

Thermodynamic work from operational principles

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2016 New J. Phys. 18 103017

(<http://iopscience.iop.org/1367-2630/18/10/103017>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 87.77.118.212

This content was downloaded on 19/12/2016 at 10:59

Please note that [terms and conditions apply](#).

You may also be interested in:

[The role of quantum information in thermodynamics—a topical review](#)

John Goold, Marcus Huber, Arnau Riera et al.

[The extraction of work from quantum coherence](#)

Kamil Korzekwa, Matteo Lostaglio, Jonathan Oppenheim et al.

[Measures and applications of quantum correlations](#)

Gerardo Adesso, Thomas R Bromley and Marco Cianciaruso

[Equilibration, thermalisation, and the emergence of statistical mechanics in closed quantum systems](#)

Christian Gogolin and Jens Eisert

[Introducing one-shot work into fluctuation relations](#)

Nicole Yunger Halpern, Andrew J P Garner, Oscar C O Dahlsten et al.

[Limits to catalysis in quantum thermodynamics](#)

N H Y Ng, L Maninska, C Cirstoiu et al.

[A measure of majorization emerging from single-shot statistical mechanics](#)

D Egloff, O C O Dahlsten, R Renner et al.

[Thermal machines beyond the weak coupling regime](#)

R Gallego, A Riera and J Eisert

[From single-shot towards general work extraction in a quantum thermodynamic framework](#)

Jochen Gemmer and Janet Anders



PAPER

Thermodynamic work from operational principles

R Gallego¹, J Eisert and H Wilming

Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany

¹ Author to whom any correspondence should be addressed.E-mail: rgallego@zedat.fu-berlin.de**Keywords:** work, quantum thermodynamics, resource theories, monotones, quantum entropies

OPEN ACCESS

RECEIVED
28 April 2016REVISED
4 August 2016ACCEPTED FOR PUBLICATION
2 September 2016PUBLISHED
12 October 2016

Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Abstract

In recent years we have witnessed a concentrated effort to make sense of thermodynamics for small-scale systems. One of the main difficulties is to capture a suitable notion of work that models realistically the purpose of quantum machines, in an analogous way to the role played, for macroscopic machines, by the energy stored in the idealisation of a lifted weight. Despite several attempts to resolve this issue by putting forward specific models, these are far from realistically capturing the transitions that a quantum machine is expected to perform. In this work, we adopt a novel strategy by considering arbitrary kinds of systems that one can attach to a quantum thermal machine and defining *work quantifiers*. These are functions that measure the value of a transition and generalise the concept of work beyond those models familiar from phenomenological thermodynamics. We do so by imposing simple operational axioms that any reasonable *work quantifier* must fulfil and by deriving from them stringent mathematical conditions with a clear physical interpretation. Our approach allows us to derive much of the structure of the theory of thermodynamics without taking the definition of work as a primitive. We can derive, for any *work quantifier*, a quantitative second law in the sense of bounding the work that can be performed using some non-equilibrium resource by the work that is needed to create it. We also discuss in detail the role of reversibility and correlations in connection with the second law. Furthermore, we recover the usual identification of work with energy in degrees of freedom with vanishing entropy as a particular case of our formalism. Our mathematical results can be formulated abstractly and are general enough to carry over to other resource theories than quantum thermodynamics.

1. Introduction

The advent of highly-controlled experiments with small-scale quantum devices compels us to explore the technological perspective to use such devices as *thermal machines* or *heat engines* [1–5]. For this reason it is becoming increasingly important to precisely understand for which tasks quantum machines are useful and how to quantify their performance in such tasks.

Within macroscopic thermodynamics, it is well established what the relevant tasks are that a thermal machine can perform. They can, for example, perform mechanical work, chemical work or heat up and cool down a system away from the environment temperature. One way to describe those tasks in general terms is to observe that the machine induces a transition on a system that is attached to it. For mechanical work, we attach to the thermal machine a weight and the task is simply to store potential energy by lifting it. In the case of cooling, we attach a system in the Gibbs state and the task is to lower its temperature. In any of these tasks, if successfully carried over, we will say that the machine performs work and accordingly, we will refer here to the system attached as the *work-storage device*.

In contrast, the situation is much less clear when considering microscopic thermal machines in physical situations in which quantum effects are expected to be relevant [6]. Even in this regime, one may readily conceive quantum analogues of a lifted weight [7–12]. At the same time, it still makes sense to think of refrigerators [13–16] in the quantum regime, but it remains conceptually more challenging to justify such notions as reflecting the

behaviour of realistic machines at the nano-scale. By taking the idea seriously that quantum effects are expected to play a role, one should take into account the possibility that the work-storage device itself may be composed of a few atoms only. At this scale and in this regime systems are neither expected to fully adequate to the model of a lifted weight, nor would they necessarily be well described by Gibbs states, so that it may be not possible to assign a temperature a system is heated up to or cooled. Indeed, it can still be considered largely unclear what kind of tasks microscopic thermal machines will one day be used for. Can we nevertheless decide on general grounds when a quantum thermal machine can be said to perform work?

In this work, we affirmatively answer this question by introducing a fresh approach towards addressing the problem of deciding when a thermal machine performs work, and of quantifying the amount by which this is done. We allow for fully general classes of work-storage devices—going beyond those considered in [7–12]—and deduce the basic set of properties that a function quantifying work has to fulfil. With this, we are establishing a novel framework of how to define useful tasks for general quantum thermal machines. In particular, the framework introduced is general enough to recover the known notions of mechanical work and the ones relating to heating and cooling tasks within our formalism.

Conceptually more importantly still, we will adopt a strictly operational perspective. Rather than aiming at defining work as an *a priori* given quantity, we will advocate the mindset of first carefully and precisely stating what operational properties any quantity reasonably measuring work for arbitrary classes of work-storage devices should satisfy. We cast these requirements into the form of basic operational axioms that we expect a measure of work to fulfil. From these elementary axioms, we derive surprisingly stringent and specific conditions to the relevant work quantifier. In this way, we approach the question of defining work from an entirely new angle.

Our approach builds upon and further develops ideas from *quantum resource theories* [17–21]; as a consequence of this, we obtain several results on resource theories interesting in their own right. In particular, our results highlight the role of correlations and so-called *catalysts* [22–24]. Our approach also draws some inspiration from the *axiomatic approach to thermodynamics* put forth in the seminal [25], in which axiomatics for thermodynamic state transformations is introduced, even though the object of study here being quite different.

2. Motivation

Consider the situation in which a complex experimental device is claimed to be a thermal machine, in the sense that it is capable of undergoing some evolution—possibly employing heat baths, fuel, or moving parts—in order to ultimately perform a task useful from a thermodynamic perspective. Within the realm of classical thermodynamics what is considered ‘useful’ is very much established, and it is also clear how one would quantify the degree to which a device is useful. Namely, without the need to know anything about the internal details of the machine, one demands that the machine is proven useful by achieving some pre-defined task: performing mechanical work and refrigerating or heating a system are the main examples of such tasks.

In order to offer a unifying perspective within the realm of classical thermodynamics, one can describe any such tasks by introducing a system that, when attached to the machine, undergoes a transition between two states of a given family. For instance, in the task of refrigerating or heating, one attaches a system that undergoes a transition between Gibbs states of different temperatures. In the task of performing mechanical work, the system can usually be described by a lifted-weight, certainly in some degree of abstraction: this is a body in a conservative force with a state described by a deterministic state-variable (for instance the height), that undergoes a transition between two values of the height. Similarly, one can also consider notions of work such as chemical work, that is, charging a battery.

To use a unifying nomenclature, we will say that when carrying out any such useful tasks—be it cooling, heating, producing mechanical work, chemical work or any other conceivable task—the machine *performs work*. Also, let us refer to the attached system that undergoes the transition as the *work-storage device*. Whenever we use the word *work* in the following, we mean it in this general sense, unless specified otherwise.

Importantly, note that those tasks mentioned above are very well motivated from a practical and operational perspective. Indeed, a realistic macroscopic machine can be put to perform work by literally hanging a weight to be lifted, charging a battery or by refrigerating or heating systems that are well-described by Gibbs states. Hence, those models are pertinent to the real experimental situations at hand.

Let us turn now to the arena of quantum thermodynamics. Consider a completely analogue situation as the one described above: we would like to define tasks that determine whether a machine performs a useful task, with the sole difference that in this case the experimental device operates at a very small scale—say where the work-storage device is composed of few atoms. This question has been widely studied in recent years, by putting forward models that capture microscopic versions of a lifted-weight [7–12] or a refrigerator or heater, respectively [13, 16]. Although these models offer a consistent way to describe tasks that a quantum machine can

perform, at the same time it is true that the motivation from a practical and operational perspective is much less obvious. The reason being that a realistic work-storage device made of a few atoms will not be well-modelled—or at least not in many conceivable experimental situations—precisely by a lifted-weight or even by systems prepared in Gibbs states. These two paradigmatic models are ubiquitous in the descriptions of macroscopic machines, but it should be clear that at the quantum scale it is certainly not sufficient to capture all those operations that a quantum thermal machine is expected to perform. Hence, it seems of utmost relevance to characterise in a general fashion how one can identify, as well as quantify, the useful tasks that a quantum machine is capable of performing for general kinds of work-storage devices.

The goal of this work is precisely to characterise those functions that quantify the usefulness of the device as a thermal machine. These functions should be applicable in all generality, and in particular be sufficient to capture general families of work-storage devices. In other words, we are identifying which functions play the role of the energy stored in the lifted-weight or battery, or for instance, the temperature change when the work-storage device is heated or cooled. These functions—as we will see throughout this manuscript—share all the relevant features that the definition of work for classical mechanics has: they serve to test the ‘usefulness’ of the machine, they are maximised in reversible processes and they fulfil a second law. Furthermore, these functions reduce to the usual notions of mechanical work when the work-storage device takes the form of a lifted-weight, but also quantify tasks such as cooling or heating. This justifies that we call these functions *work quantifiers*, not innocently, but because they can be regarded, for the reasons above, as a generalisation of the notion of work for situations where the usual models are not in place.

Lastly, let us point out that the notion of work quantifiers that is put forward not only allows for conceiving thermodynamic tasks in the quantum regime, but it also offers a unifying way of quantifying any thermodynamical task: let it be heating or refrigerating, mechanical energy production or any other conceivable task, one can find a function that serves as a suitable quantifier for all of them.

2.1. Usefulness in thermodynamics

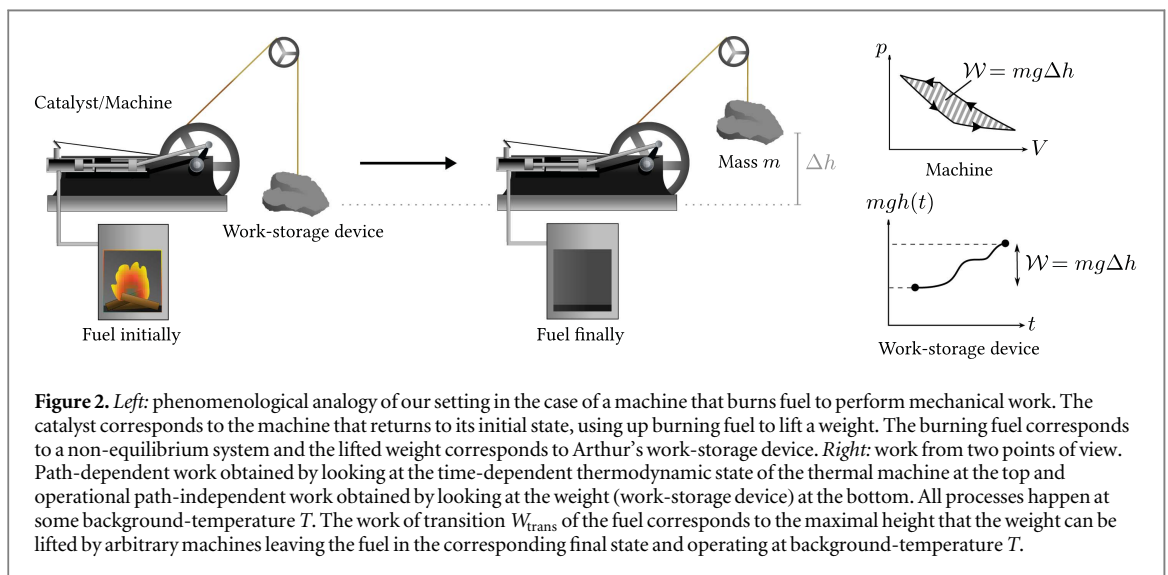
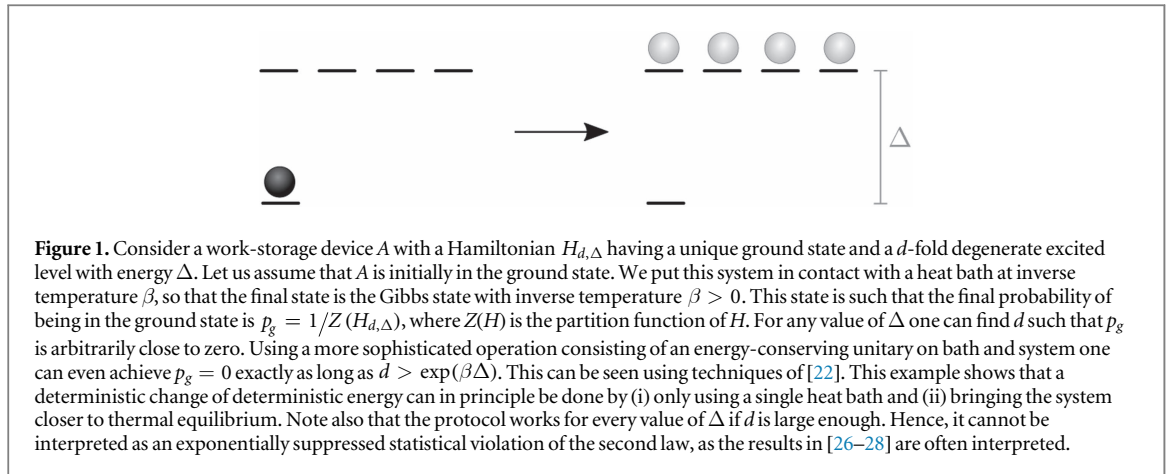
The goal in the following is to find a suitable way—given by the work quantifier—to measure the usefulness of a thermal machine in the most general case, as has been pointed out above. In this section we will present a simple example that illustrates what precisely is meant by the term ‘useful’ and why finding a quantifier for arbitrary families of states constitutes such a challenging problem.

First, let us introduce a rough definition of what is considered useful in the context of thermodynamics (for a more technical definition see sections 4 and section 3). Consider, as mentioned above, a machine that induces a transition in a work-storage device. Suppose that this machine operates in a background environment at temperature T (which can be thought as being the reference temperature of the air surrounding the machine). Then, it is reasonable to state the following.

Performing work: consider a machine inducing a transition on the work storage device. If such transition could have been also done by having the machine only using systems at equilibrium temperature T and leaving any other system involved in the process in the same state as it was at the beginning; then we conclude that the machine did *not* induce a thermodynamically useful transition. In this case, we say that the machine did not perform positive work.

This definition of usefulness paraphrases the second law of thermodynamics, which states that a machine cannot perform positive work in a cyclic process (leaving any other system in the same state in the formulation above) by using a single heat bath (systems at equilibrium temperature T above)¹. It is in this sense that it is appropriate to refer to ‘usefulness’ simply as work, as we do here. Given this definition of what performing work is and its relation with the second law, it might seem surprising that we argue here that the question at hand can be viewed as largely unsettled. One might be tempted to argue that, if not all the zoology of work quantifiers, at least it is well established that extracting energy deterministically (that is, with unit probability) must be a task that proves the machine useful, hence a valid work quantifier. We will now show that this is not necessarily the case, in that deterministic energy is not always a valid work quantifier. This is because it may violate our general principle above, and this relates to the fact that the system playing the role of the work-storage device may not be well described by a lifted-weight.

¹ Planck’s formulation understands performing work as ‘raising a weight’. However, the very same formulation is of course true if we understand performing work as ‘cooling down a system’. In this sense, it is impossible to cool down a system at temperature T by putting it in contact with a heat bath also at temperature T in a cyclic process, as well as this it is possible to raise a weight. Unifying these two forms leads to the simplest form considered here, that no machine performs work using a single heat bath.



2.2. An example

Consider a work-storage device (that we refer to as A for conciseness) which is initially in a state at zero energy and finally with deterministic energy $\Delta > 0$. Moreover this change of energy is assumed to happen each time the work-storage device is operated on by the machine, that is, it occurs with unit probability. Is the machine in this case doing something useful, in such a way we can conclude that the machine performs work? We see now at hand of a specific example of a system A , that this is not the case. Suppose that the quantum system A is described by a Hamiltonian which has a unique ground state and a d -fold degenerate excited energy-level with energy Δ . Then it is possible, as discussed in figure 1, to bring A into contact with a heat bath and map the ground state of A to a state with deterministic energy Δ —namely the maximally mixed state within that subspace. The energy Δ was stored deterministically in A , but entirely provided by a heat bath, thus we cannot conclude that the machine is doing anything useful.

This observation seems to contradict well-established notions of work that take the definition of deterministic energy as work as one of its pillars [26–28]. However, note that this example—where deterministic energy is put into the work-storage device by just using a single bath—it is only possible because the system of figure 2 does not relate to the model of the lifted-weight. It is instead a system that acts as an entropy sink at the same time it stores the energy, hence leading to unusual situations that are not present in macroscopic thermodynamics. Of course, we do not claim that the example of figure 1 is expected to model a realistic system. Nonetheless, it illustrates the difficulties emerging when trying to define what performing work is, when work-storage devices are not idealisations of a lifted weight (or its quantum versions [8–12, 16]).

3. The operational framework

The previous considerations motivate our approach to the problem of defining work, in our generalised sense, in terms of an operational viewpoint. We formulate it as a game between two players. The first one is Arthur, who possesses a quantum system which takes the role of the *work-storage device*. The system is described by a pair of a quantum state and a Hamiltonian

$$p = (\rho, H), \quad (1)$$

referred to as *object*. Typically one assigns certain properties to what is considered a valid work-storage system. For instance, as discussed in the introduction, the usual demand is to impose that it is a lifted weight or its microscopic counterparts of [8–12, 16]; or also systems in the Gibbs state if we consider the task of refrigerating or heating. However, here we are precisely interested in considering arbitrary classes of work-storage devices, hence, we will be fully general and encode such constraints by assuming that p belongs to some set \mathcal{P} , which encodes the task the machine has to perform². The system is initially described by $p^{(i)} \in \mathcal{P}$. The second player is referred to as Merlin, who has a machine capable of performing transitions between the initial object $p^{(i)}$ to a final object $p^{(f)} \in \mathcal{P}$. This machine will play the role of the thermal machine or engine, which performs a transformation on the work-storage device.

We assume that the transition $p^{(i)} \rightarrow p^{(f)}$ is performed in an environment of temperature $T = 1/\beta$ (we set $k_B = 1$), which we consider to be fixed throughout the rest of our work. In a resource-theoretic setting, this means that Merlin performs the transitions while having unlimited and free access to arbitrary heat baths at inverse temperature β . The term ‘free’ is here used in the sense of a resource theory, a notion that will be made precise later.

Arthur and Merlin, having agreed on the free character of heat baths and the properties of the work-storage device given by \mathcal{P} , would like to establish a fair way of quantifying the value of a given transition $p^{(i)} \rightarrow p^{(f)}$. That is, they aim at agreeing on a function

$$(p^{(i)} \rightarrow p^{(f)}, \beta) \mapsto \mathcal{W}(p^{(i)} \rightarrow p^{(f)}, \beta) \in \mathbb{R} \quad (2)$$

for any $p^{(i)}, p^{(f)} \in \mathcal{P}$, that establishes the prize at which Merlin sells to Arthur the transition that he has performed, where we will take the convention that $\mathcal{W}(p^{(i)} \rightarrow p^{(f)}, \beta) \geq 0$ implies that Arthur has to pay to Merlin. The prize of the transition is what we define as *work*, and the function \mathcal{W} a *work quantifier*.

Since the temperature of the free heat baths is fixed, we also often write $\mathcal{W}(p^{(i)} \rightarrow p^{(f)})$ for $\mathcal{W}(p^{(i)} \rightarrow p^{(f)}, \beta)$. Apart from the agreement on the free heat baths at inverse temperature β , the work value has to be established solely on the basis of which transition $p^{(i)} \rightarrow p^{(f)}$ is performed by Merlin, without any assumption or restriction on the internal mechanism of Merlin’s device. This is also a property inherited from the usual notion of work within phenomenological thermodynamics, where the work can be quantified by looking only at the initial and final state of the work-storage device, without knowing the details of the internal workings of the machine (also see figure 2). Lastly, we would like to stress that the language making reference to players such as Arthur and Merlin captures the usual thermodynamic setting and bounds the very same quantities usually under consideration in thermodynamics, as explained in section 3.2. Yet, the novel operational framework introduced in this work is most transparently stated in such a language, as a pedagogical tool inspired by common notions of interactive proof systems in theoretical computer science.

3.1. Free catalytic transitions

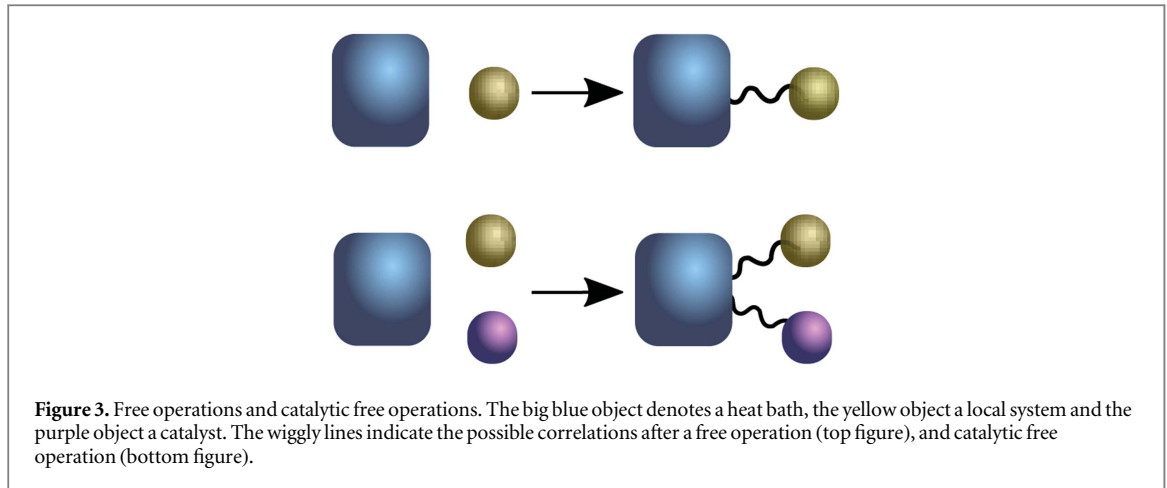
Since the notions and the use of language may be unfamiliar in the quantum thermodynamic context, we will now specify clearly what we mean by free operations in the context of a resource theory. Here, Arthur and Merlin have free access to heat baths at inverse temperature β . This will be relevant for the choice of the function \mathcal{W} , since Arthur will not pay a positive amount for a transitions that can be performed by only employing free resources. That is to say, it is important to specifically characterise the transitions that can be performed without expending valuable resources and only using baths.

Concretely, we assume that both Arthur and Merlin can pick heat baths, that is, quantum systems B prepared in a Gibbs state

$$\omega_{H_B, \beta} = \frac{\exp(-\beta H_B)}{Z_H}, \quad (3)$$

with arbitrary Hamiltonian H_B . They can also apply any global unitary U that commutes with the total Hamiltonian $H + H_B$. We use the short-hand $H_A + H_B := H_A \otimes \mathbb{1} + \mathbb{1} \otimes H_B$ whenever the support of two

² Note that the choice of \mathcal{P} has a subjective element, in the same way it amounts to a restriction to take \mathcal{P} as a lifted weight. It describes a particular choice of systems that we consider valid resources because we can handle them in a given experimental situation. To put an example, one can imagine that a given experimental setup can only handle qubits, systems of bounded entropy or energy, or systems with a fixed Hamiltonian. Our formalism allows to choose \mathcal{P} so that it encodes each of those situations.



operators is clear from the context. This amounts to the formalism of *thermal operations* first introduced in [6] and further studied in [8]. It is also meaningful to allow for more general sets of operations such as the so-called *Gibbs-preserving maps* [6, 29, 30], also see the appendix, or simply *thermalising maps* where the only possible interaction with the thermal bath is to bring the system to the Gibbs state in the spirit of [9, 31, 32]. Undoubtedly, in many thermodynamic settings, the latter one is the most realistic one capturing actual experimental situations. The final form of the work quantifier \mathcal{W} will in principle depend on which model of operations with the bath is chosen, but the formalism is general enough to be applicable in any of these situations. In the appendix we discuss in detail which are the minimal properties of the free operations that are explicitly needed to derive our results and show that the examples presented above have such properties.

More importantly, we will assume that both Arthur and Merlin, in addition to the heat bath, can also borrow any quantum ancillary system uncorrelated with the bath and the work-storage device, as long as it is returned in the same initial state and also uncorrelated with the work-storage device (see figure 3). This ancillary system is referred to as a *catalyst*, and its usage extends the set of transitions that can be performed with a bath [22, 23]. Such catalytic operations have been frequently studied in the recent literature of quantum thermodynamics, and naturally capture ‘bystanders’, so auxiliary systems that may help performing transformations. In the following, we will refer to the operations described in this section as *free operations* when done without catalyst and *catalytic free operations* when performed with catalyst. Similarly, we will refer to the transitions induced by free and catalytic free operations as *free transitions* and *catalytic free transitions*, respectively. Lastly, given any object p , we define $\mathcal{F}(p)$ as the set of objects that can be reached from p by free operations. Similarly, $\mathcal{F}_C(p)$ denotes the set of objects that can be reached from p by catalytic free operations. More precisely $p \in \mathcal{F}_C(p)$ if there exist a catalyst q so that $p \otimes q \in \mathcal{F}(p \otimes q)$. See appendix A.1 for more detailed definitions.

3.2. The thermodynamical reading of the operational framework

The game between Arthur and Merlin that we have introduced encodes a typical situation in the study of thermal machines, but does not describe it in the canonical way. The canonical analysis in the literature describes a thermal machine as composed by at least the following elements (see also figure 2).

- (i) A heat bath at inverse temperature β ,
- (ii) a system M out of equilibrium, i.e. not in the Gibbs state (3) at the temperature of the heat bath. We will refer to this system as the *fuel*, because is the resource that allows one to extract work,
- (iii) The *work-storage device* A on which the thermal machine acts,
- (iv) and possibly but not necessarily a catalyst C .

Our game formulates the problem of evaluating the work given a transition of A , that is, evaluating how much work has been performed on the work-storage device. In the language of our game, M would be any system that Merlin employs inside his machine performing the transitions and that is possibly modified (burnt) after a protocol of work extraction, i.e., it plays the role of the ‘fuel’ in traditional thermodynamics. Such three-partite (four-partite, if the catalyst is explicitly included) structure is indeed the one followed in [7, 9, 18, 29, 31, 33–40] where the function \mathcal{W} is defined *a priori* as related to the energy difference in various ways. For instance, in [7, 38] \mathcal{P} is taken as the quantum analogue of a lifted weight and the work quantifier is defined as

$$\mathcal{W}_{\text{mean}}(p_A^{(i)} \rightarrow p_A^{(f)}) = \text{tr}(\rho_A^{(f)} H_A^{(f)}) - \text{tr}(\rho_A^{(i)} H_A^{(i)}). \quad (4)$$

Another relevant model is the case of the ϵ -*deterministic work*, following the approach in [18]. It also gives rise to an instructive example of how the constraints on the sets of allowed objects \mathcal{P} come into play. The systems that are considered useful in that context are qubits such that

$$\mathcal{P}_\epsilon := \{(\rho, H) \mid H = \Delta|1\rangle\langle 1|, \|\rho - |E\rangle\langle E|\| \leq 2\epsilon\} \quad (5)$$

where $|E\rangle$ is an eigenvector of H , $\|\cdot\|$ is a norm on quantum states and $\epsilon < \frac{1}{2}$. The restriction \mathcal{P}_ϵ encodes that Arthur is interested in having states of well-defined energy or at least ϵ -close to it. Work is then given by the energy difference of the closest energy-eigenstates. Formally as

$$\mathcal{W}_{\text{det}}(p_A^{(i)} \rightarrow p_A^{(f)}) = f(\rho_A^{(f)}, H_A^{(f)}) - f(\rho_A^{(i)}, H_A^{(i)}), \quad (6)$$

with the function f being defined³ as

$$f(\rho, H) = \begin{cases} \Delta & \text{if } \|\rho - |1\rangle\langle 1|\| < 1 \\ 0 & \text{if } \|\rho - |0\rangle\langle 0|\| < 1. \end{cases} \quad (7)$$

As heuristically discussed in the introduction, both $\mathcal{W}_{\text{mean}}$ and \mathcal{W}_{det} and the limitations that they impose \mathcal{P} can be regarded as particular cases of the general framework that we put forward. Nonetheless, they serve to illustrate the mathematical objects we are concerned with. The precise way they are recovered as particular cases and the subtleties that emerge when doing so will be discussed in appendix E.

3.3. Work of transition, work cost and work value

It is important to distinguish \mathcal{W} as a work quantifier on the work-storage device from other quantities that are usually analysed in thermodynamics and referred to as *work*. Once \mathcal{W} is defined, one is then usually concerned, in the language of the present work, with the optimal amount of work that Merlin can obtain by performing a transition on his system M . This quantity has also been considered in the context of single-shot work extraction in [10, 12, 36]. Here we will refer to this function as *work of transition*.

Definition 1 (Work of transition). Given a work quantifier \mathcal{W} and inverse temperature β , a set of restrictions \mathcal{P} , and initial and final objects of M , denoted by $p_M^{(i)}$ and $p_M^{(f)}$, respectively, the work of transition $W_{\text{trans}}(p_M^{(i)} \rightarrow p_M^{(f)}, \beta)$ is defined as

$$\begin{aligned} W_{\text{trans}}(p_M^{(i)} \rightarrow p_M^{(f)}, \beta) \\ := \sup_{\substack{p_A^{(i)}, p_A^{(f)} \in \mathcal{P}; \\ p_M^{(f)} \otimes p_A^{(f)} \in \mathcal{F}_C(p_M^{(i)} \otimes p_A^{(i)})}} \mathcal{W}(p_A^{(i)} \rightarrow p_A^{(f)}, \beta). \end{aligned} \quad (8)$$

In (8), we have introduced the short hand notation

$$p_M^{(i)} \otimes p_A^{(i)} := (\rho_M^{(i)} \otimes \rho_A^{(i)}, H_M \otimes \mathbb{1}_A + \mathbb{1}_M \otimes H_A). \quad (9)$$

Also recall that the set $\mathcal{F}_C(p_M^{(i)} \otimes p_A^{(i)})$ is determined by all those free transitions that Merlin can perform, including interactions with a heat bath and the catalyst.

Notice that in contrast to \mathcal{W} , W_{trans} is evaluated on transitions on M and *not* on the work-storage system. In fact, the work associated to a given physical process can never be specified as a transition on M alone. That is, it is impossible to determine \mathcal{W} as a function of $p_M^{(i)} \rightarrow p_M^{(f)}$. One needs to either specify a particular catalytic free transition leading to $p_M^{(f)} \otimes p_A^{(f)} \in \mathcal{F}_C(p_M^{(i)} \otimes p_A^{(i)})$, or to simply define it by taking the optimal one, as we do in (8). It is precisely in this sense that work, as a function of transitions on M , is said to be a path dependent quantity when evaluated in transitions on M , and a path-independent quantity when evaluated in transitions on A . This is also the case in phenomenological thermodynamics: work can be specified by knowing only the initial and final height of the lifted weight, however it is path-dependent as function of the state of the machine (see figure 2).

One may be tempted to think at this point that the distinction between the work of transition W_{trans} and the work quantifier at the work-storage device \mathcal{W} is somehow artificial. In the end, both quantities evaluate transitions on systems and which player plays the role of Arthur or Merlin might seem at first sight arbitrary. However, let us insist that by no means Arthur and Merlin play an equivalent role. The key point is to understand that the transitions on Arthur systems are restricted so that $p_A^{(i)}, p_A^{(f)} \in \mathcal{P}$. However, transitions on Merlin are fully unrestricted. This is possibly most transparent in phenomenological thermodynamics: there, the work

³ In [18] it is assumed that $\rho_A^{(i)}$ is the ground state of H_A . Hence, we have $f^\epsilon(\rho_A^{(i)}, H_A^{(i)}) = 0$. What we introduce here as \mathcal{W}_{det} is a generalisation that follows in spirit, but lies outside the scope formally laid out in [18].

storage device undergoes a transition between two states of definite energy (a weight), however, the ‘fuel’ employed in the process may undergo arbitrary transitions.

Lastly, note that in (8) we demand that the final state of MA is uncorrelated. As discussed in [24], the creation of correlations can be a resource for performing thermodynamical transitions. Indeed, those correlations between MA will turn out to play an important role in our axiomatic formulation. Hence, for reasons of clarity of presentation, we will first consider the case where no correlations are allowed, as in (8), and study extensively the role of correlations in section 8.

Yet other relevant quantities in thermodynamics are given by the so-called *work value* and *work cost* defined as

$$W_{\text{value}}(p_M, \beta) := W_{\text{trans}}(p_M \rightarrow \omega_\beta, \beta), \quad (10)$$

$$W_{\text{cost}}(p_M, \beta) := -W_{\text{trans}}(w_\beta \rightarrow p_M, \beta), \quad (11)$$

where w_β is an object describing a thermal state. The quantity W_{value} plays a relevant role since the second law is usually put as a bound on it. It describes how much work can be extracted from the system if it is viewed as a resource. In this work we are, however, mainly concerned with the form that \mathcal{W} can take given a set of axioms. Clearly, the quantities W_{trans} , W_{value} and W_{cost} can only be defined once \mathcal{W} has been specified. We will show in section 7, though, that from the general properties of \mathcal{W} implied by the axioms, we can find a second law as $W_{\text{value}} \leq W_{\text{cost}}$.

4. Two basic axioms

We are now in the position to formulate the basis on which all of the following analysis rests. We introduce two operational axioms concerning the work quantifier \mathcal{W} . They seem as innocent as they are natural, and clearly capture features that any reasonable function of the above type quantifying work should satisfy. They are physically very intuitive. In order to precisely develop our operational framework, they will be formulated in the mindset of the game played by Arthur and Merlin, in the language of a so-called interactive proof system. In this language, they simply ensure that none of the players can get arbitrarily rich without expending valuable resources. We will carefully discuss all implications of our results, however, also in a physical language, stressing that the conclusions we draw indeed give rise to a natural framework for naturally grasping concepts of work in quantum thermodynamics.

Axiom 1 (Cyclic transitions of the work-storage device). For any cyclic sequence of transitions of the work-storage device $p_A^{(1)} \rightarrow p_A^{(2)} \rightarrow \dots \rightarrow p_A^{(n)} = p_A^{(1)}$, such that $p^{(i)} \in \mathcal{P} \ \forall i$, the sum of the work-values of the individual transitions is larger than or equal to zero,

$$\sum_{i=1}^{n-1} \mathcal{W}(p_A^{(i)} \rightarrow p_A^{(i+1)}, \beta) \geq 0. \quad (12)$$

According to our convention, if \mathcal{W} takes a negative value, then Arthur is benefiting from the transaction, i.e. Merlin pays to Arthur. Hence, the previous axiom ensures that—taking the simplest case—Arthur cannot get rich by demanding Merlin to first do a transition $p^{(1)} \rightarrow p^{(2)}$ and then asking from him to undo the transition. If this principle was violated, Arthur could get infinitely rich just by repeatedly interacting with Merlin. Note that Arthur is not even himself implementing the transition, hence, he is by definition not expending any resource.

We will now impose our second axiom, which ensures in turn that Merlin cannot get arbitrarily rich without spending resources.

Axiom 2 (Cyclic transitions of the fuel). For any cyclic sequence of transitions of the fuel (Merlin’s system) $p_M^{(1)} \rightarrow p_M^{(2)} \rightarrow \dots \rightarrow p_M^{(n)} = p_M^{(1)}$, the sum of the optimal work that Merlin can obtain in each sequence (this is given by W_{trans} in (8)) is smaller or equal to zero,

$$\sum_{i=1}^{n-1} W_{\text{trans}}(p_M^{(i)} \rightarrow p_M^{(i+1)}, \beta) \leq 0. \quad (13)$$

According to our convention and (8), if W_{trans} takes a positive value, then Merlin is benefiting from the transaction, i.e. Arthur pays to Merlin. Hence, the previous axiom ensures that—taking the simplest case—Merlin cannot get rich by the overall process of burning his fuel in a transition $p_M^{(1)} \rightarrow p_M^{(2)}$ and then restoring the fuel back to its original state $p_M^{(1)} \rightarrow p_M^{(2)}$. If this was violated, that is, the l.h.s. of (13) was positive, then Merlin would get rich while not having burnt any fuel. It is important to notice that the objects $p_M^{(i)}$ of axiom 2 are *not* restricted to be in \mathcal{P} , since these restrictions apply to the work-storage device, and here we are concerned with

transitions on the fuel (Merlin's systems) which are fully unrestricted. However, axiom 2 depends on \mathcal{P} because W_{trans} is defined as a function of \mathcal{P} and \mathcal{W} , as given by (8). Also, note that axioms 1 and 2 allow—and this will be indeed the case as discussed in section 7.1—for the l.h.s. of equation (13) to be strictly smaller than zero.

As a final remark, note that both axioms 1 and 2 encode the spirit of the second law of thermodynamics: by preventing any of the two players to become arbitrarily rich without spending resources, we are enforcing the impossibility to create a *perpetuum-mobilé*. Our approach is, however, inverse to what is usually found in phenomenological thermodynamics. Therefore, work is defined *a priori* and the second law is understood as a constraint on the possible physical processes. In contrast, axioms 1 and 2 do not impose any constraint on the allowed physical operations that Merlin is performing. They simply state that one does *not* account as work what can be generated with a bath and a catalyst with the *a priori* given physical operations. As such, in our set-up it is also *impossible to violate the second law*: if by using, say, a forthcoming post-quantum theory, someone claimed to extract work from a single heat bath, then it simply means—regardless of the details of such theory—that what it is referred to as work does not fulfil our operational principles.

5. General properties of the work quantifier

It is the key feature of the framework developed here that very basic principles already allow one to formulate stringent properties of possible work functions \mathcal{W} . In this section we will turn to discussing properties of a work-function \mathcal{W} that respects axioms 1 and 2. For conceptual clarity, we will keep the discussion rather informal in this section. For a mathematically detailed and rigorous treatment, we refer to the appendix. Nevertheless, we will have to introduce some notation and definitions first. We are looking for a function \mathcal{W} that assigns a real number to any pair of objects $p^{(i)} = (\rho^{(i)}, H^{(i)})$ and $p^{(f)} = (\rho^{(f)}, H^{(f)})$ that belongs to the given set \mathcal{P} . Such an inclusion is assumed throughout the remaining, unless explicitly mentioned otherwise. We will use Latin letters p, q, r, s, \dots to denote objects and denote the work-value of a transition $p^{(i)} \rightarrow p^{(f)}$ as $\mathcal{W}(p^{(i)} \rightarrow p^{(f)}, \beta)$ or simply $\mathcal{W}(p^{(i)} \rightarrow p^{(f)})$ if β is clear from the context. If the Hamiltonian of the two objects is identical, we will also use the notation $\mathcal{W}(\rho^{(i)} \rightarrow \rho^{(f)})$. Let us recall from previous sections that given any object p , we define $\mathcal{F}(p)$ and $\mathcal{F}_C(p)$ as the set of objects that can be reached from p by free operations and catalytic free operations, respectively. Also, in the following we always assume the existence of the *empty object* $\emptyset \in \mathcal{P}$. Physically it means that there is no work-storage device. Formally it is given by the state 1 with Hamiltonian 0 on the Hilbert-space \mathbb{C} . It therefore fulfils $p \otimes \emptyset = \emptyset \otimes p = p$ for any $p \in \mathcal{P}$.

Theorem 2 (Form of work quantifiers). *A function \mathcal{W} respects axioms 1 and 2 if and only if it can be written as*

$$\mathcal{W}(p \rightarrow q) = M(q) - M(p), \quad (14)$$

for a function M such that $M(\emptyset) = 0$ and that fulfils the following property:

- Additive monotonicity: *for all $p^{(1)}, \dots, p^{(m)}$ and $q^{(1)}, \dots, q^{(m)}$ in \mathcal{P} such that $\bigotimes_{i=1}^m q^{(i)} \in \mathcal{F}_C(\bigotimes_{i=1}^m p^{(i)})$*

$$\sum_{i=1}^m M(q^{(i)}) \leq \sum_{i=1}^m M(p^{(i)}). \quad (15)$$

In particular, the theorem implies that work, *as measured by the work-storage device*, is not path-dependent in the sense that

$$\mathcal{W}(p \rightarrow q) + \mathcal{W}(q \rightarrow s) = \mathcal{W}(p \rightarrow s), \quad (16)$$

$$\mathcal{W}(p \rightarrow q) = -\mathcal{W}(q \rightarrow p) \quad (17)$$

and that no work can be extracted in a free catalytic transition,

$$\mathcal{W}(p \rightarrow q) \leq 0, \quad \forall q \in \mathcal{F}_C(p). \quad (18)$$

Thus the work-storage device can be treated similarly to the case of a massive body under the influence of a conservative force in classical mechanics: there is a state-variable M and its difference along a transition determines the work-value of the transition. Using catalytic free operations, which generalise the concept of putting a system in contact with a heat bath in phenomenological thermodynamics, this state-variable cannot be increased. Note that this is even true when considering tasks such as heating or cooling in which the system undergoes a transition between two states with fluctuating energy and possibly large amounts of entropy.

Let us highlight that condition (17) is perfectly compatible with the well-known notion of irreversibility that emerges when considering notions of deterministic work in the spirit of [9, 31, 35–37, 41]. That is, (17) is

compatible with $W_{\text{value}} < W_{\text{cost}}$ and more generally with

$$W_{\text{trans}}(p_M^{(i)} \rightarrow p_M^{(f)}) \neq -W_{\text{trans}}(p_M^{(f)} \rightarrow p_M^{(i)}). \quad (19)$$

The validity of (19) for any nontrivial set \mathcal{P} and an extended discussion on the implications of equation (17) are discussed in section 7.1. The proof of theorem 2 can be found in appendix B.

Let us now discuss briefly the significance of *additive monotonicity*, with the following lemma.

Lemma 3 (Consequences of additive monotonicity). *If a function M fulfils additive monotonicity and $M(\emptyset) = 0$, then it fulfils also the following properties.*

- Monotonicity: $M(q) \leq M(p) \forall p, q \in \mathcal{P}$, such that $q \in \mathcal{F}_C(p)$.
- Additivity: $M(p_A \otimes p_B) = M(p_A) + M(p_B) \forall p_A, p_B, p_A \otimes p_B \in \mathcal{P}$.
- Positivity: $M(p) \geq 0 \forall p \in \mathcal{P}$.

Nonetheless, additive monotonicity is strictly stronger than demanding that M fulfils the three conditions of previous Lemma. To see this, consider for example objects $p_A^{(i)}, p_A^{(f)}, p_B^{(i)}, p_B^{(f)} \in \mathcal{P}$, but with $p_A^{(i)} \otimes p_B^{(i)}, p_A^{(f)} \otimes p_B^{(f)} \notin \mathcal{P}$. Note that monotonicity and additivity do not apply to objects that are not in the set \mathcal{P} . The condition given by (B.17) implies that

$$M(p_A^{(f)}) + M(p_B^{(f)}) \leq M(p_A^{(i)}) + M(p_B^{(i)}) \quad (20)$$

if $p_A^{(f)} \otimes p_B^{(f)} \in \mathcal{F}_C(p_A^{(i)} \otimes p_B^{(i)})$. However, this condition could not have been derived from the conditions of monotonicity and additivity, since they do not apply to objects that lie outside the set \mathcal{P} .

6. Free energies as work quantifiers

At this point a most natural question emerges: what are reasonable and natural candidates for a work quantifier fulfilling axioms 1 and 2? Clearly, the set of valid functions \mathcal{W} will crucially depend on the set of allowed states \mathcal{P} . Now, we will show that if the \mathcal{P} is fully unrestricted, then the conditions simplify to the well-known notions of monotonicity and additivity.

Theorem 4 (Work qualifiers in the unrestricted case). *If the set $\mathcal{P} = \{(\rho, H)\}$ is the set of all quantum states ρ and Hamiltonians H , then a function \mathcal{W} respects axioms 1 and 2 if and only if it can be written as*

$$\mathcal{W}(p \rightarrow q) = M(q) - M(p) \quad (21)$$

for a function M with $M(\emptyset) = 0$ such that

$$M(q) \leq M(p) \forall q \in \mathcal{F}_C(p), \quad (22)$$

$$M(p \otimes q) = M(p) + M(q). \quad (23)$$

Proof. One direction follows directly from theorem 2 and lemma 3. That is, by using theorem 2 we have that the function M fulfils additive monotonicity. Hence, by lemma. 3 one sees that if fulfils also the two properties of lemma 4. To show the inverse relation, it suffices to show that (22) and (23) imply additive monotonicity. Indeed, by taking $p = \bigotimes_{i=1}^n p^{(i)}$ and $q = \bigotimes_{i=1}^n q^{(i)}$, we have that (22) implies

$$M\left(\bigotimes_{i=1}^n q^{(i)}\right) \leq M\left(\bigotimes_{i=1}^n p^{(i)}\right) \quad (24)$$

if $\bigotimes_{i=1}^n q^{(i)} = \mathcal{F}_C(\bigotimes_{i=1}^n p^{(i)})$. Additive monotonicity follows straightforwardly applying now (23). \square

Theorem 4 implies that any function M fulfilling the properties (22) and (23) (when appropriately re-scaled to fulfil $M(\emptyset) = 0$) can be used to build valid work quantifier for every possible set \mathcal{P} . Here we present a natural family of such monotones.

Theorem 5 (Work quantifiers from Rényi divergences). *If the free transitions \mathcal{F} are given by Gibb-preserving transitions—see definition 27—or any subset of them, then the work quantifier $\mathcal{W}_\alpha(p \rightarrow q, \beta) := \Delta F_\alpha^\beta(q) - \Delta F_\alpha^\beta(p)$ with*

$$\Delta F_{\alpha}^{\beta}(p = (\rho, H)) = \frac{1}{\beta} S_{\alpha}(\rho || w_{H,\beta}), \quad (25)$$

where S_{α} is the quantum Rényi divergence [42, 43], for any $\alpha > 0$, fulfils axioms 1 and 2 for every set \mathcal{P} .

Note that although this theorem puts conditions on the free transitions \mathcal{F} , the axiom 2 imposes that this transitions can be supplied by a catalyst, through definition 1. The proof of this statement follows from the fact that Rényi divergences satisfy (22) and (23)⁴. From all the possible choices of α , the case of $\alpha = 1$ corresponding to the free energy based on the standard von-Neumann entropy plays a crucial role that will be discussed in section 8.

7. The second law and irreversibility

We will now turn to discussing the connection of our framework developed here and quantitative second laws of thermodynamics that emerge from it. In the language introduced here, such second laws are captured by the work-value of an object W_{value} being necessarily smaller than or equal to its work-cost W_{cost} (defined in (10) and (11), respectively). As already discussed after the formulation of axioms 1 and 2, the axioms already impose that the definition of work must not allow for either of the players to get arbitrarily rich, which in spirit encodes the second law of thermodynamics. Indeed, this intuition can be made explicit by noting that axiom 2, if we take $p_M^{(1)} = p = p_M^{(3)}$ and $p^{(2)} = \omega$ (where ω is any thermal object), implies

$$0 \geq W_{\text{trans}}(p \rightarrow \omega) + W_{\text{trans}}(\omega \rightarrow p) \quad (26)$$

which together with (10) and (11) imply the second law in the form

$$W_{\text{value}}(p) \leq W_{\text{cost}}(p). \quad (27)$$

We will now discuss the exact conditions when we can expect to get a strict inequality, which is a phenomenon usually referred to as *irreversibility* and that emerges in all the analyses of deterministic work (also called single-shot work extraction) [9, 31, 35–37, 41]. We will see that this will depend crucially on the restrictions that are imposed over the work-storage device given by the set \mathcal{P} .

7.1. Restrictions imply irreversibility

Let us first consider the case where no restrictions are imposed on the form of the work-storage device, that is, \mathcal{P} is the set of all pairs of states and Hamiltonians. It is then maybe not surprising that *reversibility* arises, in the sense that $W_{\text{value}} = W_{\text{cost}}$ is true. The reason for this is simple: the best strategy that Merlin can employ to extract work from an object p_M is just giving the system to Arthur. In this case the transition on A is given by $\emptyset \rightarrow p_A = p_M$ and thus the work is given simply by $M(p_M)$. The same is true in the case of the work-cost of the object. Merlin can just create p_M by taking it from Arthur. Hence, summarizing, we see that if \mathcal{P} is unrestricted we find

$$W_{\text{value}}(p_M) = M(p_M) = W_{\text{cost}}(p_M) \quad (\mathcal{P} \text{ unrestricted}).$$

At a more heuristic level, we have seen that the tasks involving thermodynamical work become trivial when no restrictions are imposed on \mathcal{P} , since the entire process reduces to Merlin giving the physical system he possesses to Arthur. This comes with no surprise if we think about the analogue situation in phenomenological thermodynamics. Take for instance the task of performing mechanical work, that is, storing potential energy on a lifted-weight. If Merlin has as a resource some instance of ‘burning fuel’, then he cannot simply give it to Arthur, expecting that the latter accepts it as a valid form of mechanical work. In other words, Arthur would demand to receive a work-storage device in \mathcal{P} , which are the set of lifted weights. Otherwise, if \mathcal{P} was completely unrestricted, giving to Arthur simply the burning fuel as such—and also all the other parts of the machine—would be indeed the best strategy for Merlin. Any other strategy would involve interactions with a heat bath, which would necessarily decrease the value of the burning fuel as measured by any monotone function.

That said, the limitations on the set \mathcal{P} , rather than being a technicality, impose the very conditions so that non-trivial thermodynamical processes take place: merlin will now have to transform resources $p_M \notin \mathcal{P}$ into resources that are in \mathcal{P} , possibly at the prize of dissipating the resource partially, which in turn yields irreversibility of the form $W_{\text{value}} < W_{\text{cost}}$. To illustrate this point, we will discuss in detail particular examples of restrictions.

⁴ See [42] for an exhaustive analysis of Rényi divergences. Indeed, for theorem 5 we only need that the function $S_{\alpha}(\cdot || \cdot)$ fulfils the data-processing inequality and additivity. The former implies $\Delta F_{\alpha}^{\beta}(\mathcal{F}_C(p)) \leq \Delta F_{\alpha}^{\beta}(p)$.

7.2. Examples of restrictions: redefining ϵ -deterministic work

Let us discuss some examples of meaningful restrictions that can be imposed on the set of states and Hamiltonians and see how they led to irreversibility in the form of $W_{\text{value}} < W_{\text{cost}}$. In section 3.2 we have already briefly introduced the notion of ϵ -deterministic work. Intuitively it describes the situation of work-storage devices which are almost in energy-eigenstates and where work is measured in terms of the energy-difference of these eigenstates. The original formulation of ϵ -deterministic work introduced in [8] does not qualify for a valid work quantifier respecting axioms 1 and 2, as it is discussed in appendix E. However, the notion of ϵ -deterministic work can be naturally integrated in our formalism modifying slightly the function \mathcal{W}_{det} , while keeping the physical constraint \mathcal{P}_ϵ . Let us therefore show how the idea can be transferred into our setting and cast into a valid work-quantifier. Consider the following set of qubit work-storage devices,

$$\mathcal{P}_\epsilon := \{(\rho, H) \mid \|\rho - |E\rangle\langle E|\|_1 < 2\epsilon, H|E\rangle = E|E\rangle\}.$$

The operational meaning of $\epsilon \geq 0$ roughly is the optimal probability to be able to distinguish the state ρ from an energy-eigenstate in a measurement. As work-quantifier we can choose any work quantifier that respects axioms 1 and 2 for the set \mathcal{P}_ϵ . We will analyse for simplicity the one induced by the von Neumann free energy, that is $\mathcal{W}(p \rightarrow p', \beta) = \Delta F_1^\beta(p') - \Delta F_1^\beta(p)$. In the case of $\epsilon = 0$, all states are energy eigenstates and, if the Hamiltonian does not change in a transition, the work-quantifier simply measures the energy-difference between the states before and after the transition. This coincides with the original definition of ϵ -deterministic work given in [8] only when $\epsilon = 0$ and p and p' have the same Hamiltonian. However, it recovers in spirit the notion of ϵ -deterministic work in a way that is compatible with our axiomatic approach.

Let us now show that one indeed obtains irreversibility in this setting. This can be shown in the simplest case of $\epsilon = 0$. It is implied by the results of [8] that if Merlin has a full-rank system described by p_M , it cannot be used to induce a transition on the work-storage device of the form $|0\rangle\langle 0|_A \rightarrow |E\rangle\langle E|_A$. Hence, $W_{\text{value}}(p_M) = 0$ for \mathcal{P}_0 . Nonetheless, p_M may be a system arbitrarily far from equilibrium, hence it is necessary to spend resources to create it and $W_{\text{cost}}(p_M) \geq \Delta F_1^\beta(p_M) > 0$.

Furthermore, we expect the phenomenon of *irreversibility* to emerge in numerous physically meaningful sets other than the ϵ -deterministic work extraction. For instance, one may imagine restrictions on \mathcal{P} that reflect work-storage devices whose Hilbert-space dimension is bounded by some finite number. Alternatively, one may consider one whose states' entropy or free energy is bounded from above. We expect that irreversibility emerges in any such setting for at least some systems, since Merlin will not be in general allowed to give his system to Arthur. The former will have to interact with the heat bath leading to unavoidable dissipation and irreversibility. We will leave the detailed investigation of such scenarios for future work.

8. The role of correlations, the second law and super-additivity

We will now turn to discuss the role of correlations between the fuel (Merlin system M) and the work-storage device A and the implications that it has for the characterisation of the work quantifier \mathcal{W} . To do this, let us first define a quantity similar to the work of transition in definition 1, but where the fuel is allowed to establish correlations with the work-storage device.

Definition 6 (Correlated work of transition). Given a work quantifier \mathcal{W} and inverse temperature β , a set of restrictions \mathcal{P} , and initial and final objects of M , denoted by $p_M^{(i)}$ and $p_M^{(f)}$, respectively, the *correlated work of transition* $W_{\text{trans}}^{\text{corr}}(p_M^{(i)} \rightarrow p_M^{(f)}, \beta)$ is defined as

$$\begin{aligned} W_{\text{trans}}^{\text{corr}}(p_M^{(i)} \rightarrow p_M^{(f)}, \beta) \\ &:= \sup_{\substack{p_A^{(i)}, p_A^{(f)} \in \mathcal{P}; \\ p_{MA}^{(f)} \in \mathcal{F}_C(p_M^{(i)} \otimes p_A^{(i)})}} \mathcal{W}(p_A^{(i)} \rightarrow p_A^{(f)}, \beta). \end{aligned} \quad (28)$$

Note that the only difference with definition 6 is that the supremum is taken over protocols that allow the final state $p_{MA}^{(f)}$ to have arbitrary correlations. We can also define the correlated work cost and value as

$$W_{\text{value}}^{\text{corr}}(p_M, \beta) := W_{\text{trans}}^{\text{corr}}(p_M \rightarrow \omega_\beta, \beta), \quad (29)$$

$$W_{\text{cost}}^{\text{corr}}(p_M, \beta) := -W_{\text{trans}}^{\text{corr}}(\omega_\beta \rightarrow p_M, \beta), \quad (30)$$

It is to be expected that axioms 1 and 2 are not sufficient to capture the second law in the case where correlations are allowed. For instance axiom 2 captures the idea that Merlin cannot get rich while returning the fuel to the same initial state. But in principle, it does not prevent Merlin from getting rich by (despite returning the fuel to the same state) establishing correlations between the fuel and the work-storage device. This is indeed the case: we can find a work quantifier \mathcal{W} fulfilling axiom 1 and 2 such that $W_{\text{value}}^{\text{corr}}(p_M) > W_{\text{cost}}^{\text{corr}}(p_M)$. Although this might

not be a surprising result we include here a specific example, because it illustrates how the notion of super-additivity will come into play: the example relies on the use of work quantifiers $\mathcal{W}(p \rightarrow p') = M(p') - M(p)$ such that M is *not* super-additive, where super-additivity means that $M(p_{AB}) \geq M(p_A) + M(p_B)$, whenever $p_{AB}, p_A, p_B \in \mathcal{P}$.

Assume now a bipartite state p_{MA} and a monotone M such that super-additivity is violated, that is, $M(p_{MA}) < M(p_M) + M(p_A)$. Let us first look at $W_{\text{cost}}^{\text{corr}}(p_M)$. One particular protocol to create p_M consists of Arthur having initially p_{MA} and giving subsystem M to Merlin, while keeping p_A . This particular protocol gives an upper bound to the work cost as

$$W_{\text{cost}}^{\text{corr}}(p_M) \leq M(p_{MA}) - M(p_A). \quad (31)$$

Secondly, we can lower bound $W_{\text{value}}(p_M)$, simply by performing the obvious protocol where Merlin gives p_M to Arthur, resulting in

$$W_{\text{value}}^{\text{corr}}(p_M) \geq M(p_M). \quad (32)$$

Combining the fact that p_{MA} violates super-additivity with equations (31) and (32) results in a strict violation of the second law $W_{\text{value}}^{\text{corr}}(p_M) > W_{\text{cost}}^{\text{corr}}(p_M)$.

Let us now discuss the implications of this example. Suppose that Merlin would like to use the fact that $W_{\text{value}}^{\text{corr}}(p_M) > W_{\text{cost}}^{\text{corr}}(p_M)$ to become arbitrarily rich, or in other words, create a *perpetuum-mobilé*. He can start by initially having p_M and obtaining $W_{\text{value}}^{\text{corr}}(p_M)$. Then he will again create p_M , having paid $W_{\text{cost}}^{\text{corr}}(p_M)$ and thus resulting in an overall benefit. Note that M is returned to its original state after each cycle, however it becomes correlated with the work-storage device. When Merlin repeats those processes, he will need fresh uncorrelated work-storage devices each time, devices that end up all being correlated with Merlin's catalyst and among them. Hence, Merlin is getting arbitrarily rich without spending resources in the sense that he is not changing his system which behaves like a catalyst, but he does spend resources, because he is establishing correlations between M and a new work-storage device at each cycle. In other words, Merlin is spending 'absence of correlations', hence it seems natural that he can obtain benefit from it. A similar, but non-equivalent, effect has been discussed in [24], where the correlations are established among different parts of the catalyst. Thus, one possible viewpoint is to state that in order to account properly for resources, correlations cannot be created. Hence, the second law would take the form (27) which is indeed fulfilled for any work quantifier satisfying axioms 1 and 2.

A complementary approach to capture the role of correlations is to take the opposite view: correlating the catalyst with the work-storage device does not spend any resource and hence, it should be considered a valid operation. Furthermore, the work quantifier has to be modified accordingly to prevent from violations of the second-law (even if correlations are created) as given by $W_{\text{value}}^{\text{corr}}(p_M) > W_{\text{cost}}^{\text{corr}}(p_M)$. For this, we introduce a reformulation of axiom 2 that accounts for correlations. We highlight that we do not regard this reformulation as being as fundamental as axiom 2. It only aims at capturing in a consistent way which are the valid work quantifiers if correlations are treated as a free resource, in the spirit of [24].

Axiom 3 (Correlated cyclic transitions of the fuel). For any cyclic sequence of transitions of the the 'fuel' (Merlin's system) $p_M^{(1)} \rightarrow p_M^{(2)} \rightarrow \dots \rightarrow p_M^{(n)} = p_M^{(1)}$, the sum of the optimal work that Merlin can obtain in each sequence when correlations with the work-storage device are allowed (this is given by $W_{\text{trans}}^{\text{corr}}$ in (28)) is smaller or equal to zero,

$$\sum_{i=1}^{n-1} W_{\text{trans}}^{\text{corr}}(p_M^{(i)} \rightarrow p_M^{(i+1)}, \beta) \leq 0. \quad (33)$$

The intuition behind axiom 3 is similar to the one of axiom 2, with the only difference that Merlin is not allowed to become arbitrarily rich even by creating correlations with the work-storage devices. Imposing axiom 3 has two important consequences. Firstly, one can easily show that if one makes use of axiom 3, then the usual second law is fulfilled, stated as

$$W_{\text{value}}^{\text{corr}}(p_M) \leq W_{\text{cost}}^{\text{corr}}(p_M). \quad (34)$$

Secondly, allowing for correlations has consequences on the allowed work quantifiers \mathcal{W} . Taking the simplest case of $n = 2$ and $p_M^{(1)} = p_M^{(2)} = p_M$, axiom 3 implies that $W_{\text{trans}}^{\text{corr}}(p_M \rightarrow p_M) \leq 0 \forall p_M$. Combining this fact with definition 6 one can easily see that \mathcal{W} , in order to respect axioms 1 and 3 has to satisfy

$$\mathcal{W}(p_A \rightarrow q_A) \leq 0$$

for all q_A, p_A in \mathcal{P} such that $q_A \in \mathcal{F}_C^{\text{corr}}(p_A)$, where we define $\mathcal{F}_C^{\text{corr}}(p)$ to be the set of objects that can be reached from p by using thermal baths and an ancillary system that is left, after the interaction with the bath, with the same marginal state and Hamiltonian, but possibly correlated with the system. We will refer to this transitions as *correlated catalytic free transitions*. It is easy to see that $p_A \otimes p_B \in \mathcal{F}_C^{\text{corr}}(p_{AB})$ for $p_A, p_B, p_{AB} \in \mathcal{P}$. Together with additivity, this implies that in order to respect axioms 1 and 3, the work quantifier is written as

$\mathcal{W}(p \rightarrow p') = M(p') - M(p)$, where

$$M(p_{AB}) \geq M(p_A) + M(p_B). \quad (35)$$

As a consequence, the following is true.

Theorem 7 (Von-Neumann free energy in the unrestricted case). *Under axioms 1 and 3, from all the Rényi free energies, only the von Neumann free energy*

$$\Delta F_1^\beta(\rho, H) := \frac{1}{\beta} S(\rho || \omega_{H,\beta}) \quad (36)$$

remains to be a valid monotone to define a work-quantifier for arbitrary sets \mathcal{P} (up to a constant). It gives rise to a second law in the form

$$W_{\text{value}}^{\text{corr}}(p_M) \leq \Delta F_1^\beta(p_M) \leq W_{\text{cost}}^{\text{corr}}(p_M). \quad (37)$$

Note that the von Neumann free energy can also be written as

$$\Delta F_1^\beta(\rho, H) = F_1^\beta(\rho, H) - F_1^\beta(\omega_{H,\beta}, H) \quad (38)$$

with $F_1^\beta(\rho, H) = \text{tr}(\rho H) - S(\rho)/\beta$. It therefore closely resembles the phenomenological free energy $U - TS$, or more precisely the exergy with respect to an environment of temperature $T = 1/\beta$. We hence recover the statement that the maximum amount of work that can be extracted by a working system with access to a heat bath of temperature T is given by the exergy of the working system with respect to the temperature T —but using reasoning very different from that of phenomenological thermodynamics. It is also interesting to see that on a formal level in the framework developed here, the von-Neumann free energy does not arise from considering an asymptotic setting, but rather arises from the way correlations are taken into account.

We have seen that super-additivity and the von Neumann free energy emerge naturally once we allow for the creation of correlations between the catalyst and the system. A similar result was obtained in [24], where it was shown that, for classical states, the change of von Neumann free energy decides whether a transition between two objects is possible if multiple catalyst can be used, which can become correlated with each other, but not with the system.

In the light of the previous discussions one might wonder whether super-additivity already singles out the von Neumann free energy as the *unique* valid monotone to define a work quantifier in the case of correlated catalysis. This is true in the case of vanishing Hamiltonians but otherwise unrestricted sets \mathcal{P} , which we state in the following theorem.

Theorem 8 (Von-Neumann free energy as a work quantifier for vanishing Hamiltonians). *Consider the set of all finite-dimensional quantum states and the vanishing Hamiltonian $\mathcal{P} = \{(\rho, \mathbb{1})\}$ and free operations given by thermal operations or Gibbs-preserving maps. Then the unique work quantifier with continuous monotone M and fulfilling axioms 1 and 3 is given, up to a constant factor, by*

$$\mathcal{W}(p \rightarrow p', \beta) = \Delta F_1^\beta(p') - \Delta F_1^\beta(p) \quad (39)$$

where ΔF_1^β is the von Neumann free energy.

Proof. Without loss of generality, consider the candidates for a work-quantifier defined as $M(\rho, \mathbb{1}) := \alpha(\log d(\rho) - f(\rho))$, where $d(\rho)$ is the dimension of the Hilbert-space of ρ , α is some positive constant and $f(\rho)$ is a yet unspecified continuous (on states of fixed dimension) function. We will show that f has to be given by the von Neumann entropy. Since $S(\rho || \mathbb{1}_{d(\rho)}/d(\rho)) = \log d(\rho) - S(\rho)$ this implies the claim. Using additivity, super-additivity we immediately obtain that f has to be additive and sub-additive. From monotonicity under thermal operations we obtain that a) $f(U\rho U^\dagger) = f(\rho)$ for any unitary and b) $f(\sum_i p_i U_i \rho U_i^\dagger) \geq f(\rho)$ for any probability-distribution p_i over unitaries U_i . Property a) implies that f only depends on the eigenvalues of ρ and is therefore equivalent to a function \tilde{f} on probability distributions, which fulfills additivity and sub-additivity. Property b) implies that \tilde{f} is *Schur-concave*, i.e., can only increase under random permutations. In [44] it has been shown that for probability-distributions without zeros, such a function is of the form $\tilde{f}(p) = cH(p) + c_{d(p)}$, where H is the Shannon-entropy, $c \geq 0$ and $c_{d_1 d_2} = c_{d_1} + c_{d_2}$. By continuity, this form extends to arbitrary probability-distributions and we obtain $f(\rho) = cS(\rho) + c_{d(\rho)}$, where S is the von Neumann entropy. From $M(\mathbb{1}_d/d, \mathbb{1}_d) = 0$, we obtain $c \log d + c_d = \log d$. This implies

$$M(\rho, \mathbb{1}) = c\alpha(\log d(\rho) - S(\rho)) = \alpha'(\log d(\rho) - S(\rho)),$$

which finishes the proof. \square

One might wonder whether the result could also hold in infinite-dimensional systems. However, in such systems the vanishing Hamiltonian does not have a well-defined thermal state for any temperature, so that it

should not be considered as a physical Hamiltonian on such systems. To extend this result to more general classes of Hamiltonians constitutes an interesting open problem. Importantly, if true, this does still not imply that one can only make use of the von Neumann free energy as a work quantifier. There are many situations of physical relevance where the set of \mathcal{P} is restricted, where one could still conceive other work quantifiers. Indeed, we have seen in section 7 how imposing constraints \mathcal{P} , rather than a technicality, is crucial to recover several commonly discussed regimes in which thermodynamics is expected to operate.

9. Summary

In this work, we have approached the subtle and much discussed question of how to quantify the performance of thermal machines at the small scale, where fluctuations and quantum effects play a relevant role. This problem has so far been addressed in the literature by focusing on finding quantum analogues of the relevant tasks in classical thermodynamics: compressing a gas, lifting a weight or cooling a system below the environment temperature. Here, we allow instead for arbitrary classes of physical systems, intended to realistically account for the transitions that a quantum thermal machine is expected to perform. The problem that we have tackled here is precisely how to quantify the performance of a quantum machine in these tasks lacking of a classical analogue. We have taken a strictly operational approach and have posed the problem of identifying reasonable functions that evaluate the value of a given transition. These functions are supposed to have basic properties analogue to the familiar notion of work in phenomenological thermodynamics. These properties have been stated in the form of strictly operational axioms that capture minimum reasonable conditions that meaningful work-quantifiers are expected to satisfy. This is a distinct deviation in mindset: we do not define quantities ad-hoc based on its classical counterparts, but aim at clarifying those characteristic features that any work quantifier should fulfil, providing a general framework.

Remarkably, simple and elementary as these axioms may appear, they provide sufficient mathematical structure to give rise to surprisingly detailed and stringent properties that any function quantifying work has to fulfil and which can be rigorously derived from the axioms.

One of the advantages of the formalism is that it is general enough to allow us to derive central concepts in thermodynamics without taking the definition of work as an *a priori* given element. For instance, our generalised work quantifiers give rise to quantitative versions of the second law. Similarly, one can precisely discuss notions of irreversibility in this framework, in the sense that in order to obtain useful work, it is necessary to dissipate the resources provided by the ‘fuel’, concomitant to familiar notions in thermodynamics.

When the system is taken to be an analogue of a lifted weight in the quantum regime, our general framework recovers the usual definition of work as the energy difference as a particular case. At an more heuristic level, this can be summarised by the insight that the task of extracting work is nothing but the transfer of free-energy from an arbitrary system (the fuel) to another system which has to fulfil a set of given restrictions (the work-storage device). In the specific situation in which those restrictions are such that the work-storage device is a lifted weight, then the free energy coincides with the energy.

For coherence of the presentation, we have focused on work quantifiers in quantum thermodynamics in the main text. It should be clear, however, that the technical results achieved are general enough to capture also other quantum resource theories, beyond the quantum thermodynamic context. The arguments laid out in main text and the supplementary material clearly highlight the role that catalysts and their correlations play in such resource theories. Furthermore, our results show that there is a close connection between catalysis, the built-up of correlations, and of reversibility. In particular, we have shown in what precise way a restriction of the state-space of work-storage devices is necessary in order to obtain irreversibility. It is our hope that the approach taken here can be seen as a further invitation to revisit notions derived from classical thermodynamics and taking an operational perspective when aiming at clarifying in what precise way they can be extended to the quantum regime.

Acknowledgments

This work has been supported by the EU (SIQS, AQuS, RAQUEL), the ERC (TAQ), the DFG (EI 519/7-1, CRC 183), the Alexander von Humboldt-Foundation and the Studienstiftung des Deutschen Volkes.

Appendix A. Scenario and definitions

A.1. Transitions and free transitions

Let us consider a pair of a quantum states and a Hamiltonian $p = (\rho, H)$. In the following we, will call such pairs *objects* and denote the associated Hilbert space by $\mathcal{H}(p)$, which for most of this work is taken to be finite-dimensional.

Definition 9 (Transition). A transition is defined by a pair of objects $p^{(i)}, p^{(f)}$ and an ordering between them. We will refer to a transition as $p^{(i)} \rightarrow p^{(f)}$.

Definition 10 (State transition). This is a transition in which the Hamiltonian remains constant. That is, if $(\rho^{(i)}, H) \rightarrow (\rho^{(f)}, H)$, we will refer to a state transition and denote it simply, if the Hamiltonian is clear from the context, by $\rho^{(i)} \rightarrow \rho^{(f)}$.

Definition 11 (Sequence). A set of $n - 1$ transitions of the form $\{p^{(k)} \rightarrow p^{(k+1)}\}_{k=1}^{n-1}$ is referred to as *sequence*. We will simply denote it by $p^{(1)} \rightarrow p^{(2)} \rightarrow \dots \rightarrow p^{(n)}$.

Such transitions are to be interpreted, in the context of the present work, as changes on the system and state Hamiltonian of the battery of Arthur as implemented by Merlin.

Definition 12 (Free image). A free image is a function \mathcal{F} that maps $p^{(i)}$ and a parameter β into sets of objects $\{p_k\} = \mathcal{F}(p^{(i)}, \beta)$. When F is such that the Hamiltonian remains constant, that is,

$$\mathcal{F}(\rho^{(i)}, \beta) = \{(\rho_k, H)\}, \quad (\text{A.1})$$

we will refer to it as *free state-image*.

Definition 13 (Free transition). A free transition is defined as any transition $p^{(i)} \rightarrow p^{(f)}$, where $p^{(f)} \in \mathcal{F}(p^{(i)}, \beta)$. When the parameter β is clear from the context, we will denote a free transition simply as $p^{(i)} \rightarrow \mathcal{F}(p^{(i)})$.

Definition 14 (Tensoring objects). Given two objects $p = (\rho, H)$ and $p' = (\rho', H')$, we define the tensor product

$$p \otimes p' := (\rho \otimes \rho', H \otimes \mathbb{1}_{\mathcal{H}(p')} + \mathbb{1}_{\mathcal{H}(p)} \otimes H'). \quad (\text{A.2})$$

In the definition we explicitly indicated on which tensor-factor the identity maps act. In the following, we will omit such indications when the information is clear from the context.

Definition 15 (Non-interacting objects). If an object based on a bipartite system of parts A and B has the form

$$p = (\rho_{AB}, H_A \otimes \mathbb{1}_B + \mathbb{1}_A \otimes H_B) \quad (\text{A.3})$$

we refer to it as *non-interacting object*.

Non-interacting objects are those objects on which we define a partial trace.

Definition 16 (Partial traces). Given any two objects $p_S = (\rho_S, H_S)$ and $p_{|S} = (\rho_{|S}, H_{|S})$, we define the trace $\text{tr}_{|S}$ as an operator acting on objects p of the form

$$p = p_S \otimes p_{|S} = (\rho_S \otimes \rho_{|S}, H_S \otimes \mathbb{1}_{|S} + \mathbb{1}_S \otimes H_{|S}), \quad (\text{A.4})$$

such that $\text{tr}_{|S}(p) = p_S$. We extend this definition to all *non-interacting* objects by the partial trace on quantum states.

At this point a remark about Hamiltonians is in order. When we consider non-interacting objects, the local Hamiltonians are not well-defined: we can always change their traces by adding a global zero of the form $(\lambda \mathbb{1}_A) \otimes \mathbb{1}_B - \mathbb{1}_A \otimes (\lambda \mathbb{1}_B)$ to the global Hamiltonian. Therefore we will from now call two Hamiltonian operators equivalent if they differ by a multiple of the identity, $H \sim H + \lambda \mathbb{1}$. For simplicity, we will, however, not indicate this in our notation and will just refer to the equivalence classes as Hamiltonians. We could also just fix the trace of the Hamiltonians. It will become clear later, why we do not follow this path.

Definition 17 (Catalytic free image). Given the free image \mathcal{F} , we define the *catalytic free image* \mathcal{F}_C as

$$\mathcal{F}_C(p^{(i)}, \beta) := \{p \mid \exists q; p \otimes q \in \mathcal{F}(p^{(i)} \otimes q, \beta)\}. \quad (\text{A.5})$$

Definition 18 (Catalytic free transition). A catalytic free transition is defined as any transition $p^{(i)} \rightarrow p^{(f)}$ with $p^{(f)} \in \mathcal{F}_C(p^{(i)}, \beta)$. When the parameter β is clear from the context, we will denote a free state-transition simply as $p^{(i)} \rightarrow \mathcal{F}_C(p^{(i)})$.

Definition 19 (Assisted transitions and sequences). Two objects $p^{(1)}, p^{(2)}$ form a *transition assisted by* (c_1, c_2) if

$$p^{(2)} \otimes c_2 \in \mathcal{F}(p^{(1)} \otimes c_1, \beta), \quad (\text{A.6})$$

Now consider a sequence of transitions $p^{(i)} \rightarrow p^{(i+1)}$ for $i = 1, \dots, n-1$. If each transition is a free transition assisted by (c_i, c_{i+1}) , respectively, we say the sequence is assisted by (c_1, c_n) .

In other words, an assisted sequence is a sequence on objects that can be performed by using free operations and an ancilla that is at the end uncorrelated with the system but might have changed its state.

Although we would like to keep this definition fully general, let us anticipate that $\{c_i\}_i$ are going to play the role of the fuel employed by Merlin, which enables (assists) a transition or sequence of transitions, by changing its state (by being burnt).

Definition 20 (Free sequence). We call a sequence assisted by (c, c) a *free sequence*.

Following with the interpretation of c as the fuel, a free sequence is then a sequence of transitions that can be implemented while not spending any fuel.

A.2. Basic assumptions on the free transitions

In the main text we have focused on the *resource theory of α -thermality*, where the free operations are, loosely speaking, defined as the energy preserving joint operations on system and bath. These are mathematically characterised by the GP-maps, or strictly contained subsets of operations, such as the thermal operations. However, our results apply potentially to widely different resource theories defined by other classes of free operations, not motivated by the thermodynamic context. In this endeavour, we aim at contributing to the emerging understanding of general resource theories [19, 20]. We state below the first assumptions on the free operations that are needed in order to derive the results of section 5 in the main text, in particular theorem 2 (restated as theorem 25 in this appendix).

Property 1 (Composability). If $p^{(3)} \in \mathcal{F}(p^{(2)}, \beta)$ and $p^{(2)} \in \mathcal{F}(p^{(1)}, \beta)$, then $p^{(3)} \in \mathcal{F}(p^{(1)}, \beta)$.

Property 2 (Swapping products). Given an object of the form $p^{(1)} \otimes \dots \otimes p^{(n)}$, then

$$P(p^{(1)} \otimes \dots \otimes p^{(n)}) \in \mathcal{F}(p^{(1)} \otimes \dots \otimes p^{(n)}, \beta), \forall \beta, \quad (\text{A.7})$$

where P permutes the labels $(1, \dots, n)$ into $(\sigma(1), \dots, \sigma(n))$.

Note that property 2 implies that the identity is a catalytic free transition, that is, $p \in \mathcal{F}_C(p, \beta)$ for all β . This follows since one can take as catalyst $q = p$ and perform a swap between the system and the catalyst.

Property 3 (Tracing as free operation). For any subsystem S of A_1, \dots, A_N of a product object, tracing out is in the free image. That is,

$$\text{tr}_S(p_{A_1} \otimes \dots \otimes p_{A_N}) \in \mathcal{F}(p_{A_1} \otimes \dots \otimes p_{A_N}, \beta). \quad (\text{A.8})$$

In the case where

$$S = \bigcup_{i=1}^N A_i, \quad (\text{A.9})$$

the entire system is traced out. In this case we introduce the notation $\text{tr}_S(p) := \emptyset$. In this instance proposition 3 is also fulfilled and we denote it by $\mathcal{W}(p \rightarrow \emptyset, \beta) \leq 0$. The object \emptyset can be seen as the pair $(1, 0)$ on $\mathcal{H} = \mathbb{C}$. Note that it therefore fulfills $p \otimes \emptyset = p$ for every object p . It is therefore a free object independent of β .

The next lemma will turn out to be very useful in the subsequent sections.

Lemma 21 (Mapping time to space). Suppose \mathcal{F} fulfills properties 1 and 2 and let $p \rightarrow p'$ be an assisted transition by (c, c') and $q \rightarrow q'$ be an assisted transition by (c', c'') . Then the transition $p \otimes q \rightarrow p' \otimes q'$ is an assisted transition by (c, c'') .

Proof. Note that by definition 19 of assisted transition and property 1, the transition $p_1 \otimes c \rightarrow p_m \otimes c'$ is free. Therefore, also the transition $p_q \otimes q_1 \otimes c \rightarrow p_m \otimes q_1 \otimes c'$ is free. An equivalent argument implies that $p_m \otimes q_1 \otimes c' \rightarrow p_m \otimes q_n \otimes c''$ is also free transition. Composing these two transitions yields that $p_1 \otimes q_1 \otimes c \rightarrow p_m \otimes q_n \otimes c''$ is also a free transition. \square

A.3. Work quantifiers

Once we have specified the transitions and the free transitions, we will define a quantifier of the *value* of a given transitions within the set of allowed work-storage devices \mathcal{P} . We will always assume that the empty object \emptyset is an element of \mathcal{P} .

Definition 22 (Work quantifier). We define the work quantifier as a function \mathcal{W} that maps a transition within \mathcal{P} and parameter $(p^{(i)} \rightarrow p^{(f)}, \beta)$ into the real numbers. If β is clear from the context, we will simply write $\mathcal{W}(p^{(i)} \rightarrow p^{(f)})$.

Appendix B. General axioms

We will now present the axioms 1 and 2 of the main text, restated in a more precise manner by making use of the mathematical definitions of section A.1.

Axiom 1 (Cyclic transitions of the work storage device). Given a collection of objects of the work-storage device $\{p^{(1)}, \dots, p^{(n)}\} \subset \mathcal{P}$ such that $p^{(n)} = p^{(1)}$, then

$$\sum_{i=1}^{n-1} \mathcal{W}(p^{(i)} \rightarrow p^{(i+1)}, \beta) \geq 0. \quad (\text{B.1})$$

Axiom 1 ensures that if a set of states can be arranged in a cyclic sequence, the total work, given by the l.h.s. of (B.1), cannot be negative. Otherwise, Arthur, who receives at the end the same object he possessed at the beginning, can repeat the protocol an arbitrarily number of times and obtain an arbitrarily large benefit.

Axiom 2 in the main text is however formulated in terms of $\mathcal{W}_{\text{trans}}$. However, as this quantity is given as a function of \mathcal{W} and \mathcal{P} we can reformulate axiom 2 as being directly expressed in terms of \mathcal{W} for transitions of the work-storage device, which makes it a more comfortable formulation to work in the following proofs of this appendix.

Axiom 2 (Reformulation of ‘cyclic transitions of the fuel’). Let $\{p_A^{(k)} \rightarrow q_A^{(k)}\}_{k=1}^{n-1}$ be a collection of assisted transitions of the work-storage device, assisted by (c_k, c_{k+1}) respectively, with $c_n = c_1$. Then

$$\sum_{k=1}^{n-1} \mathcal{W}(p_A^{(k)} \rightarrow q_A^{(k)}, \beta) \leq 0. \quad (\text{B.2})$$

Importantly, note that the objects $p_A^{(k)}$ and $q_A^{(k)} \forall k$ in this formulation describe the work-storage device, contrary to the main text formulation of axiom 2. A schematic depiction of the transitions involved in this Axiom is given by figure 4. Now we will show, that although formulated in seemingly unrelated terms, both formulations are equivalent.

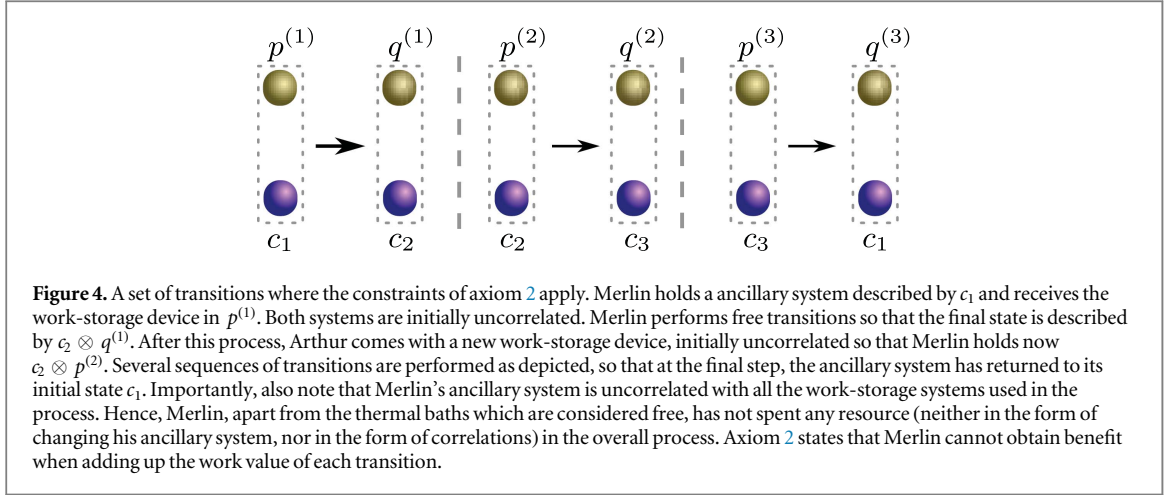
First, let us state a corollary of axiom 2 that will be useful in further proofs.

Corollary 23 (Cyclic free sequences). Let $p_A^{(1)} \rightarrow p_A^{(2)} \rightarrow \dots \rightarrow p_A^{(n)}$ a free sequence, then,

$$\sum_{k=1}^{n-1} \mathcal{W}(p_A^{(k)} \rightarrow p_A^{(k+1)}, \beta) \leq 0. \quad (\text{B.3})$$

Corollary 23 follows simply by the definition of free sequence, which is a particular case of the conditions of axiom 2, in the case where $q_A^{(k)} = p_A^{(k+1)}$.

B.0.1. Equivalence between formulations: We will now show that indeed the version of the axiom 2 given in the main text is equivalent to the one given above. Let us first assume the version given in the main text. That is, we assume that for any sequence of the fuel (Merlin’s system), where we $p_M^{(1)} \rightarrow \dots \rightarrow p_M^{(n)} = p_M^{(1)}$, then



$$\sum_{i=1}^{n-1} W_{\text{trans}}(p_M^{(i)} \rightarrow p_M^{(i+1)}, \beta) \leq 0. \quad (\text{B.4})$$

Now consider a set of assisted transitions $\{p_A^{(k)} \rightarrow q_A^{(k)}\}_{k=1}^{n-1}$ of the work-storage device, assisted by (c_k, c_{k+1}) respectively, with $c_n = c_1$, as axiom 2 states. Using definition 1, we have that

$$\mathcal{W}(p_A^{(k)} \rightarrow q_A^{(k)}) \leq W_{\text{trans}}(c_k \rightarrow c_{k+1}), \quad (\text{B.5})$$

for all $k \in \{1, \dots, n-1\}$. But then by identifying $c_i = p_M^{(i)}$ for all i , we obtain equation (B.2).

Let us now show the converse direction. We have to show, that given a sequence $p_M^{(1)} \rightarrow \dots \rightarrow p_M^{(n)} = p_M^{(1)}$, equation (B.2) implies equation (B.4). Each transition $p_M^{(k)} \rightarrow p_M^{(k+1)}$ will also induce a transition on the marginal of the work-storage device, given by $p_A^{(k)} \rightarrow q_A^{(k)} \in \mathcal{S}_k$, where \mathcal{S}_k is the set of all marginal transitions on the work-storage device that can happen together with $p_M^{(k)} \rightarrow p_M^{(k+1)}$ on the fuel, and equivalently for all k . More explicitly,

$$\mathcal{S}_k := \{p_A^{(k)} \rightarrow q_A^{(k)} \mid q_A^{(k)} \otimes p_M^{(k+1)} \in \mathcal{F}_C(p_A^{(k)} \otimes p_M^{(k)})\}. \quad (\text{B.6})$$

That said, all the $p_A^{(k)} \rightarrow q_A^{(k)} \in \mathcal{S}_k$ are an assisted transition by $p_M^{(k)} \rightarrow p_M^{(k+1)}$. By our assumption (equation (B.2)) this implies that the total work-value fulfills

$$\sum_{k=1}^{n-1} \mathcal{W}(p_A^{(k)} \rightarrow q_A^{(k)}) \leq 0 \quad (\text{B.7})$$

for all $p_A^{(k)} \rightarrow q_A^{(k)} \in \mathcal{S}_k$. Then, this implies trivially

$$\sum_{k=1}^{n-1} \sup_{p_A^{(k)} \rightarrow q_A^{(k)} \in \mathcal{S}_k} (\mathcal{W}(p_A^{(k)} \rightarrow q_A^{(k)})) \leq 0. \quad (\text{B.8})$$

Now notice that the supremum in equation (B.8) is the same as the one in the def. 1 of W_{trans} , which concludes the proof.

B.1. Implications for the work definition

We now turn to exploring implications for the work quantifiers. Since the two Axioms have been reformulated in appendix B in such a way that they only refer to objects of the work-storage device and not of the fuel, we will drop the labels M and A . Unless explicitly mentioned, we will use the letters p, q to refer to the work-storage device.

Lemma 24 (Properties of work quantifiers). Consider a free image \mathcal{F} fulfilling properties 1–3. In this case, axioms 1 and 2 are fulfilled if and only if \mathcal{W} satisfies the following properties,

1. For all $p^{(1)}, \dots, p^{(m)}$ and $q^{(1)}, \dots, q^{(m)}$ in \mathcal{P} such that $\bigotimes_{i=1}^m q^{(i)} \in \mathcal{F}_C(\bigotimes_{i=1}^m p^{(i)})$,

$$\sum_{i=1}^m \mathcal{W}(p^{(i)} \rightarrow q^{(i)}) \leq 0. \quad (\text{B.9})$$

2. For all $p, q, r \in \mathcal{P}$

$$\mathcal{W}(p \rightarrow q) = -\mathcal{W}(q \rightarrow p), \quad (\text{B.10})$$

$$\mathcal{W}(p \rightarrow q) + \mathcal{W}(q \rightarrow r) = \mathcal{W}(p \rightarrow r). \quad (\text{B.11})$$

Proof. We will first show that the axioms imply the properties, beginning with properties (B.10) and (B.11). The two properties follow immediately once we have shown that any cyclic sequence $p_1 \rightarrow p_2 \rightarrow \dots \rightarrow p_n = p_1$ has a total work-value equal to zero. Given axiom 1, which already implies that it is larger than zero, this only requires us to show that such a sequence has a work-value smaller or equal to zero. This will be done by showing that any cyclic sequence is a free sequence, which is enough to show the claim given corollary 23. We will show that any cyclic sequence is a *free sequence*, where following the notation of definition 20, $c_1 = c_n := c$ is given by $c = \bigotimes_{i=2}^{n-1} p_i$. To see that c assists any cyclic sequence from p_1 to $p_n = p_1$, consider the object $p_1 \otimes c = \bigotimes_{i=1}^{n-1} p_i$. By swapping, which is a free operation, we arrive at state $p_2 \otimes c'$ with $c' = p_1 \otimes p_3 \otimes p_4 \otimes \dots \otimes p_{n-1}$. Repeating the swapping sequentially we see that the first system goes through the transitions $p_1 \rightarrow p_2 \rightarrow \dots \rightarrow p_{n-1}$. Applying a final swap the fuel is returned to c and the system returns to object p_1 , proving the claim and thus, equations (B.10) and (B.11).

Let us now show property (B.9) from the axioms. The premise of (B.9) is that, there exists a catalytic free transition $\bigotimes_{i=1}^n p_i \rightarrow \bigotimes_{i=1}^n q_i$. Here we are taking $m = n$ without loss of generality. The other cases follow by tensoring a suitable number of empty objects \emptyset . Then the transition $p_1 \rightarrow q_1$ is an assisted transition by

$$\left(c_1 = \bigotimes_{i=2}^n p_i, \quad c'_1 = \bigotimes_{i=2}^n q_i \right). \quad (\text{B.12})$$

Secondly, the transition $p_2 \rightarrow q_2$ is an assisted transition by $(c_2 = \bigotimes_{i=2}^n q_i, \quad p_2 \bigotimes_{i=3}^n q_i)$. This can be seen by just performing a swap between the work-storage system in p_2 and the first element of the fuel in q_2 . An equivalent swapping can be used to show that $p_j \rightarrow q_j$ is an assisted transition by

$$\left(c_j = \bigotimes_{i=2}^{j-1} p_i \bigotimes_{k=j}^n q_k, \quad c'_j = \bigotimes_{i=2}^j p_i \bigotimes_{k=j+1}^n q_k \right) \quad (\text{B.13})$$

for $j = 3, \dots, n-1$. Lastly, $p_n \rightarrow q_n$ is assisted by $(c_n = \bigotimes_{i=2}^{n-1} p_i \otimes q_n, \quad c'_n = \bigotimes_{i=2}^{n-1} p_i)$. Altogether, this implies that the set of sequences $\{p_i \rightarrow q_i\}_{i=1}^n$ can be each performed with free operations assisted by (c_i, c'_i) as described previously. Note, that $c'_i = c_{i+1}$ and $c_1 = c'_n$, hence, it meets the conditions of axiom 2 which by equation (B.2) implies

$$\sum_{i=1}^n \mathcal{W}(p_i \rightarrow q_i) \leq 0.$$

Finally, let us show that the properties (B.9)–(B.11) imply the axioms. Axiom 1 is trivially satisfied since properties (B.10) and (B.11) imply that for any cyclic sequence the total amount of work is zero. Let us move to axiom 2, which has as a premise that one has $n-1$ assisted transitions $p^{(j)} \rightarrow q^{(j)}$, assisted by (c_j, c_{j+1}) with $j = 1, \dots, n-1$ and $c_n = c_1$. Then, we can use lemma 21 and see that the transition

$$\bigotimes_{j=1}^{n-1} p^{(j)} \rightarrow \bigotimes_{j=1}^{n-1} q^{(j)} \quad (\text{B.14})$$

is an assisted transition, assisted by $(c_1, c_n = c_1)$. That is, the system c is returned unchanged, hence $\bigotimes_{j=1}^{n-1} p^{(j)} \rightarrow \bigotimes_{j=1}^{n-1} q^{(j)}$ is indeed a catalytic free transition and (B.9) implies that

$$\sum_{j=1}^{n-1} \mathcal{W}(p^{(j)} \rightarrow q^{(j)}) \leq 0, \quad (\text{B.15})$$

proving (B.2) and thus axiom 1. □

Let us now show that axioms 1 and 2, or equivalently equations (B.9)–(B.11), imply that the work function \mathcal{W} must take a very particular form.

Theorem 25 (theorem 2 in the main text). *Given a free image \mathcal{F} that fulfils properties 1-3, the function \mathcal{W} fulfils axioms 1 and 2 if and only if it can be written as*

$$\mathcal{W}(p \rightarrow q) = M(q) - M(p), \quad (\text{B.16})$$

for a function M such that $M(\emptyset) = 0$ and that fulfils the following property:

Additive monotonicity: for all $p^{(1)}, \dots, p^{(n)}$ and $q^{(1)}, \dots, q^{(n)}$ in \mathcal{P} such that $\bigotimes_{i=1}^n q^{(i)} \in \mathcal{F}_C(\bigotimes_{i=1}^n p^{(i)})$

$$\sum_{i=1}^n M(q^{(i)}) \leq \sum_{i=1}^n M(p^{(i)}). \quad (\text{B.17})$$

Proof. We will prove it by showing an equivalence with conditions (B.9)–(B.11), which in turn are equivalent with axioms 1 and 2. Consider the function $M(p) := \mathcal{W}(\emptyset \rightarrow p)$. By properties (B.10) and (B.11) we have

$$\mathcal{W}(p \rightarrow q) = M(q) - M(p) \quad (\text{B.18})$$

and $M(\emptyset) = 0$ is true by definition. Clearly, (B.9) is fulfilled if and only if additive monotonicity holds for M . \square

Appendix C. Gibbs-preserving and thermal operations

In this section, we will turn to two classes of operations that can be used to model meaningful classes of thermodynamic operations in the quantum regime, namely Gibbs-preserving operations (GPO) [6, 29, 30] and thermal operations (TO) [8]. We will first introduce the necessary objects, then define what state transitions are possible, and finally show that all the necessary properties are indeed fulfilled.

Both GPO and TO have the same sets of free objects, induced by Gibbs-states.

Definition 26 (Gibbs objects). The free objects of GPO and TO are given by

$$w = (\omega_H, H), \quad \omega_H = \frac{\exp(-\beta H)}{Z_H}, \quad (\text{C.1})$$

with any Hamiltonian H , and called *Gibbs objects*.

Since $\omega_H = \omega_{H+\lambda \mathbb{1}}$ for any $\lambda \in \mathbb{R}$, Gibbs objects are well-defined. To every object $p = (\rho, H)$ we can associate the Gibbs object

$$w(p) = (\omega_H, H). \quad (\text{C.2})$$

Let us now define Gibbs-preserving transitions.

Definition 27 (Gibbs-preserving transition). A transition $p = (\rho, H) \rightarrow q = (\sigma, K)$ is *Gibbs-preserving* if there exists a quantum channel \mathcal{G} such that

$$\sigma = \mathcal{G}(\rho) \text{ and } \omega_K = \mathcal{G}(\omega_H). \quad (\text{C.3})$$

Clearly, any Gibbs-object is mapped to another Gibbs-objects under Gibbs-preserving transitions, hence the name. In the case that the Hamiltonian H does not change in a Gibbs-preserving transition we call the corresponding quantum channel \mathcal{G} a Gibbs-preserving channel with respect to H . An operational way to think about the change of Hamiltonian in Gibbs-preserving transitions is given by Gibbs-preserving operations.

Definition 28 (Gibbs-preserving operations). Any operation composed of taking thermal objects (at the fixed inverse temperature β), applying Gibbs-preserving channels and tracing out subsystems is called a *Gibbs-preserving operation* (GPO).

GPO are closed under composition since the set of Gibbs-objects is closed under tensor products and a composition of two Gibbs-preserving channels is again Gibbs-preserving. Let us discuss some examples of GPO. A particular way to describe them is through maps from objects to objects which induce Gibbs-preserving transitions.

Example 29. Suppose G maps objects to objects, such that

$$G(\rho, H) = (\mathcal{G}_H(\rho), \tilde{\mathcal{G}}(H)) \quad (\text{C.4})$$

with \mathcal{G}_H a quantum channel and $\tilde{\mathcal{G}}$ a map that maps Hamiltonians onto Hamiltonians. Furthermore, suppose that the maps $\mathcal{G}_H, \tilde{\mathcal{G}}$ fulfill the consistency relation

$$\mathcal{G}_H(\omega_H) = \omega_{\tilde{\mathcal{G}}(H)} \quad (\text{C.5})$$

for all objects (ρ, H) . Then $p \rightarrow G(p)$ is a Gibbs-preserving transition for any object p and can be written as a Gibbs-preserving operation.

To see that the above construction can be seen as Gibbs-preserving operations let \mathcal{T}_H be the channel which acts as $\mathcal{T}_H(\rho) = \omega_H$ for any ρ , \mathcal{S} the swap channel $\mathcal{S}(\rho \otimes \sigma) = \sigma \otimes \rho$ and $\omega_H \otimes$ the channel that tensors in ω_H , i.e., $\omega_H \otimes (\rho) = \omega_H \otimes \rho$. Now let G be any Gibbs-preserving operation. It is easy to check from the consistency condition that $\mathcal{S} \circ \mathcal{G}_H \otimes \mathcal{T}_H$ is a Gibbs-preserving channel with respect to the Hamiltonian $H \otimes \mathbb{1} + \mathbb{1} \otimes \tilde{\mathcal{G}}(H)$. A simple calculation furthermore shows that on the level of quantum states we have

$$\text{tr}_2 \circ \mathcal{S} \circ (\mathcal{G}_H \otimes \mathcal{T}_H) \circ \omega_{\tilde{\mathcal{G}}(H)} \otimes (\rho) = \mathcal{G}_H(\rho), \quad (\text{C.6})$$

while on the level of Hamiltonians we have the mapping $H \mapsto \tilde{\mathcal{G}}(H)$.

In the following examples we use the notation of the previous example.

Example 30. Suppose $\mathcal{G}_H = \text{id}$. Then $\mathcal{G}(H) = H$ (as equivalence classes) must hold true for the pair to be a Gibbs-preserving operation. Conversely, if $\mathcal{G}(H) = H$ then automatically $\mathcal{G}_H(\omega_H) = \omega_H$ has to be valid.

This example implies that $\mathcal{G}_H(\omega_H) = \omega_H$ if and only if $\mathcal{G}(H) = H$ (as equivalence class).

Example 31. Fix an n -dimensional unitary U_n for every $n \in \mathbb{N}$. Then letting $\tilde{\mathcal{G}}(H) = U_n H U_n^\dagger$ and $\mathcal{G}_H(\rho) = U_n H U_n^\dagger$, with $n \in \mathbb{N}$ being the dimension corresponding to the Hilbert space of ρ , defines a Gibbs-preserving operation.

This example implies that the *swap-operation* $p \otimes q \mapsto q \otimes p$ is a Gibbs-preserving operation.

Example 32. For any Gibbs object w , the map $p \mapsto p \otimes w$ is a Gibbs-preserving operation.

Example 33. For any non-interacting object p_{A_1, \dots, A_N} and any subset $S \subseteq \{A_1, \dots, A_N\}$, the partial trace $p_{A_1, \dots, A_N} \mapsto p_S$ is a Gibbs-preserving operation.

Example 34. To every Hamiltonian H choose a unitary U_H such that

$$[U, H] = 0. \quad (\text{C.7})$$

Then the map T which acts as

$$T(\rho, H) = (U_H \rho U_H^\dagger, H) \quad (\text{C.8})$$

is a Gibbs-preserving operation.

Definition 35 (Thermal operations). A thermal operation is any operation that can be composed from the operations in the examples 31–34.

By definition, thermal operations are closed under composition.

Definition 36 (Catalysis). A transition $p \rightarrow r$ is a *catalytic Gibbs-preserving transition* if there exists an object q such that

$$p \otimes q \rightarrow r \otimes q \quad (\text{C.9})$$

is a Gibbs-preserving transition. The transition is a *catalytic thermal transition* if it is induced by a thermal operation.

We will see later, corollary 37, that also catalytic Gibbs-preserving transitions always map Gibbs objects to Gibbs objects. This is even true if correlations are allowed to built up between the system and the catalyst.

Proposition 1. *Gibbs-preserving transitions and transitions induced by thermal operations fulfil properties 1–3.*

Proof. This immediately follows from the examples and the definition of thermal operations. \square

Appendix D. Free energy

In this section we give an example for a valid work-quantifier if we choose as free state transitions Gibbs-preserving transitions or those induced by thermal operations. As customary, we define the von Neumann *free energy* of an object $p = (\rho, H)$ as the function

$$\Delta F_1^\beta(\rho, H) := F((\rho, H), \beta) - F((\omega_H, H), \beta), \quad (\text{D.1})$$

where

$$F(p, \beta) := \text{tr}(\rho H) - \frac{1}{\beta} S(\rho), \quad (\text{D.2})$$

and $S(\rho)$ denotes the von Neumann entropy of ρ . We can also express ΔF_1^β using the quantum relative entropy, which is for states ρ and σ defined as

$$S(\rho||\sigma) = \text{tr}(\rho \log \rho - \rho \log \sigma) \quad (\text{D.3})$$

if $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$ and is equal to ∞ otherwise. It is well known that

$$\Delta F_1^\beta(\rho, H) = \frac{1}{\beta} S(\rho||\omega_H). \quad (\text{D.4})$$

This equation also directly shows that $\Delta F^\beta(\rho, H)$ is a well-defined function on objects: it is invariant under maps of the form

$$H \mapsto H + \lambda \mathbb{1} \quad (\text{D.5})$$

for any $\lambda \in \mathbb{R}$. In this section we will prove the following proposition.

Proposition 2 (Properties of the von-Neumann free energy monotone). *The function ΔF_1^β is a monotone under catalytic Gibbs-preserving transitions and fulfils the following properties:*

1. Normalisation: $\Delta F_1^\beta(w) = \Delta F_1^\beta(\emptyset) = 0$ for any w being a Gibbs object.
2. Extensivity:

$$\Delta F_1^\beta(p_A \otimes p_B) = \Delta F_1^\beta(p_A) + \Delta F_1^\beta(p_B). \quad (\text{D.6})$$

3. Strong generalised super-additivity: if $p_{AB}^{(f)}, p_{AB}^{(i)}$ are non-interacting objects on AB ,

$$\begin{aligned} \Delta F_1^\beta(p_A^{(f)}) - \Delta F_1^\beta(p_A^{(i)}) &\geq \Delta F_1^\beta(p_{AB}^{(f)}) \\ &\quad - \Delta F_1^\beta(p_{AB}^{(i)}) \end{aligned} \quad (\text{D.7})$$

if $p_{AB}^{(f)}$ can be reached from $p_{AB}^{(i)}$ by only acting on subsystem A .

Note that this proposition also implies that ΔF_1^β is a monotone for catalytic thermal transitions, since these constitute a strict subset of catalytic Gibbs-preserving transitions. Hence ΔF^β defines a valid work-quantifier for both Gibbs-preserving transitions and thermal operations. The property of strong generalised super-additivity furthermore implies the usual super-additivity $\Delta F_1^\beta(p_{AB}) \geq \Delta F_1^\beta(p_A) + \Delta F_1^\beta(p_B)$, if p_{AB} is non-interacting.

We will separate the proof into several propositions. We will frequently use the following well-known properties of the relative entropy:

1. Positivity: $S(\rho||\sigma) \geq 0$ and $S(\rho||\sigma) = 0$ if and only if $\rho = \sigma$,
2. Data-processing inequality: $S(T(\rho)||T(\sigma)) \leq S(\rho||\sigma)$ for any quantum channel T .
3. Mutual information: for any bipartite state $\rho_{A_1 A_2}$ we have

$$S(\rho_{AB}) = S(\rho_A) + S(\rho_B) - S(\rho_{AB}||\rho_A \otimes \rho_B). \quad (\text{D.8})$$

Positivity directly implies that, for a fixed Hamiltonian H , the Gibbs-state ω_H at inverse temperature β is the unique minimum of the function $\rho \mapsto \Delta F_1^\beta(\rho, H) \geq 0$. Thus we already know that $\Delta F_1^\beta(w) = 0$ for any Gibbs object and that $\Delta F_1^\beta(p) > 0$ if p is not a Gibbs object.

Proposition 3 (Extensivity of the free energy difference). *The function ΔF_1^β is extensive.*

Proof. The proof follows immediately from Property 3. of the relative entropy. □

Proposition 4 (Super-additivity of the free energy difference). *The function ΔF_1^β fulfils strong generalised super-additivity.*

Proof. Assume that two objects

$$p_{AB}^{(i)} = (\rho_{AB}^{(i)}, H_A^{(i)} \otimes \mathbb{1}_B + \mathbb{1}_A \otimes H_B^{(i)}) \quad (\text{D.9})$$

and

$$p_{AB}^{(f)} = (\rho_{AB}^{(f)}, H_A^{(f)} \otimes \mathbb{1}_B + \mathbb{1}_A \otimes H_B^{(f)}) \quad (\text{D.10})$$

are related through a local operation on A . Then

$$\rho_{AB}^{(f)} = (\mathcal{T}_A \otimes \mathbb{1})(\rho_{AB}^{(i)}) \quad (\text{D.11})$$

for some quantum channel \mathcal{T}_A acting on system A and therefore $\rho_B^{(i)} = \rho_B^{(f)}$ and $H_B^{(i)} = H_B^{(f)}$. We need to show that

$$\Delta F_1^\beta(p_A^{(f)}) - \Delta F_1^\beta(p_A^{(i)}) \geq \Delta F_1^\beta(p_{AB}^{(f)}) - \Delta F_1^\beta(p_{AB}^{(i)}). \quad (\text{D.12})$$

If ω_A, ω_B are two Gibbs-states, then it is easy to prove, using locality, that for any state ρ_{AB} on AB we have

$$\begin{aligned} S(\rho_{AB} \| \omega_A \otimes \omega_B) &= S(\rho_{AB} \| \rho_A \otimes \rho_B) \\ &\quad + S(\rho_A \otimes \rho_B \| \omega_A \otimes \omega_B). \end{aligned} \quad (\text{D.13})$$

Using this relation we can rewrite the r.h.s. of equation (D.12) as

$$\begin{aligned} &\frac{1}{\beta} [S(\rho_{AB}^{(f)} \| \rho_A^{(f)} \otimes \rho_B^{(f)}) - S(\rho_{AB}^{(i)} \| \rho_A^{(i)} \otimes \rho_B^{(i)})] \\ &\quad + \frac{1}{\beta} [S(\rho_A^{(f)} \otimes \rho_B^{(f)} \| \omega_{H_A^{(f)}} \otimes \omega_{H_B^{(f)}}) \\ &\quad - S(\rho_A^{(i)} \otimes \rho_B^{(i)} \| \omega_{H_A^{(i)}} \otimes \omega_{H_B^{(i)}})]. \end{aligned} \quad (\text{D.14})$$

Using $\rho_B^{(f)} = \rho_B^{(i)}$, $H_B^{(i)} = H_B^{(f)}$ and extensivity, we find that the second term in brackets reduces to

$$\frac{1}{\beta} [S(\rho_A^{(f)} \| \omega_{H_A^{(f)}}) - S(\rho_A^{(i)} \| \omega_{H_A^{(i)}})] = \Delta F_1^\beta(p_A^{(f)}) - \Delta F_1^\beta(p_A^{(i)}). \quad (\text{D.15})$$

But from the data-processing inequality we get that

$$\frac{1}{\beta} [S(\rho_{AB}^{(f)} \| \rho_A^{(f)} \otimes \rho_B^{(f)}) - S(\rho_{AB}^{(i)} \| \rho_A^{(i)} \otimes \rho_B^{(i)})] = C \leq 0. \quad (\text{D.16})$$

We thus have

$$\text{r.h.s.} = \Delta F_1^\beta(p_A^{(f)}) - \Delta F_1^\beta(p_A^{(i)}) + C \leq \Delta F_1^\beta(p_A^{(f)}) - \Delta F_1^\beta(p_A^{(i)}). \quad (\text{D.17})$$

□

What is left to be proven is that ΔF_1^β is a monotone under free (catalytic) transitions.

Proposition 5 (Monotonicity under Gibbs-preserving transitions). *The function ΔF_1^β is a monotone under Gibbs-preserving transitions.*

Proof. Consider a Gibbs-preserving transition

$$p = (\rho, H) \rightarrow r = (\mathcal{G}(\rho), K) \quad (\text{D.18})$$

with $\omega_K = \mathcal{G}(\omega_H)$. Then we get

$$\begin{aligned} S(\mathcal{G}(\rho) \| \omega_K) &= S(\mathcal{G}(\rho) \| \mathcal{G}(\omega_H)) \\ &\leq S(\rho \| \omega_H), \end{aligned} \quad (\text{D.19})$$

where the last inequality is the data-processing inequality. □

Proposition 6 (Monotonicity under catalytic Gibbs-preserving transitions). *The function ΔF_1^β is a monotone under catalytic Gibbs-preserving transitions.*

Proof. Consider a catalytic transition $p \otimes q \rightarrow r \otimes q$. From monotonicity and extensivity of ΔF_1^β , we obtain

$$\begin{aligned} \Delta F_1^\beta(r) &= \Delta F_1^\beta(r \otimes q) - \Delta F_1^\beta(q) \\ &\leq \Delta F_1^\beta(p \otimes q) - \Delta F_1^\beta(q) \\ &= \Delta F_1^\beta(p). \end{aligned} \quad (\text{D.20})$$

□

A similar proof can also be given in the setting where the catalyst is allowed to become correlated with the system. In this case, we need to use super-additivity of ΔF_1^β . The same applies for the next corollary.

Corollary 37 (Mapping Gibbs objects to Gibbs objects). *Catalytic Gibbs-preserving transitions map Gibbs objects to Gibbs objects.*

Proof. Consider a transition $w \otimes q \rightarrow r \otimes q$. Then $\Delta F^\beta(r) \leq \Delta F^\beta(w) = 0$. But $\Delta F^\beta \geq 0$ and ΔF^β vanishes only on Gibbs-objects. Hence r has to be a Gibbs object. \square

This finishes the proof of proposition 2.

Appendix E. The usual notions of work as a particular case in our formalism

In this section we will review the common definitions of work that have been considered in the literature and recast them as particular cases of our formalism. That is, we will show that the energy stored in the work-storage device is a valid work quantifier fulfilling axioms 1 and 2, where the catalytic free operations and the set of restrictions \mathcal{P} encode the behavior of a lifted weight.

E.1. The average energy of the lifted weight

Here we will discuss the model of work considered in [7, 38]. In this case, the restrictions \mathcal{P} are taken as the quantum analogue of a lifted weight:

$$\mathcal{P}^{\text{mean}} = \{(\rho_A, H_A); H_A = mgX\}, \quad (\text{E.1})$$

where X is the position operator associated to one continuous degree of freedom, m is the mass of the weight and g is the gravitational constant. Note that no restrictions are put on the state ρ_A but the Hamiltonian H_A is fixed throughout the protocol. The work quantifier is defined as

$$\mathcal{W}_{\text{mean}}(p_A^{(i)} \rightarrow p_A^{(f)}) = \text{tr}(\rho_A^{(f)} H_A) - \text{tr}(\rho_A^{(i)} H_A). \quad (\text{E.2})$$

Importantly, the treatment of the work-storage device as a lifted weight is encoded in the set of catalytic free operations \mathcal{F}_C . Following the formalism of [7] and adding to it the notion of a catalyst, we have that the free operations, that we denote by $\mathcal{F}_C^{\text{mean}}$ are given by

$$\mathcal{F}_C^{\text{mean}}(\rho_A) = \{\rho'_A; \rho' = \text{tr}_{BC}(U\omega_B \otimes \sigma_C \otimes \rho_A U^\dagger)\}, \quad (\text{E.3})$$

where ω is a Gibbs state (reflecting a heat-bath), the mean-energy is preserved

$$\text{tr}((U\omega_B \otimes \sigma_C \otimes \rho_A U^\dagger - \omega_B \otimes \sigma_C \otimes \rho_A) H_{BCA}) = 0, \quad (\text{E.4})$$

the catalyst C is left in the same final state,

$$\sigma_C = \text{tr}_{BA}(U\omega_B \otimes \sigma_C \otimes \rho_A U^\dagger), \quad (\text{E.5})$$

and finally, that the unitary U commutes with the space-translation operator on A (see [7, 39] for details).

Given all the conditions, it is shown in [7, 39] that

$$\text{tr}(\rho_A H_A) \leq \text{tr}(\rho'_A H_A) \quad \forall \quad \rho'_A \in \mathcal{F}_C^{\text{mean}}(\rho_A). \quad (\text{E.6})$$

That is, the function $M(\rho, H) := \text{tr}(\rho H)$ is a monotone under $\mathcal{F}_C^{\text{mean}}$ catalytic free operations. Lastly, one can easily show that the average energy fulfils the properties of additivity and super-additivity, hence, using theorem 4 we see that $\mathcal{W}_{\text{mean}}$ fulfills axioms 1 and 2⁵.

As a final remark, notice that the free operations $\mathcal{F}_C^{\text{mean}}$ impose a limitation in comparison to what is usually allowed when thermal operations or Gibbs preserving maps are considered. The condition that the unitary has to commute with the translation operator of the work-storage device prevents one from employing the lifted weight as an entropy sink in the spirit of the example of figure 1. At the same time, it is obvious from the model of the work-storage device and the conditions on $\mathcal{F}_C^{\text{mean}}$ that this idealisation will not represent the realistic behaviour of a nano-machine. A work-storage device made of a few atoms certainly will not have a Hamiltonian of the form (E.1) neither one can expect the operations performed in a real experimental device to, even approximately, commute with the translation operator on the work-storage device, even if its Hilbert-space allows for such operators.

⁵ Strictly speaking, the mean energy does not fulfil the property of being positive for all states in $\mathcal{P}^{\text{mean}}$, as it is also demanded by theorem 4. However, this can be tackled by simply taking $\mathcal{P}^{\text{mean}}$ so that the mean energy of the states is bounded from below. In this way, we can take the origin of energies (which does not alter the definition of work) in such a way all the states have positive energy.

E.2. The *whit* and ϵ -deterministic work extraction

Now we will consider the model of a *whit* and the notion of ϵ -deterministic work as it has been put forward in [8]. The restrictions on the work-storage device are such they are qubits with

$$\mathcal{P}_\epsilon := \{(\rho, H) \mid H = \Delta|1\rangle\langle 1|, \|\rho - |E\rangle\langle E|\|_1 \leq 2\epsilon\}, \quad (\text{E.7})$$

where $|E\rangle$ is an eigenvector of H , $\|\cdot\|_1$ is the 1-norm on quantum states and $\epsilon < \frac{1}{2}$. The restriction \mathcal{P}_ϵ encodes that Arthur is interested in having states of well-defined energy or at least ϵ -close to it. Work is then given by the energy difference of the closest energy-eigenstates, formally as

$$\mathcal{W}_{\text{det}}(p_A^{(i)} \rightarrow p_A^{(f)}) = f(\rho_A^{(f)}, H_A^{(f)}) - f(\rho_A^{(i)}, H_A^{(i)}), \quad (\text{E.8})$$

with the function f being defined as

$$f(\rho, H) = \begin{cases} \Delta & \text{if } \|\rho - |1\rangle\langle 1|\| < 1 \\ 0 & \text{if } \|\rho - |0\rangle\langle 0|\| < 1 \end{cases}. \quad (\text{E.9})$$

It is easy to see that strictly speaking, this model of the *whit* respects axioms 1 and 2 if $\epsilon = 0$. In that case, $\mathcal{P}_{\epsilon=0}$ is given only by pure energy eigenstates. Hence, we find that \mathcal{W}_{det} evaluated on $\mathcal{P}_{\epsilon=0}$ coincides with the work-quantifier defined by the non-equilibrium free-energy on any transition where the Hamiltonian is constant. However, for $\epsilon > 0$ we find that this model of the lifted weight does not satisfy the axioms 1 and 2. Indeed, one can simply check that it does not fulfil equation (18), or in other words, it is possible to store work in the *whit* by simply putting it in contact with a single thermal bath. Indeed, it has been shown in [30] that for any value of $\epsilon > 0$ and β one can find a value of $\Delta > 0$ such that there exists a thermal operation that brings a qubit work-storage system initially in the ground state to a final state $\rho^{(f)} = (1 - \epsilon)|1\rangle\langle 1| + \epsilon|0\rangle\langle 0|$. Hence, for any $\epsilon > 0$, there exist $p^{(i)}, p^{(f)} \in \mathcal{P}_\epsilon$, such that $\mathcal{W}_{\text{det}}(p^{(i)} \rightarrow p^{(f)}) = \Delta > 0$, while $p^{(f)} \in \mathcal{F}_C(p^{(i)})$, in contradiction with (18). Thus, \mathcal{W}_{det} only defines a work quantifier that is compatible with axioms 1 and 2 if $\epsilon = 0$. A discussion on how to define a work quantifier that incorporates the notion of ϵ -deterministic work without running into contradictions is presented in section 7.2.

As a final remark, let us note again that the incompatibility of ϵ -deterministic work with our axioms is unrelated with issues related to reversibility or the fact that $\mathcal{W}(p \rightarrow q) = -\mathcal{W}(q \rightarrow p)$. Indeed, also ϵ -deterministic work fulfils this property. The reason that makes it violate the axioms is the same as the one given in the example of figure 1: when $\epsilon > 0$, the *whit* can act as an entropy sink. Thus, identifying work with energy as \mathcal{W}_{det} does, allows one to extract work by using a single heat bath. Nonetheless, as we propose in section 7.2 in the main text, it is possible to keep the spirit of the ϵ -deterministic work (that the work-storage devices are ϵ -close to pure energy eigenstates) and put forward a proper work quantifier that satisfies the axioms.

Appendix F. Probability distributions of work

We will now discuss how our formalism is perfectly compatible with the notion of work as a classical random variable and the well-known results that pertain to the fluctuations of the probability distribution of work of [26, 27]. In the setting in the focus of attention in so-called fluctuation theorems one considers a system S on which an energy measurement is performed both at the beginning and the end of a given unitary evolution. That is, the initial and final energies $E^{(i)}$ and $E^{(f)}$ are random variables, and so is the work given by $w = E^{(f)} - E^{(i)}$, which occurs with probability $P_W(w)$. Let us assume that the energy difference is bounded so that $P_W(w) \neq 0$ only if $w_{\min} \leq w \leq w_{\max}$. One may then always assume the presence of a work-storage device A that stores the energy lost by S and that is—in each event—in an energy eigenstate.

More explicitly, we consider $\mathcal{P} = \{|x\rangle\langle x|, mgX\}$ where $|x\rangle\langle x|$ is an eigenstate of the truncated position operator

$$X = \int_{w_{\min}}^{w_{\max}} |x\rangle\langle x| dx \quad (\text{F.1})$$

(taking $mg = 1$ for simplicity)⁶. Then, in each event—that is, conditioned on a specific value of the initial and final measurement—the work-storage device undergoes the transition

$$p^{(i)} = (|0\rangle\langle 0|_A, X_A) \rightarrow p^{(f)} = (|w\rangle\langle w|_A, X_A). \quad (\text{F.2})$$

⁶ Here, $\int_{w_{\min}}^{w_{\max}} |x\rangle\langle x| dx$ should be understood as a finite-dimensional Hamiltonian with non-degenerate spectrum within $[w_{\min}, w_{\max}]$, as dense as necessary to reflect all the possible work-values.

Then, we simply take

$$\mathcal{W}(p^{(i)} \rightarrow p^{(f)}) = f(p^{(f)}) - f(p^{(i)}) = w$$

with $f(|x\rangle\langle x|_A, X_A) = x$. Clearly, this work quantifier fulfills axioms 1 and 2, since it coincides, for the case of $\mathcal{P} = \{|x\rangle\langle x|, mgX\}$, with taking $f = \Delta F_1^\beta$, which by theorem 5 fulfils the axioms.

Nonetheless, note that it is crucial to include as part of the work extraction scheme the step in which a measurement is performed at the beginning and at the end. This is necessary even if S is a classical system. Otherwise, the process will result in a mixed state of the work-storage device.

To be more precise, it is important to appreciate that the two following notions are not equivalent:

- (i) $|0\rangle\langle 0|_A, X_A \rightarrow |w\rangle\langle w|_A, X_A$ occurs with probability $P_W(w)$.
- (ii) $|0\rangle\langle 0|_A, X_A \rightarrow (\sum_w P_W(w) |w\rangle\langle w|_A, X_A)$ takes place.

The interpretation of work as a random variable in [26, 27] corresponds to a process of the type (i), where the probability distribution of work $P_W(w)$ encodes our *a priori* knowledge or capability to make predictions about which transition of the form (E.2) is going to take place. The transition given by (ii) is a situation that is not covered by $\mathcal{P} = \{|x\rangle\langle x|, mgX\}$ and \mathcal{W} given simply by (E.2). In order to quantify work for a transition of the form (ii), one has to properly account for the fact that the work-storage device might act as an entropy sink. In other words, in order to account for transitions of the form (ii), one cannot identify work simply with the energy difference, and one has to define a work quantifier that fulfils axioms 1 and 2 for extended sets \mathcal{P} that contain $\sum_w P_W(w) |w\rangle\langle w|_A$ as a valid state.

Although this discussion between the differences of (i) and (ii) is rather obvious, we would like to stress that it plays an important role in the interpretation of work in quantitative terms, since (i) requires to perform an energy measurement, which is not a free operation within any sensible thermodynamic framework. Hence, it should be kept in mind that when referring to work as a random variable, one is effectively quantifying the work extracted/invested in the process plus the work extracted/invested in the measurements, which of course is a perfectly valid approach.

In contrast, in the formalism of [8, 33], deterministic values of energy are obtained, not by conditioning on the value of an energy measurement, but by engineering the protocol in such a way the work-storage device ends in a deterministic state of energy.

Lastly, let us point out that our axiomatic framework can incorporate the notion of work as a random variable in situations more general than the one considered above. For example, one may consider probability distributions of work for arbitrary processes between two measurements described by POVMs $\{M_\alpha\}_\alpha$, provided that all the post-measurement states are valid work-storage devices, i.e., are fairly included in the set \mathcal{P} . Then, for any work quantifier \mathcal{W} that fulfils axioms 1 and 2 for the set \mathcal{P} , the work $\mathcal{W}(p_{\alpha_i}^{(i)} \rightarrow p_{\alpha_f}^{(f)})$ occurs with probability $P(\alpha_i, \alpha_f)$, where $p_{\alpha_i}^{(i)}$ is the initial state conditioned on having obtained outcome α_i initially (and equivalently for the state $p_{\alpha_f}^{(f)}$ after the final measurement) and $P(\alpha_i, \alpha_f)$ is the joint probability distribution of obtaining the pair α_i, α_f . In this way, we see that the framework laid out in this work and the picture of capturing work as a probability distribution are compatible.

References

- [1] Balzani V, Credi A and Venturi M 2006 *Molecular Devices and Machines* (New York: Wiley)
- [2] Cerefolini G 2009 *Nanoscale Devices* (Berlin: Springer)
- [3] Kurizki G, Bertet P, Kubo Y, Mølmer K, Petrosyan D, Rabl P and Schmiedmayer J 2015 *Proc. Natl Acad. Sci.* **112** 3866–73
- [4] Kosloff R and Levy A 2014 *Ann. Rev. Phys. Chem.* **65** 365
- [5] Alicki R 1979 *J. Phys. A* **12** 103
- [6] Janzing D, Wocjan P, Zeier R, Geiss R and Beth T 2000 *Int. J. Th. Phys.* **39** 2717
- [7] Skrzypczyk P, Short A J and Popescu S 2014 *Nature Comm.* **5** 4185
- [8] Horodecki M and Oppenheim J 2013 *Nature Comm.* **4** 2059
- [9] Aberg J 2013 *Nature Comm.* **4** 1925
- [10] Renes J 2014 *Eur. Phys. J. Plus* **129** 153
- [11] Halpern N Y and Renes J M 2016 *Phys. Rev. E* **93** 022126
- [12] Faist P, Dupuis F, Oppenheim J and Renner R 2015 *Nature Comm.* **6** 7669
- [13] Brunner N, Linden N, Popescu S and Skrzypczyk P 2012 *Phys. Rev. E* **85** 051117
- [14] Levy A and Kosloff R 2012 *Phys. Rev. Lett.* **108** 070604
- [15] Kolar M, Gelbwaser R A D and Kurizki G 2012 *Phys. Rev. Lett.* **109** 090601
- [16] Linden N, Popescu S and Skrzypczyk P 2010 *Phys. Rev. Lett.* **105** 130401
- [17] Brandao F G S L, Horodecki M, Oppenheim J, Renes J M and Spekkens R W 2013 *Phys. Rev. Lett.* **111** 250404
- [18] Horodecki M and Oppenheim J 2013 *Int. J. Mod. Phys. B* **27**
- [19] Coecke B, Fritz T and Spekkens R W 2014 A mathematical theory of resources (arXiv:1409.5531)

- [20] Brandao F G S L and Gour G 2015 *Phys. Rev. Lett.* **115** 070503
- [21] del Rio L 2015 Resource theories of knowledge *PhD Thesis*
- [22] Brandao F G S L, Horodecki M, Ng N H Y, Oppenheim J and Wehner S 2015 *Proc. Natl Acad. Sci.* **112** 3275
- [23] Ng N H Y, Mancinska L, Cirstoiu C, Eisert J and Wehner S 2015 *New J. Phys.* **17** 085004
- [24] Lostaglio M, Müller M P and Pastena M 2015 *Phys. Rev. Lett.* **115** 150402
- [25] Lieb E H and Yngvason J 2013 *Proc. Roy. Soc.* **469** 20130408
- [26] Jarzynski C 1997 *Phys. Rev. Lett.* **78** 2690
- [27] Crooks G E 1999 *Phys. Rev. E* **60** 2721
- [28] Talkner P, Lutz E and Hänggi P 2007 *Phys. Rev. E* **75** 050102
- [29] Faist P, Oppenheim J and Renner R 2015 *New J. Phys.* **17** 043003
- [30] Wilming H, Gallego R and Eisert J 2016 *Phys. Rev. E* **93** 042126
- [31] del Rio L, Aberg J, Renner R, Dahlsten O and Vedral V 2011 *Nature* **474** 61
- [32] Anders J and Giovannetti V 2013 *New J. Phys.* **15** 033022
- [33] Aberg J 2014 *Phys. Rev. Lett.* **113** 150402
- [34] Cwiklinski P, Studzinski M, Horodecki M and Oppenheim J 2015 *Phys. Rev. Lett.* **115** 210403
- [35] Dahlsten O C O *et al* 2011 *New J. Phys.* **13** 053015
- [36] Egloff D, Dahlsten O C O, Renner R and Vedral V 2015 *New J. Phys.* **17** 073001
- [37] Faist P, Dupuis F, Oppenheim J and Renner R 2015 *Nat. Commun.* **6** 7669
- [38] Skrzypczyk P, Short A J and Popescu S 2013 Extracting work from quantum systems (arXiv:1302.2811)
- [39] Malabarba A S L, Short A J and Kammerlander P 2015 *New J. Phys.* **17** 045027
- [40] Lostaglio M, Korzekwa K, Jennings D and Rudolph T 2015 *Phys. Rev. X* **5** 021001
- [41] Gemmer J and Anders J 2015 *New J. Phys.* **17** 085006
- [42] Müller–Lennert M, Dupuis F, Szehr O, Fehr S and Tomamichel M 2013 *J. Math. Phys.* **54** 122203
- [43] Wilde M M, Winter A and Yang D 2014 *Comm. Math. Phys.* **331** 593–622
- [44] Mueller M P and Pastena M 2016 *IEEE Trans. Inf. Th.* **62** 1711–20