) to have a foundation in fundamental physics (discussed in Section 8).

(2) In Sections 5, 6 and 7 we re-examine the compatibility of Landauer's principle with *fundamental* physics. This line of enquiry is different from the standard treatment of this principle, carried out in terms of the high-level theories of thermodynamics and statistical mechanics. We discuss our reasons for preferring this line of thinking in Section 5.

3. Introducing the multiple-computations theorem

In 1988, in an appendix to his book *Representations and Reality*, Hilary Putnam suggested an idea that, too, was quite astonishing when it was first presented: "Every ordinary open system is a realization of every abstract finite automaton."[15] (Hemmo & Shenker, 2013, p. 121) Putnam originally illustrated this very strong thesis as follows. Consider a system that, following the laws of physics, evolves from time 12:00 to time 12:07, changing its physical state every minute, such that its states at these moments are S0, S1, S2, S3, S4, S5, S6 and S7.[16] If (with Putnam) we assign the value "1" to the disjunction S0 ∨ S2 ∨ S4 ∨ S6 and the value "0" to the disjunction S1 ∨ S3 ∨ S5 ∨ S7, then the sequence of physical states may be seen as implementing the sequence of symbols 1010101. To see the nature of the multiple-computations thesis, notice that a different value assignment, in which the value "1" is assigned to the disjunction S0 ∨ S1 ∨ S2 ∨ S3 and the value "0" to the disjunction S4 ∨ S5 ∨ S6, results in the system implementing the sequence of symbols 1111000, as it undergoes exactly the same physical evolution as before. The reader can easily construct implementations of additional sequences by the same system during the same physical evolution.[17] Thus, as the system undergoes one and the same microphysical evolution it implements *all* the computations resulting from all the possible value assignments, not only *potentially* but in *actuality*. This is an illustration of Putnam's idea.

Putnam's insight is usually taken to mean that "syntax is not intrinsic to physics" (Searle, 1992, p. 208), or that the physics is "blind" to the syntax, in the sense that, since one physical matter of fact gives rise to a multiplicity of computations, *all* of which are equally and simultaneously carried out, it turns out that the physical matters of fact do not fix one of these as a computations that is actually carried out.[18] Searle illustrated this idea with an example that since then became paradigmatic: "Thus for example the wall behind my back is right now implementing the Wordstar program, because there is some pattern of molecular movement that is isomorphic with the formal structure of Wordstar." More generally, "For any program and for any sufficiently complex object, there is some description of the object under which it is implementing the program." (Searle, 1992. pp. 208-9).

Since its formulation by Putnam (1988), a number of criticisms were mounted against the strong form of this thesis, arguing that *not every* system is a computer, and those that are, do *not* implement *every* computation: various kinds of constraints (modal,[19] causal,[20] semantic,[21] functional,[22] and others) have to be satisfied if we are to say that a given system implements a certain computation (for an overview see Piccinini, 2017). Still, a *modest version* of the thesis remains non-controversial, namely, that *some* systems

[9] In the terms of Earman and Norton (1998, 1999).

[10] Because of this role of Landauer's principle, grounding its proof in the second law is viciously circular, since the second law is defended against the counter example of Maxwell's Demon by relying on Landauer's principle; we return to this point below.

[11] See introductory overview and papers in (Leff & Rex, 2003). But see also the argument that Maxwell's Demon is compatible with fundamental physics in (Hemmo & Shenker, 2010, 2011, 2012, 2013, 2016) for the classical case, and in (Hemmo & Shenker, 2017) for the quantum mechanical case.

[12] Recent examples of attempts to establish the empirical *falsity* of Landauer's principle are: (Cottet et al., 2017; Chida, Desai, Nishiguchi, & Fujiwara, 2017); Masuyama et al., 2018).

[13] Ladyman and Robertson (2014, p. 2287) write: "[W]e could take LP as a regulative principle that is somehow constitutive of the theory and assumed a priori."

[14] One reason for why this principle cannot be a tautology (that is not discussed in the literature) is the fact that phase space regions can be measured using a variety of measures.

[15] Adding the physical input and output adds some constraints, but even in that case Putnam's claim is quite strong. We address the approach of computational externalism in Section 8.

[16] The *Si* states are macrostates, a notion explained below (since its analysis is pivotal in understanding the limitations of Landauer's principle).

[17] Copeland (1996), in his criticism of Putnam's idea, claims that examples in which the computation is given first, and the physical states are chosen so as to fit it, do not support pan-computationalism. We don't address this claim here, since it is not relevant to our criticism on Landauer's principle in Section 4.

[18] This understanding is not universal. Some say that Putnam's thesis does *not* imply that syntax isn't intrinsic to physics, since one may still say that "*all* these implemented structures are intrinsic" (Shagrir, 2001, p. 379). Others say that Putnam's thesis does *not* imply that *computation* is not intrinsic to physics, since "computation" should be associated with the basic physical process itself (see Shagrir, 2018, Section 4, for a critical presentation of this idea and for references).

[19] For example (Chalmers, 1996; Copeland, 1996); see more on this in Section 6.

[20] For example, in order for a system to implement a computation it has to have a certain causal organization; see (Chalmers, 1996, 2012, Section 5).

[21] For example, some argue that computations involve representations, see (Shagrir, 2001; Sprevak, 2010); Egan, 2012). For a recent defense of the semantic approach to individuation of a computation, see (Shagrir, 2018); for criticism, see (Dewhurst, 2018).

[22] For example, see (Coelho Mollo, 2017; Egan, 2017).

that are taken to implement computations (and *ipso facto* satisfy the requirements for doing so) do *in fact* implement more than one computation as they undergo one and the same microphysical process (e.g. Shagrir, 2012). For our argument in this paper the most modest non-controversial version of the thesis is all we need. Hereafter, we use the name "*the multiple-computations theorem*" to refer to this modest version, and it is the one for which we provide examples below. (Whether or not this is a strictly speaking theorem, and in this case, what entails it, is a matter that we shall come back to at Sections 5 and 8, as we consider its physical underpinning.)

Why is the multiple-computations thesis astonishing, and what makes it significant? The (modest) multiple-computations thesis, by implying that "syntax is not intrinsic to physics" (Searle, 1992, p. 208), challenges physicalism about computation. Some think that, *ipso facto*, it also challenges "*realism about computation*", which is the idea that "whether or not a particular physical system is performing or implementing a particular computation is at least sometimes a fact that obtains independently of human beliefs, desires and intentions." (Ladyman, 2009, p. 377). Therefore, the (modest) multiple-computations theorem has a number of significant implications.

One implication is that the multiple-computations theorem presents a challenge to the computational theory of mind (Rescorla, 2017), since it implies that the brain (assuming that it implements computations) simultaneously implements two (or more) different computations, which may be two (or more) different entire minds (see discussions in Chalmers, 1996, 2012; Shagrir, 2001, 2012; Piccinini, 2015, 2017).[23] To face this challenge, there is an ongoing debate in the philosophy of mind and cognitive science about the nature of computing systems and the computations that can be implemented in them, and about the nature of computations that can give rise to minds. The present paper is not part of this debate: *the computing systems that we shall discuss in the context of Landauer's principle needn't be brains, and needn't be minds*. Still, some insights gained in this debate will be important for us, and will be addressed in Section 8.

Another implication, that is directly relevant to the first task of our paper, is that if determining which computation is carried out is not physical, it is hard to see how the properties of this computation (e.g. its logical (ir)reversibility) can have implications with respect to its physical properties (e.g., energy dissipation, as in Landauer's principle). Therefore, examining Landauer's principle in terms of fundamental physics requires that we provide an underpinning, in terms of fundamental physics, of the multiple-computations theorem. This task is carried out in Sections 5-8 below.

To see how the multiple-computations theorem challenges Landauer's principle, let us consider an illustration of this theorem, provided by Shagrir (2012), that we will extend later (in Section 4).

Consider a physical system (call it L) that consists of three elements A, B, and D; L is prepared at t_0 such that A and B are in certain physical states; and it evolves (according to the laws of nature) such that at some later time t_1 element D is in a certain physical state; see Fig. 2. Call the initial states (at t_0) of A and B "*input states*" and call the final state (at t_1) of D "*output state*". Alternatively, one may think of system L as consisting of two elements, that change their states according to some dynamical law Q_L. In this case, illustrated in Fig. 3, A and B are the initial states of these two elements (called "*input states*"), and their final states are C and D; but only the state D is called "*output state*".[24] We shall want to use system L as a computer, that implements symbol manipulation, and will now consider some minimal conditions for doing so.

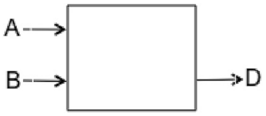

Fig. 2. Configuration of a computing system with three elements.

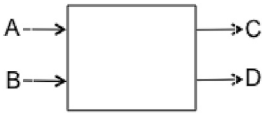

Fig. 3. Configuration of a computing system with two elements in four states.

Let X, Y and Z be three possible values of some physical magnitude pertaining to A, B and D (it is convenient to think of L as tri-stable in this context). For example, X, Y and Z may be three voltage ranges, or three position ranges, etc.[25] The dynamical rule Q_L that governs the evolution of system L is this:

(i) If both inputs A and B are Z, then the output D is Z.
(ii) If both inputs A and B are X, then the output D is X.
(iii) In all other cases, the output D is Y.

System L is built (or conveniently found in nature) such that, due to its structure and parameters, the laws of nature (described by an equation of motion) are such that the rule Q_L obtains. The three first columns of Table 1, with header "*physical states*", describe all possible evolutions of L, given all its possible input states, according to Q_L.

Of course, if we want L to implement the computation of a certain logical function, we need to define a mapping that associates the physical states of L with computational states, 0 and 1, so that given this mapping the change of L's physical states could be seen as corresponding to the desired logical function. On the question of the nature of this mapping, let us make at this stage only two important and related comments:

(i) To avoid the triviality result of the Putnam-Searle argument, according to which almost every microphysical evolution implements almost any computation, critics have proposed that further constraints should be imposed on this mapping if it is to be considered a "computation".[26] But these

[23] The strong thesis (but not the modest one) also entails that the computational theory of mind is vacuous since every system, including Searle's famous wall, implements the computations that are allegedly associated with minds.

[24] Regarding this second case, it is well known that, taking into account additional degrees of freedom, and treating them as information bearing, can affect the logical properties of the computed function (by making all computations "logically reversible"). We address this point later, and for now focus on Shagrir's (2012) example in which the output is D only.

[25] In (Shagrir, 2012) X is voltage in the range [0, 2.5), Y is voltage in the range [2.5, 5), and Z is voltage in the range [5, 10]. In this case, since the voltage ranges are unequal, one may say that the states A and B may have different probabilities or different entropies. This case is compatible with Landauer's principle, as Landauer (1961) already noticed, but for simplicity of presentation it is better to think of examples in which A and B have the same probability and entropy.

[26] Constraints (e.g., modal, causal, or others) must be added to the formal mapping to avoid the triviality result of the multiple computations theorem in its strong version. Which kinds of constraints is a debated topic in the literature, that we don't explicitly address here, since (as we said) our focus is the weaker *modest* multiple computations theorem which is presumably immune to all of the proposed constraints. For causal constraints, see (Chrisley, 1994); Melnyk, 1996; Chalmers, 1996, 2011); for dispositional ones, see (Klein, 2008); for mechanistic, see (Piccinini, 2008, 2015; Miłkowski, 2013); for modal, see (Chalmers, 1996, 2011; Copeland, 1996); Shenker, 2000); and for pragmatic constrains, see (Egan, 2012); Matthews & Dresner, 2016).

Table 1
System L with dynamics Q_L implementing *two* logical functions under *different* value assignments.

	Physical states according to Q_L			Value assignment 1			Value assignment 2		
	Input A	Input B	Output D	Input A	Input B	Output D AND	Input A	input B	Output D OR
Voltage ranges & Value Assign-ments	X	X	X	0 = X	0 = X	0 = X	0 = XorY	0 = XorY	0 = XorY
	X	Y	Y	0 = X	1 = YorZ	1 = YorZ	0 = XorY	0 = XorY	0 = XorY
	X	Z	Y	0 = X	1 = YorZ	1 = YorZ	0 = XorY	1 = Z	0 = XorY
	Y	X	Y	1 = YorZ	0 = X	1 = YorZ	0 = XorY	0 = XorY	0 = XorY
	Y	Y	Y	1 = YorZ	1 = YorZ	1 = YorZ	0 = XorY	0 = XorY	0 = XorY
	Y	Z	Y	1 = YorZ	1 = YorZ	1 = YorZ	0 = XorY	1 = Z	0 = XorY
	Z	X	Y	1 = YorZ	0 = X	1 = YorZ	1 = Z	0 = XorY	0 = XorY
	Z	Y	Y	1 = YorZ	1 = YorZ	1 = YorZ	1 = Z	0 = XorY	0 = XorY
	Z	Z	Z	1 = YorZ	1 = YorZ	1 = YorZ	1 = Z	1 = Z	1 = Z

constraints are not enough to avoid the much weaker *modest* multiple computations theorem in cases like Shagrir's (2012; see also Shagrir, 2018).[27]

(ii) At this stage we are neutral with respect to the meaning of a "mapping that associates physical states with computational states." This is part of the subject matter of intensive discussions in the philosophy of cognitive science, where the question of which features of a computing system "individuate the computation" that it carries out is still open (see Piccinini, 2017 for overview, Shagrir, 2018 for a recent critical discussion). Since Landauer's principle is allegedly proved within physics, it is important that the mapping relation, as well as the individuation of a computation, will be given in physical terms. We shall address this issue, i.e., of which physical facts determine the mapping and which physical facts individuate the computation, in Section 8.[28]

Let us now go back to the construction of the physical-to-computational mapping. Since each of A, B and D can be in any of *three* possible physical states,[29] the association of the physical states with the symbols is not trivial. Shagrir (2012) considers two options of *value assignment*. In value assignment 1, the symbol 0 is implemented by the physical state X, and the symbol 1 is implemented by any of the physical states Y or Z. In assignment 2, the symbol 0 is implemented by any of the physical states X or Y, and the symbol 1 is implemented by the physical state Z. In short:

(Value assignment 1) X is 0; {Y or Z} is 1.

(Value assignment 2) {X or Y} is 0; Z is 1.

In Table 1, the columns headed "*Value assignment 1*" and "*Value assignment 2*" describe these cases. For example, consider the second row, in which A is in the physical state X, B is in the physical state Y, and D is (according to the dynamical rule Q_L) in the physical state Y. According to value assignment 1, X implements 0 and Y implements 1, and therefore this row stands for the case A = 0, B = 1, D = 1, that is, the mapping (0,1) → 1; and when applying value assignment 2, X implements 0 and Y implements 0, and therefore this row stands for the case A = 0, B = 0, D = 0, and the mapping is (0,0) → 0.[30]

Considering all the nine input options (described in Table 1), it turns out that under value assignment 1 the computed function is: (0,0) → 0, (0,1) → 1, (1,0) → 1, (1,1) → 1, which is known as the OR function; and under value assignment 2 the computed function is: (0,0) → 0, (0,1) → 0, (1,0) → 0, (1,1) → 1, which is known as the AND function. Notice that the implementation of these logical functions is an outcome of the *combination* of *both* the value assignments *and* the dynamics.

If both value assignments are possible, then both computations are carried out simultaneously by the same microphysical evolution Q_L. Hence, the physics does not determine which of these two computations is carried out; both are! And so Shagrir's (2012) example is an illustration of the multiple-computations theorem.[31]

The last two statements depend, of course, on the meaning, and in particular the physical meaning, of "*value assignment*". What kind of fact makes it the case that a given physical state, say Y in our example, "corresponds to" or "is associated with" or "stands for" or "realizes" or (finally) "implements" (etc.) the symbol 1 under one "value assignment" and 0 under another? (We shall not address the differences between these notions. We take this to be essentially the problem sometimes called in the literature "individuation of computation".) This topic is under debate in contemporary thinking in both philosophy of physics and philosophy of cognitive science, and is also important for the philosophy of computation in general. Our aim in this paper is to offer an account of computation in terms of *fundamental physics*, and therefore our task is to describe physical facts that will carry out the roles of the above notions that are relevant for Landauer's thesis and for the multiple-computations theorem (we do not examine whether or not our fundamental physical account fully replaces all of these terms; this discussion is beyond the scope of this paper). We return to this subject in Section 8.

4. The multiple-computations challenge for Landauer's principle

In the example of the multiple-computations theorem by Shagrir (2012), presented in Section 3, both computations (AND and OR) are logically irreversible, since one cannot infer the input state uniquely from the output state. This is always the case where

[27] See (Dewhurst, 2018), for a recent criticism, and an attempt to individuate computations (but not logical functions) on the basis of macroscopic physical features *without* representation (and without syntax). But this proposal too faces trivialization: since it entails that almost every sequence of macroscopic states is a computation (regardless of which logical functions it computes) and this leads immediately to a problem of *unlimited* multiple computations of a somewhat different sort than the standard problem.

[28] The multiple-computations theorem may be understood as implying that the physics is *not sufficient* for fixing the mapping. The prevalent view is that the physics is also *not necessary* for fixing the mapping, since computations are multiply realizable.

[29] Other examples require a continuum of physical states.

[30] Notice that in Shagrir's (2012) example both value assignments, and hence both simultaneous computations, are implemented at the same level of organization; others have described simultaneous implementations at different levels of organization, e.g., (Chalmers, 1996), and Section 7 below.

[31] Note that this conclusion, i.e. that both computations are equally real, is part of the *modest* multiple-computations theorem; see Section 8 for more on this issue. Also, here we first have the dynamics and value assignment, and then we discover which logical functions they implement; rather than the other way around. So, Copeland's (1996) criticism does not hold.

Table 2
System K with dynamics Q_K implementing a *reversible* computation under Value assignment 1.

	Physical states for dynamics Q_K				Value assignment 1			
	Input A	Input B	Output C	Output D	Input A	Input B	Output C	Output D
Voltage ranges & Value Assign-ments	X	X	Y	Y	0 = X	0 = X	1 = YorZ	1 = YorZ
	X	Y	Y	X	0 = X	1 = YorZ	1 = YorZ	0 = X
	X	Z	Y	X	0 = X	1 = YorZ	1 = YorZ	0 = X
	Y	X	X	Y	1 = YorZ	0 = X	0 = X	1 = YorZ
	Y	Y	X	X	1 = YorZ	1 = YorZ	0 = X	0 = X
	Y	Z	X	X	1 = YorZ	1 = YorZ	0 = X	0 = X
	Z	X	X	Y	1 = YorZ	0 = X	0 = X	1 = YorZ
	Z	Y	X	X	1 = YorZ	1 = YorZ	0 = X	0 = X
	Z	Z	X	X	1 = YorZ	1 = YorZ	0 = X	0 = X

there are two inputs and one output, and it is well known that such logically irreversible computations can be embedded within logically reversible ones if additional ports (inputs and especially outputs) are added to the system. Let us now consider another system, call it K, in which there are two inputs A and B and two outputs C and D (as in Fig. 3 above). System K works under *different* dynamical rules Q_K, but is seen under the *same* two value assignments as above. As we shall see, just like system L in the example by Shagrir (2012), our system K carries out (at least) two computations during one and the same physical evolution (governed by the dynamics Q_K), where the difference between these computations is only in the assignment of values. However, in our example one computation (implemented by the dynamics Q_K under value assignment 1) is logically *reversible*, while the other computation (implemented by the same dynamics Q_K under value assignment 2) is *genuinely* logically *irreversible*.

As before, the abstract terms X, Y and Z stand for three possible physical states of the ports A, B, C, and D. The dynamical rule Q_K that governs the evolution of system K is this:

(i) If input A is X, then output C is Y.
(ii) If input A is either Y or Z, then output C is X.
(iii) If input B is X, then output D is Y.
(iv) If input B is either Y or Z, then output D is X.

The value assignments are the same as before:

(1) X is 0; {Y or Z} is 1.
(2) {X or Y} is 0; Z is 1.

In each of Tables 2 and 3, the first four columns headed "*physical states for dynamics* Q_K" describe all the 9 possible pairs of inputs and the corresponding physical outputs according to the dynamical rules Q_K. Therefore, these four columns are identical in the two Tables 2 and 3. The next four columns (in each table) describe, for each physical state, the value 0 or 1 assigned to it according to the value assignment. In Table 2 value assignment 1 is applied, with the result that the physical transformation (A,B)→(C,D) implements the mapping (0,0)→(1,1); (0,1)→(1,0); (1,0)→(0,1); (1,1)→(0,0). In Table 3 value assignment 2 is applied, and accordingly the physical transformation (A,B)→(C,D) implements the mapping (0,0)→(0,0); (0,1)→(0,0); (1,0)→(0,0); (1,1)→(0,0). According to the multiple-computations theorem, as discussed above, there is no objective physical matter of fact in the world as to which of these mappings is carried out; both are. (Once again, this depends on the meaning of "value assignment," see Section 8 below.)

It is of utmost importance to distinguish between two ways of using system K as a computer, as follows.

(I) Focusing on only some of the ports (in Figs. 2 and 3) as the information bearing ones. Suppose that the user wishes to focus, for her practical needs, on only some of the ports and ignore others (as in the example discussed in the previous section) (We discuss the subtle term "user" in Section 8, and for now use it informally and intuitively.). Consider, for example, Tables 2 and 3: If she focuses on only A, B and D and ignores C, then under value assignment 1, the physical transformation (A,B)→D may be interpreted as a computation of the logically irreversible mapping (0,0)→1; (0,1)→0; (1,0)→1; (1,1)→0; and under value assignment 2, the same transformation may be seen as a computation of a different logically irreversible mapping: (0,0)→0; (0,1)→0; (1,0)→0; (1,1)→ 0. (That both are logically irreversible is not surprising since we have two inputs and one output.) Alternatively, she may focus on the physical transformation B→D, ignoring both A and C, then under value assignment 1, this transformation may be seen as a computation of the logically *reversible* mapping sometimes called "*not*", 0 → 1 and 1 → 0, and under value assignment 2, the same transformation may be seen as a computation of the logically *irreversible* function of "*erasure*", 0 → 0 and 1 → 0. These two computations implemented by the B→D physical transformation are already problematic for Landauer's principle, but we do not focus on this case for the following reason.

While, in the practice of implementing computations, ignoring some elements of the computer may be very useful, it is well known that in the context of Landauer's principle one must carefully consider all the physical elements of the system, ignoring none, since only by taking all of them into account can one assess whether or not the process is *genuinely* irreversible, and whether or not it is *genuinely* dissipative. (Bennett's, 2003 analysis and responds to critics includes examples of such a meticulous search for the missing degrees of freedom.) Hence from now on we shall follow *all* the system's elements. (This will be our guiding line also in the analysis based on fundamental physics, in Section 5 onwards.)

(II) Taking all the ports (in Figs. 2 and 3) into account. As Tables 2 and 3 show, the microphysical evolution (A,B)→(C,D) governed by Q_K implements two computations, of which one (value assignment 1, Table 2) is logically reversible and the other (value assignment 2, Table 3) is logically irreversible. This is similar to the two computations implemented by the physical evolution B→D, as we saw above, but in this case, we have taken *all* the ports into account.

According to Landauer's principle, the logically irreversible computation is *necessarily* accompanied by energy dissipation by *at least* *k*log2 per bit of lost information, while the logically reversible computation need not be accompanied by any minimal amount of dissipation. What does Landauer's principle predict for the physical process described by the dynamical rule Q_K? Is there or is there not a minimal amount of dissipation that must accompany the physical process in question? If we focus on one value assignment the answer is that there is; if we focus on another value assignment the

Table 3
System K with dynamics Q_K implementing an *irreversible* computation under Value assignment 2.

	Physical states for dynamics Q_K				Value assignment 2			
	Input A	Input B	Output C	Output D	Input A	Input B	Output C	Output D
Voltage ranges & Value Assign-ments	X	X	Y	Y	0 = XorY	0 = XorY	0 = XorY	0 = XorY
	X	Y	Y	X	0 = XorY	0 = XorY	0 = XorY	0 = XorY
	X	Z	Y	X	0 = XorY	1 = Z	0 = XorY	0 = XorY
	Y	X	X	Y	0 = XorY	0 = XorY	0 = XorY	0 = XorY
	Y	Y	X	X	0 = XorY	0 = XorY	0 = XorY	0 = XorY
	Y	Z	X	X	0 = XorY	1 = Z	0 = XorY	0 = XorY
	Z	X	X	Y	1 = Z	0 = XorY	0 = XorY	0 = XorY
	Z	Y	X	X	1 = Z	0 = XorY	0 = XorY	0 = XorY
	Z	Z	X	X	1 = Z	1 = Z	0 = XorY	0 = XorY

answer is that there isn't. Assuming that the amount of energy dissipation is a *fact* in the world,[32] it seems that the multiple-computations theorem, as illustrated in the above example, entails that Landauer's thesis makes contradictory statements concerning this fact. If that is the case, then the multiple-computations theorem entails that Landauer's principle is false.[33] (This challenge to Landauer's principle is new and comes from a direction hitherto unexamined in the literature.)

This result – if true – is important, and has quite astonishing consequences: it entails that the multiple-computations theorem is crucial for determining whether or not the second law of thermodynamics is a theorem of contemporary physics.[34] The reason is that, as we said above, the contemporary dominant argument for the universal validity of the second law of thermodynamics relies on Landauer's principle, since this principle plays a central role in "exorcising" a potential counter example for the second law, namely *Maxwell's Demon*. And so, since this result is quite dramatic, let us examine it carefully. Can the above problem be solved? Two lines of thinking come to mind.

(1) **A solution from rejecting the (applicability for this case of the) multiple-computations theorem**. There may be a criterion for preferring one computation as the one that – in some sense – *actually* takes place. In that case, the physical process described above implements either a logically reversible computation or a logically irreversible one (exclusive or), and no contradiction arises. The search for such a criterion is central in the contemporary research in the philosophy of cognitive science and philosophy of mind, and huge efforts are made to come up with it. In Section 8 we examine this issue in terms of fundamental physics. It is of interest to notice that the validity of Landauer's thesis hinges on this matter, so that a topic that is usually taken to be in the field of cognitive science and its philosophy has highly important implications for physics. This will become understandable once we provide the physical underpinning for the multiple-computations theorem, in Sections 5 and 8.

If one assumes the universal truth of Landauer's principle, then one can use it as a constraint on possible computations: Dynamical processes that allow for both logically reversible and logically irreversible computations under suitable value assignments are physically prohibited. However, since this entails radical conclusions, concerning the *impossibility* of certain sequences of microstates, where the only reason for their prohibition comes from the "top down" constraint of Landauer's principle, we think that this response is unacceptable.

(2) **A solution by amending Landauer's principle.** Another option that comes to mind is to embrace (a modest version of) the multiple-computations theorem, as indeed implying that more than one computation takes place during the same physical evolution, and revise Landauer's principle in a way that will be consistent with this fact. The key here is to notice that Landauer's thesis posits a *minimum* of *k*log2 of dissipation per lost bit, rather than a fixed amount of dissipation. One may then say that the minimum amount of dissipation in a given physical process should be one of the following.

Amended Landauer's principle, Option 1: The minimum amount of dissipation (according to the *amended* principle) should be equal to the sum of the dissipations (according to the *original* principle) in all the irreversible computations that are implemented by a given system during a given microphysical evolution (this may result in a huge amount of dissipation, since the number of computations may be unbounded).

Amended Landauer's principle, Option 2: The minimum amount of dissipation (according to the *amended* principle) should be the dissipations (according to the *original* principle) of *N*klog2, where *N* is the number of bits lost in the computation for which this number is the largest, among those implemented by that microphysical process.

Notice that in order for Options 1 or 2 to be empirically (or at least ontologically) significant there must be a matter of fact in the world concerning which implemented computations take place during a given microevolution, and consequently there must be a matter of fact in the world concerning which set of coarse grained degrees of freedom are information bearing, so that the dissipation takes place (in each such case, according to the original Landauer's principle) in the rest of the degrees of freedom.[35] However, a given set of coarse-grained degrees of freedom can be *both* information bearing in the framework of one implemented computation, *and* non-information bearing in the framework of another computation that is implemented simultaneously by the same microevolution. Here is a simple example: if the voltage ranges implement the computation, then the position of the system on the laboratory table is a set of non-information bearing degrees of freedom; but ranges of positions on the laboratory table may implement a computation, and relative to this computation, the voltage ranges are non-information bearing degrees of freedom. Consequently, Options 1 and 2 concern the sum or the maximum of the dissipation in the computations that are implemented by *every* set of

[32] While this assumption may prima facie seem almost trivial, this is not the case, since as we shall see the question of which degrees of freedom (in a given microevolution of a given system) are the information bearing ones is subtle. We address this point below.

[33] Note that our two computations are *compatible* with the restrictions on implementation proposed by Ladyman, Presnell, Shrot, and Groisman (2007); Ladyman (2009).

[34] But see also (Hemmo & Shenker, 2010, 2011, 2012) for proofs that Maxwell's Demon is compatible with fundamental physics.

[35] We are grateful to an anonymous referee for this journal for encouraging us to present this result.

degrees of freedom, under *every* possible coarse graining, and in *both* roles – as information bearing and as non-information bearing.

Options 1 and 2 are consistent with the letter of Landauer's principle, even though they do not seem to have been envisioned in the standard arguments by Landauer and others. Let us point out two implications of these options.

(a) Reversible Computing. Famously, *if* the *original* Landauer's principle (i.e., not the *amended* principle as in Options 1 or 2) is true *then* Reversible Computing algorithms, e.g., as in (Bennett, 1973, 1982) or (Fredkin & Toffoli, 1982) could have entropic advantage. However, it follows from the multiple-computations theorem that a physical evolution that implements a Reversible Computing algorithm could, in the general case, under different value assignments, also implement *irreversible* computations, and then, according to Options 1 or 2, the logical reversibility associated with the said algorithms may not always have the entropic advantage they are often thought to have.

(b) Maxwell's Demon. Great efforts have been made to "exorcise" Maxwell's Demon using Landauer's principle, where this principle was implemented given some specific value assignment (starting with Bennett's, 1982). But if the abovementioned Options 1 or 2 are accepted then, even if no dissipation occurs under a given value assignment (as for example in Hemmo & Shenker, 2010, 2011), other value assignments may provide the necessary dissipation to save the second law of thermodynamics. On this latter line of thinking it turns out that the question of Maxwell's Demon cannot be decided conclusively on the basis of any particular value assignment.

While (again) the amended versions 1 and 2 of Landauer's principle agree with the letter of the original principle, and avoid the threat of inconsistency following the multiple computations theorem, we think these amended versions are *so* far from the arguments by Landauer (1961), Bennett (1982, 2003), and many others, that they cannot be taken as established by those arguments. The force of the original arguments has been in making a surprising connection between *logical* irreversibility and dissipation of energy, which is a *physical* matter of fact (perhaps even measurable); but in the amended versions 1 and 2 this connection between logic and entropy is so weak, as to be practically lost. Finally, while the empirical predictions of the original Landauer's principle were not tested due to technological limitations (because the dissipation involved was too small to be uncontroversially detected), the amended principle verges on being *unfalsifiable*, as one can always bring in the hypothesis that additional hidden computations take place on some additional coarse-grained sets of degrees of freedom. In our view these are signs that we are on the wrong track, and that a much simpler and down-to-earth approach is called for.

Indeed, as we shall see below, the fact that Landauer's principle seems to imply inconsistent predictions (due to the multiple computations theorem) is not surprising, since there are *additional* strong reasons to think that this principle is not a theorem of fundamental physics, reasons that are *independent* of the multiple-computations thesis. In the next sections we shall explore the physical basis of Landauer's principle, and then return to the physical explanation of the notion of value assignment, that is: to the physical basis of the multiple-computations theorem, thus solving the above problems.

5. The principles of fundamental physics that are in play in Landauer's principle and the multiple-computations theorem: microstates, macrovariables, macrostates

If Landauer's principle were a theorem of physics, then its threat by the multiple-computations theorem would have been not only surprising, but also worrying. However, Landauer's principle has been challenged in a number of ways. The criticisms mounted against the universal validity of this principle can be divided into two kinds: from high level theories and from fundamental physics.

Criticism from high level theories. In his ground-breaking paper, Landauer (1961) described his principle as follows. "Consider a statistical ensemble of bits in thermal equilibrium. If these are all reset to ONE, the number of states covered in the ensemble has been cut in half. The entropy therefore has been reduced by $klog_e 2 = 0.6931$ per bit.[36] The entropy of a closed system, e.g. a computer with its own batteries, cannot decrease; hence this entropy must appear elsewhere as a heating effect, supplying 0.6931 kT per restored bit to the surroundings." (Landauer (1961). p. 265) This line of thinking, according to which the principle is grounded in the second law of thermodynamics (or its statistical mechanical counterparts) is the prevalent one, *both* in arguments supporting Landauer's principle (including e.g., Bennett, 1982, 2003; Feynman, 1996; Bub, 2001; Ladyman, 2009; Ladyman & Robertson, 2014) and in arguments *criticizing* it (including e.g., Earman & Norton, 1998, 1999; Maroney, 2005; Norton, 2005, 2011).

One problem in grounding Landauer's principle in the second law is that Landauer's principle is itself central in contemporary defenses of the universality of the second law; thus, relying on the second law to establish Landauer's principle is *viciously circular*. Here are a few more details. There are two kinds of ways to establish the universal truth of the second law of thermodynamics. One is empirical evidence: the second law enjoys enormous empirical support, and the overwhelming empirical evidence makes it uncontroversial that there are no perpetual motion machines *in our world*. The second kind of way to establish the universal truth of the second law of thermodynamics is by showing that this universal truth is a *theorem* of fundamental physics (which is, in turn, taken to be fundamentally universally true). Maxwell's Demon is a thought experiment that challenges this latter grounding of the second law. Importantly, Maxwell's Demon is not in conflict with the empirical evidence, because the available proofs that Maxwell's Demon is compatible with fundamental physics leave open the possibility that *both* second law behavior and Maxwellian Demon behavior are compatible with fundamental physics; they may hold for different initial conditions of the

[36] In the literature concerning Landauer's principle the notion of "entropy" is usually applied without explaining it. However, as is well known, there are two "theoretical frameworks" (see e.g., Frigg, 2008; Werndl & Frigg, 2017) both called statistical mechanics, that offer two different notions of entropy that are supposed to account (at least approximately) for the thermodynamic notion of entropy: one follows the work of Boltzmann and the other of Gibbs. The notion standardly used in the literature in our context is a Boltzmannian one: the entropy of a system in a given microstate is a function of the (Lebesgue) measure of the macrostate to which this microstate belongs. The Gibbs entropy is defined for systems in equilibrium, where "equilibrium" is understood as a measure that is invariant under the dynamics. On this account, "entropy" is a function of the entire phase space, that may be seen as some sort of weighted averages calculated given the appropriate measures over that space (for the appropriate ensembles). One consequence of this is that this notion of Gibbsian entropy remains constant, and cannot account for the approach to thermodynamic equilibrium. To solve this problem Gibbs introduced the idea of successive coarse graining. Unlike the Boltzmannian coarse graining into macrostates, the Gibbsian coarse graining is not associated with macrovariables, and moreover, it needs to change constantly (in our terms: constantly replacing one measuring device with another as it were) in a way that makes the graining finer and finer, and in the limit much finer than the capabilities of any measuring device. We find it difficult to see how this notion of entropy can be applied in the context of Landauer's principle. Landauer's principle concerns the evolution of the computing system from an initial macrovariable to a final one, and on our view the "translation" of this idea to Gibbsian terms leaves out the essential magnitudes of Landauer's principle and the arguments for it. Let us also remark that the Gibbsian approach is known to be conceptually very problematic; see (Callender, 1999; Ridderbos & Redhead, 1998). One explanation for its usefulness in practice is that it can be explained in terms of Botlzmann's macrostates, if the dynamics is taken into account; this idea is described in (Hemmo & Shenker, 2012, Ch. 11).

universe, for example. Nevertheless, for many thinkers this last option is not satisfactory, and they strive to prove that Maxwellian Demons are incompatible with fundamental physics. One way of doing so, in fact the most prevalent way in contemporary research, is by relying on Landauer's principle (Leff & Rex, 2003). Clearly, however, relying on the second law in establishing Landauer's principle is circular if that law itself is defended by relying on Landauer's principle. Not every circularity is vicious; we think this one is (see also Earman & Norton, 1998, 1999). In order to defend the universality of the second law, Landauer's principle should be grounded, not in the second law, but in independent arguments, such as those of fundamental physics.[37] Whether or not this can be done is precisely the subject we are about to examine in what follows.

Criticism from fundamental physics. The notion of "*physical computation*" has been studied extensively in recent philosophy of cognitive science, and the question of which physical processes can justifiably be said to "*implement*" computations[38] has been greatly clarified, although it is still under debate (see overview in Piccinini, 2017). Taking from these insights some minimal requirements that a physical process ought to satisfy in order for it to implement a computation, we describe what "*implementing a computation*" may mean in terms of *fundamental* physics, and then show how Landauer's principle might arise in this framework. This approach, which is far more reductive than the standard studies of Landauer's principle (that are couched in high level theories, as mentioned above), strengthens the status of Landauer's principle where it holds, but – at the same time – exposes the cases for which it does not hold, thus showing that it is not a universal principle of physics.[39]

Rendering Landauer's principle in terms of fundamental physics will help us bring out more strongly the connection between this principle and the multiple-computations theorem, and provide a basis for the latter in fundamental physics. We shall also see (in Section 8 below) that our approach will give physical basis for the attempts to counter the triviality result of the multiple computations theorem in its strong version by adding to the formal physical-to-computational mapping some physical conditions on the association of physical states with functional and computational states (see e.g. Godfrey-Smith, 2009; Schuetz, 2012). Even more interestingly, we shall propose a way to counter on the basis of physics the modest multiple computations theorem, but as we shall see this will require a reductive mind-brain identity theory (see Section 8).

Let us begin in the remainder of this section with clarifying the notions of "*microstate*", "*macrovariable*" and "*macrostate*" that appear in this principle and the arguments for it. We shall then address the dynamical aspects of the principle (in Section 6) and the entropic aspects (in Section 7). These notions are crucial also for the physical basis of the multiple-computations theorem (as we shall see in Section 8).

The first thing to notice is that, according to each of the fundamental theories of physics, the world[40] is at any moment[41] in a '*microstate*', that is, in a well-defined, (ideally) fully (or maximally) describable in terms of that theory (e.g. as personified by "*Laplace's Demon*"). In other contexts, the term '*microscopic*' sometimes means "small", or "part of a whole", but in our context the term '*microstate*' denotes the *complete* state of the system (according to a given theory). Although we use classical physics for illustration, we employ the notion of "microstate" in a way that is general and applicable to any fundamental theory.[42] In classical mechanics the microstate is the 3-dimensional-positions plus 3-dimensional-velocities (or momenta) of each of the *n* particles that comprise the world, so that to describe the microstate we need 6*n* numbers.[43] Accordingly we say that the system has 6*n* "*degrees of freedom*", that is, 6*n* ways of changing the microstate. The equations of motion of each theory describe the evolution of the world, that is, the sequence of microstates through which the world evolves, given its parameters and constraints; this sequence is called "*trajectory*".

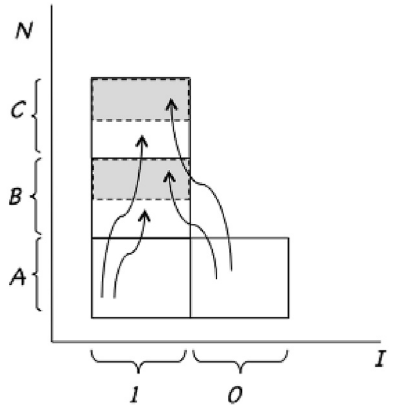

Fig. 4. State space and trajectories of a system implementing *erasure*.

Fig. 4 depicts the *state space* of system S, in which every *point* represents a microstate of S and every *axis* (or dimension) stands for one degree of freedom (The details of Fig. 4 will be explained as we proceed.). For obvious reasons, in Fig. 4 we cannot depict all the axes of the state space, and therefore we present two degrees of freedom, called I and N; The generalization to more degrees of freedom is conceptually straightforward. For example, in classical mechanics, I (and similarly N) might be the position or the velocity of some particle of S along some direction.

System S is meant to be the most general case of a physical system that implements the logical function of *erasure*: 1 → 1, 0 → 1. (Other logically irreversible operations can be seen as combining logically reversible ones plus erasures, so that this case is quite general.) What does this mean in terms of fundamental physics? Since, according to physics, the world, including system S, is at any moment in a microstate (*and that is all that there is in the world*), our first task is to explain how can the *microstate* of S at some given instant t_0 implement a given logical symbol, either 0 or 1 as the case may be. To achieve this task, we introduce the notion of "*macrovariable*".[44]

Consider a microstate x that is represented by a point somewhere in the region (1,A) in Fig. 4. Being a member of the set (1,A) of microstates of S, our microstate x shares with all the other microstates in this set the following feature. The projection of x (and of each of the other microstates in this set) onto axis I is onto the interval marked "1" in Fig. 4, in which the degree of freedom I has a certain range of values. Whenever the microstate x of system S has this feature, we say that S is in state 1 or that it implements the symbol 1; and similarly, for region (0,A) and the symbol 0. Since by

[37] We thank James Ladyman for a correspondence about this point.

[38] Or "corresponds to" or "is associated with" or "stands for" or "realizes" them, etc. We shall not address the important differences between these ideas.

[39] Our present line of thinking continues our work on Landauer's principle and Maxwell's Demon in (Hemmo & Shenker, 2010, 2011, 2012, 2013).

[40] Or the part of it to which the theory pertains.

[41] Relativistically understood.

[42] For the problem of Hempel's dilemma, see (Ney, 2008).

[43] Constraints reduce this number, but this point is not essential here.

[44] For more details on the meaning of macrovariables and macrostates in statistical mechanics see (Hemmo & Shenker, 2012, 2016; Shenker, 2017a, 2018).

looking at the I degree of freedom (without looking at the N degree of freedom) we can know which symbol is implemented by S, we call I an "*information bearing degree of freedom*" (following Landauer, 1961). As system S evolves in time, it goes through a sequence of microstates (the "trajectory"), and the projections of these microstates on the information bearing degree of freedom may change between the regions marked "0" and "1." In this way S can implement a computation (In the examples presented in Sections 3 and 4 we could say that the degree of freedom I stands for voltage and the regions "1" and "0" stand for certain values of the voltage, according to the value assignments 1 and 2.).

Importantly, while the projection of the microstates onto the regions "0" and "1" along the information bearing degree of freedom, is all we need in order to implement the corresponding symbols, from a physical perspective this projection is only an *aspect* of the microstate x, given by a *partial description* of it, since it ignores other degrees of freedom and other details concerning the microstate x. We call an aspect of a microstate, given by a partial description of it, a "*macrovariable*". For example, all the microstates within the region (1,A) in Fig. 4 share the macrovariable that we denote by "1" as well as the macrovariable we denote by "A", the former being a range of values of the *information bearing degree of freedom* I and the latter being a range of values of the *non-information bearing degree of freedom* N. While for the purpose of computation we are interested in the I degree of freedom and its partition to the "1" and "0" regions, for other purposes we may be interested in other aspects of the microstate x of system S (e.g., the velocity of the object S when it is thrown at us). "We simply think of each bit as being located in a physical system, with perhaps a great many degrees of freedom, in addition to the relevant one. However, for each possible physical state which will be interpreted as ZERO, there is a very similar possible physical state in which the physical system represents a ONE." (Landauer, 1961, p. 265, p. 265).

It is important (for our present argument) to notice that *macrovariables have a dual nature*. On the one hand, they pertain to *individual* microstates, as we have just explained. On the other hand, they pertain to *sets* of microstates. To realize this duality, consider the following. The *actual* microstate x of S at each point of time has a variety of macrovariable, i.e. a variety of aspects, given by various partial descriptions of it, and in this sense *all* of them "exist" when that microstate x obtains. When we carry out a particular *measurement* of S, our measuring device is sensitive to a *particular* macrovariable of the microstate x (and not to other macrovariables of x), and reveals the value of that macrovariable (and not others) of the *actual* microstate of the system that obtains at the time of measurement (since, by assumption, that microstate is all there is in the world at that moment).[45] However, since a macrovariable is given by a *partial* description of the microstate, it gives rise to an *equivalence set* of microstates, consisting of all those that share the same macrovariable; this set is often called *macrostate*.[46] At each moment one of the microstates (at most) of the macrostate set is *actual* and the rest are *counterfactual*, but in all of them the system would appear to us the same, whenever our measuring device is only sensitive to the shared macrovariable.[47] Measurements and their connection to macrovariables, in both senses, will be central to our argument later.

(The reader may ask herself at this point the following question, that we shall address in Section 8 below. In the case illustrated by Fig. 4, the values "0" and "1" are assigned to certain regions of the degree of freedom I. Since nothing in Fig. 4 tells us what that degree of freedom is and what these regions are, they could be anything: certain values of voltage or other values of voltage, certain positions of certain particles or other positions of other particles, and so on. Any macrovariable, that is, any aspect of the microstate x of S at *t*, given by any partial description of this microstate, could be assigned the values 0 and 1 (as long as it can have two sufficiently well-defined regions of values). However, when we actually use a system as a computer we in general know which macrovariables to focus our attention on, in order to read out the information. The question arises: what fact fixes the assignment of values 0 and 1 to certain values of certain macrovariables, and not others? We discuss this important issue below, in Section 8.).

Finally, let us also make a comment that will make the connection between our approach and the literature concerning the multiple computations theorem more transparent. As we mentioned earlier, to avoid the triviality result of the multiple computations theorem (in its strong reading) according to which almost every microphysical evolution implements almost any computation, some authors have proposed to add constraints on the value assignment (the physical-to-computational mapping), in particular on what counts as the right physical states for implementing a computation and the way physical states are joined together by the mapping. For example, Godfrey-Smith (2009, p. 292) proposed that the realization of a function should involve not just a mapping between physical and formal states, but that "microstates grouped into coarse-grained categories be physically similar.". And similarly, Schuetz (2012, p. 104) argued that: "If physical states are not given, CVI/SVIs run into insurmountable difficulties: following the construction of the Slicing Theorems, physical states supporting counterfactuals can be defined, for which the system implements almost any computation.". It seems to us that these authors propose that certain physical conditions need be satisfied in order for one to say that a certain computation (or a function) is implemented. Perhaps this approach might be workable to avoid the triviality result of the strong version of the multiple computations theorem. If so, as we shall see, our approach provides a naturalized physical basis for these proposals in terms of macrovariables. The idea is that a given kind of computation (or a function) may be implemented by a microphysical process if and only if the implementing physical system evolves through a *certain sequence of macrovariables*.

However, although as we shall show below, such conditions may be formulated in terms of physics, it is unclear what would be the *physical* justifications of these conditions. Why is it that only certain macrovariables (physical conditions) may implement a computation, or even a certain kind of computation, and not others? We shall see that even if one just had a list of macrovariables that may be said to implement a given kind of computation, this will presumably *not* be enough to avoid the *modest* multiple computations theorem and defend Landauer's principle: The two value assignments associated with different ranges of voltage in the computations of Section 4 are equally based on the *same* kind of physical macrovariables (that would presumably appear in our list), and so in this sense they seem to have equal fit for realizing computations. What one needs, in addition, is a criterion for *selecting* as physically preferred the macrovariables that give rise to *only one of the two computations* (and more generally a physical criterion for selecting only one computation for a given microphysical evolution). We

[45] This is a useful idealization. For a discussion of the idea that measurements take time, see (Hemmo & Shenker, 2012, Ch. 11; Shenker, 2017a).

[46] This is a generalization of the famous partition of the Gamma space on the basis of coarse graining of the μ space, described by Ehrenfest and Ehrenfest (1912) in explaining Boltzmann's view: that partitioning is an example of partitioning the state space into macrostates on the basis of macrovariables. See detailed account in (Hemmo & Shenker, 2012, Section 5.6).

[47] In statistical mechanics this set gives rise to the notions of probability and entropy, see (Hemmo & Shenker, 2012, 2016; Shenker, 2017a,b).

shall briefly examine some proposals in Section 8, and see their limitations, and then propose a criterion that results in *a selection of macrovariables*, which we think is unavoidable in a physicalist approach.

6. Landauer's principle in fundamental physics: the dynamics

Computers are built to handle *all* possible inputs: they are expected to carry out the computation correctly for *any* input. This point has been strongly emphasized in the context of both theses discussed in this paper. In the context of the multiple-computations theorem:

> "The conditionals involved in the definition of implementation … have modal force, and in particular are required to support counterfactuals: *if* the system were to be in state p, *then* it would transit to state q. This expresses the requirement that the connection between connected states must be reliable and lawful, and not simply a matter of happenstance." (Chalmers, 1996, p. 312-3).

In the context of Landauer's thesis:

> "[A] computer pushes information around in a manner that is independent of the exact data which are being handled, and is only a function of the physical circuit connections." (Landauer, 1961 p. 262 p. 262)

This important requirement – which we accept – is seen by many to lead almost directly to Landauer's principle (e.g., Ladyman, 2009). This, however, is not the case, and our next task is to prove this. But let us first show how to account, *in terms of fundamental physics*, for the following two statements:

1. **The particular case**: As system S evolves from its particular initial microstate at a particular time t_0 to its particular final microstate at time t_1, it implements (a token of) a given computation. For example, if the computation is of a logical function such as the ones presented in Sections 3 and 4, each individual "run" of the computation implements *one row* of the Tables only.

2. **The set of counterfactual cases**: Had system S evolved from a different initial microstate (or: if it does evolve from a different initial microstate at some other time t_2) to its final microstate, it would also implement (a token of) the same computation. Here, for example, all the rows in the Tables in Sections 3 and 4 are considered.

(In Sections 7 and 8 we discuss the physics of associating a macrovariable with a symbol, that is, the *physics* of value assignment.)

This duality may remind the reader of the duality, pointed out above, of the notion of *macrovariable*. And here lies the key for achieving this desideratum. In terms of the physics of our computing system, the equations of motion that govern the evolution of S over time must describe its evolution for *all* possible initial microstates. In our example of Fig. 4, since both values 1 and 0 along the I degree of freedom are possible initial macrovariables, *all* the microstates in both sets (1,A) and (0,A) are *possible initial microstates*. (We assume here that along the N axis only A is possible; B and C are not possible input states in this system; those are considered later.) And so, if S is to implement logical erasure (1 → 1, 0 → 1) then the equation of motion must describe what happens to *all* of those microstates: *all* of the trajectory segments, that start out at the input time t_0 in either the region (1,A) (implementing the input 1) or the region (0,A) (implementing the input 0), must end up at the output time t_1 in microstates that implement the symbol 1. *Prima facie* one might think that this means that all the final microstates should be in (1,A). However, such an evolution is impossible, due to a *theorem of mechanics* (called *Liouville's theorem*) that places a certain limitation on possible mechanical evolutions. The limitation is that the total volume (by Lebesgue measure) of the final set of microstates cannot be smaller than the total volume (by Lebesgue measure) of the initial set of microstates; but the region (1,A) is smaller (by Lebesgue measure) than region (1,A) plus region (0,A). Since the transformation from region (1 + 0,A) to region (1,A) violates Liouville's theorem of classical mechanics, this is an *impossible evolution* according to this theory.[48]

In order to implement logical erasure in a way that satisfies Liouville's theorem, we need to distinguish between the information bearing degrees of freedom I and the non-information bearing degrees of freedom N. Suppose, for example, that the equations of motion bring about the transformation (1 + 0,A)→(1,B + C) (in terms of Fig. 4), then:

(i) Looking at the projection of the mapping on the information bearing degrees of freedom I, the transformation is from region "1 + 0" to region "1", implementing the computation 1 → 1, 0 → 1.
(ii) If one takes into account all the degrees of freedom, in both I and N, then Liouville's theorem is satisfied.

There are a number of dynamical rules (i.e. equations of motion) that satisfy both requirements (i) and (ii). In the simplest case (call it *Dynamics 1*) trajectory segments that start out in microstates in either (1,A) or (0,A) end up in microstates *anywhere* in the region (1,B + C). In Section 7 we describe other examples of dynamical rules that satisfy requirements (i) and (ii), and have interesting implications for Landauer's principle; but for now, the simple case of Dynamics 1 suffices.

Remark. It may be (as Landauer, 1961 noticed) that the Lebesgue measures of the "1" and "0" projections on I (or those of regions (1,A) and (0,A)), are not the same. Everything we say applies to this generalized case as well.

Remark. Above (in Section 5) we mentioned that the prevalent way of thinking about Landauer's principle is not in terms of fundamental physics, but rather in terms of thermodynamics and statistical mechanics; not in terms of Liouville' theorem but in terms of the second law of thermodynamics. Landauer (1961) wrote that "the entropy of a closed system, e.g. a computer with its own batteries, cannot decrease; hence this entropy must appear elsewhere as a heating effect, supplying 0.6931 kT per restored bit to the surroundings." (Landauer (1961). p. 265). It might seem, prima facie, that the ideas are essentially the same, if entropy is associated with the Lebesgue measure of the macrostate, as is usual in Boltzmannian statistical mechanics. However, the difference between the two ways of thinking is crucial, as will become clear as we proceed, in Section 7.

7. Landauer's principle in fundamental physics: the entropy

Now that we have the *dynamics* of implementing logical erasure in place, we can turn to see what is the *entropic* behavior during such an implementation. Here, let us begin with a very important distinction that Bennett (2003) introduced to the literature with respect to Landauer's principle:

[48] Obviously, other physical theories may impose other constraints.

> "If a logically irreversible operation like erasure is applied to *random data*, the operation still may be thermodynamically reversible …. But if, as is more usual in computing, the logically irreversible operation is applied to *known data*, the operation is thermodynamically irreversible …. " (Bennett, 2003, p. 502, our italics)

The thermodynamic irreversibility alluded to here is the one stemming from Landauer's principle; and for Bennett, the argument for this principle is grounded in the second law of thermodynamics (or its statistical mechanical counterpart).[49] However, Bennett's distinction between *random data* and *known data* is meaningful and significant also in the context of fundamental physics that we address here. The exact nature of this distinction is a bit subtle, and this subtlety is the same whether one works in Bennett's theoretical framework or in ours. Let us explain this distinction in two stages:

1) We shall start (in the present section 7) by explaining what "*known data*" and "*random data*" mean by understanding them as pertaining to two possible *measuring devices*. Our main point agrees with Ladyman's (2009 p. 382) remark that "In practice of course it is only possible to use a system as a computer if the relevant physical states are distinguishable by us, with our measurement devices; and it is possible for us to put the system into a chosen initial state so as to compute the function in question for it".
2) In Section 8 we shall examine what these "*measuring devices*" might mean in terms of fundamental physics, and we will see how important the notion of "measuring device" is for understanding the physics of "value assignment" (or "implementation" or "individuation" of computation, mentioned in Sections 3 and 4 above). In this section we shall talk about "measuring devices" and avoid terms like "observers", in order to avoid reference to notions like "subjectivity" and "agency" that might come up if the latter are used; compare Searle, 1992, pp. 208-9; but we shall come back to "observers" in Section 8.

Suppose that the state of system S is measured by some measuring device that measures whether the input is 0 or 1. (If S is an element in a computer then the measuring device may be some other element in the computer that uses S's state as its own input; or it may be something external to the computer. We address the computational theory of mind in Section 8.) In order for the measuring device to be suitable for its task, its physical interaction with the computing system S must be such that the device is *sensitive* to the value of the degree of freedom I of the microstate x of system S.[50] Specifically, if the microstate x of S is in the region "1" (of degree of freedom I) then the pointer of the measuring device should point at the symbol "1" (say, engraved on its plate), and if the microstate x of S is in the region "0" along I then the pointer of the measuring device should point at the symbol "0". Since the state of the pointer is correlated with the microstate of S being in *either* the region (1,A) *or* the region (0,A) (exclusive or), we say (metaphorically!) that (after the measurement has been completed) the measuring device "*knows*" the input of the computation. This is the case of "*known data*" in Bennett's (2003) terms. (We stress that this "knowledge" is merely the state of the pointer of a measuring device, no agency is involved. We return to this point in Section 8.) Accordingly, alluding to Bennett's (2003) terminology, we call this device "*known data measuring device*".

We could also have another measuring device, which is *not* sensitive to the distinction between the ranges "0" and "1" of axis I: the pointer of this insensitive device points in the same way regardless of whether the microstate x of S is in region "0" or "1" along I. The only thing that one can read from the pointer of this device, once the measurement is completed, is that the microstate x of S is somewhere in the region (1 + 0,A). We shall say (metaphorically!) that this device "*does not know*" the input; and again, alluding to Bennett's (2003) terminology of "*random data*", we shall call this device "*random data measuring device*".

(Remark. Sometimes the term "random data" implies, in addition, that the unknown values "0" and "1" have equal probabilities (in some appropriate sense of this term). Our discussion can be generalized for the case of unequal probabilities, as well as to the case in which (1,A) and (0,A) differ in their Lebesgue measures. We do not discuss this issue here as it does not contribute to our main point.)[51]

(Notice that these two measuring devices could be said to pertain to different "levels of organization", unlike the two computations in the examples in Section 3 and 4 which pertain to the same "level of organization". We submit that this difference is not important for the assessment of Landauer's principle, nor to other implications of the multiple computations theorem, due to the fundamental-physical understanding of these theses that we propose here.)

Bennett (2003) writes that in the case of "known data" the implementation of logical erasure is necessarily dissipative, and in the case of "random data" it is not. What can this mean? While Bennett (2003) takes Landauer's principle to be grounded in the second law of thermodynamics and not in Liouville's theorem, with respect to the distinction between these two cases there is an important similarity between the two ways of thinking, which is this.

Suppose that the equation of motion that governs the evolution of S is *Dynamics 1* (in which – recall – trajectory segments that start out in microstates in either (1,A) or (0,A) end up in microstates *anywhere* in the region (1,B + C)), and assume further that both of our measuring devices (both the "known data device" and the "random data device") are *physically insensitive* to the difference between regions B and C along the non-information bearing degrees of freedom N.[52] Thus, for *both* devices the region that contains the *final* microstate of S is (1,B + C). For the "known data device", the transformation is (1,A)→(1,B + C), (0,A)→(1,B + C); and for the "random data device" the transformation is (1 + 0,A)→(1,B + C).

[49] Bennett (2003, p. 502) continues: "… because the environmental entropy increase is not compensated by any decrease of entropy of the data. This wasteful situation, in which an operation that could have reduced the data's entropy is applied to data whose entropy is already zero, is analogous to the irreversibility that occurs when a gas is allowed to expand freely, without doing any work, then isothermally compressed back to its original volume." As we said, grounding the proof of Landauer's principle in the second law is circular, since the second law is defended against the counter example of Maxwell's Demon by relying on Landauer's principle.

[50] Recall that at the end of Section 4 we said that the questions: which degrees of freedom are "information bearing; " and what is the fact that makes it the case that (in a certain context) a certain degree of freedom is "information bearing," are highly non-trivial. We set aside now these questions; and will address their far-reaching consequences in Section 8.

[51] In general, the notions of "entropy" and "probability" are not only conceptually different, but may not even coincide quantitatively. Probability is determined by the relation between the partition of the phase space to macrostates and the dynamics which determines the evolution of the bundle of trajectories. See (Hemmo & Shenker, 2012, 2016).

[52] The random data user, that is insensitive to the distinction between "1" and "0", can nevertheless infer that the output is "1" from knowing that the non-information bearing state is B or C and not A.

To see how these cases fit Landauer's principle we need to look at their entropy and at their logical reversibility.

Entropy is associated with *k*logW where W is the volume by Lebesgue measure of the macrostate of the system.[53] Assume (as is usual, and for simplicity) that each of the regions (1,A), (0,A), (1,B), and (1,C) has volume (by Lebesgue measure) that we shall denote "1 rectangle". Thus, for the "known data device" the entropy increases by *k*log2 (from *k*log (1 rectangle) to *k*log (2 rectangles)); and for the "random data device" the entropy remains constant (*k*log (2 rectangles)) throughout the computation.

For the known data device, the evolution is logically irreversible, since from the output (1,B + C) one cannot infer whether the input was (1,A) or (0,A); but for the random data device the evolution is logically reversible, since from the output (1,B + C) one can infer that the input was (1 + 0,A).

Notice that in both cases the mapping along the information bearing degrees of freedom I is the logically irreversible erasure 1 → 1, 0 → 1, as an artifact of ignoring the non-information bearing degrees of freedom N. Ignoring N is useful for pragmatic purposes, of using S as a computer; but since the entropy of S is determined by the volume taking into account all the degrees of freedom, in considering logical reversibility in the context of Landauer's principle we need to consider all of them (Recall that we stressed a similar point in Section 4.).

The result is that Dynamics 1 is in line with Landauer's principle for both kinds of measuring devices: In the "known data" case one bit of information is lost, and the minimum dissipation is *k*log2; In the "random data" case no information is lost, and there is no minimum dissipation.

However, the case of *Dynamics 1* is not the only possible one for implementing logical erasure, and the "known data" and "random data" are not the only relevant measuring devices. The following are also cases of implementing logical erasure while satisfying Liouville's theorem:

Dynamics 2: (1,A)→(1,B), (0,A)→(1,C).

Dynamics 3: (0,A)→{(1,Btop)+(1,Ctop)}, (1,A)→{(1,Bbottom)+(1,Cbottom)}, see colored regions and trajectory segments in Fig. 4.

The following are also cases of measuring devices that can be of either "known data" or "random data" with respect to the distinction between "1" and "0":

"*Known BC data measuring device*" and "*random BC data measuring device*" are (correspondingly) sensitive (or not) to the difference between the regions B and C along the N degrees of freedom, in exactly the same sense that the "known 1,0 data" and "random 1,0 data" – considered above – are sensitive (or not) to the difference between the regions "0" and "1" along the I degrees of freedom. The argument so far focused (implicitly) on "random BC data measuring devices"; we shall describe here the "known BC data" ones. With respect to the I degrees of freedom, we shall only consider the case of "*known 1,0 data*" (the reader can complete the picture by examining the case of "random 1,0 data" by herself).

Dynamics 2 is in line with Landauer's principle for both measuring devices, and the case of Dynamics 3 is in line with this principle for the "random BC data" device. But the combination of the "known BC data measuring device" and Dynamics 3 is not in line with Landauer's principle. The entropy is *k*log (1 rectangle) throughout the process, since the input (being a case of "known data" concerning "0" and "1" in the sense of Bennett, 2003) is either (1,A) or (0,A) and the output (being, again a case of "known data" concerning "B" and "C") is either (1,B) or (1,C). Nevertheless, the process is *genuinely* logically irreversible, since given Dynamics 3 knowing that the output is either (1,B) or (1,C) is not sufficient to entail whether the input was (1,A) or (0,A). This logical irreversibility is due to the "*blending*" of the trajectories of Dynamics 3 as seen by the "known BC data" measuring device (for more on "blending" see Landauer, 1992, 1996; Hemmo & Shenker, 2012, 2013).

This case is a counter example for Landauer's principle.[54]

7.1. Landauer's principle: empirical generalization or theorem?

There are a number of possible objections to the above conclusion, and to the argument that leads to it. One of them is this. It is sometimes said that the vast empirical support of the second law of thermodynamics entails the falsity of arguments that challenge the universal validity of this law; in other words, the idea is that if the conclusion of an argument challenges this universal validity, then so much the worse for its argument: this is a sign that something is wrong with it. In particular, in our context, if the "known BC data measuring device" entails that Landauer's principle is not universally true, and if this entails that Maxwell's Demon is possible, threatening the universal truth of the second law, then – on this line of thinking – this is a good enough reason to say that the "known BC data measuring device" is not acceptable.

We beg to differ, on logical grounds and on philosophical grounds.

While we do not doubt that the second law of thermodynamics (and some of its statistical mechanical counterparts) enjoys enormous empirical support, we do stress that its proof from fundamental physics should be non-circular. This is a crucial point of logic and of philosophy. Therefore, we insist that the only way to save Landauer's principle from the counter example based on the "known BC data measuring device" is to prove a *no-go theorem*, that precludes this hypothetical device as impossible, and that is derived from *fundamental* physics. We are not aware that such a no-go theorem of fundamental physics exists. (We doubt that this is possible, given our analysis of the concept of macrovariable, here and elsewhere.)

In order to consider this possible objection, we need to give some thought to the notion of a "*possible measuring device*". Recall that we introduced this idea in order to clarify Bennett's (2003) distinction between "known data" and "random data". This investigation will take us back to the multiple-computations theorem and its physical foundations. We now turn to this.

8. Measuring devices, value assignments, and the physics of the multiple-computations theorem

The challenges to Landauer's principle (in Sections 4 and 7 above) boil down to two questions, and we shall now see that they are essentially the same question, and try to answer it.

Section 4 **question**: Can one **value assignment** be preferred, e.g. as the "actual" one? If so, what sort of fact fixes this preference? In

[53] This is entropy in a Boltzmannian framework, which is the one used in this context. See footnote above concerning Gibbsian entropy. In a Gibbsian framework it is not trivial to account for the entropy of Dynamics 1, but on our view given an acceptable account the result is the same. We do not go into this here.

[54] Notice that while this counter example is of entropy *conserving* genuine erasure, one can come up with counter examples in which the erasure will be even entropy *decreasing*. These stronger counter examples involve partitioning the phase space in a different way (that expresses the resolution power of other measuring devices) and possibly also different dynamical rules. This has been shown in (Hemmo & Shenker, 2010, 2011, 2012, 2013). We do not expand here on the implications of this idea for Maxwell's Demon and the second law of thermodynamics; these topics exceed the framework of this paper.

other terms, sometimes used in the literature, what fact fixes the "individuation of computation" implemented by a given physical system during a given process? (If no fact fixes one value assignment as *preferred*, then all assignments equally coexist, including those that give rise to logically reversible computation and those that give rise to logically irreversible ones, and then Landauer's principle makes contradictory predictions.)

Section 7 **question**: Can certain hypothetical **measuring devices** be ruled out, e.g. as "impossible" ones or (conversely) be "preferred"? If so, what sort of fact makes a hypothetical measuring device actual, or possible, or impossible? (If no fact makes certain measuring devices impossible then all of them are possible, including the "known BC data measuring device" that is associated with a dissipationless logically irreversible computation that Landauer's principle deems impossible.) Let us begin by showing the connection between value assignment (or individuation of computation) and measuring devices.

What is "value assignment" with respect to a physical system that implements a computation? What sort of *fact* "value assignment" is? Let's start with Searle's (1992) proposal. For Searle, the fact of value assignment is a fact about an "agent" or "observer", that assigns the values to the states of the observed system. He writes: "the ascription of syntactical properties is always relative to an agent or an observer who treats certain physical phenomena as syntactical" (Searle's (1992)., p. 208). On this view, to understand what sort of fact "value assignment" is, we need to understand what sort of facts "agent" or "observer" are.

The problem is that notions like "agent" or "observer" are vague, and clarifying them for our present purpose (of addressing Landauer's principle and its connection to the multiple computations theorem, in terms of fundamental physics) may require that we offer a naturalistic-physicalist solution to the mind-body problem, a task that is far beyond the scope of this paper. Therefore, if possible, we would like to treat Landauer's principle in a way that does not depend on one's views concerning the mind-body problem. How, then, should we go about explaining "value assignment"?

Our first attempt at accounting for the notion of "value assignment" while sidestepping the mind-body problem is by using the notion of a "*measuring device*" to replace (Searle's) "agent" or "observer". The idea is to understand the phrase "*value assignment w* for system S" as the phrase "*measuring device w* that measures the state of system S".[55] Let us explain this idea, and then see whether our attempt (at explaining "value assignment" while bypassing the mind-body problem) is successful. We shall shortly see its limits.

(**Remark**. Notice that prima facie, this idea may not work for the computational theory of mind, since the "agent" or "observer" is taken to *observe* the computing system S, rather than to *be* system S.[56] But we will come a bit closer to the theory of mind later on.)

What is a measuring device, in this context? Very roughly, we take a measuring device to be a physical system, that interacts with the system of interest S, such that the final physical state of the "*pointer*" element of the measuring device reflects the physical state of the measured system S as it was at the interaction time.[57] For example, consider a measuring device in which the pointer ends up pointing at either the symbol "1" or the symbol "0" engraved on its plate, following an interaction with system S. We *call* our measuring device "value assignment 1" if it interacts with system S in such a way that if the state of S is X, then the pointer ends up pointing at "1", and if the state of S is either Y or Z then the pointer points at "0". We *call* the device "value assignment 2" if the interaction is slightly different, so that the pointer points at "1" if S is in either X or Y, and points at "0" if S is in Z. Thus, to say that we assign the value 0 or 1 to the state of S according to value assignment w, means to say that we describe the state of the measuring device called "value assignment w" following its interaction with S, rather than the state of S directly.

On this way of thinking about value assignment it seems that we have a physical criterion for preferring one **value assignment** for the states of a computing system S, e.g. as the "actual" one: the preference is fixed by the **measuring device** that is *actually at work* measuring the states of S.

Here the next question arises: What makes it the case that, at a given occasion, measuring device 1 rather than 2 (say) is *actually at work*, so that computation 1 rather than 2 (say) is actually implemented? In the context of physics (and more particularly in the context of providing the physical foundations of the multiple computations theorem and Landauer's principle), it is preferable to avoid (if possible) talk about "choice" of a measuring device by some "agent", since otherwise we would need to provide a physicalist account of these terms, a task that is far beyond our present scope. Our attempt at avoiding the need to "choose" between two optional measuring devices is to endorse a framework in which there is only one possible measuring device, and our way to do this is to take the entire relevant *actual environment* of system S (possibly its entire unlimited environment) as the measuring device: In this way there is no distinction between the measuring device and a possible "user" that may "choose" it: the "user" is part of the physical environment as well. Thus, the *actual state of affairs* that happens to obtain in the universe, including the way in which the environment interacts with S, fixes the computation that is actually implemented by S, rendering the other computations merely *counterfactual*.[58]

This line of thinking has been suggested in the philosophy of cognitive science: it is a form of "*externalism about computation*," the view that value assignment (computational individuation) is based on features that are external to the system (see Shagrir, 2018 Section 5 for a critical discussion of this idea). Externalism about computation implies that if a given physical system S, undergoing a given micro-physical process, is transferred to a different environment, it can change its computational identity.[59] In our terminology this means that if S is measured by a different measuring device, it is given a different syntactic value assignment, and accordingly implements a different computation. This idea is under debate in contemporary philosophy of cognitive science (Shagrir, 2018.[60]), and we would now like to address it from a perspective that is relevant for our interest.

[55] Ladyman (2009 p. 382) writes: "In practice of course it is only possible to use a system as a computer if the relevant physical states are distinguishable by us, with our measurement devices; and it is possible for us to put the system into a chosen initial state so as to compute the function in question for it".

[56] In the context of the semantic view concerning the individuation of computation, Shagrir (2018) writes: "Presumably, the content of the computations that take place in our brain is not defined by the interpretation of an external observer".

[57] We do not address the stability requirement involved here.

[58] We assume for simplicity that the state of affairs is stable in a way that enables us to speak of the same value assignment obtaining for the entire duration of the computation. Relaxing this assumption adds complications that do not contribute to understanding the heart of the matter.

[59] For example, according to Piccinini (2008) the computational identity changes if this transfer involves a change of a functional task.

[60] Shagrir's (2001) main aim is to support the claim that content has a role in individuating cognitive computations. We do not address this topic here.

Suppose that we have a good measuring device E (for "environment"), which is of the "value assignment 1" kind. This means that if system S happens to be in a microstate x in which the value of the particular macrovariable, to which the device E is sensitive, is X, then the pointer of E ends up pointing at the symbol "0" engraved on its place (for short: "*E ends up in state 0*"). (And similarly, for values Y or Z of S and engraved symbol "1".)

But what do we mean by saying that "E is in state 0"?

If we want "E is in state 0" to have a *physical* meaning, then it has to be the same kind of physical meaning as that of "S is in state X". We explained the latter above, recall: the microstate x of S has infinitely many macrovariables (infinitely many aspects, given by infinitely many possible partial descriptions), and the way to prefer one of them as obtaining (and thus the corresponding computation as implemented) is by noticing that the environment E is sensitive to that macrovariable, and reflects its value. For instance, E's pointer state responds to S's state in the way we call "value assignment 1" and not "value assignment 2".

Now, the environment E is also a physical system, and as such it is (at any moment) in some microstate y of its own, which *also* has infinitely many macrovariables (infinitely many aspects, given by the infinitely many possible partial descriptions). And just like system S, the way to prefer one of the macrovariable of system E is by noticing that *its* environment acts as a measuring device that is sensitive to one of *its* macrovariables, and not to the others; call that extra-environment E'. In our example, the environment E′ should prefer the macrovariable of pointing at "0" or "1", over other options such as more fine-grained macrovariables (e.g. pointing at the leftmost part of the region marked "0") or more coarse-grained ones (e.g. pointing at "0 or 1") or qualitatively different ones (e.g. having temperature T). And so, in order to say (in a physically meaningful way) that E is in a state in which its pointer points at the engraved symbol "0" we need E to be measured by an extra-environment E′, that is sensitive to precisely that macrovariable. For example, we would *say* that E is in state "0_1" (or that we assign the value "0_1" to E) just in case the extra environment E′ is in state "0_1", and we would assign the value "1_1" to E just in case the extra environment E′ is in state "1_1"; and so on, *ad infinitum* (See Shagrir, 2018; Section 5 for an example in the same spirit.).

As long as this infinite regress is not halted within the framework of fundamental physics, externalism about computation cannot solve the multiple-computations problem, and hence cannot solve the challenges for Landauer's principle in Section 4 and Section 7.

And so, the question is, finally: Can this infinite regress be stopped using the resources of fundamental physics? To do so one would need to show that some level in the regress has a special status, due to which no further regress is needed.

One option here might be seeing S itself as the extra-environment of E, so that each of them is the environment of the other. However, the reader can easily see for herself that in this case it is not clear which system implements the computation and what is the value assignment, so the very claim that computation is being implemented becomes unclear.

The other option is that the particular macrovariable of E (in which its pointer points at "0" or "1" engraved on its plate) has some *physically-describable internal feature* in virtue of which we can justifiably say that no further regress is needed. (It could be the macrovariable of the extra-environment E′ or any other one along the line, as long as the regress is halted somewhere.) What might this be? Some ideas in the philosophy of mind can be understood along these lines.

One proposal is that the preferred macrovariable (but not the others) is associated with a (preferred) *internal* semantic contents.[61] For example, with respect to the simultaneous implementation of AND and OR in Section 3, if it happens to be a fact that the state X of system L has as its content the number 0 and states Y and Z have as their content the number 1, then system L falls under the computational kind AND. We are not sure how to understand "content" in physical terms, and hence find it difficult to incorporate this idea into our analysis. Indeed, the semantic view is not committed to naturalism. Notice that if the notion of contents is understood in an *externalist* way, then a version of the above infinite regress problem recurs. (For a defense of the semantic view, see Shagrir, 2001, 2012, 2018). Another proposal is to revert to "functional tasks" as fixing the value assignment (e.g., Piccinini, 2008, 2015), possibly by being co-extensive with, or even explanatory of, "semantic tasks". We are not sure how to understand "function" or "task" or "functional task" in physical terms, and whether it can be naturalized, and hence find it difficult to incorporate this idea, too, into our analysis. We don't expand on this point here.

On our view, the only fully physical solution, i.e., a solution that can (in principle) be fully described in physical terms, is to say that the preferred macrovariable is (identical with) "mental states" of (elements of) E, in which case the values "0" and "1" are (identical with) experiences of an observer, so that S carries out the corresponding computation only relative to that observer. Some observers may be "value assignment 1 observers", and others may be "value assignment 2 observers"; which is which is determined, entirely, by the physics of E and its interaction with S. This – the reader may recall from the beginning of this section – takes us back to Searle's (1992) brief remark, but we do not endorse Searle's views on the mind-body problem (see Searle, 2002). The view we put forward here, as a solution to the multiple-computations problem, as well as a potential way to face the challenges to Landauer's principle, is in the framework of a reductive physicalist identity theory of mind. We do not argue for this view here (see Shenker, 2017c): our point is that it turns out that (a) the multiple-computations theorem in cognitive science has a sound basis in fundamental physics; and (b) the challenge to Landauer's principle in physics is strongly connected to the question of whether the mental is physical; (c) The way to meet both challenges is by adopting an identity theory of mind and brain.

Funding

This research was supported by the Israel Science Foundation (ISF), grant number: 1148/18 and by a grant from Lockheed-Martin.

Acknowledgements

We thank: Eli Drezner, Nir Fresco, James Ladyman, and Oron Shagrir for valuable comments on earlier drafts of this paper; Erez Firt, Violetta Rosemberg, Aviv Spector and Gal Vishne from our ISF research group, for discussions; and an anonymous reviewer, for raising helpful points and questions. We thank the Israel Science Foundation and Lockheed-Martin for supporting this research.

[61] In the literature in cognitive science on this topic one does not encounter explicitly the notion of "macrovariables" that we propose here, but the salient features of this notion are present, suggesting that indeed this is the right way to think of the physical underpinning of multiple computations. For example, the very idea that the contents determine grouping of states (e.g. creating the two different value assignments in Section 3 and 4 above) is essentially the idea that the groupings are of (sets of micro-)states that share the same macrovariable. A 3 element configuration.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.shpsb.2019.07.001.

References

Ben-Menahem, Y. (2018). *Causation in science*. Princeton University Press.

Bennett, C. (1973). Logical reversibility of computation. *IBM Journal of Research and Development, 17*, 525–532.

Bennett, C. (1982). The thermodynamics of computation: A review. *International Journal of Theoretical Physics, 21*, 905–940.

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

Bub, J. (2001). "Maxwell's Demon and the thermodynamics of computation. *Studies in History and Philosophy of Modern Physics, 32*, 569–579.

Callender, C. (1999). Reducing thermodynamics to statistical mechanics: The case of entropy. *Journal of Philosophy, XCVI*, 348–373.

Chalmers, D. (1996). Does a rock implement every finite-state automaton? *Synthese, 108*, 309–333.

Chalmers, D. (2011). A computational foundation for the study of cognition. *Journal of Cognitive Science, 12*, 323–357.

Chalmers, D. (2012). The varieties of computation: A reply. *Journal of Cognitive Science, 13*, 211–248.

Chida, K., Desai, S., Nishiguchi, K., & Fujiwara, A. (2017). Power generator driven by Maxwell's demon. *Nature Communications, 8*, 15310.

Chrisley, R. L. (1994). "Why everything doesn't realize every computation. *Minds and Machines, 4*, 403–420.

Coelho Mollo, D. (2017). Functional individuation, mechanistic implementation: The proper way of seeing the mechanistic view of concrete computation. *Synthese*. https://doi.org/10.1007/s11229-017-1380-5.

Copeland, B. J. (1996). What is computation? *Synthese, 108*, 335–359.

Cottet, Nathanaël, Jezouin, Sébastien, Bretheau, Landry, Campagne-Ibarcq, Philippe, Ficheux, Quentin, Anders, Janet, Auffèves, Alexia, Azouit, Rémi, Rouchon, Pierre, & Huard, Benjamin (2017). *Observing a quantum Maxwell demon at work*. PNAS. https://doi.org/10.1073/pnas.1704827114.

Dewhurst, J. (2018). Individuation without representation. *The British Journal for the Philosophy of Science, 69*(1), 103–116.

Earman, J., & Norton, J. (1998). Exorcist XIV: The wrath of Maxwell's demon. Part I. From Maxwell to Szilard. *Studies in History and Philosophy of Modern Physics, 29*(4), 435–471.

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

Egan, F. (2012). Metaphysics and computational cognitive science: Let's not let the tail wag the dog. *Journal of Cognitive Science, 13*, 39–49.

Egan, F. (2017). Function-theoretic explanation and the search for neural mechanisms. In D. M. Kaplan (Ed.), *Explanation and integration in mind and brain science* (pp. 145–163). Oxford University Press.

Ehrenfest, P., & Ehrenfest, T. (1912). *The conceptual foundations of the statistical approach in mechanics* (p. 1990). New York: Dover.

Feynman, R. (1996). In J. G. Hey, & W. Allen (Eds.), *Feynman lectures on computation*. Reading, MA: Addison-Wesley.

Fredkin, E., & Toffoli, T. (1982). Conservative logic. *International Journal of Theoretical Physics, 21*, 219–253.

Frigg, R. (2008). A field guide to recent work on the foundations of statistical mechanics,. In D. Rickles (Ed.), *The Ashgate Companion to contemporary Philosophy of physics* (pp. 99–196). London: Ashgate.

Godfrey-Smith, P. (2009). Triviality arguments against functionalism. *Philosophical Studies, 145*, 273–295.

Hemmo, M., & Shenker, O. (2010). Maxwell's demon. *The Journal of Philosophy, 107*(8), 389–411.

Hemmo, M., & Shenker, O. (2011). Szilard's perpetuum mobile. *Philosophy of Science, 78*(2), 264–283.

Hemmo, M., & Shenker, O. (2012). *The road to Maxwell's Demon*. Cambridge: Cambridge University Press.

Hemmo, M., & Shenker, O. (2013). Entropy and computation: The Landauer-Bennett thesis reexamined. *Entropy, 15*, 3387–3401.

Hemmo, M., & Shenker, O. (2016). *Maxwell's Demon*. Oxford Handbooks Online. http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199935314.001.0001/oxfordhb-9780199935314-e-63.

Hemmo, M., & Shenker, O. (2017). A quantum mechanical Maxwellian Demon. In B. Loewer, B. Weslake, & A. Winsberg (Eds.), *Time's arrow and the origin of the universe: Reflections on time and chance: Essays in honor of David Albert's work*. Cambridge, MA: Harvard University Press (in press).

Hemmo, M., & Shenker, O. (2019). *Two kinds of objective probabilities. The Monist* (in press).

Hoefer, C. (2016). Causal determinism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of philosophy*. Spring 2016 Edition https://plato.stanford.edu/archives/spr2016/entries/determinism-causal/.

Kim, J. (1990). Supervenience as a philosophical concept. *Metaphilosophy, 21*, 1–27.

Kim, J. (2012). The very idea of token physicalism. In S. Gozzano, & C. Hill (Eds.), *New Perspectives on type identity* (pp. 167–185). Cambridge University Press.

Klein, C. (2008). Dispositional implementation solves the superfluous structure problem. *Synthese, 165*, 141–153.

Ladyman, J. A. C. (2009). What does it mean to say that a physical system implements a computation? *Theoretical Computer Science, 410*, 376–383.

Ladyman, J. A. C., Presnell, S. M., Shrot, A. J., & Groisman, B. (2007). The connection between logical and thermodynamic irreversibility. *Studies in History and Philosophy of Modern Physics, 38*(1), 58–79.

Ladyman, J. A. C., & Robertson, K. (2014). Going round in circles: Landauer vs. Norton on the thermodynamics of computation. *Entropy, 16*, 2278–2290.

Landauer, R. (1961). Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development, 3*, 183–191.

Landauer, R. (1992). Information is physical. In *Proceedings of PhysComp, workshop on physics and computation* (pp. 1–4). Los Alamitos, CA, USA: IEEE Computers Society Press.

Landauer, R. (1996). The physical nature of information. *Physics Letters A, 217*, 188–193.

Leff, H. S., & Rex, A. (2003). *Maxwell's Demon 2: Entropy, classical and quantum information, computing*. Bristol, UK: Institute of Physics Publishing.

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

Masuyama, Y., Funo, K., Murashita, Y., Noguchi, A., Kono, S., Tabuchi, Y., Yamazaki, R., Ueda, M., & Nakamura, Y. (2018). Information-to-work conversion by Maxwell's demon in a superconducting circuit quantum electrodynamical system. *Nature Communications, 9*, 2018. Article number: 1291.

Matthews, R. J., & Dresner, E. (2016). Measurement and computational skepticism. *Noûs, 51*, 832–854.

Melnyk, A. (1996). Searle's abstract argument against strong AI. *Synthese, 108*, 391–419.

Miłkowski, M. (2013). *Explaining the computational mind*. MIT Press.

Ney, A. (2008). Physicalism as an attitude. *Philosophical Studies, 138*, 1–15.

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

Norton, J. (2011). Waiting for Landauer. *Studies in History and Philosophy of Modern Physics, 42*, 184–198.

Piccinini, G. (2008). Computation without representation. *Philosophical Studies, 137*, 205–241.

Piccinini, G. (2015). *Physical computation*. Oxford University Press.

Piccinini, G. (2017). In E. N. Zalta (Ed.), *Computation in physical systems*. The Stanford Encyclopedia of Philosophy (Summer 2017 Edition) https://plato.stanford.edu/archives/sum2017/entries/computation-physicalsystems/.

Putnam, H. (1988). *Representations and reality*. Cambridge, Mass: MIT Press.

Rescorla, M. (2017). In E. N. Zalta (Ed.), *The computational theory of mind*. The Stanford Encyclopedia of Philosophy (Spring 2017 Edition) https://plato.stanford.edu/archives/spr2017/entries/computational-mind/.

Ridderbos, K., & Redhead, M. (1998). The spin echo experiments and the Second Law of thermodynamics. *Foundations of Physics, 28*(8), 1237–1270.

Schuetz, M. (2012). What is it not to implement a computation: A critical analysis of Chalmers' notion of implementation. *Journal of Cognitive Science, 13*, 75–106.

Searle, J. (1992). *The rediscovery of the mind*. Cambridge, Mass: MIT Press.

Searle, J. (2002). Why I am not a property dualist. *Journal of Consciousness Studies, 9*, 57–64.

Shagrir, O. (2001). Content, computation and externalism. *Mind, 110*, 369–400.

Shagrir, O. (2012). Can a brain possess two minds? *Journal of Cognitive Science, 13*, 145–165.

Shagrir, O. (2018). *In defense of the semantic view of computation* (in press).

Shenker, O. (2000). *Logic and entropy*. http://philsci-archive.pitt.edu/documents/disk0/00/00/01/15/index.html.

Shenker, O. (2017a). Foundations of statistical mechanics: Mechanics by itself. *Philosophy Compass*. https://doi.org/10.1111/phc3.12465.?.

Shenker, O. (2017b). Foundations of statistical mechanics: The auxiliary hypotheses. *Philosophy Compass*. https://doi.org/10.1111/phc3.12464.

Shenker, O. (2017c). Flat physicalism: Some consequences. *Iyyun: The Jerusalem Philosophical Quarterly, 66*, 211–225.

Shenker, O. (2018). Foundations of quantum statistical mechanics. In E. Knox, & A. Wilson (Eds.), *Routledge companion to the philosophy of physics*. Oxford: Routledge (in press).

Sprevak, M. (2010). Computation, individuation, and the received view on representation. *Studies in History and Philosophy of Science, 41*, 260–270.

Szabo, L. E. (2012). Mathematical facts in a physicalist ontology. *Parallel Processing Letters, 22*. https://doi.org/10.1142/S0129626412400099.

Werndl, C., & Frigg, R. (2017). Mind the gap: Boltzmannian versus Gibbsian equilibrium. *Philosophy of Science, 84*(5), 1289–1302.