An additive Hamiltonian plus Landauer's Principle yields quantum theory

Chris Fields

528 Zinnia Court
Sonoma, CA 95476 USA

fieldsres@gmail.com

March 25, 2015

Abstract

It is shown that no-signalling, a quantum of action, unitarity, detailed balance, Bell's theorem, the Hilbert-space representation of physical states and the Born rule all follow from the assumption of an additive Hamiltonian together with Landauer's principle. Common statements of the "classical limit" of quantum theory, as well as common assumptions made by "interpretations" of quantum theory, contradict additivity, Landauer's principle, or both.

1 Introduction

Let $\mathbf{U}$ be a closed system with a self-interaction described by a Hamiltonian operator $H_{\mathbf{U}}$. Then $H_{\mathbf{U}}$ is *additive* if the following two conditions hold: 1) for any arbitrarily-selected two-way partition of $\mathbf{U}$ into two subsystems $\mathbf{A}$ and $\mathbf{B}$ with self-interactions $H_{\mathbf{A}}$ and $H_{\mathbf{B}}$ respectively, $H_{\mathbf{U}} = H_{\mathbf{A}} + H_{\mathbf{B}} + H_{\mathbf{AB}}$, where $H_{\mathbf{AB}}$ represents the $\mathbf{A}$ - $\mathbf{B}$ interaction; and 2) $H_{\mathbf{U}}$ is invariant under arbitrary iterative two-way partitions of either $\mathbf{A}$ or $\mathbf{B}$ into two smaller subsystems. In particular, further partitioning $\mathbf{B}$ into subsystems $\mathbf{B1}$ and $\mathbf{B2}$ and writing $H_{\mathbf{B}} = H_{\mathbf{B1}} + H_{\mathbf{B2}} + H_{\mathbf{B1B2}}$ has no effect on $H_{\mathbf{U}}$; we can, for example, write $H_{\mathbf{U}} =$

$H_{\mathbf{A}} + (H_{\mathbf{B1}} + H_{\mathbf{B2}} + H_{\mathbf{B1B2}}) + H_{\mathbf{AB}}$. Provided $H_{\mathbf{U}}$ is additive in this sense, the invariance of $H_{\mathbf{U}}$ under iterated decompositions of $\mathbf{U}$ into subsystems is a trivial consequence of the associativity of addition.

It is shown here that the additivity of $H_{\mathbf{U}}$, together with Landauer's Principle that "information is physical," i.e. that irreversible encodings of classical information require the expenditure of free energy [1, 2], has a surprising consequence: any complete, self-consistent physics that satisfies these constraints is quantum theory, specifically *unitary* and hence *standard, linear* quantum theory. "Complete" here means that the theory applies in exactly the same way to all systems, including, in particular, any observers embedded in $\mathbf{U}$; "self-consistent" entails an absence of logical contradictions. As Hamiltonians in classical physics can be added associatively, this result implies that classical physics fails to satisfy at least one of the other conditions given above, i.e. that classical physics either fails to allow arbitrary iterative decompositions of $\mathbf{U}$, is incomplete, so that $H_{\mathbf{U}}$ is not well-defined, or else violates Laundauer's Principle. As will be shown below, interpretations of quantum theory in which classicality objectively "emerges" from quantum dynamics or is regarded as objectively "given" by observation similarly fail to satisfy at least one of these conditions.

T. F. Jordan has shown that if separate (unentangled) quantum systems must interact to influence each other's behavior, then quantum theory must be linear [3], and conversely, that if quantum theory is modified to make it non-linear (specifically, if the Hamiltonian of a system is allowed to depend on its state), then separate (unentangled) quantum systems can influence each other's behavior without interacting [4]. The present result complements this earlier work by showing that the additivity of the Hamiltonian, which is itself a consequence of linearity, is sufficient to produce standard, linear quantum theory. Any empirical demonstration supporting a non-linear quantum theory would, therefore, require abandoning the assumption that systems, in particular the universe $\mathbf{U}$, can be iteratively decomposed in any arbitrarily-chosen way without physical consequences as well as abandoning the assumption that systems must interact to influence each other's behavior. It is clear, intuitively, why this would be the case. If the Hamiltonian of a system depends on its state, the behavior of that system can be expected to be sensitive, and possibly extremely sensitive, to how the system is defined. Small differences in the definition of the system, i.e. small differences in the decomposition of $\mathbf{U}$ by which the system was defined could, at least in principle, generate large differences in the system's behavior. This would be surprising: we expect similarly-defined systems to behave in similar ways, and systems defined to be nearly identical to behave in nearly-identical ways. The Hamiltonian being additive for any arbitrarily-chosen iterative decomposition of $\mathbf{U}$ assures that this expectation is met.

The tactic employed here is straightforward: it is to show that quantum theory provides the only self-consistent description of the experimental outcomes obtainable by any observer who obtains experimental outcomes only by physically interacting with the observed system. As $\mathbf{U}$ is by assumption closed, any "observer" must clearly be a subsystem of $\mathbf{U}$. The observer is chosen arbitrarily and is assumed to have no capabilities or characteristics that differ in any way from those of any other system, i.e. any other subsystem of $\mathbf{U}$ under consideration. In particular, the observer is specifically disallowed from making any *a priori*

2

assumptions about the system being observed; *all* information that the observer has about the observed system is obtained by physically interacting with the observed system and by no other means. "Observation" is therefore merely a way of describing any arbitrarily-chosen physical interaction in a way that focuses on the classical information transferred by the interaction. It is shown that with these assumptions, observation is interaction with a "black box" as defined by classical information theory [5, 6, 7]; it is then shown that the constraints on information transfer from a black box to an observer are sufficient to derive unitary quantum theory. The derivation presented is entirely constructive and employs no specifically quantum-theoretic concepts. It makes fully explicit the assumptions that distinguish classical from quantum theory and demonstrates how these assumptions render classical theory either incomplete or inconsistent as a formal description of the physical world.

2 Observation as interaction with a black box

Let $\mathbf{U}$ be an isolated, closed system ("the universe") and assume that the Hamiltonian $H_{\mathbf{U}}$ is additive in the sense defined above. No other assumption about the mathematical form of $H_{\mathbf{U}}$ is made. Consider some arbitrary two-way partition of $\mathbf{U}$ into subsystems $\mathbf{O}$ and $\mathbf{B_O}$: these will be referred to as the "observer" and the observer's "box" respectively. Let $H_{\mathbf{OB_O}}$ represent the observer-box interaction; hence $H_{\mathbf{U}} = H_{\mathbf{O}} + H_{\mathbf{OB_O}} + H_{\mathbf{B_O}}$. Note that as they are defined here, both $\mathbf{O}$ and $\mathbf{B_O}$ are open systems; each is open to the other but to nothing else.

The assumption that $H_{\mathbf{U}}$ is additive is, as described above, the assumption that $H_{\mathbf{U}}$ is invariant under arbitrary further partitioning of the "box" $\mathbf{B_O}$. This invariance is very strong: it allows $H_{\mathbf{B_O}}$ to be written as a sum $H_{\mathbf{B_O}} = H_{\mathbf{C}} + H_{\mathbf{CD}} + H_{\mathbf{D}}$, where $\mathbf{C}$ and $\mathbf{D}$ are the results of any arbitrary two-way partition of $\mathbf{B_O}$, without any effect on either $H_{\mathbf{O}}$ or $H_{\mathbf{OB_O}}$. The additivity of $H_{\mathbf{U}}$, together with the associativity of addition, thus renders both $H_{\mathbf{O}}$ and $H_{\mathbf{OB_O}}$ invariant under arbitrary iterative partitioning of $\mathbf{B_O}$. Note that this is true, without restriction, for any arbitrary two-way partition of $\mathbf{U}$ into components $\mathbf{O}$ and $\mathbf{B_O}$. Consistent with the requirement of completeness, the choice of the "observer" $\mathbf{O}$ has no consequences whatsoever for the theory that is obtained.

The invariance of $H_{\mathbf{OB_O}}$ under arbitrary iterative partitioning of $\mathbf{B_O}$ has a straightforward information-theoretic interpretation: $H_{\mathbf{OB_O}}$ can transfer no information about partitions of $\mathbf{B_O}$, and therefore no information about the physical interactions between partitions of $\mathbf{B_O}$, from $\mathbf{B_O}$ to $\mathbf{O}$. The information channel defined by $H_{\mathbf{OB_O}}$ cannot, therefore, provide $\mathbf{O}$ with any information about the internal self-interaction $H_{\mathbf{B_O}}$ of $\mathbf{B_O}$. It is natural to describe this situation as one in which $\mathbf{O}$ has no access, via the channel $H_{\mathbf{OB_O}}$, to either the "internal structure" or the "internal dynamics" of $\mathbf{B_O}$. In classical information theory, a system $\mathbf{B_O}$ and information channel $H_{\mathbf{OB_O}}$ with these characteristics define a "black box," a system with observerable external behavior, but with undiscoverable internal structure or dynamics. Shannon [5] introduced classical information theory to describe finite transmis-

sions received via a channel of fixed capacity from an unknown, uncharacterized, distant and therefore inaccessible source, i.e. from a black box. Using the wartime metaphor of a sealed device that could be manipulated and its overt behavior observed but could not be opened for internal inspection, Ashby [6] and Moore [7] independently proved that finite observations at finite resolution are insufficient to uniquely specify the "machine table," the set of all possible input-output transitions, of a black box. The proof is straightforward: if the number of internal states of the box is unknown and unknowable, there can be no assurance that the finite number of manipulations that have been performed have accessed all of the internal states of the box. Without such assurance, there can be no assurance that all possible outputs of the box have been observed, and hence no assurance that the resulting machine table is complete. An observer who interacts with a black box can, therefore, place a lower limit on the complexity of the box's behavior, but cannot, by finite observation, place an upper limit on the complexity of the box's behavior. The very next output from any black box can be a complete surprise.

The discussion that follows shows that quantum systems can be considered to be black boxes in this classical information-theoretic sense, by showing that the only self-consistent description of the experimental outcomes obtainable from a black box is unitary quantum theory. It is assumed throughout that observational outcomes are finitely encoded and can, therefore, be considered to be finite bit strings. The Schrödinger equation and the Born Rule provide, it will be shown, the only self-consistent description of how such bit strings change over time.

A core assumption of classical physics is that multiple observers share a world of independently-manipulable, causally-independent, and hence separable objects. Treating $\mathbf{B_O}$ as a black box obviously contradicts this classical assumption. The fact that doing so generates quantum theory shows that this "white box" assumption of classical physics is empirically incorrect, and that its re-introduction by interpretations of quantum theory is bound to generate paradox.

3   No-signalling and its corollaries

Consider now an arbitrary partition of $\mathbf{B_O}$ into subsystems $\mathbf{O'}$ and $\mathbf{B}$; these will be referred to as the "second observer" and the "shared box" respectively. Assume the $\mathbf{O}$-$\mathbf{O'}$ interaction is negligible, i.e. $H_{\mathbf{OO'}} \sim 0$. The Hamiltonian $H_{\mathbf{U}}$ can then be written:

$$H_{\mathbf{U}} \sim H_{\mathbf{O}} + H_{\mathbf{O'}} + H_{\mathbf{B}} + H_{\mathbf{OB}} + H_{\mathbf{O'B}} \tag{1}$$

It is clear from the reasoning above that $\mathbf{B}$ must also be a black box. The additivity of $H_{\mathbf{U}}$ allows $\mathbf{B}$ to be arbitrarily partitioned without affecting either $H_{\mathbf{OB}}$ or $H_{\mathbf{O'B}}$; hence neither $\mathbf{O}$ nor $\mathbf{O'}$ can be regarded as having observational access to the "interior" of $\mathbf{B}$, and neither can place upper limits on the behavioral complexity of $\mathbf{B}$.

Theorem 1 (no-signalling): A shared black box does not provide a classical communication channel.

Proof: In order for $\mathbf{O'}$ to send classical information to $\mathbf{O}$ through $\mathbf{B}$, $\mathbf{O'}$'s actions on $\mathbf{B}$ via $H_{\mathbf{O'B}}$ must be classically correlated with outcomes received by $\mathbf{O}$ from $\mathbf{B}$ via $H_{\mathbf{OB}}$. However, if $H_{\mathbf{OO'}} \sim 0$, $H_{\mathbf{OB}} \sim H_{\mathbf{OB_O}}$ and $H_{\mathbf{O'B}} \sim H_{\mathbf{O'B_{O'}}}$, where $\mathbf{B_{O'}}$ comprises all of $\mathbf{U}$ except $\mathbf{O'}$. As $H_{\mathbf{OB_O}}$ (similarly, $H_{\mathbf{O'B_{O'}}}$) is invariant under any arbitrary alternative partitioning of $\mathbf{B_O}$ that redefines the labels '$\mathbf{O'}$' and '$\mathbf{B}$' (similarly, any arbitrary alternative partitioning of $\mathbf{B_{O'}}$ that redefines the labels '$\mathbf{O}$' and '$\mathbf{B}$'), $H_{\mathbf{OB_O}}$ (similarly, $H_{\mathbf{O'B_{O'}}}$) must be invariant under any redefinition of $H_{\mathbf{O'B}}$ (similarly, of $H_{\mathbf{OB}}$) that reflects such an alternative partitioning. Observational outcomes obtained via $H_{\mathbf{OB_O}}$ ($H_{\mathbf{O'B_{O'}}}$) are, therefore, independent of how $\mathbf{B_O}$ ($\mathbf{B_{O'}}$) is partitioned and thus independent of how $H_{\mathbf{O'B_{O'}}}$ ($H_{\mathbf{OB_O}}$) is defined. The outcomes obtained by $\mathbf{O}$ via $H_{\mathbf{OB}} \sim H_{\mathbf{OB_O}}$ cannot, therefore, be dependent upon $\mathbf{O'}$'s actions via $H_{\mathbf{O'B}} \sim H_{\mathbf{O'B_{O'}}}$. If $\mathbf{O}$'s measured outcomes are independent of $\mathbf{O'}$'s actions, they cannot be classically correlated with $\mathbf{O'}$'s actions, i.e. classical communication between $\mathbf{O}$ and $\mathbf{O'}$ cannot occur. $\square$

From a physical perspective, the lack of classical correlation between $H_{\mathbf{OB}}$ and $H_{\mathbf{O'B}}$ demonstrated here corresponds to the inability of $\mathbf{O}$ to determine by finite observation (i.e. via $H_{\mathbf{OB_O}}$) either how $\mathbf{B_O}$ has been partitioned into $\mathbf{O'}$ and $\mathbf{B}$ or which degrees of freedom of $\mathbf{B}$ have been used by $\mathbf{O'}$ to encode classical information. It is impossible, in other words, for $\mathbf{O}$ to distinguish by observation the aspects of $\mathbf{B_O}$'s behavior that encode signals from $\mathbf{O'}$ from the aspects of $\mathbf{B_O}$'s behavior that are independent of or side-effects of such signals; hence $\mathbf{O}$ cannot, even in principle, distinguish "signals" from "noise." Note that this is true not just for $\mathbf{O}$, but also for $\mathbf{O'}$ and indeed for any observer. No third party $\mathbf{X}$ can "fix" the partitions that define $\mathbf{O}$ and $\mathbf{O'}$ and observe a classical correlation between them, as any such $\mathbf{X}$ interacts solely with their own black box $\mathbf{B_X}$ via an interaction $H_{\mathbf{XB_X}}$ that is independent, due to its additivity, of any partitions drawn within $\mathbf{B_X}$.

It is standardly assumed that ancillary information that allows both the identification and decoding of signals, e.g. public keys, parity bits, state-preparation data, or other information about the signal or method of encoding to expect, is communicated from $\mathbf{O'}$ to $\mathbf{O}$ via a separate classical channel; all Local Operations, Classical Communication (LOCC) protocols, for example, make this assumption [8]. If the *only* system shared by $\mathbf{O'}$ and $\mathbf{O}$ is a black box, however, no such ancillary channel can be assumed. An intervening black box being the *only* available channel is, from a communications protocol perspective, equivalent to the existence of an intermediary agent (a "Mallet" or man-in-the-middle attacker) who can and unpredictably does make arbitrary, undetectable changes to any message and arbitrary, undetectable substitutions of one message for another. Under such circumstances, $\mathbf{O'}$'s state changes (i.e. "encoding actions") and $\mathbf{O}$'s state changes (i.e. "receipts of data") are uncorrelated and hence communication between $\mathbf{O'}$ and $\mathbf{O}$ does not occur. The data received by $\mathbf{O}$ are, of course, correlated with state changes of $\mathbf{B_O}$ as will be described in detail below; but because $H_{\mathbf{OB_O}}$ is the *only* interaction that affects $\mathbf{O}$'s state, this is all that can be said.

As a specific example, consider the case in which Alice sends a qubit and Bob receives

one, but neither Alice nor Bob has any information other than that they themselves are interacting with a quantum channel. In particular, Bob has no information specifying that Alice sent a qubit at time t, no information about how Alice prepared the qubit, and no information about the integrity of the channel; indeed, Bob does not know that Alice is interacting with the channel, or even that Alice exists. Bob not only does not have this information; Bob cannot, even in principle, obtain this information; all Bob can do is insert qubits into the channel or extract qubits from the channel. In such a case, Bob has received and measured a qubit, but nothing correlates Bob's measurement of the qubit with Alice's preparation. The qubit, therefore, carries no information from Alice to Bob. It carries information from *the world* (i.e. the black box $\mathbf{B_{Bob}}$) to Bob; Bob obtained the qubit from the world and, given Landauer's Principle, his measurement of the qubit affects his physical state. But it carries no information *from Alice*; neither Bob, Alice, nor any third party interacting solely with their own terminus of some quantum channel can establish any classical correlation between Alice's preparative actions on the qubit and Bob's measured outcome.

This no-signalling theorem corresponds to the standard, quantum-theoretic no-signalling condition required for compliance with special relativity [9] if it is assumed that $\mathbf{O}$ and $\mathbf{O'}$ are spacelike-separated observers who can only communicate by encoding information in a shared *quantum* environment, i.e. who can only communicate via a quantum channel. Theorem 1 suggests, therefore, a functional analogy between a black box and a quantum channel: neither enables signalling between observers if it is the *only* channel available. This functional analogy will be borne out in more detail below. However, the present Theorem 1 replaces the idea that $\mathbf{O}$ and $\mathbf{O'}$ must be separated by *spatial* degrees of freedom to be subject to a restriction on classical communication through the intervening environment with the idea that *any* inaccessible and hence uncharacterizable intervening degrees of freedom constitute a black box and hence a communication-limiting environment.

It is not, however, necessary to assume that the $\mathbf{O}$-$\mathbf{O'}$ interaction is negligible. Even if $\mathbf{O}$ and $\mathbf{O'}$ are in direct contact, unless they together compose a *complete* two-way partitioning of $\mathbf{U}$ they cannot exchange classical information.

Theorem 2 (strong no-signalling): Unless $\mathbf{O'} = \mathbf{B_O}$, $H_{\mathbf{OO'}}$ does not implement a classical information channel.

Proof: Because $\mathbf{B_O}$ is a black box, $H_{\mathbf{OB_O}}$ can transfer no classical information about partitions of $\mathbf{B_O}$. In particular, $H_{\mathbf{OB_O}}$ is invariant under arbitrary redefinitions of the partitioning of $\mathbf{B_O}$ into $\mathbf{O'}$ and $\mathbf{B}$; hence $H_{\mathbf{OB_O}}$ is invariant under arbitrary redefinitions of both $H_{\mathbf{O'B}}$ and $H_{\mathbf{OO'}}$. Outcomes of interactions via $H_{\mathbf{OB_O}}$ are, therefore, independent of interactions via $H_{\mathbf{OO'}}$ and so cannot be classically correlated with their outcomes. $\square$

This strong no-signalling theorem follows trivially from the fact that $\mathbf{O}$'s state depends only, via $H_{\mathbf{OB_O}}$, on the state of $\mathbf{B_O}$. Physically, $\mathbf{O}$ cannot determine by finite observation which degrees of freedom of $\mathbf{B_O}$ are included within $\mathbf{O'}$; hence $\mathbf{O}$ cannot determine by finite observation what, if any, classical information has been encoded by the action, via $H_{\mathbf{O'B_{O'}}}$, of $\mathbf{O'}$ on $\mathbf{B_{O'}}$, of which $\mathbf{O}$ itself is by assumption a proper part. As in the case of

Theorem 1, neither $\mathbf{O}'$ nor any third party can observe a classical correlation either. This strong no-signalling theorem is stronger than the standard no-signalling condition in that it restricts classical communication even between observers who are in contact and are hence not only not spacelike separated, but not separated by any degrees of freedom. Under this stronger no-signalling condition, the mere mutual exposure of two distinct systems to a common black box, i.e. to an environment containing inaccessible, uncharacterizable degrees of freedom, restricts classical signalling between them.

The presence of an environment containing inaccessible, uncharacterizable degrees of freedom is a commonplace assumption of decoherence [10, 11]. However, in environment-driven decoherence this environment is either assumed to be "random" or "thermal" and hence characterizable as a classical statistical ensemble or, in quantum Darwinism [12, 13], assumed to specifically encode pointer-state information generated by its interactions with a specific, identified system of interest. Neither classical randomness nor specific encoding of pointer-state information about a specific distant system can, however, be inferred solely from finite observations of the environment [14, 15]. Either feature must, instead, by assumed *a priori*. Environmental decoherence theory cannot, therefore, characterize observations made by any arbitrarily-chosen observer that is assumed to have only information obtained by observation; it fails, therefore, to satisfy the completeness requirement stated earlier. We will return to this issue in the discussion of Theorem 6 below.

The strong no-signalling theorem has four immediate corollaries.

Corollary 2.1: An observer $\mathbf{O}$ cannot determine by finite observation whether it is separated by intervening degrees of freedom of $\mathbf{B_O}$ from a second observer $\mathbf{O}'$ fully contained within $\mathbf{B_O}$.

Proof: Because $\mathbf{B_O}$ is a black box, $\mathbf{O}$ can obtain no information about partitions of $\mathbf{B_O}$ into $\mathbf{O}'$ and $\mathbf{B}$. □

In particular, $\mathbf{O}$ cannot observationally distinguish between the three situations shown in Fig. 1, in which as before, $\mathbf{B}$ and $\mathbf{O}'$ are the components of an arbitrary two-way partition of $\mathbf{B_O}$. Physically, $\mathbf{O}$ cannot determine whether outcomes are being received from components of $\mathbf{B_O}$ that are "inside" or "outside" of $\mathbf{O}$ and hence cannot distinguish between the two.
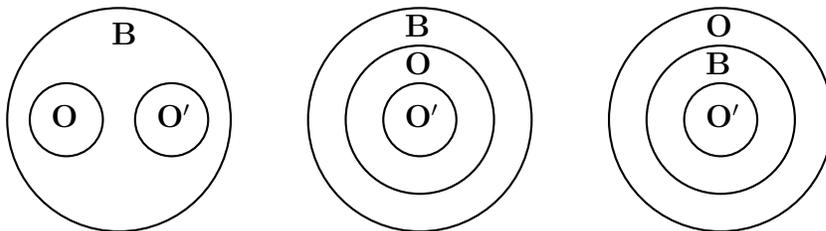

Fig. 1: Corollary 2.1 forbids $\mathbf{O}$ from observationally determining whether it surrounds or is surrounded by either $\mathbf{O}'$ or $\mathbf{B}$.

Corollary 2.2: Unless $\mathbf{O'} = \mathbf{B_O}$, $\mathbf{O'}$ cannot function as a quantum reference frame for $\mathbf{O}$.

Proof: By Theorem 2, there is no classical information channel from $\mathbf{O'}$ to $\mathbf{O}$; therefore $\mathbf{O'}$ cannot provide $\mathbf{O}$ with classical outcomes. $\square$

A quantum reference frame is a physical system, such as a meter stick, a balance or a clock, that enables an observer to obtain classical, arbitrarily-encodable and hence "fungible" information from the world [16]. Corollary 2.2 says that such a reference frame cannot be embedded in a black box such as $\mathbf{B_O}$. Intuitively, this is obvious. Using a reference frame such as a meter stick requires *identifying* it as a meter stick. If $\mathbf{O'}$ is embedded in $\mathbf{B_O}$, $\mathbf{O}$ can only identify $\mathbf{O'}$ if $\mathbf{O}$ has access to at least some internal states of $\mathbf{B_O}$, specifically states of $\mathbf{O'}$ and its immediate surroundings. If $\mathbf{O'}$ is a meter stick, for example, $\mathbf{O}$ must have observational access to *its* states, for example its length and position states, in order to identify it as a meter stick. If $\mathbf{B_O}$ is a black box, however, $\mathbf{O}$ has no such access; no partitions of $\mathbf{B_O}$ are observationally accessible to $\mathbf{O}$, because $H_{\mathbf{OB_O}}$ is strictly invariant across arbitrary partitionings of $\mathbf{B_O}$. Quantum reference frames are sometimes explicitly considered to be components of the observer, not of the observer's environment, for example in [17]. Corollary 2.2 shows that this view of quantum reference frames is the only one consistent with additivity of $H_{\mathbf{U}}$. A consequence of this view is that quantum reference frames cannot, even in principle, be exchanged between observers.

One special case of corollary 2.2 is particularly significant for the discussion that follows. A clock is a quantum reference frame, so no proper subsystem of $\mathbf{B_O}$ can function as a clock for $\mathbf{O}$. Outcomes obtained by $\mathbf{O}$ from $\mathbf{B_O}$ are not, therefore, time-stamped by any *external* clock. As will be seen below, such outcomes can be time-stamped by an internal, $\mathbf{O}$-specific clock. Corollary 2.2 thus requires "time" to be strictly observer-relative in any physics characterized by an additive Hamiltonian.

Corollary 2.3 (no-memory): An observer cannot employ a black box as a read-write memory.

Proof: To employ $\mathbf{B_O}$ as a read-write memory, $\mathbf{O}$ must be capable of encoding classical information on some particular collection $\mathbf{M}$ of degrees of freedom of $\mathbf{B_O}$ and then retrieving that same classical information by some future interaction with $\mathbf{M}$. However, by Corollary 2.2, no such $\mathbf{M}$ can function as a source of classical information for $\mathbf{O}$. $\square$

Physically, $\mathbf{O}$ cannot identify by finite observation which states or degrees of freedom of $\mathbf{B_O}$ have been affected by one of $\mathbf{O}$'s actions. Nor can $\mathbf{O}$ identify by finite observation the source within $\mathbf{B_O}$ of any particular outcome. Either of these putative observational identifications of specific states or degrees of freedom of $\mathbf{B_O}$ would require observational access by $\mathbf{O}$ to the internal structure of $\mathbf{B_O}$, access that is ruled out in principle if $\mathbf{B_O}$ is a black box. In particular, since $\mathbf{O}$ can place no upper limit on the number of accessible states or the number of degrees of freedom of $\mathbf{B_O}$, any interaction with $\mathbf{B_O}$ may be an interaction with previously unaccessed states or degrees of freedom. Hence $\mathbf{O}$ cannot, in principle, establish by finite observation that any particular outcome has the same source within $\mathbf{B_O}$ as any other particular outcome, including any previously-obtained outcome. Finite observations of $\mathbf{B_O}$, for example, do not permit $\mathbf{O}$ to infer that what appears to be a piece of paper with

8

writing on it is *the same* piece of paper that **O** wrote on at some previous time, nor do they permit **O** to infer that what is written on the paper is *the same message* that was written on it at some previous time. Nothing observable by **O**, in particular, enables **O** to rule out the possibility that Mallet the man-in-the-middle attacker did not modify the message, or substitute one piece of paper for another.

That observers have no access to a read-write memory is obviously counter-intuitive. Zurek, for example, considers the ability of observers to "readily consult the content of their memory" as distinguishing them from apparatus ([10] p. 759). Corollary 2.3 shows that observers defined in this way do not exist in any universe in which observers interact only with black boxes, i.e. in any universe characterized by an additive Hamiltonian. Zurek's criterion for picking out observers is, moreover, inconsistent with the definition of **O** as an *arbitrary* subsystem of **U** and hence inconsistent with the requirement of completeness. If the Hamiltonian is additive, nothing distinguishes observers from apparatus. The state of any system and hence the future behavior of any system is dependent upon its past interactions with its environment; **O**'s state and future behavior depend on its past interactions with **B_O** and **B_O**'s state and future behavior depend on its past interactions with **O**. This dependence of current state and future behavior on past interactions is the only sense in which any system has a "memory" in the current setting.

Corollary 2.4 (no-cloning): An observer cannot determine by finite observation that any subsystem $\mathbf{S}'$ of **B_O** is a clone of any other subsystem **S**.

Proof: By Theorem 2, **O** does not have a classical information channel to either **S** or $\mathbf{S}'$; hence **O** can obtain no information with which to determine that either **S** or $\mathbf{S}'$ is a "copy" of the other (cf. [18]). □

Here again, the physical interpretation is clear: **O** and **B_O** comprise a complete, two-component partition of **U**, **O** interacts with **B_O** via $H_{\mathbf{OB_O}}$, and there are no additional, specific interactions between **O** and specific components of **B_O**. Hence **O** can obtain no information about specific components of **B_O**, and in particular no specific information about **S** and $\mathbf{S}'$. It is this inability of **O** to obtain specific information about components of **B_O** that defines **B_O** as a black box.

To summarize, an additive Hamiltonian imposes a stronger constraint on inter-observer signalling than does special relativity. An observer that interacts with a black box has no classical information channel to any proper component of that black box, including any "second observer" contained within the box. Wigner is thus correct, in his famous thought experiment, to say that on walking into his friend's laboratory he queries the joint friend-apparatus system; indeed he queries the joint friend-apparatus-environment system, i.e. the black box **B_Wigner**. His error is to infer from his friend's response that the state of the apparatus "collapsed" when his friend observed it, and thus to impute to his friend's "consciousness" a wave-function collapsing power [19]. This inference is, clearly, based on Wigner's presumption that he in fact queried not the friend-apparatus-environment system but rather just his friend, specifically. If Wigner's world is a black box, such a friend-specific query is not possible: all queries are addressed to, and answered by, the entire world outside the observer.

9

4 Local and global thermodynamics

It has been assumed that physical interactions convey information (i.e. Hamiltonians define information channels) and that observers obtain information *only* via such interactions. Let us now further clarify these assumptions by explicitly adopting Landauer's principle, variously stated as the claim that the irreversible receipt of finite information has a finite free energy cost [1] or that "information is physical" and must be at all times physically encoded [2]. Compliance with this principle requires that any system that receives finite information undergoes a state transition as a result, and that this state transition burns free energy. If we assume that $H_\mathbf{O} \ll H_{\mathbf{OB_O}}$, i.e. that $\mathbf{O}$'s state transitions are results of receiving outcomes from $\mathbf{B_O}$ in compliance with Landauer's principle, then these state transitions can be regarded, from an external, theoretical perspective, as counting the outcomes received. As this count increases monotonically, $\mathbf{O}$ can be regarded, from this external perspective, as a clock with period $\Delta t_\mathbf{O} = 1$ by definition. Let a coordinate $t_\mathbf{O}$ count ticks of this clock; with respect to this $t_\mathbf{O}$, the interaction $H_{\mathbf{OB_O}}$ is clearly irreversible. The acquisition of classical information by $\mathbf{O}$ can, therefore, be regarded as costing at least $0.7\ kT_\mathbf{O}$ per bit, where the local temperature $T_\mathbf{O}$ is taken to be defined by the integral:

$$0.7\ kT_\mathbf{O} \Delta t_\mathbf{O} = \int_{\Delta t_\mathbf{O}} H_{\mathbf{OB_O}} dt_\mathbf{O}. \tag{2}$$

Acquiring an N-bit outcome then requires an action of at least $0.7\ NkT\Delta t_\mathbf{O}$; hence the value $0.7\ kT_\mathbf{O}\Delta t_\mathbf{O}$ can be regarded as the "quantum of action" $h_\mathbf{O}$ that $\mathbf{O}$ must expend to acquire one bit. This value must be finite for any $\mathbf{O}$, and hence must have a finite minimal value across all partitions of $\mathbf{U}$, provided Landauer's principle holds.

Conformational changes in macromolecules following photon absorption provide an experimentally tractable and conceptually relatively uncontroversial model system for the acquisition and storage of information by unengineered physical systems. As the observer $\mathbf{O}$ can be any physical system, a macromolecule $\mathbf{m}$ can be regarded as an observer. For rhodopsin at $T_\mathbf{m} = 310$ K (i.e. 37 C, physiological temperature), $\Delta t_\mathbf{m} \sim 200$ fs [20]. In this case $kT_\mathbf{m} \sim 4.3 \cdot 10^{-21}$ J and the action $h_\mathbf{m} = 0.7\ kT_\mathbf{m}\Delta t_\mathbf{m} \sim 6.0 \cdot 10^{-34}$ J·s per bit, a value remarkably close to that of Planck's constant $h \sim 6.6 \cdot 10^{-34}$ J·s. In what follows, therefore, h will be regarded as the minimal quantum of action required to receive one bit of classical information, as is justified by countless experiments.

This local action is defined, however, only relative to the $\mathbf{O}$-$\mathbf{B_O}$ partition. Because this partition is arbitrary, it can have no effect on the overall state of $\mathbf{U}$.

Theorem 3: The physical state of $\mathbf{U}$ is invariant under local actions.

Proof: The physical state of $\mathbf{U}$ depends only on the self-interaction $H_\mathbf{U}$. This $H_\mathbf{U}$ is invariant under arbitrary partitionings of $\mathbf{U}$; in particular, it is invariant under arbitrary redefinitions of the $\mathbf{O}$-$\mathbf{B_O}$ partition and hence of $H_{\mathbf{OB_O}}$. Therefore the physical state of $\mathbf{U}$ cannot depend on $H_{\mathbf{OB_O}}$. $\square$

This statement is clearly counter-intuitive; it is natural to think that the actions of observers - or local physical interactions in general - affect the state of the universe. This is not, however, correct. Any local interaction $H_{\mathbf{OB_O}}$ is simply a re-labeling, via the $\mathbf{O}$-$\mathbf{B_O}$ partition, of some components of $H_{\mathbf{U}}$, in particular, those components given by $H_{\mathbf{OB_O}} = H_{\mathbf{U}} - (H_{\mathbf{O}} + H_{\mathbf{B_O}})$. The additivity of $H_{\mathbf{U}}$ guarantees that such partition-dependent relabelings have no effect on $H_{\mathbf{U}}$ and hence no effect on the state of $\mathbf{U}$.

As $\mathbf{U}$ is by definition a closed system, its overall physical state is unobservable in principle; any observer $\mathbf{O}$ is embedded in $\mathbf{U}$ and can only obtain information, via $H_{\mathbf{OB_O}}$, about the physical state of the box $\mathbf{B_O}$. Theorem 3 shows that the overall physical state of $\mathbf{U}$ is independent of $H_{\mathbf{OB_O}}$ for any $\mathbf{O}$ and $\mathbf{B_O}$ and thus independent of the outcome information that any observer can obtain. As no upper limit can be placed on the complexity of $\mathbf{U}$ by any observer, an arbitrarily large "multiverse" of distinct observer-box partitions is consistent with the information obtainable by any observer.

Because no observer can observe the overall physical state of $\mathbf{U}$, no observer can observe the action of the state propagator $U_{\mathbf{U}}$ associated with $H_{\mathbf{U}}$. As $\mathbf{U}$ is by definition closed, it interacts with nothing; in particular, it is open to no environmental source of or sink for free energy. Hence $\mathbf{U}$ itself is not an observer; there is nothing for $\mathbf{U}$ to observe and no source of free energy to enable observations. There is, therefore, no clock that defines a universal time period $\Delta t_{\mathbf{U}}$ for the time coordinate $t_{\mathbf{U}}$ conjugate to $H_{\mathbf{U}}$. In the absence of such a clock, it is consistent with all possible observations by all possible observers that $\Delta t_{\mathbf{U}} = \infty$, i.e. that the evolution of $\mathbf{U}$ in $t_{\mathbf{U}}$ is described by the Wheeler-DeWitt equation $H_{\mathbf{U}}|\mathbf{U}\rangle = 0$, where $|\mathbf{U}\rangle$ is the in-principle unobservable overall physical state of $\mathbf{U}$.

Corollary 3.1 (unitarity): The state propagator $U_{\mathbf{U}}$ associated with $H_{\mathbf{U}}$ must be unitary.

Proof: Were $U_{\mathbf{U}}$ asymmetric in $t_{\mathbf{U}}$, a finite $\Delta t_{\mathbf{U}}$ would be definable by $\Delta t_{\mathbf{U}} = h/\Delta E_{\mathbf{U}}$ for any finite energy increment $\Delta E_{\mathbf{U}}$. In this case $E_{\mathbf{U}}$ would increase or decrease in increments of $\Delta E_{\mathbf{U}}$ as $U_{\mathbf{U}}$ propagated $\mathbf{U}$'s state forward in units of $\Delta t_{\mathbf{U}}$, contradicting the definitional assumption that $\mathbf{U}$ is closed against sources of or sinks for free energy. The propagator $U_{\mathbf{U}}$ must, therefore, be symmetric in $t_{\mathbf{U}}$, i.e. unitary. $\square$

As is well known, the unitarity of $U_{\mathbf{U}}$ is equivalent to the physical state of $\mathbf{U}$ being a pure quantum state with a von Neumann entropy of zero, and hence gives rise to the "problem of time" in quantum theory. Corollary 3.1 shows that any physics with an additive Hamiltonian has this problem of time. Corollary 3.1 also closes an important loophole in Corollary 2.2. Were $U_{\mathbf{U}}$ asymmetric in $t_{\mathbf{U}}$, any $\mathbf{O}$ both sufficiently small that $\mathbf{B_O} \sim \mathbf{U}$ and able to measure $\Delta E_{\mathbf{B_O}}$ over an interval $\Delta t_{\mathbf{O}}$ would then have an external time reference frame. The unitarity of $U_{\mathbf{U}}$ prevents this.

If the physical state of $\mathbf{U}$ is independent of $H_{\mathbf{OB_O}}$ for every choice of $\mathbf{O}$ and $\mathbf{B_O}$, Newton's third law must hold for local actions.

Theorem 4 (detailed balance): The action of $\mathbf{O}$ on $\mathbf{B_O}$ via their interaction $H_{\mathbf{OB_O}}$ is exactly balanced by the action of $\mathbf{B_O}$ on $\mathbf{O}$.

Proof: The labels '$\mathbf{O}$' and '$\mathbf{B_O}$' are arbitrary and may be exchanged; hence from the perspective of $\mathbf{B_O}$, $\mathbf{O}$ is a black box from which $\mathbf{B_O}$ obtains a finite number of finite-resolution

outcomes. In particular, $\mathbf{B_O}$ obtains an outcome each time $\mathbf{O}$ makes an observation and hence "asks a question" [21] of $\mathbf{B_O}$. Enforcing Landauer's principle as before, $\mathbf{B_O}$'s state transitions can be regarded, from an external perspective, as counting outcomes received from $\mathbf{O}$ with a clock period $\Delta t_{\mathbf{B_O}}$. As $\mathbf{O}$ delivers no outcome without a "question" from $\mathbf{B_O}$ - from $\mathbf{O}$'s perspective, without a previous outcome from $\mathbf{B_O}$ - it is clear that $\mathbf{B_O}$'s and $\mathbf{O}$'s outcome counts and hence the periods of their counters must be the same, i.e. that $\Delta t_{\mathbf{B_O}} = \Delta t_{\mathbf{O}}$. If $H_{\mathbf{OB_O}}$ is to be well-defined, its integral over $\Delta t_{\mathbf{B_O}} = \Delta t_{\mathbf{O}}$ must be well-defined; therefore,

$$\int_{\Delta t_{\mathbf{B_O}}} H_{\mathbf{OB_O}} dt_{\mathbf{B_O}} = \int_{\Delta t_{\mathbf{O}}} H_{\mathbf{OB_O}} dt_{\mathbf{O}} = 0.7 \ NkT\Delta t_{\mathbf{O}} = 0.7 \ NkT\Delta t_{\mathbf{B_O}} \tag{3}$$

at a common temperature T. $\square$

Physically, detailed balance requires the classical information content of $\mathbf{O}$'s "question" to $\mathbf{B_O}$ to equal the classical information content of $\mathbf{B_O}$'s "answer." The no-memory corollary (2.3) shows that such information, once transferred, becomes inaccessible to its source, thus justifying the assumption of local, observer-relative irreversibility and hence of Landauer's principle. In particular, $\mathbf{O}$ can never determine that $\mathbf{B_O}$ has returned to the "same" state or vice-versa.

5 Hilbert-space representation

Theorem 4 requires that in any particular interaction, $\mathbf{O}$'s input to $\mathbf{B_O}$ transfers the same number N of classical bits as does $\mathbf{B_O}$'s response to $\mathbf{O}$. Theorem 4 does not, however, require $\mathbf{O}$ to be capable of *counting* these outgoing bits. Hence Theorem 4 places no restriction on $\mathbf{O}$'s "internal representation" of the "questions" $\mathbf{O}$ poses to $\mathbf{B_O}$. These questions can be described, from $\mathbf{O}$'s perspective, in any arbitrary way.

Let $\{\alpha_{ij}\}$ be the complete set of outcomes recordable by $\mathbf{O}$, assuming for convenience that each of the α_{ij} is encoded as an N-bit string. Note that if $\mathbf{O}$ is restricted to finite observations, i.e. to finite total energy consumption, this set $\{\alpha_{ij}\}$ must be finite. Freedom to describe $\mathbf{O}$'s questions to $\mathbf{B_O}$ in any arbitrary way is then freedom to choose the maximal values of i and j in any way that preserves their finite product, e.g. as 1 and 100, 100 and 1 or 10 and 10. Whatever the value chosen for the maximal value of i, let $\{\zeta_i\}$ be a set of functions:

$$\zeta_i : \mathcal{H}_{\mathbf{B_O}} \to \{\alpha_{ij}\}, \tag{4}$$

where $\mathcal{H}_{\mathbf{B_O}}$ is an as-yet uncharacterized abstract space representing the physical states of $\mathbf{B_O}$, and where each of the α_{ij} is in the image of at least one of the ζ_i. Note that nothing requires these functions ζ_i to be injections; hence the cardinality of the underlying set of

$\mathcal{H}_{\mathbf{B_O}}$ can be arbitrarily larger than the finite product ij. Call these maps ζ_i the "reference frames" of $\mathbf{O}$.

Consider a particular observation of $\mathbf{B_O}$ by $\mathbf{O}$ in which the Hamiltonian $H_{\mathbf{OB_O}}$ implements some particular reference frame ζ_i; in this case, $\mathbf{O}$ obtains as an outcome some particular α_{ij}. From an external, theoretical perspective, what is known is that $H_{\mathbf{OB_O}}$ acts over a time interval $\Delta t_{\mathbf{O}}$ and implements an action 0.7 $NkT\Delta t_{\mathbf{O}}$, which can be assumed to be minimal and hence identified as Nh. As there is no external time reference frame to assign an "objective" value to $\Delta t_{\mathbf{O}}$, we can treat the interaction as effectively instantaneous and write, for the nth interaction between $\mathbf{O}$ and $\mathbf{B_O}$,

$$H_{\mathbf{OB_O}}(t_{\mathbf{O}} = n\Delta t_{\mathbf{O}}) = (1/Nh)\delta(t_{\mathbf{O}} - n\Delta t_{\mathbf{O}}). \tag{5}$$

This representation is consistent with $\mathbf{O}$'s in principle ignorance of the complexity of $\mathbf{B_O}$, i.e. with the possibility that $\mathbf{B_O}$ undergoes an arbitrarily large number of state changes between interactions with $\mathbf{O}$.

Intuitively, a reference frame such as a meter stick acts on only a limited collection of the degrees of freedom of $\mathbf{U}$; when one measures a length with a meter stick, one interprets it as the length of *something*. No single measurement, however, reveals *which* particular collection of degrees of freedom is being measured; an outcome of 0.6 m, for example, indicates that *something* is 0.6 m long, but does not reveal *which* something is 0.6 m long. Any given outcome could, in principle, have been obtained from any of an arbitrarily large number of 0.6 m long somethings. If we give the measured something a name, e.g. '$\mathbf{S}$' on the basis of its measured 0.6 m length, that name is arbitrarily referentially ambiguous [22], i.e. it can equally-well refer to any something that is 0.6 m long. Describing $\mathbf{S}$ formally as a subsystem within a black box $\mathbf{B_O}$ merely enforces this referential ambiguity via Corollary 2.3. With this in mind, we can choose to view $H_{\mathbf{OB_O}}$ as an interaction between $\mathbf{O}$ and an otherwise uncharacterized something $\mathbf{S}$ when we consider observations in which $H_{\mathbf{OB_O}}$ implements ζ_i; we can then rewrite (5) as:

$$H_{\mathbf{OS}}(t_{\mathbf{O}} = n\Delta t_{\mathbf{O}}) = (1/Nh)\delta(t_{\mathbf{O}} - n\Delta t_{\mathbf{O}}). \tag{6}$$

We can, similarly, choose to regard the something $\mathbf{S}$ as occupying a "state" $|\mathbf{S}(t)\rangle$ at $t_{\mathbf{O}}$, bearing in mind that for consistency with Corollary 2.3, the *physical* meaning of $|\mathbf{S}(t)\rangle$ is simply a state of $\mathbf{B_O}$, one of possibly arbitrarily many, that yields some particular α_{ij} as an outcome when queried with ζ_i at $t_{\mathbf{O}}$.

Consider now a finite sample $\mathcal{S}$ of observations of $\mathbf{B_O}$ by $\mathbf{O}$ in which $H_{\mathbf{OB_O}}$ implements the particular reference frame ζ_i. These observations can be thought of as picking out a particular subcollection of the degrees of freedom of $\mathbf{B_O}$, i.e. a particular properly-contained subsystem $\mathbf{S}$ of $\mathbf{B_O}$ as discussed above. From $\mathbf{O}$'s perspective, the prior probability distribution for the α_{ij} over all of $\mathbf{B_O}$ can only be uniform, as any non-uniform prior distribution would require placing an upper limit on the complexity of $\mathbf{B_O}$, an upper limit that, as noted earlier, is unobtainable by any observer. From a practical perspective, however, one would

like to know the priors over just the subsystem **S** picked out by the sample of observations $\mathcal{S}$, i.e. one would like to calculate the probability of observing some particular outcome α_{ik} in an experiment $\mathcal{S}$ that employs the reference frame ζ_i.

Theorem 5: If the sample $\mathcal{S}$ picks out a properly-contained subsystem **S** of $\mathbf{B_O}$, it cannot be treated as a classical ensemble.

Proof: Assume that $\mathcal{S}$ is treated as a classical ensemble. It is then a specification, for **O**, of classical information, namely the prior probability distribution of the α_{ij} over the subsystem **S**. If $\mathcal{S}$ indeed picks out such a proper subsystem **S** of $\mathbf{B_O}$, however, this **S** must be regarded as the source of classical information for **O**. By Theorem 2, no subsystem of $\mathbf{B_O}$ can be a source of classical information for **O**; hence **S** cannot be a source of classical information. The sample $\mathcal{S}$ cannot, therefore, transfer classical information from **S**; in particular, $\mathcal{S}$ cannot specify a prior probability distribution of the α_{ij} over **S**. Hence $\mathcal{S}$ cannot be treated as a classical ensemble. $\square$

Physically, $\mathcal{S}$ being a classical ensemble corresponds to ζ_i being a reference frame that rules out, *a priori*, any potential effects of the degrees of freedom of $\mathbf{B_O}$ that are not in **S** on the outcomes of observations made with ζ_i. It corresponds, in other words, to ζ_i perfectly isolating **S** from the rest of $\mathbf{B_O}$. Any such isolation clearly violates the additivity of $H_\mathbf{U}$.

Corollary 5.1: If **S** and **E** jointly partition $\mathbf{B_O}$, **E** cannot be a classical ensemble.

Proof: By Theorem 5, substituting the label '**E**' for '**S**'. $\square$

In the terminology of decoherence, **E** is the "environment" of **S** [10, 11]. Corollary 5.1 shows that the environment cannot be treated as a classical ensemble in any theory in which $H_\mathbf{U}$ is additive. Assuming a classical environment for the purposes of decoherence calculations, or equivalently, assuming that **O** does not interact with and hence is isolated from **E**, is therefore formally inconsistent with quantum theory, not merely an example of circular reasoning as has been shown previously [14, 15].

Corollary 5.2 (Bell's theorem): If **S** is a proper subsystem of $\mathbf{B_O}$, then either the state $|\mathbf{S}\rangle$ is counterfactually indefinite or the dynamics $|\mathbf{S}(t)\rangle \to |\mathbf{S}(t + \Delta t_\mathbf{O})\rangle$ depend on degrees of freedom outside **S**.

Proof: Assume the contrary. In this case **S** can be considered to be a classical ensemble, contradicting Theorem 5. $\square$

From a physical perspective, the non-separability of the state $|\mathbf{B_O}\rangle$ implied by Bell's theorem follows immediately from the additivity of $H_\mathbf{U}$; additivity requires that *no* boundary partitioning degrees of freedom of **U** into subsystems can have any physical consequences. In particular, the **S**-**E**, **S**-**O** and **O**-**E** boundaries can have no physical consequences; therefore drawing such boundaries cannot disrupt quantum entanglement to produce "separable states" $|\mathbf{S}\rangle$, $|\mathbf{E}\rangle$ or $|\mathbf{O}\rangle$. No flow of classical information across these boundaries can, therefore, have any physical consequences, consistent with Theorem 3 above.

It remains now only to characterize the abstract state space $\mathcal{H}_{\mathbf{B_O}}$ on which the reference frames ζ_i act. Over the course of some experiment $\mathcal{S}$ employing ζ_i, a particular outcome α_{ik}

will occur, if it occurs at all, for some observations but not, in general, for all observations. We can therefore think of $H_{\mathbf{OS}}$ as an interaction that on some occasions "picks out" degrees of freedom of $\mathbf{B_O}$ that yield α_{ik} as an outcome and that on other occasions picks out degrees of freedom that yield other outcomes. In this case we can write:

$$H_{\mathbf{OS}}(t_{\mathbf{O}}) = \sum_j b_j e^{-(\varphi_j t_{\mathbf{O}})} \delta(t_{\mathbf{O}} - n\Delta t_{\mathbf{O}}), \tag{7}$$

where $e^{-(\varphi_j t_{\mathbf{O}})}$ is the periodic function that picks out a collection of degrees of freedom of $\mathbf{B_O}$ that yields the jth outcome when $\mathbf{B_O}$ is acted upon by ζ_i and b_j measures how many such collections there are. Making the action of $H_{\mathbf{OS}}(t_{\mathbf{O}})$ on $|\mathbf{S}\rangle$ explicit, (7) becomes:

$$H_{\mathbf{OS}}(t_{\mathbf{O}})|\mathbf{S}\rangle = \sum_j b_j e^{-(\varphi_j t_{\mathbf{O}})} \delta(t_{\mathbf{O}} - n\Delta t_{\mathbf{O}})|\mathbf{S}\rangle, \tag{8}$$

which can equally well be written:

$$H_{\mathbf{OS}}(t_{\mathbf{O}})|\mathbf{S}\rangle = (\sum_j b_j e^{-(\varphi_j t_{\mathbf{O}})}|\mathbf{S}_j\rangle)\delta(t_{\mathbf{O}} - n\Delta t_{\mathbf{O}}), \tag{9}$$

where physically $|\mathbf{S}_j\rangle$ is, as above, just a state of $\mathbf{B_O}$ that yield α_{ij} as an outcome when acted upon by ζ_i.

Physically, the superposition (9) expresses $\mathbf{O}$'s uncertainty about *which* of the potentially arbitrarily many collections of degrees of freedom of $\mathbf{B_O}$ that could yield α_{ij} as an outcome when acted upon by ζ_i are actually being acted upon by ζ_i and yielding the outcome α_{ij} at $t_{\mathbf{O}}$. It expresses, in other words, $\mathbf{O}$'s uncertainty about which 0.6 m long something is currently being measured. Bell's theorem tells us, very explicitly, that $\mathbf{O}$ cannot resolve this uncertainty: the phase φ_j that determines when α_{ij} is produced as an outcome is unmeasurable in principle. This phase cannot, therefore, be real-valued; it must be the case that $\varphi_j = i\phi_j$ for some real value ϕ_j. The abstract space $\mathcal{H}_{\mathbf{B_O}}$ must, therefore, be a Hilbert space, not a real configuration space; the reference frames ζ_i must correspond to measures on this Hilbert space, i.e. projections or more generally, POVMs. To see that this is the case, note that the inverse image $Im^{-1}(\alpha_{ij})$ of each possible outcome α_{ij} is, by definition, that set of states of $\mathbf{B_O}$ that yields α_{ij} when acted upon by some one (or more) of the reference frames ζ_i; because the ζ_i are by definition functions from $\mathcal{H}_{\mathbf{B_O}}$, the sum $\oplus_{ij} Im^{-1}(\alpha_{ij})$ must equal $\mathcal{H}_{\mathbf{B_O}}$. For each ζ_i, consider the collection of states $Im^{-1}(\alpha_{ij})$ a "system" $|\mathbf{S}_j\rangle$; it is then clear that $\sum_j |\mathbf{S}_j\rangle\langle\mathbf{S}_j| = \mathbf{Id}$. Hence $\sum_j b_j^2 = 1$, allowing the b_j^2 to be interpreted as Born-rule probabilities. As the $|\mathbf{S}_j\rangle$ and hence the b_j are defined only relative to $H_{\mathbf{OB_O}}$ and hence relative to the $\mathbf{O}$-$\mathbf{B_O}$ partitioning of $\mathbf{U}$, these probabilities are observer-relative and hence "subjective" in the Bayesian sense.

15

6 Conclusion

Assuming an additive Hamiltonian and hence the invariance of the Hamiltonian under decompositions of the state space is equivalent, clearly, to assuming an associative product to represent state space decomposition. The Hilbert-space tensor product $\otimes$ is associative; hence standard quantum theory assumes the invariance of the Hamiltonian under state-space decompositions explicitly. The in-principle non-uniqueness of state space decompositions within standard quantum theory has been widely noted, particularly under the rubric of "entanglement relativity" [23, 24, 25, 26, 27, 28, 29, 30]. What has been shown here is that the invariance of the Hamiltonian under decompositions of the state space is not merely an interesting *feature* of quantum theory, but is rather the *source* of quantum theory.

If merely assuming an additive Hamiltonian and Landauer's principle yields quantum theory as shown here, what makes classical physics, in which the Hamiltonian is invariant under state-space decomposition as noted earlier, "classical" and not "quantum"? The answer, clearly, is that classical physics standardly assumes that observers can obtain outcome information both *specifically* and *exclusively* from particular, identified subsystems of $\mathbf{B_O}$, i.e. from particular identified subsystems of the universe $\mathbf{U}$ minus the observer. This is not a formal, axiomatic assumption, but is rather an implicit metaphysical assumption. This implicit assumption contradicts Theorem 2, and it takes various forms. For example, it may be assumed that $\mathbf{O}$ interacts so quickly with a particular, identified subsystem $\mathbf{S}$ that the measurement resolution is effectively infinite (i.e. $c \to \infty$), or that $\mathbf{O}$ interacts so subtly with $\mathbf{S}$ that no energy is transferred and hence no finite action is required (i.e. $h \to 0$). Alternatively, it may be assumed that $\mathbf{S}$ is a classical ensemble, i.e. that $\mathbf{S}$ is completely isolated from all other systems and hence characterizable by a known prior probability distribution. As this is *a fortiori* an assumption that $\mathbf{S}$ has been and remains isolated from any *physical* $\mathbf{O}$, it amounts to the assumption that $\mathbf{O}$ is not a physical system and hence that observation is not a physical interaction. Classical physics is, therefore, either inconsistent with the thermodynamics of observation, or else fails to be complete by failing to describe observation as a physical interaction.

Common "interpretations" of quantum theory make similar assumptions. Decoherence theory, for example, typically assumes that $\mathbf{O}$ is "not observing" the environment $\mathbf{E}$ and hence is unaffected by it. This is effectively an assumption that *no* state change of $\mathbf{E}$ can transfer information to $\mathbf{O}$ that will affect the outcome of $\mathbf{O}$'s observations of $\mathbf{S}$, an assumption that is readily falsified by walking into any laboratory and turning off the lights. The "environment as witness" formulation of decoherence (e.g. [31]) assumes that interactions with $\mathbf{E}$ transfer information *specifically* about $\mathbf{S}$ to $\mathbf{O}$, i.e. that $\mathbf{E}$ is a classical information channel from $\mathbf{S}$ to $\mathbf{O}$, contradicting Theorem 2. While the Copenhagen interpretation in its purest form concerns only the observer's knowledge, it is often presented to both students and the public as requiring quantum states to *physically* collapse, a requirement that clearly contradicts additivity. Even QBism (e.g. [17]) assumes that observers who are ignorant of a system's physical state - who might not, in principle, know where the system is or what it

looks like - can nonetheless re-identify that very same system for subsequent observations, again contradicting Theorem 2.

The spectacular empirical success of quantum theory tells us that assuming an additive Hamiltonian yields excellent and enormously useful predictions about the world. What has been shown here is that the formalism of quantum theory also presents us with a clear and simple message: we are physical systems, and the world that we observe is, to us, a black box.

Acknowledgements

References

[1] R. Landauer. Irreversibility and heat generation in the computing process. *IBM J. Research Development*, 5: 183-195, 1961.

[2] R. Landauer. Information is a physical entity. *Physica A*, 263: 63-67, 1999.

[3] T. F. Jordan. Assumptions that imply quantum dynamics is linear. *Phys. Rev. A*, 73, 022101, 2006.

[4] T. F. Jordan. Fundamental significance of tests that quantum dynamics is linear. *Phys. Rev. A*, 82, 032103, 2010.

[5] C. E. Shannon. A mathematical theory of communication. *Bell System Tech. J.*, 27: 379-423, 1948.

[6] W. R. Ashby. *An Introduction to Cybernetics*. Chapman and Hall, London, 1956.

[7] E. F. Moore. Gedankenexperiments on sequential machines. In C. W. Shannon and J. McCarthy, editors, *Autonoma Studies*, pages 129-155. Princeton University Press, Princeton, NJ, 1956.

[8] E. Chitambar, D. Leung, L. Mančinska, M. Ozols and A. Winter. Everything you always wanted to know about LOCC (but were afraid to ask). *Commun. Math. Phys.*, 328, 303-326 (2014).

[9] C. Simon, V. Bužak and N. Gisin. The no-signaling condition and quantum dynamics. *Phys. Rev. Lett.*, 87: 170405, 2001. arxiv:quant-ph/0102125v1.

[10] W. H. Zurek. Decoherence, einselection, and the quantum origins of the classical. *Rev. Mod. Phys.*, 75: 715-775, 2003. arXiv:quant-ph/0105127v3.

[11] M. Schlosshauer. *Decoherence and the Quantum to Classical Transition.* Springer, Berlin, 2007.

[12] R. Blume-Kahout and W. H. Zurek. Quantum Darwinism: Entanglement, branches, and the emergent classicality of redundantly stored quantum information. *Phys. Rev. A*, 73, 062310.

[13] W. H. Zurek. Quantum Darwinism. *Nature Phys.*, 5, 181-188.

[14] R. Kastner. Einselection of pointer observables: The new H-theorem? *Stud. Hist. Phil. Mod. Phys.*, 48 Part A, 56-58, 2014. arXiv:1406.4126v2 [quant-ph].

[15] C. Fields. On the Ollivier-Poulin-Zurek definition of objectivity. *Axiomathes*, 24: 137-156, 2014. arXiv:1102.2826v5 [quant-ph].

[16] S. D. Bartlett, T. Rudolph, and R. W. Spekkens. Reference frames, superselection rules, and quantum information. *Rev. Mod. Phys.*, 79:555-609, 2007. arXiv:quant-ph/0610030v3.

[17] C. A. Fuchs and R. Schack. Quantum Bayesian coherence. *Rev. Mod. Phys.*, 85: 1693-1715, 2013. arXiv:0906.2187v1 [quant-ph].

[18] C. Fields. Bell's theorem from Moore's theorem. *Int. J. Gen. Syst.*, 42: 376-385, 2013. arXiv:1201.3672v6 [quant-ph].

[19] E. P. Wigner. Remarks on the mind-body question. In I. J. Good (Ed.) *The Scientist Speculates.* Basic, New York, pp. 284-302 (1962).

[20] Q. Wang, R. W. Schoenlein, L. A. Peteanu, R. A. Mathies and C. V. Shank. Vibrationally coherent photochemistry in the femtosecond primary event of vision. *Science*, 266: 422-424, 1994.

[21] J. A. Wheeler. Information, physics, quantum: The search for links. W. H. Zurek (Ed.) *Complexity, Entropy, and the Physics of Information* (pp. 309-336). Redwood City, CA: Addison-Wesley, 1990.

[22] C. Fields. Equivalence of the symbol grounding and quantum system identification problems. *Information*, 5: 172-189, 2014.

[23] P. Zanardi. Virtual quantum subsystems. *Phys. Rev. Lett.*, 87:077901, 2001. arXiv:quant-ph/0103030v2.

[24] P. Zanardi, D. A. Lidar, and S. Lloyd. Quantum tensor product structures are observable-induced. *Phys. Rev. Lett.*, 92:060402, 2004. arXiv:quant-ph/0308043v1.

[25] M. Dugić and J. Jeknić. What is "system": Some decoherence-theory arguments. *Int. J. Theor. Phys.*, 45: 2249-2259, 2006.

[26] M. Dugić and J. Jeknić - Dugić. What is "system": The information-theoretic arguments. *Int. J. Theor. Phys.*, 47: 805-813, 2008.

[27] A. C. de la Torre, D. Goyeneche, and L. Leitao. Entanglement for all quantum states. *Eur. J. Phys.*, 31:325–332, 2010. arXiv:1002.2893v1.

[28] N. L. Harshman and K. S. Ranade. Observables can be tailored to change the entanglement of any pure state. *Phys. Rev. A*, 84:012303, 2011. arXiv:1102.0955v1 [quant-ph].

[29] W. Thirring, R. A. Bertlmann, P. Köhler, and H. Narnhofer. Entanglement or separability: The choice of how to factorize the algebra of a density matrix. *Eur. Phys. J. D*, 64:181–196, 2011. arXiv:1106.3047v2 [quant-ph].

[30] M. Dugić and J. Jeknić - Dugić. Parallel decoherence in composite quantum systems. *Pramana* 79:199–209, 2012. arXiv:1011.5919v6 [quant-ph].

[31] H. Ollivier, D. Poulin, and W. H. Zurek. Objective properties from subjective quantum states: Environment as a witness. *Phys. Rev. Lett.*, 93:220401, 2004. (arXiv:quant-ph/0307229v2).