Journal of Physics Communications

PAPER • OPEN ACCESS

The information-theoretic foundation of thermodynamic work extraction

To cite this article: Chiara Marletto 2022 *J. Phys. Commun.* **6** 055012

IOP Publishing J. Phys. Commun. 6 (2022) 055012 https://doi.org/10.1088/2399-6528/ac70a7

Journal of Physics Communications

OPEN ACCESS

RECEIVED
20 February 2022

REVISED
25 April 2022

ACCEPTED FOR PUBLICATION
17 May 2022

PUBLISHED
27 May 2022

PAPER

The information-theoretic foundation of thermodynamic work extraction

Chiara Marletto[1,2]

[1] Clarendon Laboratory, University of Oxford, United Kingdom
[2] Centre for Quantum Technologies, National University of Singapore, Singapore

E-mail: chiara.marletto@gmail.com

Keywords: information theory, thermodynamics, quantum information, constructor theory

Abstract

In this paper I demonstrate a novel distinction between work and heat, in terms of the distinguishability of states. Specifically, I show that if it is possible to extract different amounts of work deterministically from a physical system prepared in any one of a set of states, then those states must be distinguishable from one another. This result is formulated independently of scale and of particular dynamical laws; it also provides a novel connection between thermodynamics and information theory, established via the law of conservation of energy. This connection, unlike the well-known one between information and the second law, is exact, i.e., it does not depend on coarse-graining or ensemble approximation. Albeit compatible with these conclusions, existing thermodynamics approaches cannot provide a result of such generality, because they are scale-dependent (relying on ensembles or coarse-graining) or tied to particular dynamical laws. This result provides a foundation for thermodynamics that is both broader and more secure than existing ones, with implications for the theory of von Neumann's universal constructor.

Microscopic dynamical laws are time-reversal symmetric. Hence the second law of thermodynamics, intended as mandating the irreversibility of certain dynamical trajectories, is ruled out at the microscopic scale. This tension is usually tackled with statistical mechanics: Boltzmann's and Gibbs' ensemble theories, [1], and their quantum-mechanical generalisations in the hotly investigated area of quantum thermodynamics [2–4]. These powerful methodologies derive the second law from classical or quantum reversible dynamics supplemented with additional assumptions.

Despite their tremendous success in many regimes, these schemes have problems at their foundations. First, some such schemes traditionally rely on approximations such as ensembles and coarse-graining, which make the ensuing second laws *scale-dependent*, [1], and only applicable at a certain macroscopic scale, which is never exactly defined. Examples of scale-dependent laws are those about ferromagnetic phase transitions, which become exact only in the thermodynamic limit (and are not even intended to be exact for realistic systems). I shall designate as '*scale-independent*' any law whose applicability to a system does not depend on the system's scale. Most fundamental laws are scale-independent, e.g. conservation laws or Einstein's equations.

Furthermore, some formulations of the second law are tied to a particular class of dynamical laws: for instance, quantum thermodynamics is formulated within quantum theory. Hence, they are less general than traditional thermodynamics, which consists of a set of meta-laws largely independent of the details of the dynamical laws they constrain. I shall call laws which can be expressed without reference to the details of any particular dynamics, '*dynamics-independent*'.[1] The power of dynamics-independent principles has long been known in fundamental physics. They can be used in lieu of specific dynamical laws, for instance when solving the dynamical equations is an intractable problem—e.g., to study the behaviour of a complex system. They can also be used to make predictions when known laws of motion may not apply to a given regime: e.g. consider

[1] The term dynamics here is used in a broad sense. It refers to a law of motion, including both its formal kinematic elements (e.g. the algebra of observables in quantum theory) and its dynamical ones (e.g. the equations of motion). It is also intended as opposed to 'principle', intended as a meta-law that is not exhausted by a single dynamical equation.

Bekenstein's derivation of black hole entropy formula, [5]; and the Bondi-Wheeler's derivation of redshift from conservation of energy, [6, 7], where thermodynamics principles are used instead of a particular theory of coupled matter and gravity. In this paper, I propose a new information-theoretic characterisation of work, based on distinguishability, which is independent of scale (hence it refers to no particular length or time or complexity) and of dynamics (i.e. refers to no particular equations of motion). The aim here is to conjecture a new definition of work, which must be formulated independently of particular dynamical formalisms. I shall however use examples from classical and quantum theory to illustrate the concepts I shall introduce. This result expands the reach of current approaches to thermodynamics, putting them on more general and secure foundations. Such advancement is also useful to employ thermodynamics principles to conjecture future laws of motion that will supersede current ones.

The key result will be obtained by relying on a set of general principles (which I will discuss in detail later), some of which are part of the recently proposed *constructor theory of information* [8–10]. The main principles are: P1) The principle of locality, intended here in the customary sense of no-action at a distance; P2) The interoperability of information—a general axiom that is conjectured to be obeyed by theories of physics that can support information, [9], and expresses the idea that information attributes are copiable from any information medium to any other; P3) The principle of asymptotic distinguishability which, as I will discuss, is a generalisation of the idea of quantum tomography: that by repeatedly performing many different measurements on identically prepared systems, one can estimate the quantum state of these systems and hence distinguish it from a different state, to arbitrarily high precision,[2]; P4) The principle of conservation of energy. Assuming these principles, I will prove a theorem that can be informally stated as follows:

If it is possible to extract deterministically different amounts w_x of thermodynamic work, each from a single given physical system prepared in any one of a set of attributes x, then the attributes in that set are all distinguishable from one another, where 'attribute' hereinafter indicates a set of states. Crucially, I shall define 'extracting thermodynamic work' and 'distinguishable' in a *scale-independent* and *dynamics-independent* way. Definitions of these concepts already exist, expressed within particular dynamics. For instance: in quantum information (which relies on quantum dynamical evolution) two qubit states are distinguishable if and only if they are orthogonal; in quantum thermodynamics the work deterministically extractable, asymptotically, in a process taking a quantum state ρ_1 to ρ_2 is given by: $F(\rho_1) - F(\rho_2)$, where $F(\rho) = U(\rho) - \kappa_B TS(\rho)$, and $S(\rho) = -\mathrm{Tr}\{\rho \ln \rho\}$ while $U(\rho) = \mathrm{Tr}\{\rho H\}$, H being the Hamiltonian of the isolated system. These propositions use quantum theory's formalism, hence they are dynamics-dependent. My results will be consistent with these notions of distinguishability and work extraction, but will be formulated in a strictly scale- and dynamics-independent way.

Specifically, I shall propose a more general definition of thermodynamic *work extraction*. It includes as special cases the classical and quantum-thermodynamics definitions, but it works at any scale and it also holds for a vast class of dynamical laws. I shall also establish a further unexpected connection between thermodynamics and information theory, by showing that the possibility of extracting different amounts of work *deterministically* from a system prepared in any one of a set of states implies that those states must all be distinguishable (in the information-theoretic sense, which is far more general than the quantum one) from one another. Surprisingly, this link between information theory and thermodynamics goes via the law of *conservation of energy*, instead of the second law (as usually thought). This result poses a fundamental limitation on any quantum thermodynamics protocol for extracting work from systems with quantum coherence, e.g. [2, 11].

Constructor theory of information

I now summarise informally the basics of constructor theory (CT) (see appendix A and [9, 10, 12] for the formal details). The fundamental concept in CT is that of a task. A task is the specification of a transformation expressed as a set of ordered pairr of input/output attributes. Attributes are sets of states of a given physical system. A physical system on which tasks can be performed we call 'substrate'.[3] If **a** and **b** are attributes, the attribute (**a**, **b**) of the composite system $\mathbf{S_1} \oplus \mathbf{S_2}$ is defined as the set of all states of the composite system where $\mathbf{S_1}$ has attribute **a** and $\mathbf{S_2}$ has attribute **b**. I shall assume throughout the principle of locality (**P1**), which requires that if a transformation operates only on substrate $\mathbf{S_1}$, then only the attribute **a** changes, not **b**. It is well-know that this principle is satisfied by non-relativistic unitary quantum theory (see e.g. the discussion in [13], as well as by quantum field theories).

[2] Note that throughout the paper I shall consider quantum mixed states as states of a single system, not as representing statistical ensembles.

[3] I shall still use the term 'physical system' throughout the paper with the usual informal meaning it has in physics, whenever its being or not a substrate does not need to be emphasised.

In quantum theory, for instance, a qubit is a substrate; one of its attributes is a set of states such that a given projector is sharp with value 1 in each state of that set. Denoting by 0 the attribute for the qubit's state to be in a given subspace and by 1 the attribute for the qubit's state to be in its orthogonal complement, an example of a task is $\{0 \to 1, 1 \to 0\}$, negating the qubit in a particular basis. A *variable* is a set of disjoint attributes. Given a task T, define its *transpose* as the task obtained from T by swapping each input attribute with the corresponding output attribute.

A *constructor* for a task T is a system which whenever presented with the substrate of T in any state belonging to one of the input attributes, it delivers it in one of the states of the allowed output attributes, and *retains the ability to do that again*,[4]. In quantum theory (see appendix A), a constructor is modelled by a set C with the following property: the substrate undergoes the transformation specified by T, whenever it is coupled to the environment in a state belonging to C, and C is invariant under the action of the overall unitary evolution of the joint system of substrates and environment.

A task is *impossible* if the laws of physics impose a finite limitation on how accurately it can be performed by a constructor. Otherwise, the task is *possible*. Note that a task being possible does not require there to be a perfect constructor in reality. Rather, the requirement is that the behaviour of such a constructor can be approximated to arbitrarily high accuracy, short of perfection, by a sequence of processes each of which can be realised in nature. As explained in the appendix, and in [15], if a task is possible, then for any given finite accuracy, it is possible to find a suitable environment which (under an allowed dynamical interaction) approximates a constructor for the task by delivering the substrates in the desired output attributes (within the specified accuracy). Hence, if the task is possible, there is an (infinite) sequence of approximate constructors, each one implementing the task in question to ever improving accuracies. In accordance with the results on error correction in quantum computing, logically reversible computational tasks are all possible under the unitary quantum model of quantum computation, and quantum gates are example of constructors, which can be approximated to arbitrarily high accuracy. On the other hand, the task of cloning two non-orthogonal states is impossible, because there is a finite limit to the accuracy with which it can be performed under unitary quantum theory. It is interesting to point out that approximate constructors for a possible task are typically implemented by open quantum systems, in line with what is expected in the theory of quantum controls, [16]. For the interested reader, in [15] there is a model of a possible task and approximate constructors in quantum theory, where it is highlighted that approximate constructors can be dissipative open systems.

CT consists of general newly-conjectured principles expressed solely in terms of possible/impossible tasks, intended to constrain laws of motion (such as quantum theory's or general relativity's), which are called *subsidiary theories*. The full explanation of a physical situation is given by the principles of CT and by the compatible subsidiary theories. The principles are formulated in a scale- and dynamics-independent way, so they underlie a number of subsidiary theories; they don't refer to constructors, rather to the possibility or impossibility of certain tasks. Here I shall confine attention to subsidiary theories with a space of allowed states endowed with a topology assigning a meaning to states being arbitrarily close to each other. For present purposes it is not necessary to model a possible task within a given subsidiary theory, because I shall take 'possible' as a primitive concept (just like in the theory of quantum and classical computation). However, I discuss a simple quantum model for it in appendix A, following [17].

Distinguishability

The base of my construction is a recently proposed definition of distinguishability, [9]. Existing definitions of distinguishability rely on the details of given dynamical laws (e.g. two attributes are distinguishable if they are representable as orthogonal subspaces); or rely on scale-dependent notions (e.g. the concept of macroscopic or classical limit, where all states are distinguishable). Here instead we generalise the quantum-information notion of states that can be distinguished arbitrarily well from each other with a single-shot, projective measurement (without referring to orthogonality).

First one defines a class of substrates, *information media*, by requiring that some tasks are possible on them—tasks that are conjectured to be sufficient for them to be capable of carrying information. In short, information media must have a variable X with the property that it is possible to perform all the permutation tasks on X, and that it is possible to perform the task of copying all attributes in X from one substrate to its replica (see appendix A for the formal definitions).

Any variable X that can be copied and permuted in all possible ways is called an *information variable*. An example of information medium is a qubit with an information variable being any set of two orthogonal states.

[4] The notion of a catalyst in resource theory [14] could be considered as a model for special cases of constructors—catalysts must stay in exactly the same state (as opposed to the same attribute) and their definition is dynamics-dependent.

Any two different information media (e.g. a neutron and a photon) must satisfy an *interoperability principle* (**P2**), which expresses elegantly the intuitive property that classical information must be copiable from one information medium to any other (having the same capacity), irrespective of their physical details, [9]. Specifically, if S_1 and S_2 are information media, respectively with information variable X_1 and X_2, their composite system $S_1 \oplus S_2$ is an information medium with information variable $X_1 \times X_2$, where × denotes the Cartesian product of sets.

Now I define distinguishability using information media, as follows. A variable Y is *distinguishable* if the task

$$\bigcup_{y \in Y} \{\mathbf{y} \to \mathbf{q_y}\} \tag{1}$$

is possible, where the variable $\{\mathbf{q_y}\}$, of the same cardinality as Y, is an information variable. Informally, if the above task is possible, then it is possible to implement a physical process that maps the attributes of the variable Y 1:1 onto the attributes of an information variable.[5] Hence, a set of orthogonal quantum states for which the above task is possible is a distinguishable variable—but we have expressed this fact without referring to quantum theory's specific formalism, in a scale- and dynamics-independent way.

Another principle of CT that I shall deploy is the *principle of asymptotic distinguishability* (**P3**). Informally, it requires that N copies of an attribute $\mathbf{x}$ and N copies of another attribute that is disjoint from $\mathbf{x}$ are asymptotically distinguishable. In other words, the task of distinguishing them is possible when the number N of copies goes to infinity. In quantum theory, this corresponds to the fact that any two different quantum states are tomographically distinguishable (in this case, having one of a specified set of density matrices counts as an attribute).

Work media

In traditional thermodynamics there is a general consensus, following Planck, on identifying a work repository with a system behaving 'in the same way' as a weight in a uniform gravitational field, which can be smoothly raised or lowered to different heights, [1]. In quantum thermodynamics, it is common practice to define a work repository as a system in any eigenstate of its free Hamiltonian, such as a set of bound states in an atom, utilisable as a battery; there are also other proposed notions of work repositories (see [3] for a review). Here my intention is to generalise the class of work repositories to that of *work media*, [12]. I shall define work media as a particular class of substrates satisfying an operational criterion (just like information media): certain tasks must be possible on a substrate for it to qualify as a work medium. This will provide a conjectured scale- and dynamics-independent generalisation of the notion of work repository, building on the classical definition of Planck's and Clausius'.

First, one needs to express the *principle of conservation of energy* (**P4**) in CT. Following [8], it is possible to show that in the presence of the conservation of energy the tasks on a given substrate are partitioned into *equivalence classes*, arising from the impossibility of a task that corresponds to changing the energy of a given substrate: tasks belonging to the same equivalence class violate energy conservation by the same amount (the details of how to define these equivalence classes are in the appendix B). I shall call these classes 'energy-equivalence-classes', and I shall assume for simplicity (and also in conformity with the rest of thermodynamics) that energy conservation is the only conservation law.

One can define a work medium as a substrate **Q** having a variable $W = \{\mathbf{w_+}, \mathbf{w_0}\}$ with the property that:

1. The task $T_{+,0} = \{\mathbf{w_+} \to \mathbf{w_0}\}$ belongs to an energy-equivalence class such that $T_{+,0}$ is impossible and so is its transpose.
2. There exists an attribute $\mathbf{w_-}$ of **Q**, disjoint from $\mathbf{w_0}$ and $\mathbf{w_+}$, such that the task:

$$\{(\mathbf{w_+},\ \mathbf{w_0}) \to (\mathbf{w_0},\ \mathbf{w_+}),\ (\mathbf{w_0},\ \mathbf{w_0}) \to (\mathbf{w_+},\ \mathbf{w_-})\} \tag{2}$$

is possible.

Such a variable W is a *work variable*.

An example of a substrate possessing a work variable is an atom Q with three different equally-spaced energy levels, in decreasing order of energy as follows: $\mathbf{w_+}, \mathbf{w_0}, \mathbf{w_-}$, obeying energy conservation. In the presence of finite resources, because of the conservation of energy, it is impossible to perform the task $T_{+,0}$: the task requires the energy of the atom to change. Indeed, due to energy conservation, any finite-dimensional environment coupled to the atom would have to modify its energy by an amount that is equal and opposite to the amount by which $T_{+,0}$ changes the energy of the atom, hence it cannot act as a constructor for the task. Thus condition (1) is

[5] Note that the task may not be physically reversible (i.e., its transpose need not be possible, see [15] for an example.)

satisfied under our assumptions. Also, the task (2) is possible on Q and its replica, under our assumptions. That task is implementable by a suitably engineered unitary U that is energy-preserving—i.e., U: $[U, H_1 + H_2] = 0$, where H_i is the energy of each of the replicas of Q. For instance, this unitary can be implemented to arbitrarily high accuracy by an exchange interaction between each of two atoms (representing Q and its replica) and a photon field in a cavity, see [18] for a detailed discussion. So, a quantum system with at least 3 equally spaced energy levels satisfies the definition of work medium, hence this definition is compatible with existing classical and quantum notions of work repository. Note also that the definition (2) includes not only a 'quantum battery' with its energy eigenstates, but also the states of a classical weight suspended at different heights in a gravitational field, as defined in standard thermodynamics textbooks [19]. It also includes the classical limit of existing quantum models for such classical systems. For instance, one could consider N replicas of a quantum system, each prepared in the same mixture of energy eigenstates, chosen among a set of mixtures each with a different non-zero value of free energy: the classical limit would be obtained for N that tends to infinity. Another example of a work variable is a set of coherent states each labelled by a different value of α, in the limit where the distance $|\alpha - \alpha'|$ goes to infinity. These limiting cases all qualify as work variables because task (2) being possible means that a constructor for it can be approximated to arbitrarily high accuracy. In other words, if there are no limitations to the accuracy to which the task can be performed by an approximate constructor, then the task is possible. In the case of the classical limit of a quantum model, the task defining a work variable is exactly performable by a constructor only in the asymptotic limit. In [15] the case of approximating a possible task has been described carefully, and also demonstrated with a quantum simulation.

The key fact about condition (2) is that it is *not* satisfied by purely thermal attributes such as having a particular temperature, in line with traditional thermodynamics: as is well-known, a single thermal state cannot be used to do work. For example, let's assume $\mathbf{w}_\alpha = \mathbf{T}_\alpha$, where the attributes $\mathbf{T}_+, \mathbf{T}_-, \mathbf{T_0}$ of, say, a volume of water each correspond to a thermal state with given temperature T_α. In order to satisfy the first requirement (equation (2)), an equilibrium state $(\mathbf{T_0}, \mathbf{T_0})$ should be transformed into the temperature attribute $(\mathbf{T}_+, \mathbf{T}_-)$, with no other side effects. This is impossible according to the second law in classical and quantum thermodynamics. Thus, systems endowed with thermal degrees of freedom within the standard definitions of thermodynamics do not qualify as work media.

The above definition identifies precisely the attributes that can be used to acquire energy from another system, or deliver energy to it, *reversibly*, with no other side-effects. It advances existing definitions, such as those declaring eigenstates of energy to be work repositories by fiat, because it applies independently of formalism and scale. It is wide enough to be consistent with the traditional classical notion of 'work repository' or 'mechanical means', [1], but it is also applicable to general systems that need not be mechanical in the classical sense, e.g. a quantum atom in an excited state. It also includes the classical limit of quantum models, for example those using a large number of replicas or large-amplitude coherent states. So it is a solid foundation to build a scale- and dynamics-independent notion of deterministic work extraction.

A deterministic work extractor

I will now give a conjectured definition of work extraction, using the notion of a work variable just defined. The task of deterministically extracting work from a substrate **S** in regard to a variable X of **S** is:

$$\bigcup_{x \in X} \{(\mathbf{x},\ \mathbf{w_0}) \rightarrow (\ \mathbf{f_x},\ \mathbf{w_x})\} \tag{3}$$

where $\{\mathbf{f_x}\}$ is some variable of **S** and the pairs $\{\mathbf{w_x},\ \mathbf{w_{x'}}\}$, for all $x,\ x' \in X$, are each a work variable of a work medium **Q**. Note that (by definition of work variable) all the attributes in $\bigcup_{x \in X}\{\mathbf{w_x}\}$ have to be disjoint attributes. For example, **Q** here could be an atom with several levels of energy that gets excited or de-excited by interaction with **S**.

A constructor for the above task is deterministic because it delivers with certainty one and only one output attribute for any particular input attribute, retaining the ability to do that again, and without any other side-effects. Such reliable behaviour is expected of an ideal classical heat engine and of an ideal quantum deterministic work extractor, [2], so this requirement is well-grounded in existing theories of thermodynamics. By continuity, one could also consider probabilistic work extractors, in which case what follows would still hold, with a certain probability set by the reliability of the probabilistic work extractor. Investigating the probabilistic case is outside of the scope of this paper as I am aiming at providing an exact, qualitative (not quantitative) distinction between work and heat.

The information-theoretic foundation of deterministic work extraction

I can now state the key result of the paper more formally:

Theorem 1. *A work variable is a distinguishable variable.*

This follows straightforwardly from the fact that the task (2) is possible on a work medium. Consider a work variable W and the following task, generalising (2) to having n substrates as target:

$$\begin{aligned}&\{(\mathbf{w}_{+},\ (\mathbf{w}_{\mathbf{0}})^{(2n)})\ \rightarrow\ (\mathbf{w}_{+},\ (\mathbf{w}_{+},\ \mathbf{w}_{-})^{(n)});\\ &(\mathbf{w}_{\mathbf{0}},\ (\mathbf{w}_{\mathbf{0}})^{(2n)})\ \rightarrow\ (\mathbf{w}_{\mathbf{0}},\ (\mathbf{w}_{-},\ \mathbf{w}_{+})^{(n)})\}.\end{aligned} \qquad (4)$$

When n tends to infinity, $(\mathbf{w}_{+},\ \mathbf{w}_{-})^{(n)})$ is asymptotically distinguishable from $(\mathbf{w}_{-},\ \mathbf{w}_{+})^{(n)}$, by the asymptotic-distinguishability principle. Thus, the attributes $\mathbf{w}_{+}$ and $\mathbf{w}_{\mathbf{0}}$ of a work medium are distinguishable from one another, by definition of distinguishability. The proof is expressible in quantum theory, by modelling the attributes as non-intersecting linear subspaces and using standard results from state tomography (see [17], a summary of which is in appendix C).

Given that work variables are distinguishable variables, any variable X for which the task (3) of deterministically extracting different amounts of work is possible must also be distinguishable. This is because each pair of attributes in X is distinguishable, given that the task (3) is possible and because a variable containing attributes that are all pairwise distinguishable from one another is a distinguishable variable, [9].

This concludes the proof that a deterministic work extractor is also a perfect distinguisher: the attributes of a variable from which different amounts of work can be extracted reliably, in the deterministic sense defined in (3), must be distinguishable from one another. Specialising to quantum theory, this implies that different amounts of work can only be extracted (in the sense defined in equation (3)) from a system prepared in one of a set of orthogonal subspaces. Note that in each of the attributes $\mathbf{x}$ in X one could find two states that are not distinguishable from each other and from which the same amount of work can be extracted. This is not in contradiction with my result, because the latter applies to attributes with the counterfactual ability to produce at least two different amounts of work in output.

It is also important to notice that the key aspect of this theorem is that it is proven to hold for a class of theories that are much more general than quantum theory. In this general setting, it establishes a previously unnoticed connection between information theory (via the distinguishability of attributes) and thermodynamics (via the possibility of extracting different amounts of work from those attributes, deterministically).

This result is compatible with the claims that one can extract work *probabilistically* from a superposition of different energy eigenstates, [20]: for such states, deterministic work extraction is achieved in the limit of having an ensemble of identically prepared systems in the same quantum state. Indeed, in the asymptotic limit different quantum states are distinguishable (this is what permits quantum tomography). This fact is also true for coherent states, in the large-amplitude limit. The overlap of two coherent states labelled by complex numbers A and B goes exponentially to zero as $|A - B|$ goes to infinity. When we say that it is possible to extract a given amount of work from a coherent state deterministically, we have in mind the classical limit of a quantum model where the amplitude of the coherent state is large, or where there are many copies of the coherent state available to us. In the infinite amplitude limit (or in the case of infinitely-many copies being available) the extraction of work is deterministic and the states are perfectly distinguishable. Hence a set of coherent states in this limit qualifies as a work variable.

The significance of this result becomes clear when one fully embraces the consequences of quantum theory. In quantum theory, in order to perform a task deterministically, one does not need perfectly distinguishable input states. For example, there is a (deterministic) unitary process that swaps two non-orthogonal states—see e.g. [9]. But here I suggest a novel insight—that in order to have different amounts of thermodynamic work done reliably, then the attributes involved in the work extraction must be perfectly distinguishable. This holds true independently of specific dynamical models and scale. Hence, unlike existing formulations of the distinction between work and heat (see e.g. [21]), this formulation does not rely on statistical mechanics concepts such as ‘entropies’. It would be an interesting development of this paper to apply its results to providing a reinterpretation for the concept of entropy.

A new scale- and dynamics-independent foundation for the second law

This theorem provides new methodologies to tackle the problem of formulating the second law of thermodynamics in a scale- and dynamics-independent way. I can illustrate how by discussing how it can be applied to the important issue of generalisi the concept of *adiabatic accessibility* (epitomised by the famous Joule's experiments, [1]), which is the core of the axiomatic approach to thermodynamics, [1, 19, 22, 23]. Following the axiomatic approach, an attribute **b** is adiabatically accessible from the attribute **a** if it is possible to construct a thermodynamic cycle that transforms **a** into **b** with the only side-effect being the raising or lowering of a weight in a gravitational field. So for instance the second law in traditional thermodynamics says that the state of a volume of water at a given temperature is adiabatically accessible from one at a lower temperature (because mechanical stirring can heat up an otherwise isolated volume of water); but it is not adiabatically accessible from a state at a higher temperature (because mechanical stirring cannot by itself cool an otherwise isolated volume of water).

Using work media, one can propose a variant of the definition of adiabatic accessibility, appealing to the notion of *adiabatic possibility*, with the crucial advantage of being scale- and dynamics-independent. A task $\{\mathbf{x} \rightarrow \mathbf{y}\}$ is adiabatically possible if the task:

$$\{(\mathbf{x},\ \mathbf{w}_1) \rightarrow (\mathbf{y},\ \mathbf{w}_2)\}$$

is possible for some two work attributes $\mathbf{w}_1$, $\mathbf{w}_2$ belonging to a work variable. The latter generalises the ad hoc weight-in-a-gravitational-field criterion invoked in the traditional definition, making the notion of adiabatic accessibility dynamics-independent thanks to the definition of work media, and emancipating it from statistical approximation schemes such as coarse-graining. Therefore it allows one to formulate the *second law*, [12], as the requirement that:

There are tasks that are adiabatically possible, whose transpose is not adiabatically possible.

This statement is fully compatible with traditional thermodynamics laws, but it extends as it is scale- and dynamics- independent. This provides the basis for an exact distinction between work and heat. Note that it is not possible to use this framework to derive equalities such as Jarzynski's [24] or to express quantitative versions of the second law, and their recent generalisations (see e.g. [25–27]), unless one adds more assumptions. The concept of probability, in particular, is needed to talk quantitatively about mixed states and fluctuations. The fact that the results in this paper are not quantitative is the price to pay for having a very general formalism, which is meant to provide the conceptual physical foundations for a scale- and dynamics-independent thermodynamics. However, I expect it will be possible to extend these results to the probabilistic work extraction case using the formalism of superinformation theories—theories that support a probabilistic structure and comply with constructor theory's principles (as defined in [10]). I leave this to future work.

Discussion

My theorem establishes a novel foundation for thermodynamics, based on constructor-information theory, which is scale- and dynamics-independent. In quantum theory, this result implies that if one can extract different amounts of work deterministically from a system prepared in a set of states, these states must be orthogonal to each other. The theorem I proved is similar in logic to the no-cloning theorem in quantum information: it is a no-go theorem, stating that one cannot extract different amounts of work reliably from a system prepared in any one of a set of states unless they are perfectly distinguishable. However, it is more far-reaching than the no-cloning theorem, because it is dynamics-independent, so it is more general than, but compatible with, quantum theory. For instance, it could apply to the potential successors of quantum theories—e.g. theories of coupled gravity and quantum matter. It therefore provides a promising basis for constraining future subsidiary theories, including those describing exotic objects such as black holes or closed time-like curves. It also connects information theory and thermodynamics in an unexpected way, not regarding the second law, but the conservation of energy.

An interesting parallel between a programmable quantum computer (whose admissible programs must belong to the computational basis [28, 29]) and a deterministic work extractor emerges here. I proved that variables that can serve as input to a deterministic work extractor must be a set of distinguishable attributes. This constitutes the only possible 'work basis', which, like the computational basis, has to consist of distinguishable, orthogonal subspaces. These could be either a set of sharp energy states; or a set of states that are not diagonal in the energy basis, each provided with orthogonal labels. This poses a fundamental limit on the work that can be extracted deterministically from quantum systems with coherence in the energy basis, by a machine that can extract work individually from each of the states in the energy basis. One can for instance envisage a process that extracts work optimally and deterministically from a system prepared in a particular, known, quantum state that is the coherent superposition of different energy eigenstates, [11], as compared to the corresponding thermal state with the same mean energy. Supposing the work extracted from the superposition is different from that extractable from each of the energy basis states, the results of this paper imply that such a process would have to be a special-purpose machine, which requires to know *a priori* which state has been prepared. Therefore, it is not a proper work-extractor in the thermodynamic sense, not more than a Szilard engine without its memory is.

This work provides the foundation for formulating thermodynamics in an information-theoretic, dynamics-independent and scale-independent way: hence, it can inform new experimental schemes to test this proposed scale- and dynamics-independent reformulation of the second law, see e.g. [15]. It is also a first step towards a theory of programmable constructors in quantum theory, which will generalise the theory of quantum computation to general tasks, in a way already envisaged in von Neumann' theory of the universal constructor [30]. The full development of this theory will require one to merge the theory of classical and quantum computation with thermodynamics, in a dynamics- and scale-independent way.

Acknowledgments

The author thanks David Deutsch and Vlatko Vedral for discussions and comments on earlier versions of this manuscript; Benjamin Yadin and Paul Raymond-Robichaud for helpful suggestions. This publication was made possible through the support of the ID 61 466 grant from the John Templeton Foundation, as part of the The Quantum Information Structure of Spacetime (QISS) Project (qiss.fr). The author's research was also supported by the Eutopia Foundation and by the grant number (FQXi FFF Grant number FQXi-RFP-1812) from the Foundational Questions Institute and Fetzer Franklin Fund, a donor advised fund of Silicon Valley Community Foundation.

Data availability statement

No new data were created or analysed in this study.

Appendix A. Constructor theory

Constructor theory is a meta-theory with its own physical principles that are intended to supplement and constrain dynamical theories, such as quantum theory and general relativity, which therefore we call *subsidiary theories*, [8]. Every subsidiary theory that is constructor-theory compliant must provide a set of allowed states of the allowed substrates, endowed with a topology.

An *attribute* **x** is a set of states all having a property x. For instance, in quantum theory, the set of all quantum states of a qubit where a given projector Π is sharp with value 1 is an attribute. A *variable* is a set of disjoint attributes, where 'disjoint' has to be intended in the sense of set theory.

If **a** is an attribute of substrate $\mathbf{S}_1$ and **b** is an attribute of substrate $\mathbf{S}_2$, the attribute $(\mathbf{a}, \mathbf{b})$ of the composite substrate $\mathbf{S}_1 \oplus \mathbf{S}_2$ is defined as the set of all states where $\mathbf{S}_1$ has attribute **a** and $\mathbf{S}_2$ has attribute **b**. As I stated in the main text, the principle of locality requires that if a transformation operates only on substrate $\mathbf{S}_1$, then only the attribute **a** changes, not **b**. It is well-know that this principle of no-action at a distance is satisfied by non-relativistic unitary quantum theory, as well as by quantum field theories.

In quantum theory, assuming for instance a two-qubit Hilbert space $\mathcal{H}_{ab}$, we can consider the class of attributes defined as

$$(\mathbf{a},\ \mathbf{b})\ \doteq \{\rho_{ab} \in \mathcal{H}_{ab}\colon\ \mathrm{Tr}\{\rho_{ab}\Pi_a \otimes \Pi_b\} = 1\}$$

where Π_a and Π_b are given projectors defined on each qubit's Hilbert space. (In the general case this attribute may include quantum states where the subsystems are entangled.)[6] A *task* is the abstract specification of a physical transformation, represented as a finite set of ordered pairs of input/output attributes: $T = \{\mathbf{a_1} \to \mathbf{b_1}, \mathbf{a_2} \to \mathbf{b_2},\ \cdots, \mathbf{a_n} \to \mathbf{b_n}\}$.

A *constructor* for a task T is a system which whenever presented with the substrate of the task T in one of the input attributes, it delivers it in one of the states of the allowed output attributes, and *retains the ability to do that again*. A task is *impossible* if the laws of physics impose a limit on how accurately it can be performed by a constructor. Otherwise, the task is *possible*. Constructor-theoretic statements never refer to specific constructors, only to the fact that tasks are possible or impossible. This is what allows them to be scale- and dynamics-independent.

Tasks close an algebra, [17]. Two tasks T_1 and T_2 can be composed in series (whenever the output set of attributes of T_1 includes the input set of attributes of T_2), or in parallel, with the usual informal meaning of parallel and serial composition, [9]. I denote the serial composition of two tasks as T_1T_2; the parallel composition as $T_1 \otimes T_2$. The transpose of a task T, denoted by $T^{\sim}$, is the task with the input/output pairs of T inverted: $T^{\sim} \doteq \{\mathbf{b} \to \mathbf{a}\}$. One requires that $(T^{\sim})^{\sim} = T$; and that $(T_1 \otimes T_2)^{\sim} = T_1^{\sim} \otimes T_2^{\sim}$.

[6] A more general definition of attributes should be given using the Heisenberg picture of quantum theory and requiring the expected values of given observables to have a particular value, but for present purposes this definition is sufficient.

A cardinal principle of constructor theory, called the *composition law*, is that the composition of two possible tasks is a possible task.

A model of possible tasks and constructors in quantum theory

I will now provide a quantum model for a constructor, following [17]. Consider the composite system of two quantum systems, C and S, with total Hilbert space $\mathcal{H} = \mathcal{H}_C \otimes \mathcal{H}_S$. Fix a unitary law of motion U describing their interaction. I denote by $\Sigma(X)$ the $+1$-eigenspace of the projector X; also, for a general operator B, define $B^{(C)} = B \otimes 1$ and $B^{(S)} = 1 \otimes B$. Fix two attributes of S, defined as $\mathbf{x} = \Sigma(X^{(S)})$ and $\mathbf{y} = \Sigma(Y^{(S)})$, where X and Y are two projectors. Each of these attributes can be thought of as the set of states of S in which the corresponding projector is sharp with value 1. Fix a task $T = \{\mathbf{x} \to \mathbf{y}\}$.

Consider now the set of states of C defined as follows:
$V_T = \{|\psi\rangle \in \mathcal{H}_C\colon\ \forall |x\rangle \in \Sigma(X^{(S)}),\quad U(|\psi\rangle|y\rangle) \in \Sigma(Y^{(S)})\}$. This is the set of states of C with the property that, when C is initialised in one of those states, and presented with the substrate S in the state $|x\rangle \in \Sigma(G)$ with the attribute $\mathbf{g}$, it delivers the substrate in a state with attribute $\mathbf{y}$. Note that in the final state C and S could be entangled. Note also that C may no longer be able to cause the transformation again once it has performed it once. It is straightforward to check that the V_T is a vector space. A necessary set of conditions for C to be a constructor for the task A_t are:

- V_T is non empty;
- There exists a subspace $W_T \subseteq V_T$ such that

$$U(W_T \otimes \Sigma(X^{(S)})) \subseteq W_T \otimes \Sigma(Y^{(S)}).$$

These states of C retain their property of being capable to cause the transformation T over and over again. I shall denote by Π_{W_T} the projector onto the smallest subspace $W_T \subseteq \mathcal{H}_C$ with that property.

If the above two conditions are satisfied, we can define the (non-empty) set
$V_{C_T} = \{|\psi\rangle \in \mathcal{H}_C\colon\ \forall |x\rangle \in \Sigma(X^{(S)}),\quad U(|\psi\rangle|x\rangle) \in \Sigma(\Pi_{W_T}^{(C)} Y^{(S)})\}$, which is easily proven to be a vector space. States in this subspace either belong to W_T or they are brought into that space after one application of U. The projector Π_{C_T} onto this subspace is the projector for being a constructor for the task A_T.

That the task T is possible implies in quantum theory that there exists a subspace V_{C_T} with the above properties. (See also reference [15] where a more general set of necessary conditions for a constructor are given).

Constructor theory of information

Define the cloning task for the variable X as:

$$C(X) \doteq \bigcup_{x \in X} \{(\mathbf{x}, \mathbf{x_0}) \to (\mathbf{x}, \mathbf{x})\} \tag{5}$$

where $\bigcup$ is the set-union symbol and x_0 is a fixed attribute. That a set X is copiable means that the task $C(X)$ is possible, for some x_0. In quantum theory, this task is possible whenever all elements in X are orthogonal to one another; otherwise, if X consists of non-orthogonal states, it is impossible. For example, when X is a boolean variable, $X = \{0, 1\}$, and $\mathbf{x_0} = 0$, the task $C(X)$ can be implemented by a controlled-not gate.

An information medium is a substrate with the property that the cloning task $C(X)$ and the permutation task:

$$\Pi(X) \doteq \bigcup_{x \in X} \{\mathbf{x} \to \Pi(\mathbf{x})\}, \tag{6}$$

are all possible, for all permutations Π on the set of labels of the attributes in X and some attribute $\mathbf{x_0} \in X$. Once more, as an example, a set of orthogonal states in quantum theory, without additional symmetries or super-selection rules, qualifies as an information variable.

The task $C(X)$ corresponds to *copying*, or cloning, the attributes of the first substrate onto the second, target, substrate; $\Pi(X)$, for a particular Π, corresponds to a logically reversible computation. For example, a qubit is an information medium with any set of two orthogonal quantum states, $X = \{0, 1\}$, as defined above.

As explained in the main text of this paper, a variable Y is *distinguishable* if the task

$$\bigcup_{y \in Y} \{\mathbf{y} \to \mathbf{q_y}\} \tag{7}$$

is possible, where the variable $\{\mathbf{q_y}\}$, of the same cardinality as Y, is an information variable.

Let me define $S(n) \doteq \underbrace{S \oplus S \oplus \dots S}_{n}$, a substrate consisting of n instances of the substrate S, and $x(n) \doteq \underbrace{(x, x, \dots, x)}_{n}$, attribute of $S(n)$. Denote by $x(\infty)$ the attribute of $S(\infty)$, an unlimited supply of instances of S. Consider any two disjoint attributes x and x'.[7]

Asymptotic distinguishability requires the attributes $x(\infty)$ and $x'(\infty)$ of $S(\infty)$, whenever they are defined, to be distinguishable (hence, the variable $Y = \{x(\infty), x'(\infty)\}$ is distinguishable as in (7)).

Appendix B. Conservation of energy

I shall now express the requirement imposed by the law of conservation of energy in a scale- and dynamics-independent way, by formalising the observation that that a conservation law formulated by a given subsidiary theory induces a specific assignment of possible and impossible tasks, [8]. One can express this assignment with scale- and dynamics-independent statements, i.e. without appealing to any particular formalism such as Hamiltonian dynamics.

Consider the set Σ of all tasks on a substrate S consisting of only one input/output ordered pair. A conservation law for an additive quantity of the system S (energy for instance) induces a partition of Σ into equivalence classes, defined as follows. Each equivalence class X_E has the property that for any two tasks T_1 and T_2 belonging to X_E:

- Either the tasks $\{T_1, T_2\}$ and their transposes $\{T_1^{\sim}, T_2^{\sim}\}$ are all *impossible*; or they are all possible.
- The task $T_1 \otimes T_2^{\sim}$, and its transpose, are both possible tasks.

By using the properties of serial and parallel composition and the definition of transpose, [12], one can check that the two above conditions define an equivalence relation between tasks.

Using the properties of equivalence classes, one can introduce a real-valued function F, with the property that for any two pairwise tasks $T_1, T_2, F(T_1) = F(T_2)$ if and only if they belong to the same equivalence class.

There are infinitely-many possible functions F that could label the equivalence classes. How does one choose F so that it expresses a given conservation law? By the properties of parallel and serial composition, one first notices that there is only one class where both T_1 and T_2 and their transposes are all possible. So, to express a conservation law with this approach, the key step is to select a function F labelling the classes with the property that $F(T) = 0$ for all tasks T in that class. In all the other classes, any task T and its transpose are both impossible: these classes can each be labelled by a non-zero value of $F(T)$. This is physically motivated as follows: upon this choice, the label $F(T)$ represents the amount by which the task T violates the conservation law. In the class labelled by $F = 0$, all tasks are possible as they do not require to modify the conserved quantity. In all the other classes, by our definition above, each task T is impossible, but the task $T \otimes T^{\sim}$ is in turn possible. So one can interpret the label $F(T)$ as the non-zero amount by which the task T requires the conserved quantity to be changed: while T is impossible, $T \otimes T^{\sim}$ is possible given that it requires to change F by equal and opposite amounts on each substrate.

Given that in this paper we are assuming for simplicity that the only conservation law is the conservation of energy, I shall call each equivalence class an *energy-equivalence class*; if two tasks T_1 and T_2 belong to the same energy-equivalence class, I will write: $T_1 \sim T_2$; which means that the two tasks T_1 and T_2 violate the conservation law, they do so by the same amount.

Hence, the conservation of energy induces the constraint that the possible and impossible tasks on substrates S obey the two conditions listed above, with F chosen as described. The choice of the specific function F and any further constraint on it are up to each particular dynamical theory to specify, and are not relevant for present purposes, because the theorems expressed in this paper hold at the meta-level of principles, which are intended to underlie all subsidiary theories that conform to them: from classical Hamiltonian mechanics and quantum theory, to other theories that we may yet have to discover.

By noticing that each task in Σ is an ordered pair of attributes, the partition of tasks in Σ into equivalence classes induces a partition into classes of the set of their input/output attributes. One can choose a function E that labels each class by a real number, with the property that $E(\mathbf{a}) = E(\mathbf{b})$ if and only if the two attributes belong to the same class, and if $T = \{\mathbf{a} \rightarrow \mathbf{b}\}$, then the function F labelling the equivalence class of tasks is related to the function E by the following relation: $F(T) = E(\mathbf{b}) - E(\mathbf{a})$. The labelling of attributes defined by E can be thought of, in this case, as an energy function (defined, as usual, up to a constant). Thus I shall say that an attribute has a particular value of energy if it belongs to one of these classes labelled by that particular value of energy, under a

[7] Note that in the paper *disjoint* sets has to be intended in the sense of set theory, i.e., as indicating 'sets with empty intersection'. So, crucially, it does not mean 'orthogonal': the attributes $\{|+\rangle\}$ and $\{|0\rangle, |1\rangle\}$ are disjoint.

fixed labelling E compatible with the partition into equivalence classes of the set of all pairwise tasks. It is also possible to show that the function E has to be bounded both from above and from below, [8, 12].

Appendix C. Theorem 1 in quantum theory

In quantum theory, theorem 1 (see main section of this paper) can be proven by considering the general properties of programmable constructors—as outlined in [17], which generalises a proof proposed in [29]. I shall now summarise the key steps of the proof.

Under laws of motion represented by a unitary U, a system C may be a constructor for different tasks on the same substrate S initialised to a generic, fixed attribute $G^{(S)}$ (that can be thought of as a blank state). For instance, let Π_1 be the projector for being a constructor for the task A_{t_1} defined by the projector T_1, and Π_2 be the projector for being a constructor for the task A_{t_2} associated with the projector T_2. In this case, C can be considered as a programmable constructor with two kinds of programs in its repertoire, one to produce objects with the property T_1, the other to produce objects with the property T_2. (For example, C could be the register of a quantum computer, S its workspace.) Indeed, programs are (abstract) constructors.

Suppose that the two tasks are specified by unambiguous attributes, i.e, $\Sigma(T_1) \cap \Sigma(T_2) = \{0\}$. Then, one can prove that the projectors for the programs for each of those tasks must be orthogonal to each other: $\Pi_1\Pi_2 = 0$.

By hypothesis, U has the property that, for states $|P_i\rangle \in \Sigma(\Pi_i)$

$$U(|P_1\rangle|g\rangle) \in \Sigma(\Pi_1^{(C)}T_1^{(S)}) \tag{8}$$

$$U(|P_2\rangle|g\rangle) \in \Sigma(\Pi_2^{(C)}T_2^{(S)}). \tag{9}$$

Consider using the same program on several copies of the substrate $S^{(n)} = \underbrace{S \oplus S \oplus \ldots \oplus S}_{n}$, each initialised in the legitimate input attribute:

$$\begin{aligned} |P_1\rangle\ |g\rangle^{\otimes n} &\to |\psi_1^{(n)}\rangle \\ |P_2\rangle\ |g\rangle^{\otimes n} &\to |\psi_2^{(n)}\rangle \end{aligned} \tag{10}$$

where $|\psi_i^{(n)}\rangle \in \Sigma(\Pi_i^{(C)}\hat{T}_i^{(n)})$, where $\hat{T}_i^{(n)} = \mathbb{1} \otimes \underbrace{T_i \otimes T_i \otimes \ldots \otimes T_i}_{n}$. The above property must be true because, at the end of each transformation, the property of being a constructor for that specific task is preserved. Let us introduce the operator norm, $\|A\| = \mathrm{Sup}\{|A|v\rangle| : ||v\rangle| = 1\}$. On the one hand this norm is a cross-norm: $\|\hat{T}_1^{(n)}\hat{T}_2^{(n)}\| = \|T_1T_2\|^n$. On the other hand, $0 \leqslant \|T_1T_2\| < 1$ because the intersection between $\Sigma(T_1)$ and $\Sigma(T_2)$ is empty. This in turn follows from the theorem that the projector onto $\Sigma(T_1) \cup \Sigma(T_2)$ is $\lim_{n\to\infty}(T_1T_2)^n$.

Which implies that, if the intersection is empty, there can be no non-zero states $|v\rangle$ with the property that $T_1T_2|v\rangle = |v\rangle$ (otherwise they would be in the intersection). This fact, together with the fact that $\|T_1T_2\| \leqslant \|T_1\|\,\|T_2\| = 1$ implies that $\|T_1T_2\| < 1$.

Hence, in the limit $n \to \infty$ one has that

$$\|\hat{T}_1^{(n)}\hat{T}_2^{(n)}\| = \|T_1T_2\|^n \to 0$$

which implies that

$$\lim_{n\to\infty}\hat{T}_1^{(n)}\hat{T}_2^{(n)} = 0.$$

This means that the states $\lim_{n\to\infty}|\psi_1^{(n)}\rangle$, $\lim_{n\to\infty}|\psi_2^{(n)}\rangle$ are orthogonal, and so must be $|P_1\rangle$ and $|P_2\rangle$, because for arbitrary n the transformation performed by that network is unitary. Picking the two pure states $|P_1\rangle$ in the $+1$-eigenspace of Π_1 and $|P_2\rangle$ in the $+1$-eigenspace of Π_2 with the property that $|\langle P_1|P_2\rangle|^2$ is maximal, the above result shows that $|\langle P_1|P_2\rangle|^2 = 0$, thus proving that $\Pi_1\Pi_2 = 0$. In other words, the network asymptotically works as a distinguisher between the two constructor subspaces.

Specialising this general result to the case analysed in the main section of this paper, one can simply take the attribute $\mathbf{w_0}$ appearing in the proof of theorem 1 to be $\Sigma(\Pi_1)$ and $\mathbf{w}_+$ in that proof to be $\Sigma(\Pi_2)$, with $\Sigma(T_1)$ corresponding to $(\mathbf{w}_+, \mathbf{w}_-)$, $\Sigma(T_2)$ corresponding to $(\mathbf{w}_-, \mathbf{w}_+)$, and $\Sigma(G)$ corresponding to $(\mathbf{w}_0, \mathbf{w}_0)$. Then the quantum-theory proof just outlined, showing that $\Pi_1\Pi_2 = 0$, implies that theorem 1 is true in quantum theory, as it implies that that the projector associated with the attribute $\mathbf{w}_+$ is orthogonal to the projector associated with the attribute $\mathbf{w_0}$, hence that the two attributes are distinguishable.

ORCID iDs

Chiara Marletto https://orcid.org/0000-0002-2690-4433

References

[1] Uffink J 2001 *Stud. Hist. Phil. Mod. Phys.*, B **32** 305–94
[2] Goold J, Huber M, Riera A, del Rio L and Skrzypczyk P 2016 *J. Phys. A: Math. Theor.* **49** 143001
[3] Alicki R and Kosloff R 2018 *Thermodynamics in the Quantum Regime. Fundamental Theories of Physics* vol 195 ed F Binder *et al* (Cham: Springer)
[4] Gemmer J, Michel M and Mahler G 2009 *Quantum Thermodynamics* (Berlin: Springer)
[5] Bekenstein J D 1972 *Lett. Nuovo Cimento* **4** 737
[6] Misner C W, Thorne K S and Wheeler J A 2017 *Gravitation* (Princeton, NJ: Princeton University Press)
[7] Bondi H 1986 *Eur. J. Phys.* **7** 14
[8] Deutsch D 2013 *Constructor Theory, Synthese* **190** 18
[9] Deutsch D and Marletto C 2015 *Proc. R. Soc.* A **471** 20140540
[10] Marletto C 2016 *Proc. R. Soc.* A **472** 20150883
[11] Korzekwa K, Lostaglio M, Oppenheim J and Jennings D 2016 *New J. Phys.* **18** 023045
[12] Marletto C 2018 *Constructor Theory of Thermodynamics* arXiv:1608.02625
[13] Deutsch D and Hayden P 2000 *Proc. Roy. Soc.* A **456** 1999
[14] Coecke B, Fritz T and Spekkens R 2016 *Inf. Comput.* **250** 59–86
[15] Marletto C *et al* 2022 *Phys. Rev. Lett.* **128** 080401
[16] Nha H and Carmichael H J 2005 *Phys. Rev.* A **71** 013805
[17] Marletto C 2013 Issues of control and Causation in quantum information theory *Thesis* Bodleian Library
[18] Cirac J I, Zoller P, Kimble H J and Mabuchi H 1997 *Phys. Rev. Lett.* **04** 78
[19] Buchdahl H A 1966 *The Concepts of Classical Thermodynamics* (Cambridge: Cambridge University Press)
[20] Skrzypczyk P, Short A and Popescu S 2014 *Nat Commun.* **5** 4185
[21] Dahlsten O *et al* 2011 *New J. Phys.* **13** 053015
[22] Carathéodory C 1909 *Mathematische Annalen* **67** 355–86
[23] Lieb E and Yngvason J 1999 *Phys. Rept* **310** 1–96
[24] Jarzynski C 2008 *Eur. Phys. J.* B **64** 331–40
[25] Nielsen MA and Chuang I L 1997 *Phys. Rev. Lett.* **79** 041017
[26] Łobejko M *et al* 2021 *Nat. Comm.* **12** 918
[27] Horodecki M and Oppenheim J 2013 *Nat. Comm.* **4** 2059
[28] Myers J M 1997 *Phys. Rev. Lett.* **78** 1823
[29] Nielsen M and Chuang C 1997 *Phys. Rev. Lett.* **79**
[30] von Neumann J and Burks A W 1966 *Theory of Self-Reproducing Automata* (Champaign, IL: University of Illinois Press)