

On the thermodynamical cost of some interpretations of quantum theory

Carina E. A. Prunkl* ¹ and Christopher G. Timpson[†] ²

¹*Balliol College, University of Oxford*

²*Brasenose College, University of Oxford*

November 24, 2017

Abstract

Recently, Cabello et al. (2016) claim to have proven the existence of an empirically verifiable difference between two broad classes of quantum interpretations. On the basis of three seemingly uncontentious assumptions, (i) the possibility of randomly selected measurements, (ii) the finiteness of a quantum system’s memory, and (iii) the validity of Landauer’s principle, and further, by applying computational mechanics to quantum processes, the authors arrive at the conclusion that some quantum interpretations (including central realist interpretations) are associated with an excess heat cost and are thereby untenable—or at least—that they can be distinguished empirically from their competitors by measuring the heat produced. Here, we provide an explicit counterexample to this claim and demonstrate that their surprising result can be traced back to a lack of distinction between system and external agent. By drawing the distinction carefully, we show that the resulting heat cost is fully accounted for in the external agent, thereby restoring the tenability of the quantum interpretations in question.

Keywords: Quantum Interpretations, Computational Mechanics, Landauer’s Principle, Thermodynamics

*carina.prunkl@balliol.ox.ac.uk

†christopher.timpson@bnc.ox.ac.uk

Contents

1	Introduction	3
2	Computational Mechanics and the Foundations of Quantum Mechanics	4
2.1	Computational Mechanics: Input-Output Processes	4
2.2	Foundations: Division of Interpretations into two Groups	5
2.3	Three Assumptions	7
2.4	Heat Dissipation between Successive Measurements	8
3	Limitations of Computational Mechanics	11
3.1	Interlude: Landauer's Principle and Irreversibility	11
3.2	A Counterexample: Type I without Heat Dissipation	13
3.3	Some Remarks on Computational Mechanics	18
4	Conclusion	18

1 Introduction

For nearly a century, physicists and philosophers alike have puzzled over how to interpret quantum theory, unable to decide unambiguously between a variety of more or less promising candidates. In a recent publication, Cabello et al. (2016) put forward an argument which seeks to demonstrate the existence of a real, physical—as opposed merely to a metaphysical—difference between various interpretations of quantum mechanics. Moreover, the authors assert that it is in principle possible to measure this difference experimentally. Their argument is based on methods derived from computational mechanics—a growing field that is concerned with the simulation and prediction of stochastic processes. Interestingly, when applied to certain physical processes, computational mechanics is able to provide us with thermodynamical limitations on these processes (Wiesner et al., 2012; Garner et al., 2017). Cabello et al.’s argument is a concrete, foundationally motivated, application of computational mechanics which suggests that there is a thermodynamical cost to bear for a subset of quantum interpretations: perhaps a pathological one.

The link between thermodynamics and computational mechanics can be understood as follows: depending on the complexity, i.e., randomness, of a pattern that is to be simulated, greater or fewer resources are needed in order either to create the pattern or to predict its future, given observations of past data sequences. By ‘pattern’ we simply mean a time-series of data points, for example measurement outputs. We can take the computational system we are interested in simulating to be a black box, with the only accessible empirical data being its input and output variables. It can then be proven that there exists a machine, called an ϵ -machine, which is predictively optimal and uses the minimum resources, while simulating the input-output behaviour of the target system (Crutchfield and Young, 1989). For some thermodynamic systems this method shows up the limitations for work extraction via physical processes. Given a resource-theoretic understanding of thermodynamics (i.e., an understanding which conceives thermodynamics primarily to be a theory about what tasks one can perform when furnished with certain resources¹), one might say that computational mechanics can be considered a useful tool for the task of understanding and enhancing the foundations of thermal physics.

Cabello et al. begin by dividing the set of quantum interpretations into two subsets, what they term *Type I* and *Type II* interpretations. They then argue that Type I interpretations are associated with a thermodynamical cost, rendering such interpretations highly problematic: either one of the (very plausible) three assumptions must be given up, or there exists a surprising heat generation which could be ruled in or out experimentally (and one

¹Modern accounts include (Horodecki and Oppenheim, 2013; Wallace, 2014; Brandão et al., 2015; Gour et al., 2015), however, the underlying idea that thermodynamics is to be understood relative to an agent and her means goes back to Maxwell (1871) (c.f. Myrvold (2011)) and was later famously promoted by Jaynes (1965).

would be surprised indeed if such heat generation were in fact to be found). If correct, this result would seem an outstanding breakthrough whose far reaching consequences might not only force us to abandon some of the most popular interpretations of quantum mechanics (Type I interpretations include such favourites as de Broglie–Bohm theory, Everett, and dynamical collapse theories such as GRW, for example) but would shake the foundations of our understanding of the relationship between scientific theories and the underlying ontic structure of the world.

Our prime concern in this paper is to assess Cabello et al.’s argument and the tenability of their conclusions. But we also have their example in mind as a test-case for the application of computational mechanics in pursuit of dividends in foundations of physics.

We will begin with a brief outline of stochastic input-output processes before presenting the argument of Cabello et al. (2016), which applies this mathematical machinery to quantum systems. We will then analyse why Cabello et al.’s argument about the thermodynamical costs of some quantum interpretations fails, including offering a straightforward counterexample. We will show that the adumbrated heat cost is in fact not associated with the quantum system itself at all—it is not of quantum origin—and thus controversies over quantum interpretations are not germane to it, nor it to them. Rather, the heat cost arises with the external experimental setup stipulated by Cabello et al.

2 Computational Mechanics and the Foundations of Quantum Mechanics

We begin by introducing the most important aspects of stochastic input-output processes, as they form the backbone of the argument. More detailed discussions may be found in (Barnett and Crutchfield, 2015; Crutchfield and Young, 1989).

2.1 Computational Mechanics: Input-Output Processes

The goal behind modelling systems’ behaviour by input-output processes is to find the minimal structural requirements that produce a particular statistical pattern. To do so, one works backwards from the statistics of experimental outputs to then find the minimal amount of resources needed in order to simulate output strings that are statistically indistinguishable from the actual experimental result.

More formally: A stochastic process $\overleftrightarrow{Y}$ is described as a bi-infinite one-dimensional chain $\dots, Y_{-1}, Y_0, Y_1, \dots$ of discrete random variables $\{Y_t\}$ with values $\{y_t\}$, where t is a discrete time parameter and the direction of the arrow above the random variable indicates whether the chain extends to the past (left arrow), the future (right arrow) or to past *and* future

(left-right arrow) infinity. The $\{y_t\}$ are the particular values the random variable takes at time t and in our case we can think of them as the output values of an experiment performed on the system. For example, for a spin-measurement on a qubit—the kind of case with which Cabello et al. will be concerned—the outcome-types could be “up” and “down” for example, taken from the output alphabet $\mathcal{Y} = \{\text{“up”}, \text{“down”}\}$. If not only the output but also the input is stochastic (in our case, this will correspond to a choice of spin-measurement basis, which will be taken to be random) the effect of the input random variable on the future statistics needs to be taken into account as well. Such a process must then be modelled by a so-called stochastic *input-output process*, $\overleftarrow{Y}|\overleftarrow{X}$, with input values $\{x_t\}$ from an alphabet $\mathcal{X}$. The whole input-output process may then be described as a collection of stochastic processes $\overleftarrow{Y}|\overleftarrow{x} \equiv \{\overleftarrow{Y}|\overleftarrow{x'}\}_{\overleftarrow{x'} \in \overleftarrow{\mathcal{X}}}$, where we take each process $\overleftarrow{Y}|\overleftarrow{x}$ to correspond to all possible output sequences $\overleftarrow{Y}$ that could arise from one particular input sequence $\overleftarrow{x}$, drawn from the set of all possible input sequences $\overleftarrow{\mathcal{X}}$.

The *probability distribution*² over the set of all possible output sequences, given a particular input sequence, is then given by what is called the *channel’s distribution*:

$$\mathbf{P}(\overleftarrow{Y}|\overleftarrow{x}) = \{\mathbf{P}(\overleftarrow{Y} \in \sigma | \overleftarrow{X} = \overleftarrow{x'})\}_{\sigma \subseteq \overleftarrow{\mathcal{Y}}, \overleftarrow{x'} \in \overleftarrow{\mathcal{X}}} \quad (1)$$

The idea is now to divide the input-output sequences into *pasts* and *future* and furthermore to divide the various input-output pasts into sets that yield the same distribution over input-output futures. Two input-output pasts $\overleftarrow{z} = (\overleftarrow{x}, \overleftarrow{y})$ and $\overleftarrow{z}'$ that yield the same future input-output conditional probabilities $P(\overrightarrow{Y}|\overrightarrow{X}, \overleftarrow{Z} = \overleftarrow{z}) = P(\overrightarrow{Y}|\overrightarrow{X}, \overleftarrow{Z} = \overleftarrow{z}')$ are then said to belong to the same *causal state* s . Denote the set of causal states $\mathcal{S}$. The ϵ -map is then introduced as the mapping $\epsilon : \overleftarrow{\mathcal{Z}} \rightarrow \mathcal{S}$ from any input-output past onto its corresponding causal state (Barnett and Crutchfield, 2015). This map also induces a probability distribution over the causal states, which since the process is stationary and ϵ is time-independent, is called the process’ *stationary distribution*. The causal states contain all the relevant information for optimally predicting the future output statistics of the system and contain as much information as any of its input-output pasts. Here we take the input sequences to be uniformly distributed. The minimal amount of information needed to be stored in order to predict future outputs optimally is then given by the Shannon information $H(\mathcal{S})$ and is called the *statistical complexity*. This also quantifies the amount of resources needed in order to model the system’s future behaviour.

2.2 Foundations: Division of Interpretations into two Groups

Cabello et al. (2016) seek to use the above machinery in combination with a few plausible

²We consider only stationary probabilities, which means that the probabilities are time translation invariant.

assumptions to raise difficulties for a group of well-known quantum interpretations. Their approach is to divide the set of quantum interpretations into two broad classes, based on their respective takes on quantum probabilities: Type I interpretations are interpretations that regard probabilities as determined by “intrinsic properties of the system” (Cabello et al., 2016, p.1). These properties typically change post-measurement, depending on the choice of measurement performed on the quantum system. Examples which they mention of Type I interpretations include de Broglie-Bohm theory (Bohm, 1952; Goldstein, 2017), many worlds interpretations, e.g. (Everett, 1957; Wallace, 2012), Ballentine’s statistical interpretation (Ballentine, 1970), modal interpretations (Lombardi and Dieks, 2016) and consistent histories, as well as GRW dynamical collapse theories (Ghirardi et al., 1986) and Spekkens’ toy model (Spekkens, 2007).

Type II interpretations, on the other hand, comprise those interpretations which treat probabilities as “relational properties between an observer and the system” (Cabello et al., 2016, p.1). According to such interpretations, the quantum state corresponds to the “experiences an observer has of the observed system”(Cabello et al., 2016, p.1). To the class of Type II interpretations, the authors assign, amongst others, the Copenhagen interpretation, Wheeler’s view (Wheeler and Zurek, 2014), relational interpretations (Rovelli, 1996) and QBist interpretations (Fuchs and Peres, 2007).

We note to begin that there are some significant problems associated with such a division of interpretations. First, some of the interpretations listed as Type I do not seem unambiguously to belong in either camp: the Everett interpretation, for instance, may well be taken to imply that probabilities ought to be regarded relationally as opposed to being intrinsic properties of a quantum system. After all, the Everettian could argue, the ‘bare quantum formalism’ is fully deterministic at the fundamental level. The probabilities, by contrast, occur at an emergent level and justifications of the Born rule are typically given via decision-theoretic arguments (Wallace, 2012), very strongly suggesting that probabilities in Everett, if they need an interpretation at all, are best interpreted as concerning relational experiences of an agent within a given branch.

The Type I/ Type II distinction also seems problematic with regard to Ψ -epistemic interpretations, in which the quantum state is taken to represent information about how a system is³. For instance, the probabilities occurring in Spekkens’ toy model—which allegedly belongs to the Type I camp—are purely epistemic and result from an incomplete knowledge about the underlying ontic state⁴ (Spekkens, 2007), as opposed to being intrinsic properties of the system: they merely occur on the level of the observer whereas the underlying ontic evolution could be, for all we know, a deterministic one. It thereby seems

³For details on the Ψ -ontic/ Ψ -epistemic distinction, see Harrigan and Spekkens (2010).

⁴In analogy with Jaynes’ view of probabilities in classical statistical mechanics, where the probability distribution over the underlying ontic state equally represents ignorance (Jaynes, 1965)

somewhat misplaced to regard these probabilities as “observer-independent” (Cabello et al., 2016, p.1), in the sense of being part of the furniture of the world, a notion more commonly found in the context of propensity interpretations of probabilities (Popper, 2002), but certainly not in an epistemic context.

We may try to make clearer sense of the Type I/ Type II distinction by phrasing it differently in order to capture what Cabello et al. seem to have in mind: namely a distinction between interpretations that either endorse or reject the notion of an underlying ontic physical state. What the interpretations listed under Type II then have in common is that they all roughly follow along Niels Bohr’s line: There is no quantum world. At least, there is not a free-standing mind- and agent- independent one. Type II interpretations thereby operate on a completely different level from Type I interpretations, the latter of which consider quantum mechanics as ultimately describing a free-standing mind-independent objective reality. Drawing a distinction between these fundamentally different conceptions of scientific theory on the basis of their conception of probabilities, however, is misleading and unsatisfying. The probabilities are mere epiphytes on a deeper, more pressing issue about the nature of science.

Consequently, if Cabello et al. are indeed correct and Type I interpretations could be ruled-out experimentally, then their result would have profound consequences: it undermines one of our core traditional conceptions of scientific practice, namely that scientific theories tell us what the world independent of us is like.

2.3 Three Assumptions

As Cabello et al. explain, their argument concerning Type I interpretations rests on three assumptions, which we can state as follows:

- (i) Which measurement is performed on a system is decided randomly, and in particular independently of the state of the system.
- (ii) A (finite dimensional) quantum system has a limited memory.
- (iii) Landauer’s principle is valid.

They go on to suggest that for Type I interpretations, it follows from (i) that a system’s *intrinsic properties* typically ought to *change*, depending on which measurement is performed on the system. When the authors speak of ‘limited memory’, (ii), they seem to have in mind something like the following: when a system is being measured in some basis, it typically generates a new value of its intrinsic property or set of intrinsic properties, which will determine the quantum probabilities of the future measurements. Measurement in

a different basis will force the system to change these values in order to comply with the correct quantum probabilities, thereby deleting—or perhaps rather overwriting—the previous values. Due to its limited memory capacity, the system cannot possess all the intrinsic values that will determine the probabilities of all possible measurement series. We may assume that by ‘intrinsic properties’ Cabello et al. have in mind properties that determine the probabilities of future measurement outcomes as opposed to properties specifying pre-determined definite outcomes, the reason being that this latter interpretation of “intrinsic properties” would force us to adapt some sort of hidden variable interpretation, which, by definition is only a subgroup of the Type I class. We therefore take it that the former meaning of ‘intrinsic properties’ is intended.

Preliminary assumption (iii) is the validity of Landauer’s principle. As it is often stated, Landauer’s principle asserts that the erasure of information in a system’s information bearing degrees of freedom is accompanied by an increase of entropy in the non-information bearing degrees of freedom (Landauer, 1961). This increase of entropy will lead to the dissipation of $kT \ln 2$ of heat per deleted bit, where T is taken by Cabello et. al. to be the temperature of the system (though more usually it is taken to be the temperature of the environment) and k is, of course, the Boltzmann constant.

2.4 Heat Dissipation between Successive Measurements

We are now invited to consider a single qubit, on which an observer performs successive projective measurements in one of the two Pauli-bases, σ_x or σ_z , chosen at random with equal probability (assumption (i)). Since the state of the qubit changes after successive measurements in non-orthogonal bases, within a Type I framework the internal properties of the system must change too. Given that the system has limited memory (assumption (ii)), it needs to generate new values that determine its future behaviour and store them in its memory. To do so, the system will need—as they put it—to ‘erase information’. Applying Landauer’s principle (assumption (iii)), Cabello et al. (2016) finally conclude that during this erasure process, heat is dissipated into the environment.

To quantify the amount of information that needs to be erased, they model the system as a computational machine, a black box that generates output strings on the basis of some input and its internal memory. The optimal and minimal machine that is able to produce output strings that are statistically indistinguishable from the actual experimental outcomes, will be, as we mentioned earlier, an ϵ -machine (Crutchfield and Young, 1989). It maximises the mutual information between input-output past and output future, thus simulating the statistical process, and with minimal resources.

Applying the computational mechanical machinery to the qubit in question, we identify the input random variable X_t with the choice of measurement basis, randomly selected from

the alphabet $\mathcal{X} = \{\sigma_x, \sigma_z\}$. The output variable Y_t represents the measurement results and can take values ± 1 . The causal state after the measurement is simply taken to be the respective quantum state, $s_o = |0\rangle$, $s_1 = |1\rangle$, $s_+ = |+\rangle$ or $s_- = |-\rangle$ (with, note, no specific view here taken on the ontology or otherwise, of the quantum state).

The machine has a probability $1/2$ of changing its causal state after a measurement. Hence, half of the time, it must update its internal properties, and Cabello et al maintain that this requires the erasure of information. They say “*The average information that must be erased per measurement is the information contained in the causal state previous to the measurement, S_{t-1} , that is not contained in the causal state after the measurement, S_t .*” (Cabello et al., 2016, p.2). It should be noted that this formulation is perhaps somewhat misleading, as it suggests that *a particular* causal state itself carries a certain amount of information. In fact, it is not *the* causal state to which we assign an entropy, but it is instead the probability distribution over causal states which is associated with an entropy and thereby with a Shannon information. Such an average over causal states however is not a causal state itself.

In any case, the amount of information that needs to be erased is equal to the *conditional entropy* of

$$I_{erased} = H(S_{t-1}|X_t, Y_t, S_t). \quad (2)$$

(See Appendix.)

In the given experiment, the probability distribution over the causal states is uniform, which allows us to only consider a particular causal state s_0 , brought about by a particular measurement σ_z in order to determine I_{erased} . Together with the fact that $H(Y_t|S_t) = 0$, the average erased information for the case of a successively measured qubit is then calculated to be

$$I_{erased} = H(S_{t-1}|\sigma_z, s_0) = - \sum_{s_j \in \mathcal{S}} P(S_{t-1} = s_j|\sigma_z, s_0) \log P(S_{t-1} = s_j|\sigma_z, s_0). \quad (3)$$

The three possible causal states at time $t-1$ are s_0, s_+ and s_- , with transition probabilities $1/2, 1/4$ and $1/4$.⁵ The conditional entropy then turns out to be $I_{erased} = -\frac{1}{2} \log \frac{1}{2} - 2 \cdot \frac{1}{4} \log \frac{1}{4} = \frac{3}{2}$ bits.

Cabello et al. (2016) thereby conclude that once we accept assumptions (i)–(iii), it follows that the system on average must dissipate $\frac{3}{2}kT \ln 2$ units of heat per measurement, if understood as belonging to Type I.⁶ In principle, the above experiment could be implemented

⁵The authors write at time t , but must have intended $t-1$.

⁶The authors furthermore generalise their result to N-outcome measurements with an associated heat

in a lab. From our previous observations of measurements on quantum systems, it is however safe to say that we would be very surprised indeed to observe any such heat dissipation. Cabello et al. thereby suggest that Type I interpretations are unlikely to be representative of the world, at least if their plausible assumptions (i)–(iii) hold. The above argument supposedly does not apply to Type II interpretations, however, as for these interpretations “measurement outcomes are created randomly when the observables are measured, without any need to overwrite information in the system and therefore without the system dissipating heat due to Landauer’s principle” (Cabello et al., 2016, p.3).

In the conclusion of their paper Cabello et al. do canvass the possibility that one or more of assumptions (i)–(iii) might be thought to fail instead of Type I interpretations being lumbered with an excess heat cost, and in particular they judge that the de Broglie–Bohm theory and the Everett interpretation should not have the excess heat quantity attached to them, as both interpretations violate (ii)—the finite memory assumption. As they see it, in the de Broglie–Bohm case this is because the ontology includes a continuous field (the unitarily evolving wave function), and in the Everett case, because one has a splitting into a plurality of worlds. But it is at best obscure that either de Broglie–Bohm or Everett should be thought to violate the finite memory assumption. In the de Broglie–Bohm case, from the fact that the quantum state is taken to be real⁷ it does not follow that a system has infinite memory capacity: nearly all the state is irrelevant to the time evolution of the definite physical quantities most of the time anyway due to decoherence, whilst a particle’s motion is only guided by the value of the wavefunction assigned to the region immediately surrounding it in any case⁸. And one might note that realist collapse theories such as GRW *also* have a continuous field in them: what difference could it make to judgements of memory capacity whether the continuous field (sometimes) jumps around stochastically (GRW) or whether it instead evolves deterministically but most of it being irrelevant to the evolution of a particle (de Broglie–Bohm)? With regard to Everett: even if there is branching, each world would be one in which the finite memory condition held, so each world would be one in which the excess heat cost obtained⁹. If anything, this would look worse, rather than better, for Everett.

In fact, Cabello et al. are quite right that there is no excess heat cost which arises for de Broglie–Bohm or for Everett, but this is not for the reason they allege (the possibility of infinite memory capacity). Rather it is, as we shall see, an instance of a general proposition:

generation which scales linearly with N . Thus sufficient measurements on a single system would produce as much heat as you like. This does not sound promising for Type I interpretations.

⁷Putting aside those views which would see it as law-like—nomological rather than ontological. (Dürr et al., 1995, 2012)

⁸In the multi-particle case, similarly, the motion of the occupied point in configuration space (representing the position of the n particles) is guided by the wavefunction assigned to the immediately surrounding region of configuration space.

⁹At least in approximations where coarse-graining means that a system – perhaps an isolated, or relatively isolated, degree of freedom – is being treated effectively as a finite-dimensional system.

there is no excess heat cost for any Type I interpretation over a Type II interpretation.

3 Limitations of Computational Mechanics

Let us now analyse how Cabello et al. arrive at their surprising results. We will begin with a brief recapitulation of Landauer’s principle before delivering an explicit counterexample to Cabello et al.’s claim and then identifying the shortcomings of their argument.

3.1 Interlude: Landauer’s Principle and Irreversibility

A great deal in this argument hinges on the application of Landauer’s principle, often taken to be the claim that the implementation of a logically irreversible operation is accompanied by a dissipation¹⁰ of $kT \ln 2$ units of heat per bit into the environment. An operation is considered to be logically irreversible, if the output of the operation does not uniquely determine the input (Landauer, 1961). Often, the concept of information is invoked in order to describe this logical irreversibility. Wiesner et al. (2012) write that “logically irreversible operations forget information about the computational device’s preceding logical state.” [p.4060].

To appreciate why such characterisations of Landauer’s principle in terms of ‘forgetting’ can be misleading, however, we consider a logically irreversibly operation which can be implemented without any heat cost, the so-called RAND operation (Maroney, 2005). RAND randomises the logical state of a bit, regardless of its input state. Physically, we may think of an implementation of RAND in the standard way of considering a molecule in a box, in which a partition is included. The molecule is originally on the left side (or right side) of the box and the whole system is in contact with a heat bath at temperature T_{HB} . Implementing RAND then simply requires one to remove the partition, wait for a sufficiently long time and then re-insert the partition. The operation is logically irreversible because the output (the randomly distributed molecule) does not uniquely determine the input (the molecule being in one of the two mutually exclusive states)—but obviously there was no heat exchange with the environment during this process. In a very naive sense, the RAND operation has ‘erased information’, but this erasure of information has taken place at no heat cost. Landauer’s principle therefore cannot simply be the statement that the implementation of a logically irreversible operation leads to the dissipation of heat into the environment.

The most prominent application of Landauer’s principle is the so-called *Landauer erasure* process, a resetting operation that maps the state of a randomised bit back to some pre-defined initial state. For the above described molecule-in-a-box scenario, such an erasure

¹⁰Heat dissipation is generally taken to be thermodynamically irreversible.

can be implemented by removing the partition and then pushing it isothermally in from one side of the box, until the particle is found once again certainly in the left (or right) side of the box. For this last step, work is performed on the system and heat is transferred into the surrounding heat bath¹¹. This is done in a thermodynamically reversible fashion and therefore corresponds to a *heat transfer* and not to a *heat dissipation*. Because the system ends up in a pre-defined state, the Landauer erasure is distinct from the RAND operation described above, although both of them are logically irreversible. In general, whether or not a given logical operation can be performed in a thermodynamically reversible or irreversible depends on the choice of implementation. In principle therefore, any logical operation can be implemented in a thermodynamically reversible fashion. More details on this can be found in (Maroney, 2009), who's approach we follow closely in this section.

What Landauer's principle does is provide us with a link between logical operations and the fundamental microdynamics. Properly put, it states that a logical transformation, reversible or irreversible, must be accompanied by a minimal average heat dissipation into the environment according to

$$\langle \Delta Q \rangle \geq -T_{HB} \Delta S, \quad (4)$$

where ΔS refers to the difference in von Neumann entropy given by

$$\Delta S = \sum_{\beta} P(\beta) (S_{\beta} - k \ln P(\beta)) - \sum_{\alpha} P(\alpha) (S_{\alpha} - k \ln P(\alpha)) \quad (5)$$

and ΔQ is the heat generated in the environment (Maroney, 2009). Phrased like this, Landauer's principle not only becomes utterly unmysterious, it also becomes clear that whether or not heat is transferred into the environment solely depends on the von Neumann entropy of the system before and after the operation. This means in particular that one does not need to make any reference to information erasure or the like.

If we re-write Landauer's principle in a way that makes use of the concept of information by having it explicitly include the entropy of the information bearing degrees of freedom, the change in (Gibbs-) von Neumann entropy is related to the change in Shannon entropy by $\Delta S = \sum_{\beta} P(\beta) S_{\beta} - \sum_{\alpha} P(\alpha) S_{\alpha} + k \Delta H \ln 2$, with ΔH being the change in Shannon entropy. Making use of Equation (4), Landauer's principle then becomes

$$\Delta S_{NI} \geq -k \Delta H \ln 2, \quad (6)$$

where the entropy change of the non-information bearing degrees of freedom is given by

¹¹Similarly in the quantum case for a qubit, one needs to step-wise raise one of the two energy levels to infinity (Barnett and Crutchfield, 2015).

the entropy change in the environment and the weighted entropy changes of the sub-ensembles: $\Delta S_{NI} = \Delta S_E + \sum_{\beta} P(\beta)S_{\beta} - \sum_{\alpha} P(\alpha)S_{\alpha}$. This quantity is increasing for logically irreversible, deterministic operations. For non-deterministic operations however, it is decreasing, making Equation (4) more useful in the general context.

Now: Let us consider applying Landauer’s principle in the form of Equation (4) to the repeatedly measured quantum system. It is clear that once the measurement process is up and running, the quantum system will be in a maximally mixed state at each time step, independent of the chosen measurement basis. This means that the difference in the von Neumann entropy of the quantum system before and after each measurement is zero. Given, furthermore, that the density matrix is the appropriate entity for calculating the entropy of a quantum system—even if one’s interpretation involves further variables as is the case in the de Broglie–Bohm theory for example—it follows that Landauer’s principle does not predict a heat cost for Type I interpretations: the lower bound on heat dissipation into the environment is zero. Notice that we have had to say nothing here of the storing or deleting or erasure of information.

But now, interestingly, we seem to have arrived at a contradiction: On the one hand, applying Landauer’s principle to the successive measurements on a quantum system seems to entail a non-zero lower bound to the average heat cost per measurement. Whilst on the other it entails a zero lower bound. What needs to give?

In fact, nothing. This appearance of contradiction is misleading. The two results are not in fact in conflict. Both involve licit applications of Landauer’s principle, but it is only the second (the zero heat minimum heat cost claim) which pertains to features of the quantum system.

To explain this point and to help clarify where Cabello et al.’s argument has gone wrong, we will now construct an explicit counterexample to their argument. It will become apparent that the heat cost they calculate does not stem from the quantum system itself but from the particular setup that is chosen. The heat cost will be shown to be due to *external* matters.

3.2 A Counterexample: Type I without Heat Dissipation

Spekkens’ toy model (Spekkens, 2007) is explicitly taken to be a Type I interpretation. We will make use of this simple and transparent framework in order to illustrate how successive measurements in non-orthogonal bases do *not* lead to a predicted heat cost.

In Spekkens’ toy model, measurements of the system only yield incomplete knowledge about the underlying state in such a way that the maximal amount of knowledge about the ontic state equals the lack of knowledge about it. A Spekkens qubit can be in one of four possible

ontic states, ‘1’, ‘2’, ‘3’ or ‘4’. Measurements are identified with questions of the form: is the system in state $1 \vee 2$ or $3 \vee 4$ (measurement in the $|0\rangle / |1\rangle$ basis), or, alternatively, $1 \vee 3$ or $2 \vee 4$ (measurement in the $|+\rangle / |-\rangle$ basis)? A measurement result of $|0\rangle$ will therefore yield a state of knowledge of $1 \vee 2$. Impressively, quantum behaviour of a single qubit for measurements restricted to the x , y , and z bases is fully recovered in this model.

If a system is prepared to be in $1 \vee 3$ but then measured to be in $1 \vee 2$, thereby performing the Spekkens equivalent of a non-orthogonal measurement, we know for a fact that the system must have been in ontic state 1 *before* the measurement (Spekkens, 2007). This means that it is possible to know the precise ontic state of the system at an earlier time, but impossible to know it at the current time. A measurement in a non-orthogonal basis therefore *disturbs* the system in such a way that its values become non-definite in its previous basis.

Since we only consider measurements in the x and z directions, we can illustrate Spekkens’ toy model conveniently by considering a classical particle entrapped in a two dimensional box in contact with a heat bath, as illustrated in Figure (1a). Each section of the box corresponds to one of the four ontic states the particle can be in. To perform measurements, either a horizontal or a vertical partition can be inserted. A measurement in the $|0\rangle / |1\rangle$ basis for example corresponds to inserting a partition *vertically* (Figure 1b)) and measuring on which side of the partition the particle is found. Performing a measurement in the $|+\rangle / |-\rangle$ basis in contrast corresponds to the insertion of the partition *horizontally* before measuring the the particle’s position (Figure 1c)). There can only be one partition in the box at a given time, and so in order to perform a measurement in the $|+\rangle / |-\rangle$ basis by inserting the partition horizontally, we need to remove the partition from the previous measurement in the $|0\rangle / |1\rangle$ basis. This must happen sufficiently quickly compared to the free motion of the particle so as to ensure that two consecutive measurements in the same basis yield the same result. On the other hand, time between consecutive measurements must be sufficiently large for the particle’s position to be uniformly distributed over the region to which it is confined. Alternatively, one could imagine that the measurement disturbs the particle in such a way, that its position is randomized over the restricted region. With the above setup at hand, we now implement Cabello et al.’s experiment and include successive random measurements in non-orthogonal bases.

We begin by noting the need to distinguish between two notions of measurement which one might have in mind. On the one hand, one can consider some physical process taking place which leaves the system in some definite value of some observable. (A definite possessed value of the observable, whether or not anyone knows it.) On the other hand, one might consider an external observer or agent who comes to know what the definite value (or values) a system possesses at a given time are. (This difference is akin to the selective/non-selective measurement distinction familiar in quantum foundations.) Such an external observer need

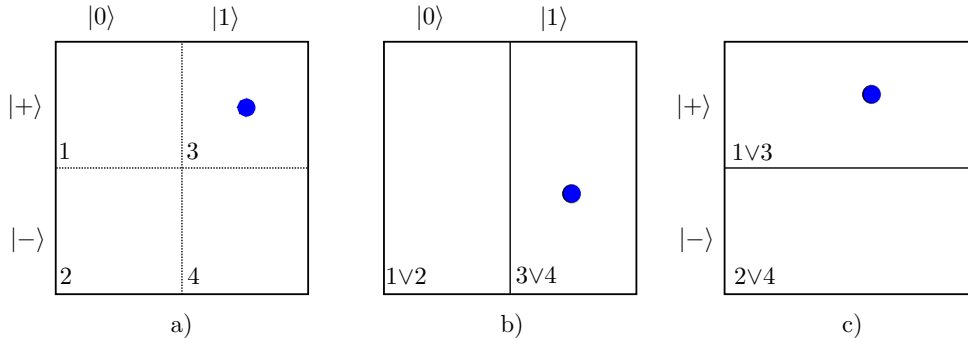


Figure 1: Illustration of standard measurements in Spekkens’ toy model. We consider a classical particle in a two dimensional box. a) illustrates how at each instant, the particle is in one of 4 mutually exclusive ontic states. Measurements in the standard bases are modeled as the b) vertical or c) horizontal insertions of a partition into the box, followed by a location measurement that allows one to identify the particle’s position as either being left/right or top/bottom. Two consecutive left/right and top/bottom measurements are non-commutative.

not be a person, but could be anything that reliably correlates with the measurement outcome, such as a memory cell.

We first consider the case in which there is no external agent. Measurements on the Spekkens particle are performed by either the horizontal or vertical insertion of the partition. In order to implement Cabello et al.’s experimental setup, we require that the orientation of the partition change randomly, in such a way that it has a probability of 50% of remaining in its previous position and a probability of 50% of changing its orientation from horizontal to vertical or vice versa. This is an implementation of Cabello et al.’s random variable X_t . At each time step, the system will be in a well defined state with definite values. It is evident *prima facie* that there is no heat exchange with the environment at any point¹². Moreover, this Spekkens setup is merely a more elaborate version of the previously introduced RAND operation. Evidently the system is in a definite state with definite values after each measurement, however, this “generation” of new values is not accompanied by a heat exchange with the environment.

One may object to the above reasoning by demanding that the partition *itself* must have a reset state, that it therefore must delete information and reset itself after each measurement. However, the described Spekkens’ setup does not require a costly *Landauer erasure*: the system *overwrites* its previous state after each time step¹³. As there is no external agent

¹²Whether or not the act of removing and re-inserting the partition is thermodynamically reversible or irreversible is debatable given to the single particle nature of the experiment. One may argue that the removal of the partition resembles a free expansion, which is thermodynamically irreversible. However, if we take thermodynamic reversibility to be equal with the claim that the (Gibbs–) von Neumann entropy of the system remains constant, then the removal and re-insertion of the partition is indeed reversible, in accordance with Maroney (2009). Either way, there is clearly no heat exchanged with the environment.

¹³The behaviour of the partition for example could be implemented by a beam splitter, followed by a NOT operation on the one hand and the identity operation on the other hand. In particular, no reference to the partition’s previous state must be made for either of those operations.

who records the measurement outcomes by correlating a measurement apparatus with the system's state, there are no resources needed for this implementation of measurements.

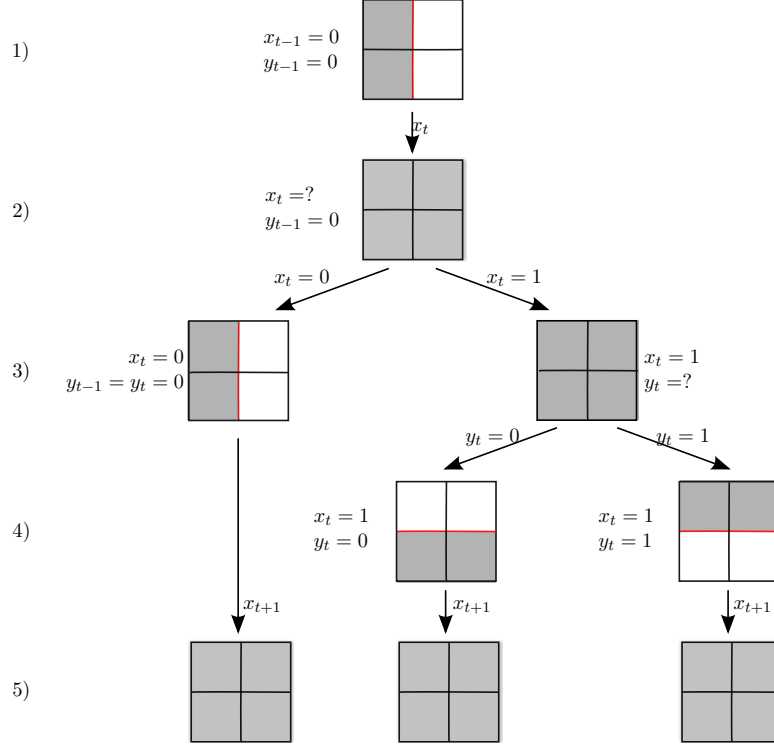


Figure 2: Illustration of the various steps in an exemplary measurement cycle.

The situation changes if we require the measurement performed and the result obtained to be determined and/or recorded by an external agent. Such an agent need not be human, but could be any system that is able to correlate itself with the Spekkens system and perform operations depending on the outcome. In the case of the experimental setup described above, a (memoryless) external agent needs to acquire at least two bits of information in order to determine which state the system is in at a given time t : one bit that determines whether the particle was measured in the x or z basis and one bit that determines the outcome of the measurement, namely whether the particle is left/right or top/bottom respectively. We imagine now a situation in which the agent is memoryless, i.e. has no access to the choice of measurement basis and the measurement result at time $t - 1$. If we allow the agent to have only two binary memory cells, one that reads out the choice of basis given by X_t , and one that reads out the position of the particle, Y_t , then once the agent has determined the state of the system at time t , she needs to reset both memory cells in order to prepare herself for the next measurement cycle at time $t + 1$. Given that resetting memory cells is costly, there will be a heat cost of $kT \ln 2$ associated with each measurement cycle.

This is where Cabello et al.'s result enters the picture. The described heat cost can in fact

be reduced to $3/2$ bits once we allow the agent to use a recording device, or a *memory*, which allows her to access the measurement results at the previous time-step¹⁴ $t - 1$. In this case, the agent does not need to perform the position measurement iff $x_t = x_{t-1}$, i.e. if she finds that the measurement basis at time t is the same as at time $t - 1$. This is due to the fact that consecutive measurements in the same basis always yield the same measurement results. Given that she has access to the previous measurement result, she can therefore skip the location measurement in 50% of the cases and hence save resources. If the agent has access to a memory that specifies the previous measurement basis and measurement outcome, the average amount of bits needed to specify the measurement and outcome at time t can thereby be reduced to $3/2$ bits. The minimal average heat cost per measurement cycle therefore becomes $\frac{3}{2}kT \ln 2$, the value Cabello et al. derived.

Figure 2 illustrates the various steps in an exemplary measurement cycle: after an input x_t , the state of the system has changed into a maximally mixed state. The first step of the agent must be to determine the measurement basis, i.e. the position of the partition. This step must be performed during *each* cycle, and so, given a finite memory for the agent and the need to create new blank memory states, there is always a heat cost of $kT \ln 2$ associated with this step. If $x_t = x_{t-1}$, nothing further happens, y_{t-1} changes into y_t , but no resources are required for this step and so the cycle is finished. If $x_t \neq x_{t-1}$ however, a position measurement must be performed that determines y_t . This once more leads to $kT \ln 2$ units of heat, since the agent only has limited memory and therefore needs to reset her memory before each measurement. Since this measurement of y_t is only performed half of the time, the average heat cost associated with it is only $\frac{1}{2}kT \ln 2$ units of heat. Adding up the various contributions leads to an average heat cost of $\frac{3}{2}kT \ln 2$ per measurement cycle, in accordance with Cabello et al.'s results. From the point of view of an external observer, however, the Spekkens quantum system itself is in a maximally mixed state from step 2) onwards.

The alleged quantum heat cost therefore merely results from the need to reset the various memory cells of the agent, that are needed to determine the measurement outcome. It is unavoidable, as Landauer (1996) showed, and rules out the physical implementation of a Maxwell's demon. The heat cost associated with Cabello et al.'s setup is thereby no more mysterious than the heat cost involved in the consecutive measurement of the outcomes of a fair coin flip, given limited resources: at each time step the measurement apparatus must be reset so as to be able to perform the consecutive measurements. Differently put, in terms of particles and boxes: performing consecutive RAND operations on a system

¹⁴In practice, we need to grant the agent at least two more memory cells as resources, such as to be used as a memory for the measurement basis and measurement result at time $t - 1$. In computational mechanics it is more common simply to supply the agent with an empty tape on which she records the measurement outcomes and at the same time allow her to access the tape on which the basis choices are written, thereby providing her with $\overleftarrow{y}_t$ and $\overleftarrow{x}_t$.

leads to a heat cost iff the system's state is recorded at each time step, in which case the measurement apparatus needs to be reset before each measurement. What Cabello et al. have shown therefore, is simply that if we allow the agent to have a *memory*, the average heat cost for repeatedly recording the outcomes can be reduced from 2 bits to $3/2$ bits, and no further.

Furthermore, in order to explain the origin of this heat cost, we did not need to mention any *intrinsic values* which supposedly determine the future behaviour of the quantum system. The resources needed for Cabello et al.'s experiment are resources that are to be provided by the observer who wants to record the measurement results, and not by the quantum system itself. We note that the observer and his or her resources could well be purely classical; and we note moreover that the very same heat cost would be incurred even for a Type II interpretation.

3.3 Some Remarks on Computational Mechanics

The above examples demonstrate that the heat cost attributed to certain types of quantum interpretation by Cabello et al. leads back to an *external* agent repeatedly performing measurements with limited information storage available. The external agent does not need to be human but could equally be a machine and the calculated heat cost indeed gives a lower bound to the energy required to run such process. Cabello et al.'s result therefore clearly plays an important role for determining the resources needed to perform certain quantum computational tasks. What their argument does not establish however, is a distinctive and perhaps problematic heat cost for Type I interpretations. We can trace back the misinterpretation of their mathematical results to a lack of discrimination between the agent who acts and the system that is acted upon. Instead, agent and system together were treated as a unified computational machine, leading to the erroneous conclusion that the individual system itself was driving the computation and was the locus of the heat cost. This suggests that we need to be more careful when we apply computational mechanics to the foundations of quantum mechanics. In particular the tempting but somewhat misleading idea that information is a *concrete* rather than *abstract* entity, which gives it the status of a physical substance that can be transferred and destroyed, leads to intuitions that must be double checked by contrasting them with real, physical situations (cf. (Timpson, 2013) for more details on the concrete vs. abstract distinction).

4 Conclusion

We have analysed the claim that there is a physical difference between two classes of quantum interpretations in terms of an excess heat generation in successive measurements,

a difference which could in principle be tested experimentally. This is not so: there is no such differential heat production. In so far as there is a heat cost associated with successive randomly selected measurements on a quantum system, it arises from accounting for the external resources needed to record the performance and results of the measurement operations on the system, and not from a putative erasure process that takes place within the quantum system itself. That there is no heat cost involved with consecutive measurements of this kind arising from the quantum system itself—regardless of how one interprets probabilities—can immediately be seen upon calculating the difference of the von Neumann entropy of the quantum system between the various time steps, which difference is zero. This fact is reconciled with Cabello et al.’s quite correct mathematical result precisely by noting that the latter only concerns—when properly understood—costs of the external record, of the external agent.

The question of which interpretations might truly represent our world thereby remains unanswered by thermodynamical considerations concerning successive measurement scenarios.

Acknowledgements

We would like to thank Adan Cabello, Mile Gu, Andrew Garner and Vlatko Vedral for helpful discussions as well as David Wallace and Harvey Brown for useful comments. This work was partially supported by a grant from the Templeton World Charity Foundation and the British Society for the Philosophy of Science.

Appendix

The quantity I_{erased} has its origin in the difference of the Shannon entropies at times t and $t - 1$ as can be seen in the following, short derivation.

$$\begin{aligned}
H(X_t Y_t S_t) - H(X_t Y_{t-1} S_{t-1}) &= H(X_t Y_t S_{t-1}) - H(X_t Y_{t-1} S_{t-1}) - H(S_{t-1} | X_t Y_t S_t) \\
&= H(X_t) + H(Y_t | S_t) + H(S_t) \\
&\quad - H(X_t) - H(Y_{t-1} | S_{t-1}) - H(S_{t-1}) - H(S_{t-1} | X_t Y_t S_t) \\
&= -H(S_{t-1} | X_t Y_t S_t) = -I_{erased},
\end{aligned}$$

where we assumed that the process is *unifilar*, namely $H(S_t | X_t Y_t S_{t-1}) = 0$, that the input is time-independent, i.e. $H(X_{t-1}) = H(X_t)$ and that the current causal state determines the output uniquely $H(Y_t | S_t) = 0$. Following Landauer's Principle, the physical implementation of a logical operation leads to a generation of heat equal to at least $-kT \ln 2$ times the total change in Shannon entropy. To calculate the total change, all involved random variables at a given time need to be taken into account. I_{erased} therefore provides a lower bound for the heat cost involved in the implementation of Cabello et al.'s experiment.

References

- L. E. Ballentine. The Statistical Interpretation of Quantum Mechanics. *Reviews of Modern Physics*, 42(4):358–381, Oct. 1970. doi: 10.1103/RevModPhys.42.358.
- N. Barnett and J. P. Crutchfield. Computational Mechanics of Input-Output Processes: Structured transformations and the ϵ -transducer. *Journal of Statistical Physics*, 161(2): 404–451, Oct. 2015. doi: 10.1007/s10955-015-1327-5.
- D. Bohm. A Suggested Interpretation of the Quantum Theory in Terms of "Hidden" Variables. I. *Physical Review*, 85(2):166–179, Jan. 1952. doi: 10.1103/PhysRev.85.166.
- F. Brandão, M. Horodecki, N. Ng, J. Oppenheim, and S. Wehner. The second laws of quantum thermodynamics. *Proceedings of the National Academy of Sciences*, 112(11): 3275–3279, 2015. doi: 10.1073/pnas.1411728112.
- A. Cabello, M. Gu, O. Gühne, J.-Å. Larsson, and K. Wiesner. Thermodynamical cost of some interpretations of quantum theory. *Physical Review A*, 94(5), Nov. 2016. doi: 10.1103/PhysRevA.94.052127.
- J. P. Crutchfield and K. Young. Inferring statistical complexity. *Physical Review Letters*, 63(2):105–108, July 1989. doi: 10.1103/PhysRevLett.63.105.
- D. Dürr, S. Goldstein, and N. Zanghi. Bohmian Mechanics and the Meaning of the Wave Function. *arXiv:quant-ph/9512031*, Dec. 1995.
- D. Dürr, S. Goldstein, and N. Zanghi. *Quantum Physics Without Quantum Philosophy*. Springer Science and Business Media, Nov. 2012. doi: 10.1007/978-3-642-30690-7.
- H. Everett. "Relative State" Formulation of Quantum Mechanics. *Reviews of Modern Physics*, 29(3):454–462, July 1957. doi: 10.1103/RevModPhys.29.454.
- C. A. Fuchs and A. Peres. Quantum Theory Needs No 'Interpretation'. *Physics Today*, 53(3):70–71, Jan. 2007. doi: 10.1063/1.883004.
- A. J. P. Garner, J. Thompson, V. Vedral, and M. Gu. Thermodynamics of complexity and pattern manipulation. *Phys. Rev. E*, 95:042140, Apr 2017. doi: 10.1103/PhysRevE.95.042140.
- G. C. Ghirardi, A. Rimini, and T. Weber. Unified dynamics for microscopic and macroscopic systems. *Physical Review D*, 34(2):470–491, July 1986. doi: 10.1103/PhysRevD.34.470.
- S. Goldstein. Bohmian mechanics. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, 2017.

- G. Gour, M. P. Müller, V. Narasimhachar, R. W. Spekkens, and N. Yunger Halpern. The resource theory of informational nonequilibrium in thermodynamics. *Physics Reports*, 583:1–58, July 2015. doi: 10.1016/j.physrep.2015.04.003.
- N. Harrigan and R. W. Spekkens. Einstein, Incompleteness, and the Epistemic View of Quantum States. *Foundations of Physics*, 40(2):125–157, 2010. doi: 10.1007/s10701-009-9347-0.
- M. Horodecki and J. Oppenheim. Fundamental limitations for quantum and nanoscale thermodynamics. *Nature Communications*, 4:2059, June 2013. doi: 10.1038/ncomms3059.
- E. T. Jaynes. Gibbs vs Boltzmann Entropies. *American Journal of Physics*, 33:391–398, May 1965. doi: 10.1119/1.1971557.
- R. Landauer. Irreversibility and Heat Generation in the Computing Process. *IBM J. Res. Dev.*, 5(3):183–191, July 1961. doi: 10.1147/rd.53.0183.
- R. Landauer. The physical nature of information. *Physics Letters A*, 217(4-5):188–193, July 1996. doi: 10.1016/0375-9601(96)00453-7.
- O. Lombardi and D. Dieks. Modal Interpretations of Quantum Mechanics. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, 2016.
- O. J. E. Maroney. The (Absence of a) Relationship Between Thermodynamic and Logical Reversibility. *Studies in History and Philosophy of Science Part B*, 36(2):355–374, 2005. doi: 10.1016/j.shpsb.2004.11.006.
- O. J. E. Maroney. Generalising Landauer’s Principle. *Physical Review E*, 79(3), Mar. 2009. doi: 10.1103/PhysRevE.79.031105.
- J. C. Maxwell. *Theory of Heat*. Longmans, Green and Co., London, UK, 1871.
- W. C. Myrvold. Statistical mechanics and thermodynamics: A maxwellian view. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 42(4):237–243, 2011. doi: 10.1016/j.shpsb.2011.07.001.
- K. Popper. *The Logic of Scientific Discovery*. Routledge, 2nd edition, Feb. 2002. doi: 10.2307/2104331.
- C. Rovelli. Relational quantum mechanics. *International Journal of Theoretical Physics*, 35(8):1637–1678, Aug. 1996. doi: 10.1007/BF02302261.
- R. W. Spekkens. Evidence for the epistemic view of quantum states: A toy theory. *Physical Review A*, 75(3):032110, Mar. 2007. doi: 10.1103/PhysRevA.75.032110.

- C. G. Timpson. *Quantum Information Theory and the Foundations of Quantum Mechanics*. Oxford University Press, Oxford, Apr. 2013. doi: 10.1086/678554.
- D. Wallace. *Emergent Multiverse: Quantum Theory According to the Everett Interpretation*. Oxford University Press, Oxford, July 2012. doi: 10.1093/bjps/axu009.
- D. Wallace. Thermodynamics as Control Theory. *Entropy*, 16(2):699–725, Jan. 2014. doi: 10.3390/e16020699.
- J. A. Wheeler and W. H. Zurek, editors. *Quantum Theory and Measurement*. Princeton University Press, Princeton, NJ, July 2014. doi: 10.1515/9781400854554.
- K. Wiesner, M. Gu, E. Rieper, and V. Vedral. Information-theoretic lower bound on energy cost of stochastic computation. *Proc. R. Soc. A*, 468(2148):4058–4066, Dec. 2012. doi: 10.1098/rspa.2012.0173.