Studies in History and Philosophy of Modern Physics xxx (2017) 1–9

Contents lists available at ScienceDirect

Studies in History and Philosophy of Modern Physics

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

On the thermodynamical cost of some interpretations of quantum theory

Carina E.A. Prunkl [a, *], Christopher G. Timpson [b]

[a] Balliol College, University of Oxford, UK
[b] Brasenose College, University of Oxford, UK

ARTICLE INFO

Article history:
Received 26 May 2017
Received in revised form
8 January 2018
Accepted 10 January 2018
Available online xxx

Keywords:
Quantum interpretations
Computational mechanics
Landauer's principle
Thermodynamics

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.

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 (Garner et al., 2017; Wiesner et al., 2012). 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. One 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 & 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 resources[1]), one might say

* Corresponding author. Balliol College, Broad Street, Oxford OX1 3BJ, UK.
E-mail addresses: carina.prunkl@balliol.ox.ac.uk (C.E.A. Prunkl), christopher.timpson@bnc.ox.ac.uk (C.G. Timpson).

[1] Modern accounts include (Horodecki & Oppenheim, 2013; Wallace, 2014; Brand 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).

https://doi.org/10.1016/j.shpsb.2018.01.002
1355-2198/

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 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 & Crutchfield, 2015; Crutchfield & 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 $\ldots, Y_{-1}, Y_0, Y_1, \ldots$ 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 $\mathscr{Y} = \{\text{"up"}, \text{"down"}\}$. 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 modeled by a so-called stochastic *input-output process*, $\overleftrightarrow{Y}\,\big|\,\overleftrightarrow{X}$, with input values $\{x_t\}$ from an alphabet $\mathscr{X}$. The whole input-output process may then be described as a collection of stochastic processes $\overleftrightarrow{Y}\,\big|\,\overleftrightarrow{X} \equiv \{\overleftrightarrow{Y}\,\big|\,\overleftrightarrow{x}\}_{\overleftrightarrow{x} \in \overleftrightarrow{\mathscr{X}}}$, where we take each process $\overleftrightarrow{Y}\,\big|\,\overleftrightarrow{x}$ to correspond to all possible output sequences $\overleftrightarrow{Y}$ that could arise from one particular input sequence $\overleftrightarrow{x}$, drawn from the set of all possible input sequences $\overleftrightarrow{\mathscr{X}}$.

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*:

$$\mathbf{P}\left(\overleftrightarrow{Y}\,\middle|\,\overleftrightarrow{x}\right) = \left\{\mathbf{P}\left(\overleftrightarrow{Y} \in \sigma \middle| \overleftrightarrow{X} = \overleftrightarrow{x}\right)\right\}_{\sigma \triangleleft \overleftrightarrow{\mathscr{Y}},\, \overleftrightarrow{x} \in \overleftrightarrow{\mathscr{X}}} \tag{1}$$

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 $P(\overrightarrow{Y}\big|\overrightarrow{X}, \overleftarrow{Z} = \overleftarrow{z}) = P(\overrightarrow{Y}\big|\overrightarrow{X}, \overleftarrow{Z} = \overleftarrow{z}')$ are then said to belong to the same *causal state* s. Denote the set of causal states $\mathscr{S}$. The ε-map is then introduced as the mapping $\varepsilon : \overleftarrow{\mathscr{Z}} \to \mathscr{S}$ from any input-output past onto its corresponding causal state (Barnett & 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 $H(\mathscr{S})$ 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 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,

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

Please cite this article in press as: Prunkl, C. E. A., & Timpson, C. G., On the thermodynamical cost of some interpretations of quantum theory, Studies in History and Philosophy of Modern Physics (2017), https://doi.org/10.1016/j.shpsb.2018.01.002