

Decompositional equivalence: A fundamental symmetry underlying quantum theory

Chris Fields

*243 West Spain Street
Sonoma, CA 95476 USA*

fieldsres@gmail.com

October 2, 2015

Abstract

Decompositional equivalence is the principle that there is no preferred decomposition of the universe into subsystems. It is shown here, by using a simple thought experiment, that quantum theory follows from decompositional equivalence together with Landauer's principle. This demonstration raises within physics a question previously left to psychology: how do human - or any - observers identify or agree about what constitutes a "system of interest"?

Keywords: Black box; Cybernetics; Information; Measurement; Objectivity; Observer

1 Introduction

The enormous empirical success of quantum theory provides strong evidence that quantum theory is true, at least in the pragmatic sense of enabling correct predictions about the observable behavior of the world. What has been unclear for almost 90 years now is *why* it is true. Despite decades of effort, we have not found simple, intuitively-compelling physical principles that lead us to *expect* quantum theory to be true. The theory is, instead, almost universally regarded and presented to students as simply a mathematical framework that makes correct if counter-intuitive predictions. Without a clear physical picture to motivate the mathematical formalism of quantum theory, we are left with a choice between weakly-constrained metaphysical speculations and "shut up and calculate."

My goal in this paper is to show that standard, unitary quantum theory follows from two simple, intuitively compelling physical principles. The first of these is Landauer's principle that "information is physical" or more precisely, that 1) classical information only "exists" when it has been encoded by a thermodynamically-irreversible state change in some physical system, and 2) any such encoding requires a finite expenditure of free energy [1, 2]. The second principle is that all fundamental physical interactions are entirely invariant under arbitrary decompositions of any physical system into subsystems. This principle of "decompositional equivalence" has been assumed in one form or another, generally implicitly, from the very beginnings of modern science; indeed the possibility of scientific investigation would seem to require a universe in which what we humans choose to call things – in particular, what we choose to designate as the "system of interest" – does not affect what is going on. What is shown here is that these two principles, taken together, require the formal description of information acquisition by an observer to be standard, unitary quantum theory. Quantum theory thus emerges, in this treatment, not as a theory with a "measurement problem," but as a clear and physically well-motivated formalism that precisely limits the information that any observer of any system can obtain by measurement.

The paper is organized as follows. The next section defines decompositional equivalence, discusses its close relationship to classical reductionism, and shows that it is incorporated implicitly into the state-space formalisms of both classical and quantum physics. Both classical and quantum physics, in other words, require decompositional equivalence. The following two sections discuss two consequences of decompositional equivalence: 1) that there can be no distinct or "preferred" class of observers and 2) that all finite observation – all observation subject to Landauer's principle – can be described as interaction between an observer and a *black box* as defined by classical cybernetics, i.e. a system with observable external behavior but an unobservable and hence unknowable interior [3, 4]. The fifth section employs a simple thought experiment to show that if Landauer's principle is assumed, the only correct description of the interaction between an observer and a black box is standard, unitary quantum theory. It shows, in particular, that the three principal axioms of quantum theory, those specifying the Hilbert space formalism, unitary evolution, and measurement by projection all follow from decompositional equivalence when a finite energetic cost of encoding observational outcomes is assumed. The sixth section discusses the physical meaning of superposition and shows how the standard quantum no-go theorems are entailed by decompositional equivalence plus Landauer's principle. The seventh section illustrates the previous results using the double-slit experiment as an example. The concluding section briefly discusses common assumptions of the "classical worldview" that violate decompositional equivalence, Landauer's principle or both and therefore contradict quantum theory. These include assumptions that the sources of observational outcomes can be precisely identified and that systems preparation or any other physical process involves local causation only.

2 Decompositional equivalence: A ubiquitous but largely implicit assumption

“Nature is relentless and unchangeable, and it is indifferent as to whether its hidden reasons and actions are understandable to man or not.”

– Galileo Galilie

Physics – indeed, all of science – is motivated by a deep intuition that *there is a way that the world works* and that this way that the world works is independent of what we humans or any other observers may say, do, or believe. This intuition underlies both the idea that Nature has laws and the idea that those laws can be given a universal, observer-independent mathematical formulation. It also underlies the idea of an experiment as a “question to the world” to which the world gives an unbiased answer.

This deep intuition can be made more precise by noting that to conduct an experiment, an observer must: 1) designate a “system of interest” on which the experiment is to be performed, 2) choose the experiment to perform, and then 3) manipulate the system of interest in some way to “prepare” it for observation. While the outcome of the experiment clearly depends on these three actions by the observer, it is universally assumed that the *processes by which the world produces that outcome* do not. These processes reflect how the world works, and how the world works is independent of what observers say, do or believe. This distinction between the overt, observable behavior of the world and the “hidden reasons and actions” that produce that behavior is a key component of the intuition that there is a way that the world works. Consistent with common usage, these hidden reasons and actions can be referred to as “fundamental interactions” and taken, for the purposes of the present argument, to be the fundamental interactions of the Standard Model plus gravity, string theory, M theory, or some future fundamental physics.

Let us focus on the first of the actions that an observer must take to perform an experiment: the designation of a system of interest. A second deep intuition in both physics and other sciences is that there are no *principled* limitations on what can be subjected to scientific investigation. At least in principle, the choice of a system of interest is *entirely arbitrary*; any system whatsoever can be chosen. This arbitrariness is nicely illustrated by the common practice of considering some arbitrary voxel when discussing classical fluid flow; the voxel boundaries can be drawn anywhere, without restriction, to focus attention on the behavior of some small sample of the fluid. The arbitrary movability of the “Heisenberg cut” in Copenhagen quantum theory provides another example; conceptually including the measurement apparatus or even the observer’s sensory organs within the “system of interest” makes no difference to the way that the world works and hence no difference to the outcome of the experiment.

The principle of decompositional equivalence combines these two deep intuitions by noting that the choice of a system of interest is an instance of *system decomposition*; in particular,

it is the decomposition of “the world” into the system of interest and “everything else.” The principle generalizes this instance to all decompositions, by stating:

Decompositional equivalence: All fundamental physical interactions are entirely invariant under arbitrary decompositions of any physical system into subsystems.

In particular, all fundamental physical interactions are entirely invariant under arbitrary decompositions of the universe \mathbf{U} into a system of interest \mathbf{S} , an observer \mathbf{O} and a surrounding environment \mathbf{E} , where \mathbf{E} is taken to include everything in \mathbf{U} not included in either \mathbf{S} or \mathbf{O} . When stated in this way, decompositional equivalence entails a form of scale invariance: if fundamental interactions are invariant under arbitrary decompositions of \mathbf{U} into \mathbf{S} , \mathbf{O} and \mathbf{E} , then they are invariant under changes in the scales at which \mathbf{S} , \mathbf{O} and \mathbf{E} are defined. Both \mathbf{S} and \mathbf{O} , in particular, can be defined to be arbitrarily small or large, subject only to the constraint that \mathbf{S} , \mathbf{O} and \mathbf{E} together decompose \mathbf{U} . Decompositional equivalence requires, therefore, that there be no preferred scale of observation and no preferred scale of observers. If \mathbf{U} satisfies decompositional equivalence, fundamental interactions in \mathbf{U} are independent of what *any* observer might choose to designate as the “system of interest.”

It is important to emphasize that this principle of decompositional equivalence concerns the decomposition of *systems into subsystems* and says nothing about the decomposition of states into substates or component states. Hence the claim that a composite system \mathbf{X} satisfies decompositional equivalence implies nothing about the state occupied by \mathbf{X} ; in particular, it does not imply that the state occupied by \mathbf{X} is separable in the quantum-theoretic sense, as discussed in §5 and §6 below.

Noting that any decomposition of a system into subsystems effectively draws boundaries, equivalent to the voxel boundaries mentioned above, around the subsystems, the principle can alternatively be stated as:

Decompositional equivalence (Variant 1): All fundamental physical interactions are entirely independent of boundaries drawn to decompose a physical system into subsystems.

or

Decompositional equivalence (Variant 2): The boundaries around, and therefore the boundaries between subsystems of a physical system have no effect on the overall behavior of the system.

This last statement of the principle is closely related to the classical “reductionist” principle that the behavior of a system is completely determined by the behavior of its “microscopic” components. The historical reductionist program of fully predicting all macroscopic system behavior from a theory of the behavior of such microscopic components plus a set of “bridge

laws” [5] has largely been abandoned and “emergent phenomena” are now widely regarded as unpredictable, perhaps even in principle, from the behavior of microscopic components (e.g. [6]). However, such phenomena are nonetheless still regarded as *implemented by* the microscopic components of a system and hence as (possibly unpredictable) outcomes of their fundamental interactions. This reductionist principle has been a mainstay of the physical sciences since the early 19th century development of the classical theory of atoms; indeed, it is the principle that motivates the continuing search for yet more microscopic physical components and yet more fundamental physical interactions. Any system to which it applies must satisfy decompositional equivalence; if it did not, macroscopic “emergent phenomena” could result from boundary-dependent alterations of the fundamental interactions and would thus be determined not just by the behavior of the microscopic components but by the observer-dependent placement of subsystem boundaries.

It is worth noting explicitly that decompositional equivalence is a *physical* principle, not a mathematical principle. As such, its theoretical role is to both motivate and constrain the choice of mathematical representations of physical systems and physical dynamics. It requires, in particular, that any correct mathematical representation of physical dynamics must be independent of the mathematical representation of subsystem decomposition and hence of subsystem boundaries. As will be shown in §5 below, the conjunction of decompositional equivalence and Landauer’s principle materially implies the mathematical representation provided by standard, unitary quantum theory; if both principles hold, unitary quantum theory must be true, and if unitary quantum theory is false, one or the other of the two principles must be false.

Conversely, both classical and quantum theories materially imply decompositional equivalence by building it into their respective representations of the space of possible physical states. Both theories employ abstract vector spaces to represent physical states, the real phase space in classical physics and the complex Hilbert space in quantum physics. Both spaces can be arbitrarily decomposed into subspaces using vector-space product operators, the Cartesian product \times in classical physics and the tensor product \otimes in quantum physics. These operators are both associative: for any decomposition of \mathbf{U} into \mathbf{S} , \mathbf{E} and \mathbf{O} , the classical phase space $\mathcal{C}_{\mathbf{U}} = \mathcal{C}_{\mathbf{S}} \times \mathcal{C}_{\mathbf{E}} \times \mathcal{C}_{\mathbf{O}} = (\mathcal{C}_{\mathbf{S}} \times \mathcal{C}_{\mathbf{E}}) \times \mathcal{C}_{\mathbf{O}} = \mathcal{C}_{\mathbf{S}} \times (\mathcal{C}_{\mathbf{E}} \times \mathcal{C}_{\mathbf{O}})$ in classical physics, and the Hilbert space $\mathcal{H}_{\mathbf{U}} = \mathcal{H}_{\mathbf{S}} \otimes \mathcal{H}_{\mathbf{E}} \otimes \mathcal{H}_{\mathbf{O}} = (\mathcal{H}_{\mathbf{S}} \otimes \mathcal{H}_{\mathbf{E}}) \otimes \mathcal{H}_{\mathbf{O}} = \mathcal{H}_{\mathbf{S}} \otimes (\mathcal{H}_{\mathbf{E}} \otimes \mathcal{H}_{\mathbf{O}})$ in quantum physics. Both classical and quantum physics, therefore, require the state space of any system to be *entirely invariant* under arbitrary product decompositions. In both theories, moreover, physical dynamics are represented by a Hamiltonian operator that is additive and therefore associative: if $\hat{H}_{\mathbf{U}}$ is the Hamiltonian representing the dynamics of \mathbf{U} , then in both theories $\hat{H}_{\mathbf{U}} = \hat{H}_{\mathbf{S}} + \hat{H}_{\mathbf{E}} + \hat{H}_{\mathbf{O}} + \hat{H}_{\mathbf{SE}} + \hat{H}_{\mathbf{SO}} + \hat{H}_{\mathbf{EO}} + \hat{H}_{\mathbf{SEO}}$ for any arbitrary choice of \mathbf{S} , \mathbf{E} and \mathbf{O} , where $\hat{H}_{\mathbf{S}}$, $\hat{H}_{\mathbf{E}}$ and $\hat{H}_{\mathbf{O}}$ are the respective self-interactions and the higher-order terms are between-system interactions. Both classical and quantum physics, therefore, require the dynamics of any system to be *entirely invariant* under arbitrary product decompositions of the system’s state space. Were decompositional equivalence shown to be false in our universe, neither state spaces nor Hamiltonians would be arbitrarily decomposable in this way and the basic mathematical formalisms of both quantum and classical

physics would have to be rejected.

As important as what these formal representations require is what they *do not* require. The classical phase-space representation does not, by itself, require that position and momentum be instantaneously and hence simultaneously measurable with arbitrary precision; it does not, in other words, require the observable state of every or even any classical system to be a point in phase space. Hence it does not require Landauer’s principle to be violated; indeed doing so would render classical statistical mechanics inconsistent. This key component of the “classical worldview” is instead a separate assumption that the classical phase space merely allows. Similarly, the state-space representations of neither classical nor quantum physics require the assumption that subsystems of every or even any composite system \mathbf{X} can be individually and independently manipulated, and neither requires the assumption that observational outcomes obtained from a composite system \mathbf{X} can, in every or even any case, be attributed to specific, uniquely identifiable subsystems of \mathbf{X} . It is shown below that these two assumptions, both of which are central components of the classical worldview, violate the conjunction of decompositional equivalence and Landauer’s principle and are therefore inconsistent with the state-space formalisms of either classical or quantum physics if Landauer’s principle is assumed. Various standard approximations and idealizations commonly employed in both classical and quantum physics make these assumptions and therefore violate decompositional equivalence, Landauer’s principle or both. Alternatives to standard, unitary quantum theory that introduce a non-unitary physical “collapse” process (e.g. [7, 8, 9]), in particular, violate decompositional equivalence if the collapse process is taken to occur at some, but not all possible, boundaries in the state space. It is the ubiquitous presence of ancillary assumptions that violate either decompositional equivalence or Landauer’s principle that render typical presentations of even standard unitary quantum theory paradoxical. When both principles are explicitly and rigorously respected, nothing collapses, nothing decoheres and there are no “multiple worlds.” The paradoxes associated with “collapse” are replaced, in this case, by deep empirical questions about the structure and functioning of observers, human or otherwise.

Demonstrating that decompositional equivalence and Landauer’s principle together imply standard, unitary quantum theory is surprisingly straightforward. The first step is to recognize that if \mathbf{O} can be chosen arbitrarily, \mathbf{O} cannot be assumed to encode any particular information about \mathbf{S} prior to making observations. Therefore the only information about \mathbf{S} that \mathbf{O} can be regarded as encoding is information obtained by observation. Given Landauer’s principle, obtaining information by observation requires physical interaction. Hence \mathbf{O} can obtain information about \mathbf{S} only through physical interaction. The second step is to note that the subsystem of \mathbf{U} with which \mathbf{O} interacts, which will be referred to as $\mathbf{B}_{\mathbf{O}}$ for \mathbf{O} ’s “box,” comprises both \mathbf{S} and \mathbf{E} . Because \mathbf{U} satisfies decompositional equivalence, \mathbf{O} ’s physical interaction with $\mathbf{B}_{\mathbf{O}}$ must be invariant under arbitrary alternative partitionings of $\mathbf{B}_{\mathbf{O}}$. This invariance of the \mathbf{O} - $\mathbf{B}_{\mathbf{O}}$ interaction under arbitrary alternative partitionings of $\mathbf{B}_{\mathbf{O}}$ renders $\mathbf{B}_{\mathbf{O}}$ a black box: if the observable behavior of $\mathbf{B}_{\mathbf{O}}$ is invariant under arbitrary rearrangements of any subsystem boundaries within $\mathbf{B}_{\mathbf{O}}$, then \mathbf{O} ’s observations of this behavior *a fortiori* can yield no information about such subsystem boundaries and

hence no information about the “internal structure” of \mathbf{B}_O . The third step is to show that if Landauer’s Principle is respected, the acquisition of information from a black box must be described by unitary quantum theory. The three sections that follow make this demonstration precise.

3 Information can be obtained only by observation

As before, let \mathbf{S} , \mathbf{E} and \mathbf{O} be physical subsystems that jointly decompose the universe \mathbf{U} . If \mathbf{U} satisfies decompositional equivalence, \mathbf{S} , \mathbf{E} and \mathbf{O} can be defined or designated arbitrarily; the labels ‘ \mathbf{S} ,’ ‘ \mathbf{E} ’ and ‘ \mathbf{O} ’ can, therefore, have no special meaning. Any subsystem whatsoever can be designated as the system of interest; similarly, any subsystem whatsoever can be considered to be an “environment” or an “observer.” For a fixed \mathbf{S} , in particular, the \mathbf{E} - \mathbf{O} boundary can be varied arbitrarily. Any restriction on the kinds of subsystems labelled with these terms, e.g. any restriction on the Hamiltonians \hat{H}_E , \hat{H}_O or \hat{H}_O beyond the requirement that $\hat{H}_U = \hat{H}_S + \hat{H}_E + \hat{H}_O + \hat{H}_{SE} + \hat{H}_{SO} + \hat{H}_{EO} + \hat{H}_{SEO}$, violates decompositional equivalence.

The statement that \mathbf{O} , the “observer” can be chosen arbitrarily naturally raises the question of what “observation” means. Intuitively, to make an observation is to obtain a particular, determinate observational outcome; to make an observation is, in other words, to obtain *classical* information. Landauer’s principle requires classical information to be encoded in a thermodynamically-irreversible way, e.g. recorded on some physical memory. Hence an “observer” must have at least one physical degree of freedom that can change in a thermodynamically-irreversible way. Setting aside the question of how thermodynamic irreversibility is physically implemented until §5 below, this requirement has two implications. First, \mathbf{O} cannot be an “empty” subsystem of \mathbf{U} ; it must contain at least one physical degree of freedom. An observer cannot, in other words, be a merely notional coordinate system as is sometimes assumed. This has the important consequence that *all observers are inside U* and hence that *U cannot be observed* even in principle. There is, therefore, no “observable state of \mathbf{U} ” even in principle. The second consequence is that the physical state of \mathbf{O} must be well-defined *in the context of the S - E - O decomposition* of \mathbf{U} . Note that this does not require that the degree(s) of freedom composing \mathbf{O} must have a well-defined state in the context of any other decomposition of \mathbf{U} , and does not require that either \mathbf{S} or \mathbf{E} have a well-defined state in any decomposition. Hence the separability, in the quantum-theoretic sense, of the unobservable “state” of \mathbf{U} is not required by the designation of some subsystem of \mathbf{U} as an observer.

While it appears trivial from a formal perspective, the arbitrariness with which \mathbf{S} , \mathbf{E} and \mathbf{O} can be chosen has important consequences for how any theory consistent with decompositional equivalence is interpreted and used. The first and most important of these is that observers cannot be assumed to have any special characteristics or to be in any particular special state at the initiation of observations. Observers cannot, in particular, be assumed to have any prior knowledge of the system being observed. If observation is characterized

in Bayesian terms, the probability distribution over the states of the observed system prior to any observations being conducted must be assumed to be uniform.

The assumption that observers have no prior knowledge of the system being observed is often stated in discussions of both classical and quantum physics. Schlosshauer, for example, includes the idea that observers can be “initially completely ignorant” when describing classical observations:

“Here (i.e. in classical physics) we can enlarge our ‘catalog’ of physical properties of the system (and therefore specify its state more completely) by performing an arbitrary number of measurements of identical physical quantities, in any given order. Moreover, many independent observers may carry out such measurements (and agree on the results) without running into any risk of disturbing the state of the system, even though they may have been initially completely ignorant of this state.”

([10] p. 16)

Ollivier, Poulin and Zurek employ similar language to operationally define *objectivity* in a quantum-theoretic context:

“A property of a physical system is *objective* when it is:

1. simultaneously accessible to many observers,
2. who are able to find out what it is without prior knowledge about the system of interest, and
3. who can arrive at a consensus about it without prior agreement.”

[11] p. 1; [12] p. 3

Both of these statements, however, leave unmentioned and unaddressed a fundamental question: how do the many independent observers *identify* the system being jointly observed? If they are *completely ignorant* of the system, having *no prior knowledge* about it, then they cannot know, for example, where it is located, what it looks like, its mass, or its velocity relative to their own. How then can they agree that they have observed the same system? Identifying the system of interest, it would seem, *requires* observers to have *a priori* knowledge of the system to be identified.

The above two passages illustrate two common and often implicit assumptions: 1) that the observations employed to identify the system of interest are unproblematic and do not need to be considered as part of a theoretical description, and 2) that multiple observers can, by observation, identify *exactly the same* system. They also illustrate a common, subtle but important equivocation in the use of the term “system.” When “system” refers to what the observers jointly identify, it refers to something, such as an item of apparatus,

about which the observers are assumed to have considerable *a priori* information. The observers are, in particular, assumed to know where the apparatus is, what it looks like, and so forth. When, on the other hand, “system” refers to what the observers do not have any *a priori* knowledge about, it refers *only* to the collection of degrees of freedom that are being probed by the experiment. This latter usage has been emphasized explicitly by Tegmark (e.g. [13, 14]) in his discussions of decoherence. As all components of the state of the “system” in this latter usage are unknown by definition, this latter system cannot, even in principle, be identified by observation. Hence the above passages also reveal a third common assumption: that identifying the apparatus, or more precisely, the prepared state of the apparatus, is sufficient to render *all and only* the degrees of freedom that the experiment is designed to probe observable. This assumption can, informally, be stated as “preparing the apparatus is preparing the system” where “system” is here used in the second of the above two senses.

As will be shown below, the three assumptions made explicit above all violate decompositional equivalence. Formal descriptions of measurement that stipulate, either explicitly from a theoretical or “god’s eye” [15] perspective or implicitly through the use of notation, that multiple observers have identified and are observing exactly the same system not only beg the question of how this could have been accomplished, but are inconsistent with the quantum formalism, which as noted above enforces decompositional equivalence through both its state space description and the additivity of the Hamiltonian. If the multiple observers can be arbitrarily chosen and cannot be assumed to have any prior knowledge of the system to be observed, all that can be said is that each observer interacted with a system – some system or other – and received one or more observational outcomes. The observational outcome(s) received constitute each observer’s entire knowledge of the state of the system. As noted in §8 below, the question of *how* multiple observers can, in practice, arrive at an agreement that they have observed the same, or at least equivalent systems becomes an empirical question that physics must answer.

4 Observation is interaction with a black box

A well-established, over half-century-old body of theory describes the situation in which an observer can obtain outcomes from an observed system, but can know nothing about the system beyond the outcomes that have been obtained. This is precisely the situation of an observer who interacts with a “black box” as described by classical cybernetics (e.g. [3, 4]) or of an observer who receives signals, via a communication channel, from an unknown source as described by classical information theory [16]. A black box is, by definition, a system with observable overt behavior but an inaccessible, unobservable and hence unknowable interior. Ashby [3] and Moore [4] independently proved that, while observations of the overt behavior of a black box can clearly place a *lower* limit on the complexity of the unknown mechanism inside the box, no finite set of observations of the box’s behavior can place an *upper* limit on its internal complexity. The very next behavior of any black box can

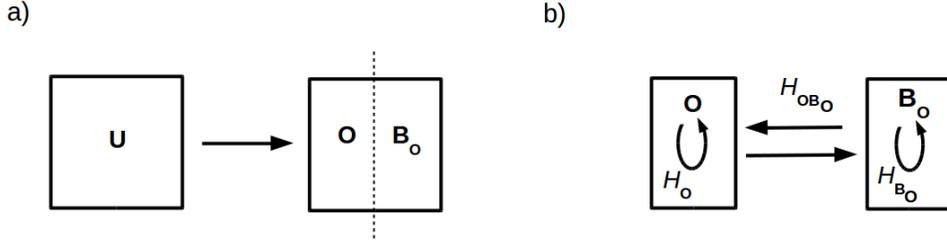


Figure 1: a) A system \mathbf{U} is partitioned into two subsystems \mathbf{O} and $\mathbf{B}_{\mathbf{O}}$. b) The systems \mathbf{O} and $\mathbf{B}_{\mathbf{O}}$ have internal dynamics $\hat{H}_{\mathbf{O}}$ and $\hat{H}_{\mathbf{B}_{\mathbf{O}}}$ respectively and interact via $\hat{H}_{\mathbf{O}\mathbf{B}_{\mathbf{O}}}$.

be a complete surprise, one that indicates the existence of unanticipated internal degrees of freedom and unanticipated internal dynamics.

It is important to note that the term “black box” implies an *epistemic* boundary between the observer and the system observed, not a spatial separation. Hence one can, for example, characterize a stock market as a “black box” for any observer who has access only to the past and current stock prices. It remains a black box even for an observer standing in the middle of its trading floor.

Consider now the behavior of a universe \mathbf{U} that has been arbitrarily partitioned into two subsystems \mathbf{O} and $\mathbf{B}_{\mathbf{O}}$ as shown in Fig. 1. Decompositional equivalence requires that this partitioning has no effect on the dynamics of \mathbf{U} and indeed, no physical consequences whatsoever; in particular, *no physical separation* between \mathbf{O} and $\mathbf{B}_{\mathbf{O}}$ is implied. Let us assume, for the purposes of argument, that the self-interaction of \mathbf{U} can be represented by an operator (notationally anticipating a Hamiltonian) $\hat{H}_{\mathbf{U}}$, making no assumption about the structure of this operator other than that it is additively decomposable. The partition between \mathbf{O} and $\mathbf{B}_{\mathbf{O}}$ then allows this interaction operator to be written as $\hat{H}_{\mathbf{U}} = \hat{H}_{\mathbf{O}} + \hat{H}_{\mathbf{B}_{\mathbf{O}}} + \hat{H}_{\mathbf{O}\mathbf{B}_{\mathbf{O}}}$, where the last term represents the interaction between the partitions.

In this case, the interaction $\hat{H}_{\mathbf{O}\mathbf{B}_{\mathbf{O}}}$ can be regarded as defining a classical information channel between \mathbf{O} and $\mathbf{B}_{\mathbf{O}}$ and \mathbf{O} can be regarded as obtaining observational outcomes by interacting with $\mathbf{B}_{\mathbf{O}}$. The channel defined by $\hat{H}_{\mathbf{O}\mathbf{B}_{\mathbf{O}}}$ is, moreover, the *only* information channel between \mathbf{O} and $\mathbf{B}_{\mathbf{O}}$. In particular, \mathbf{O} has no independent access to the internal dynamics $\hat{H}_{\mathbf{B}_{\mathbf{O}}}$ of $\mathbf{B}_{\mathbf{O}}$. The system $\mathbf{B}_{\mathbf{O}}$ is, under these conditions, a black box from the perspective of \mathbf{O} ; hence the nomenclature, “ \mathbf{O} ’s box” for $\mathbf{B}_{\mathbf{O}}$ introduced earlier.

Note that the labels ‘ \mathbf{O} ’ and ‘ $\mathbf{B}_{\mathbf{O}}$ ’ can be re-interpreted so as to designate $\mathbf{B}_{\mathbf{O}}$ the “observer” and \mathbf{O} the “observed system.” In this case $\mathbf{B}_{\mathbf{O}}$ obtains outcomes from \mathbf{O} via the information channel defined by $\hat{H}_{\mathbf{O}\mathbf{B}_{\mathbf{O}}}$; from the perspective of $\mathbf{B}_{\mathbf{O}}$, \mathbf{O} is a black box.

The system $\mathbf{B}_{\mathbf{O}}$ being a black box from \mathbf{O} ’s perspective has an immediate consequence

of interest for any physical theory constructed by \mathbf{O} : since any further partition of $\mathbf{B}_\mathbf{O}$ is internal to $\mathbf{B}_\mathbf{O}$, \mathbf{O} can obtain no information about any such internal partition. This becomes obvious when the relevant interactions are examined. If $\mathbf{B}_\mathbf{O}$ is further partitioned into \mathbf{S} plus \mathbf{E} , the internal interaction $\hat{H}_{\mathbf{B}_\mathbf{O}}$ can be written as $\hat{H}_{\mathbf{B}_\mathbf{O}} = \hat{H}_\mathbf{S} + \hat{H}_\mathbf{E} + \hat{H}_{\mathbf{SE}}$. As \mathbf{O} has no access to $\hat{H}_{\mathbf{B}_\mathbf{O}}$, \mathbf{O} can have no access to the internal self-interactions $\hat{H}_\mathbf{S}$ and $\hat{H}_\mathbf{E}$ or to the two-way interaction $\hat{H}_{\mathbf{SE}}$. Indeed decompositional equivalence guarantees that the interaction $\hat{H}_{\mathbf{OB}_\mathbf{O}}$ via which \mathbf{O} acquires observational outcomes from $\mathbf{B}_\mathbf{O}$ is entirely independent of arbitrary redefinitions of the \mathbf{S} - \mathbf{E} partition and hence arbitrary redefinitions of the interaction $\hat{H}_{\mathbf{SE}}$. Hence in any universe satisfying decompositional equivalence, *observers cannot “see” system-environment boundaries and observers cannot observationally characterize system-environment interactions.* The split between “system” and “environment” is both *merely notional* and *entirely arbitrary*; it has no effect whatsoever on the observable behavior of $\mathbf{B}_\mathbf{O}$. Observers cannot, in particular, observationally characterize individual states of either system or environment. All observers can do is obtain outcomes, via $\hat{H}_{\mathbf{OB}_\mathbf{O}}$, from $\mathbf{B}_\mathbf{O}$, the black box within which the subsystems \mathbf{S} and \mathbf{E} , however they or their interaction may be defined, are fully contained.

This limitation on the information actually obtainable by observers immediately implies that the three common assumptions about observability made explicit in the last section are illegitimate in any theory respecting both decompositional equivalence and Landauer’s principle. Observers cannot, in particular, be assumed to identify systems unproblematically; indeed as discussed in §8 below, how observers manage to identify systems in practice becomes a compelling empirical question. Multiple observers cannot be assumed to identify exactly the same system, or even to uniformly distinguish each other from the surrounding shared environment. Finally, even the identification of all of $\mathbf{B}_\mathbf{O}$ as a prepared apparatus cannot be assumed to entail any specific consequences for any unobserved degrees of freedom within $\mathbf{B}_\mathbf{O}$; as with any black box, the very next behavior of $\mathbf{B}_\mathbf{O}$ can be a complete surprise that reveals the existence of unanticipated and uncharacterized degrees of freedom. It should be emphasized, moreover, that no specifically quantum-mechanical assumptions have been made in reaching these conclusions; they follow solely from *classical* cybernetics or information theory when decompositional equivalence and Landauer’s principle are respected.

The results of this section also show that the assumptions made in standard environmental-decoherence calculations (e.g. [10, 17]) cannot be given anything other than a pragmatic, *post-hoc* justification. Observers cannot, in particular, either identify the system-environment boundary at which decoherence is taken to occur or establish that the environmental state is sufficiently random that system-environment entanglement results in decoherence. While *post-hoc* justifications of the assumptions required for decoherence calculations may be adequate in practical settings, they cannot support in-principle foundational claims; in particular, they cannot support the common claim that decoherence *explains* the “emergence of classicality” from unitary dynamics [18, 19, 20]. To claim that decoherence explains the emergence of classical system boundaries when such boundaries must be assumed to perform decoherence calculations is to beg the question. To claim

that decoherence explains the emergence of classical system states (“pointer states”) when a classical state of the environment must be assumed to perform the calculation is also to beg the question.

5 How to derive quantum theory

5.1 Overview

The primary claim of this paper is that decompositional equivalence and Landauer’s Principle, taken together, materially imply standard, unitary quantum theory. Quantum theory can be false, therefore, only if one or the other of these two principles is violated. Its secondary claim is that when quantum theory is seen as an inevitable result of decompositional equivalence plus Landauer’s Principle, it ceases to be paradoxical. In particular, the need for a paradoxical “collapse” process or for “multiple worlds” disappears. Indeed, when it is seen as a consequence of these two principles, quantum theory appears to be simple and intuitively compelling. It appears, in other words, as a theory we should *expect* to be true from first principles.

To show that quantum theory follows from decompositional equivalence plus Landauer’s principle, we consider a thought experiment that illustrates the simplest informative interaction between an observer and the world: one that yields a one-bit outcome. By working through this thought experiment, we show that action must be quantized and that the principal axioms (their statements in [17] are taken to be canonical; similar statements can be found in [21] and elsewhere) are implied by decompositional equivalence plus Landauer’s principle. In each case, the formal representation required by the axiom is first shown to arise naturally under the conditions imposed by the thought experiment. The contrapositive statement, that the axiom cannot be false without violating either decompositional equivalence or Landauer’s principle, is then demonstrated. In so doing, we show why a number of common assumptions cannot be made without violating either decompositional equivalence or Landauer’s principle and hence contradicting quantum theory. This sets the stage for understanding, in §6, why quantum theory has the structure that it does.

5.2 Thermodynamics of observation

Consider an observer \mathbf{O} capable of receiving and recording only a one-bit outcome. Suppose, for example, that \mathbf{O} is equipped with a horizontally-oriented meter stick that has been modified to produce a digital “1” signal if but only if it comes into contact with an object that is 1 m wide and to produce a “0” signal if it comes into contact with any object that is not 1 m wide. Suppose, moreover, that this observer is embedded in a world $\mathbf{B}_{\mathbf{O}}$ containing many different objects, some number $n \gg 1$ of which have a horizontal dimension of 1 m while $m \gg 1$ others have other horizontal dimensions. For simplicity, we can treat \mathbf{O} as fixed and unmoving, and imagine that some mechanism external to \mathbf{O} occasionally moves

an object into contact with \mathbf{O} 's meter stick. As before, $\mathbf{B}_{\mathbf{O}}$ is taken to include everything in the universe that is not part of \mathbf{O} . Figure 2 illustrates this situation, indicating by the observer's blindfold that *only* the "1" and "0" outcomes generated by the meter stick can be detected by \mathbf{O} . From \mathbf{O} 's perspective, $\mathbf{B}_{\mathbf{O}}$ is clearly a black box: in particular, \mathbf{O} has no observational access to the mechanism that occasionally brings objects into contact with the meter stick, no means of counting the objects, and no means other than the meter stick of distinguishing them. This situation clearly generalizes to any finite number of measurement devices yielding any finite number of finite-resolution outcomes. It represents, therefore, the situation faced, in practice, by all observers that are restricted to expending finite energy.

As \mathbf{O} explores $\mathbf{B}_{\mathbf{O}}$, the meter stick occasionally produces a "1" signal indicating the presence of a 1 m wide object; it also occasionally produces a "0" signal indicating the presence of something else. Let us now invoke Landauer's principle: recording each of these outcomes requires a thermodynamically-irreversible change in the physical state of \mathbf{O} . Each of these outcome-recording state changes must, therefore, consume at least $0.7kT_{\mathbf{O}}$ of free energy, where k is Boltzmann's constant and $T_{\mathbf{O}} > 0$ is the temperature of \mathbf{O} . For simplicity, let us assume maximal efficiency, so that each state change requires exactly $0.7kT_{\mathbf{O}}$. This free energy must be supplied either by \mathbf{O} itself, e.g. from an on-board battery that powers the generation of signals by the digital meter stick, or supplied by $\mathbf{B}_{\mathbf{O}}$, e.g. by extracting energy from each object the meter stick contacts. In the first case, each bit recorded by \mathbf{O} dissipates $0.7kT_{\mathbf{O}}$ into $\mathbf{B}_{\mathbf{O}}$; in the second, each bit recorded extracts $0.7kT_{\mathbf{O}}$ from $\mathbf{B}_{\mathbf{O}}$. The \mathbf{O} - $\mathbf{B}_{\mathbf{O}}$ interaction, which anticipating its characterization as a Hamiltonian will be written $\hat{H}_{\mathbf{O}\mathbf{B}_{\mathbf{O}}}$, can, therefore, be considered an energy observable from either the perspective of \mathbf{O} or the perspective of $\mathbf{B}_{\mathbf{O}}$; \mathbf{O} can, in particular, be regarded as "counting" energy increments of $0.7kT_{\mathbf{O}}$. There is, however, no "external" perspective from which the energy states of both systems can be simultaneously determined.

Let us now further simplify the situation by assuming that these outcome-recording state changes are the *only* state changes that \mathbf{O} undergoes. In this case, \mathbf{O} 's state changes can be viewed as implementing a *clock* that defines an observer-relative time coordinate $t_{\mathbf{O}}$. Let $\Delta t_{\mathbf{O}}$ be the period of this clock. Since \mathbf{O} 's state changes *only* when this clock "ticks," the period $\Delta t_{\mathbf{O}}$ can be defined as constant without loss of generality.

It is with respect to this internally-defined, observer-relative time coordinate $t_{\mathbf{O}}$ that \mathbf{O} 's state changes can be defined as "irreversible" and hence "classical"; indeed \mathbf{O} 's state changes are irreversible precisely *because* they can be viewed as implementing a clock. From \mathbf{O} 's "internal" perspective, this corresponds to their being both sequential and distinct from each other, both of which are intuitive characteristics of classicality. Note that in this situation, classicality has not "emerged" from the physical dynamics of the \mathbf{O} - $\mathbf{B}_{\mathbf{O}}$ interaction. There is nothing about the dynamics that *imposes* the interpretation of \mathbf{O} 's state changes as counting increments of transferred energy. This is rather a convenient theoretical description of the dynamics from the observer's own, or from a theorist's, perspective that associates bits as units of information transfer with $0.7kT_{\mathbf{O}}$ increments of energy transfer and hence defines time increments as an alternative notation for energy increments.

This intuitive connection between incremented time and classicality can perhaps be sharp-

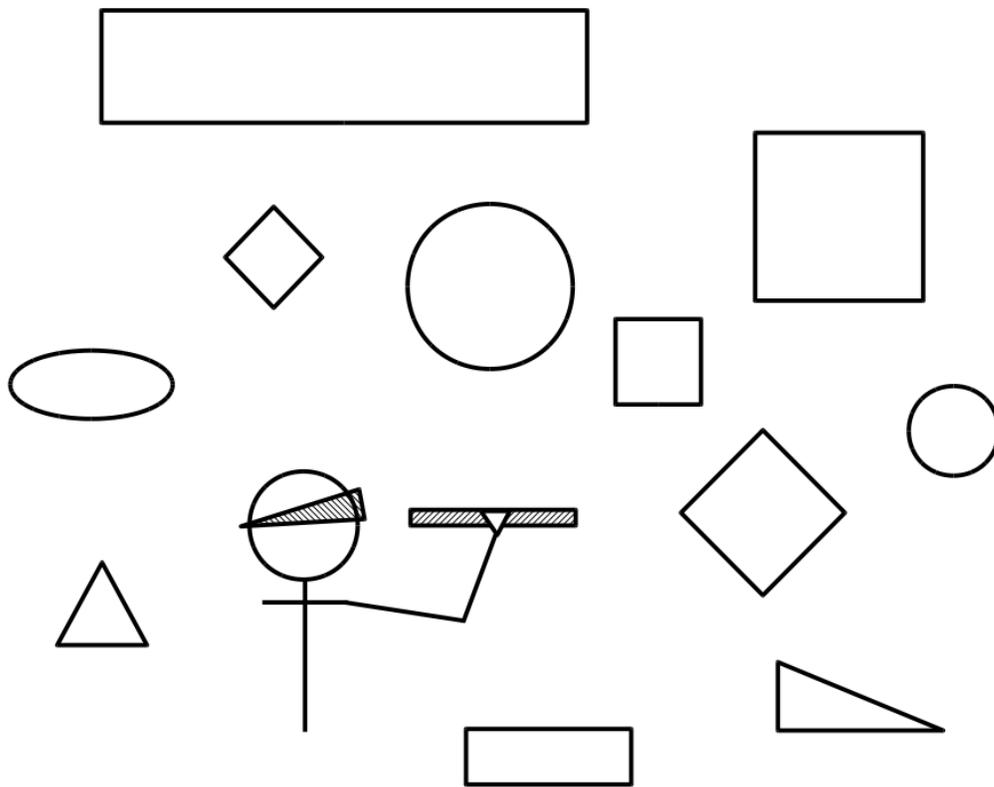


Figure 2: A blindfolded observer equipped only with a horizontally-oriented meter stick, embedded in a world containing multiple objects, some but not all of which have a horizontal dimension of 1 m.

ened by contrasting classical and quantum descriptions of a particle in a one-dimensional box. In the classical case, each collision of the particle with the wall of the box results in a reversal of the momentum of the particle; it can therefore be viewed as a measurement by the particle of the position of the wall. Each such measurement dissipates an increment of kinetic energy from the particle into the wall. The back-and-forth bounces of the particle can, therefore, be viewed as implementing a clock; as the particle loses energy to the wall, the clock slows down and finally stops. In the quantum case, on the other hand, there are no discrete events in which the momentum reverses; the particle can, therefore, be represented as a standing wave. In this case, nothing is measured, no energy is lost and from the particle’s perspective, there is no elapsed time. The t coordinate in the Schrödinger equation is, in this case, not an observable and indeed has no physical meaning.

5.3 Quantization of action

As noted above, the interaction $\hat{H}_{\mathbf{O}\mathbf{B}_\mathbf{O}}$ transfers $0.7 kT_\mathbf{O}$ between \mathbf{O} and $\mathbf{B}_\mathbf{O}$ during the time increment $\Delta t_\mathbf{O}$. The minimum action of $\hat{H}_{\mathbf{O}\mathbf{B}_\mathbf{O}}$ is then:

$$0.7 kT_\mathbf{O}\Delta t_\mathbf{O} = \int_{\Delta t_\mathbf{O}} \hat{H}_{\mathbf{O}\mathbf{B}_\mathbf{O}} dt_\mathbf{O}. \quad (1)$$

As decompositional equivalence allows \mathbf{O} to be chosen arbitrarily, let us assume that \mathbf{O} has been chosen in such a way that $0.7 kT_\mathbf{O}\Delta t_\mathbf{O}$ is minimal across all two-way partitions of \mathbf{U} into an observer and that observer’s world. We can then define $h' = 0.7 kT_\mathbf{O}\Delta t_\mathbf{O}$ as the *minimal action to receive and encode 1 bit* in \mathbf{U} and call it a “quantum” of action.

Taking \mathbf{U} to be our universe and supposing that the energy efficiency of biological photoreceptors has been optimized by evolutionary processes, it seems reasonable to estimate a numerical value of h' by considering such systems. For a molecule \mathbf{m} of rhodopsin at $T_\mathbf{m} = 310$ K (i.e. 37 C, physiological temperature), $\Delta t_\mathbf{m} \sim 200$ fs ([22]; other biological photoreceptors have similar response times). In this case $kT_\mathbf{m} \sim 4.3 \cdot 10^{-21}$ J and the 1-bit information transfer action is $0.7 kT_\mathbf{m}\Delta t_\mathbf{m} \sim 6.0 \cdot 10^{-34}$ J·s, a value remarkably close to that of Planck’s constant $h \sim 6.6 \cdot 10^{-34}$ J·s. We can, therefore, identify h' as h .

Note that this definition of h is entirely thermodynamic, uses no particularly “quantum” concepts and does not appeal to any distinction between “microscopic” and “macroscopic” objects or behaviors. Note also that what is “quantized” by this definition of h is the action of *transferring one bit of classical information*; indeed if no bits are transferred, there is no elapsed time and hence no action. Jennings and Leifer [23] have emphasized in a recent review that a quantum of action can easily be reproduced within a purely-classical model of measurement. One might speculate that had classical information theory been developed before quantum theory, rather than the reverse, not just the quantization of action but much else about the theory might not have been viewed as physically paradoxical.

5.4 Observationally-distinguishable states

Let us now adopt an explicitly theoretical, “god’s eye” perspective on the $\mathbf{O} - \mathbf{B}_\mathbf{O}$ interaction; indeed this perspective is already implicit in Fig. 2. From this perspective, the “ontic” physical state of $\mathbf{B}_\mathbf{O}$ can be described as a spatial configuration of objects, some one of which may, but need not, be in contact with \mathbf{O} ’s meter stick. The mechanism that occasionally moves an object into contact with \mathbf{O} ’s meter stick can be considered to be implemented by a self-interaction $\hat{H}_{\mathbf{B}_\mathbf{O}}$ (again notationally anticipating the Hamiltonian) that acts on these states. As this is the only mechanism that affects the state of \mathbf{O} , we can assume for simplicity that this is the only mechanism implemented by $\hat{H}_{\mathbf{B}_\mathbf{O}}$. Both the states and $\hat{H}_{\mathbf{B}_\mathbf{O}}$ can, from this perspective, be considered classical as further discussed in §5.7 below.

The “god’s eye” perspective is not, however, the perspective of any observer; no measurements can be made from this perspective. What is of interest is how the observable results of the action of the dynamics $\hat{H}_{\mathbf{B}_\mathbf{O}}$ can be described from \mathbf{O} ’s perspective, using only the information obtainable by observation to construct the description. This limited, observer’s perspective, not the theoretical, “god’s eye” perspective, is the one that we, as human observers, in fact experience. A theory inferred from this perspective is the only kind of theory that can be constructed on the basis of observational evidence. As we will see, a *physical* theory inferred from this perspective, i.e. a theory that describes transitions between *observationally-distinguishable states*, has the formal structure of quantum theory.

As the only observational outcomes that \mathbf{O} can obtain from $\mathbf{B}_\mathbf{O}$ are the values “1” and “0,” \mathbf{O} can attribute at most two observationally-distinguishable states to $\mathbf{B}_\mathbf{O}$. Let us call these two states $|1\rangle$ and $|0\rangle$. These states are clearly “observer relative” in the sense introduced by Rovelli [24]; they are defined solely on the basis of the observational outcomes that \mathbf{O} can detect. They are also “relational” in the deeper sense of being states imputed to $\mathbf{B}_\mathbf{O}$ by \mathbf{O} , where both \mathbf{O} and $\mathbf{B}_\mathbf{O}$ are themselves only defined relative to the $\mathbf{O} - \mathbf{B}_\mathbf{O}$ decomposition. By determining the interaction $\hat{H}_{\mathbf{O}\mathbf{B}_\mathbf{O}}$ and thereby determining the outcomes that \mathbf{O} can obtain, the $\mathbf{O} - \mathbf{B}_\mathbf{O}$ decomposition fully determines the observationally-distinguishable states of $\mathbf{B}_\mathbf{O}$. Decomposition itself, therefore, resolves the “choice of basis problem” (e.g. [17]). It is only when the decomposition is underspecified – typically by underspecifying the structure and hence the measurement capabilities of and outcomes obtainable by \mathbf{O} – that the interaction Hamiltonian and hence the interaction basis appears to require a decomposition-independent specification. Once the outcomes obtainable by \mathbf{O} are fixed by a full specification of $\hat{H}_{\mathbf{O}\mathbf{B}_\mathbf{O}}$, any uncertainty about the measurement basis vanishes.

From \mathbf{O} ’s perspective, then, a sequence of observed outcomes ... 1, 1, 0, 1, 0, 0, 0, ... corresponds *by definition* to a sequence of observationally-distinguishable states ... $|1\rangle, |1\rangle, |0\rangle, |1\rangle, |0\rangle, |0\rangle, |0\rangle$... of $\mathbf{B}_\mathbf{O}$. The effect on these states of the self-interaction $\hat{H}_{\mathbf{B}_\mathbf{O}}$ can be represented by a two-component function $\{\hat{E}_i\} = \{\hat{E}_0, \hat{E}_1\}$ where:

$$\hat{E}_0 : |0\rangle, |1\rangle \mapsto |0\rangle; \quad \hat{E}_1 : |0\rangle, |1\rangle \mapsto |1\rangle. \quad (2)$$

The above sequence of outcomes then corresponds to the sequence $\dots \hat{E}_1, \hat{E}_1, \hat{E}_0, \hat{E}_1, \hat{E}_0, \hat{E}_0, \hat{E}_0, \dots$ of operator actions. These two components \hat{E}_0 and \hat{E}_1 of $\{E_i\}$ are clearly orthogonal. Letting $\mathcal{H}_{\mathbf{B}_O}$ be an abstract space of to-be-characterized structure that contains $|0\rangle$ and $|1\rangle$ – we will see below that $|0\rangle$ and $|1\rangle$ are in fact basis vectors of $\mathcal{H}_{\mathbf{B}_O}$ – the components \hat{E}_0 and \hat{E}_1 clearly resolve the identity on $\mathcal{H}_{\mathbf{B}_O}$.

On the physically-reasonable assumption that the abstract state space $\mathcal{H}_{\mathbf{B}_O}$ is measurable, the function $\{E_i\}$ is a positive operator-valued measure (POVM) on $\mathcal{H}_{\mathbf{B}_O}$, as required by the measurement axiom of quantum theory (axiom iii in [17]). Hence the requirement that measurements be representable by POVMs is a consequence of just two intuitively quite natural ideas: that any operation on any space only makes sense as a representation of measurement if the space is measurable, and that the set of outcomes obtainable by any observer is well-defined and, if Landauer’s principle is to be respected, finite (that observations can have only finite numbers of recordable outcomes has been stressed previously, e.g. in [25]). A space fails to be measurable, however, only if it has no σ -algebra of subsets over which a real-valued function that commutes with subset union can be defined (e.g. [26]). Decompositional equivalence requires that subsystem boundaries have no effect on how a system works; it thus requires *any* function that characterizes a system’s overall behavior to commute with subsystem union. A space that fails to be measurable cannot, therefore, represent a system that satisfies decompositional equivalence. Hence if measurements of a system’s behavior cannot be represented by POVMs, the system cannot satisfy decompositional equivalence; decompositional equivalence materially implies, therefore, the representability of measurement by POVMs. The POVM representation and the labelling of observationally distinguishable states by outcome values together immediately imply that an observation o leaves the system in the observationally distinguishable state $|o\rangle$ (axiom iv in [17]) as indeed shown already in [17]. As all such states are observer-relative as noted above, no *physical* “collapse” is implied by this notation; indeed any boundary-dependent physical collapse would violate decompositional equivalence.

5.5 Unitarity

From the theorist’s perspective illustrated in Fig. 2, it is clear that the state designations “ $|0\rangle$ ” and “ $|1\rangle$ ” are highly ambiguous; both refer to multiple spatial configurations of objects that are distinguishable from a “god’s eye” perspective but are indistinguishable by \mathbf{O} . The state transitions represented by \hat{E}_0 and \hat{E}_1 are similarly, from the theorist’s perspective, implemented by multiple distinct changes in the spatial configuration of objects. These operators, therefore, under-represent the unknown mechanism that brings objects into contact with \mathbf{O} ’s meter stick. Note that this is true even if there are, in fact, only two distinct objects in \mathbf{B}_O , as nothing in the specification of \hat{E}_0 and \hat{E}_1 can capture this fact. Let us represent this unknown mechanism, from \mathbf{O} ’s perspective, by an operator $\hat{U}_{\mathbf{B}_O} : \mathcal{H}_{\mathbf{B}_O} \rightarrow \mathcal{H}_{\mathbf{B}_O}$ acting on the state space $\mathcal{H}_{\mathbf{B}_O}$, making at this point no assumption about the structure of $\mathcal{H}_{\mathbf{B}_O}$ other than that it is measurable.

As \mathbf{O} functions as a clock, the time coordinate t_O can be used to parameterize the dynamics

that occur in \mathbf{B}_O and hence to parameterize the action of the unknown $\hat{U}_{\mathbf{B}_O}$. Each tick of this clock and hence the two endpoints of each interval Δt_O correspond to some object being brought into contact, through the action of $\hat{U}_{\mathbf{B}_O}$, with O 's meter stick. Hence $\hat{U}_{\mathbf{B}_O}$ can be regarded, without loss of generality, as a periodic function of t_O with period Δt_O : with respect to some arbitrarily chosen initial value $t_O = 0$, one or the other of $|1\rangle$ or $|0\rangle$ is observed at each subsequent time $t_O = \Delta t_O, t_O = 2\Delta t_O, t_O = 3\Delta t_O$, etc., so one or the other of \hat{E}_0 and \hat{E}_1 must act at each of these times. Landauer's principle guarantees that the period Δt_O is finite, and hence that any such sequence requires finite observer-relative time. Whether \hat{E}_0 or \hat{E}_1 acts so that $|1\rangle$ or $|0\rangle$ is observed after N clock ticks, i.e. at $t_O = N\Delta t_O$ can, therefore, be viewed as determined by the values at $t_O = N\Delta t_O$ of two t_O -dependent, anti-correlated phase angles $\phi_1(t_O)$ and $\phi_2(t_O)$. As $\hat{U}_{\mathbf{B}_O}$ is unobservable, in principle, by O , these phase angles must be unobservable, in principle, by O . They must, therefore, be represented as complex phases, i.e. there are real functions $\varphi_1(t_O)$ and $\varphi_0(t_O)$ such that $\phi_1(t_O) = i\varphi_1(t_O)$ and $\phi_2(t_O) = -i\varphi_0(t_O)$. Hence we can write:

$$\hat{U}_{\mathbf{B}_O}(t_O) = \alpha_0 e^{-i\varphi_0(t_O)} \hat{E}_0 + \alpha_1 e^{-i\varphi_1(t_O)} \hat{E}_1 \quad (3)$$

where α_0 and α_1 are real coefficients chosen so that $\alpha_0^2 + \alpha_1^2 = 1$. As \hat{E}_0 and \hat{E}_1 resolve the identity on $\mathcal{H}_{\mathbf{B}_O}$, this $\hat{U}_{\mathbf{B}_O}(t_O)$ is unitary. The physical dynamics represented by $\hat{U}_{\mathbf{B}_O}$ is, therefore, symmetric in t_O . From the ‘‘god’s eye’’ perspective, this time symmetry corresponds to reversibility of the mechanism that brings objects into contact with the meter stick. From O 's perspective, it corresponds to O 's inability to determine whether a bit string that can only be regarded as random is being read forwards or backwards. The unitarity of $\hat{U}_{\mathbf{B}_O}$ from O 's perspective justifies the representation of the self-interaction $\hat{H}_{\mathbf{B}_O}$, again from O 's perspective, as a Hamiltonian operator.

The unitarity of the state propagator $\hat{U}_{\mathbf{B}_O}$ for any closed, i.e. unobserved system \mathbf{B}_O is a standard axiom of quantum theory (axiom ii of [17]). As before, if this axiom is false, decompositional equivalence must be violated. A failure of unitarity is a failure of time invariance; it indicates that sequences of outcomes ordered in the $+t_O$ direction can be distinguished from sequences ordered in the $-t_O$ direction. As t_O is fully defined by the O - \mathbf{B}_O partition, this asymmetry in t_O can only be a consequence of the definition of this partition. Decompositional equivalence requires, however, that partitions and hence their definitions have no physical consequences; hence such an asymmetry violates decompositional equivalence. Note that the existence of any observer-independent, ‘‘objective’’ time coordinate t in \mathbf{U} requires some boundary (e.g. an ‘‘initial condition of the universe’’) at which to be defined and similarly violates decompositional equivalence. The (in principle unobservable) evolution of \mathbf{U} is, therefore, also time-symmetric for any arbitrarily-chosen, merely parametric coordinate t and is hence unitary. The unitarity of this inferred evolution $\hat{U}_{\mathbf{U}}$ justifies the assumption of an additively-decomposable Hamiltonian self-interaction $\hat{H}_{\mathbf{U}}$ as a representation of ‘‘how the world works’’ and hence justifies the representation of the O - \mathbf{B}_O interaction as a Hamiltonian $\hat{H}_{O\mathbf{B}_O}$ for any arbitrarily chosen decomposition of \mathbf{U} into O and \mathbf{B}_O .

The objective asymmetry of time is a central feature of the classical worldview that is notoriously difficult to reproduce within quantum theory. Objective, i.e. observer-independent

decoherence is often suggested as a source of asymmetric time (e.g. [10, 17]), but as seen above, observer-independent decoherence requires the specification of an observer-independent boundary at which to act and thus violates decompositional equivalence as well as begging the question of classicality. Rovelli [27] has recently shown that an apparently asymmetric time, in any direction, can be obtained by suitably dividing a quantum universe into multiple ($N \gg 2$) subsystems and then adopting a coarse-graining that defines certain variables as effectively classical. This procedure is not only observer-relative but defines asymmetric time relative to a coarsely-observed subsystem; it therefore respects decompositional equivalence.

5.6 The Born rule

From the “god’s eye” perspective, it is clear that the ratio $\alpha_0^2/\alpha_1^2 = m/n$. One would, moreover, expect from this perspective that over a sufficiently long elapsed time, the probabilities P_0 and P_1 of observing $|0\rangle$ and $|1\rangle$, respectively, would be such that $P_0/P_1 = m/n$, i.e. one would expect the Born rule (axiom v in [17]) to hold. Because $\mathbf{B}_\mathbf{O}$ is a black box, however, the numbers m and n cannot be determined by observation, so the equation $\alpha_0^2/\alpha_1^2 = m/n$ cannot be derived by \mathbf{O} . Similarly, the ergotic assumption that the observed frequencies of $|0\rangle$ and $|1\rangle$ between $t_\mathbf{O} = 0$ and $t_\mathbf{O} = N\Delta t_\mathbf{O}$ for some large N correspond to probabilities of future observations cannot be proved valid for any black box, as its validity would contradict the Ashby-Moore theorem noted earlier. The Born rule remains, therefore, a useful *rule* for an observer, but it cannot be considered a theorem. This pragmatic view of the Born rule has previously been advanced by Fuchs [28] as the appropriate view within quantum Bayesianism.

5.7 Hilbert space

As the propagator $\hat{U}_{\mathbf{B}_\mathbf{O}}$ is defined as acting on the state space $\mathcal{H}_{\mathbf{B}_\mathbf{O}}$, \mathbf{O} can attribute a state $|\mathbf{B}_\mathbf{O}\rangle$ to $\mathbf{B}_\mathbf{O}$ between observations by construction. At the time of any observation, i.e. at $t_\mathbf{O} = N\Delta t_\mathbf{O}$ for any $N \geq 0$, this state is such that:

$$\hat{U}_{\mathbf{B}_\mathbf{O}}(t_\mathbf{O}) : |\mathbf{B}_\mathbf{O}(t_\mathbf{O} = N\Delta t_\mathbf{O})\rangle \mapsto |\mathbf{B}_\mathbf{O}(t_\mathbf{O} = (N + 1)\Delta t_\mathbf{O})\rangle. \quad (4)$$

Combining Eq. 3 and 4 to replace operators with the states they produce, we then have:

$$|\mathbf{B}_\mathbf{O}(t_\mathbf{O})\rangle = \alpha_0 e^{-i\varphi_0(t_\mathbf{O})} |0\rangle + \alpha_1 e^{-i\varphi_1(t_\mathbf{O})} |1\rangle \quad (5)$$

at any $t_\mathbf{O}$.

With the characterization of $|\mathbf{B}_\mathbf{O}(t_\mathbf{O})\rangle$ given by Eq. 5, it is clear that the postulated $\mathcal{H}_{\mathbf{B}_\mathbf{O}}$ is a finite dimensional and hence separable Hilbert space, that the observationally-distinguishable states $|0\rangle$ and $|1\rangle$ are basis vectors, and that E_0 and E_1 are von Neumann projections. The generalization from two basis vectors to any finite number is, moreover,

straightforward. As any physically-implemented measurement device has finite resolution – infinite resolution would violate Landauer’s Principle – infinite-dimensional Hilbert spaces are, as previously pointed out by Fuchs [28], merely a convenience for performing calculations.

The requirement that $\mathcal{H}_{\mathbf{B}_O}$ is a separable Hilbert space is the final standard axiom of quantum theory (axiom i in [17]). Let us consider, once again, the physical consequences of a failure of this axiom. First, were the state space of a physical system not separable, it would be impossible to enumerate the basis vectors and hence the physical degrees of freedom associated with any subsystem; the notion of “decomposing” the system into subsystems would then be ill-defined. If the state space lacked an inner product, there would be no well-defined sense in which states could be regarded as similar or dissimilar. A separable vector space is, therefore, required by the most basic intuitions regarding systems and their states. On the other hand, strengthening the axiom by limiting the state space to real-valued coordinates is naturally interpreted as implying the in-principle observability of every state. Unless additional axioms limiting the observability of some states or combinations of states are imposed, the represented system fails to be a black box, thus contradicting either decompositional equivalence or Landauer’s principle. As shown by Leifer [29], “hidden variable” theories consistent with quantum theory must impose such additional axioms, either explicitly or implicitly through the definitions of the hidden variables; the observer’s blindfold in Fig. 2 provides an example of such a constraint. The space for which the *observationally-distinguishable* states form a basis remains a Hilbert space in such theories, as shown here for the “world” of Fig. 2.

5.8 Zurek’s “axiom o” and tensor products

Zurek [17] includes among the axioms of quantum theory an “axiom o” stating that: 1) \mathbf{U} consists of systems and 2) the Hilbert spaces of composite systems are tensor products of the Hilbert spaces of their constituents. The second clause is sometimes (e.g. in [21]) considered an independent axiom. If a Hilbert-space representation of *all* systems has already been assumed, this “axiom” serves to operationally define a “constituent” system; its effect, in this case, is to limit decompositions to finite numbers of constituents, thus avoiding non-separable infinite tensor products.

The first clause of “axiom o” can be read weakly as merely stating that decompositions of $\mathcal{H}_{\mathbf{U}}$ are possible, i.e. that the notion of a “composite system” makes sense. It can also be read as the statement that all possible (finite) decompositions of $\mathcal{H}_{\mathbf{U}}$ objectively exist, i.e. exist independently of their observation by any observer. Any reading of this axiom as postulating the objective existence of only some decompositions and not others, however, violates decompositional equivalence [19]. Similarly, any reading of the axiom as postulating that all observers interact with exactly the same systems violates decompositional equivalence. Either of these readings import into quantum theory a component of the classical worldview, that the universe comprises a collection of well-defined, physically-bounded,

independently-observable *objects*, that the other axioms of quantum theory collectively contradict.

5.9 Summary

What has been shown here is that the primary axioms of quantum theory, and hence quantum theory itself, follow from decompositional equivalence and Landauer’s principle. If quantum theory is false, in particular, one or the other, or both, of these principles must be violated. Both of these principles express deep physical intuitions: the intuition that there is an observer-independent way that the world works and the intuition that effort and therefore energy must be expended to acquire information. We should, therefore, intuitively expect quantum theory to be true. Indeed, we should be utterly astonished should we ever discover that quantum theory is not true. Should we ever discover, for example, that fundamental physical interactions change at the boundaries of macroscopic objects, our deepest intuitions about the world would have to be abandoned.

Landauer’s principle and decompositional equivalence play complementary roles in quantum theory. By imposing a finite energetic cost on information acquisition, Landauer’s principle places a finite temporal interval between every pair of observations. *Every* observation is, therefore, coarse-grained in time as a matter of principle; finite resolution renders all measurements coarse-grained in the measured degree of freedom as well. Coarse-graining in time gives the world an interval between observations in which to change its state. Decompositional equivalence, on the other hand, restricts what can be observed; in particular, it denies observers access to internal boundaries and hence to bounded components of the world. The only “states” that any observer \mathbf{O} sees are states of that observer’s world $\mathbf{B}_{\mathbf{O}}$. These states are distinguished and hence individuated only by observed outcome values, i.e. only by how world affects the observer. Between observations, they can only be represented as superpositions.

By ruling out “god’s eye” perspectives across the board, decompositional equivalence and Landauer’s principle together generalize the restrictions on observational access noted for specific cases in [29]. If the world is regarded as having, like the world depicted in Fig. 2, “ontic” states and behaviors that underlie the state changes observed by \mathbf{O} , these ontic states and behaviors of the world are unobservable in principle and hence unavailable as an empirical basis for theory construction. The only states that are available as an empirical basis for theory construction are the observationally-distinguishable states, i.e. states labelled by distinct observed outcome values. Such observationally-distinguishable states and the discrete transitions between them are fully described by quantum theory.

As decompositional equivalence and Landauer’s principle together imply that any observer’s “world” is a black box, the present results can be summarized as showing that the terms “quantum system” and “black box” are co-extensive. Classical information theory can thus be regarded as a re-discovery of quantum theory from a different starting point and with a different notation. Fuchs’ use of the term “interiority” to characterize the essentially unpredictable nature of quantum systems [28] is therefore highly appropriate.

6 The physical meaning of superposition

6.1 Systems and states

The expressions in Eq. 3 and 5 are clearly superpositions; the operators E_0 and E_1 are superposed in Eq. 3 and the basis vectors $|0\rangle$ and $|1\rangle$ are superposed in Eq. 5. The physicality of superpositions is widely regarded as one of the central mysteries of quantum theory. It is, therefore, worth reflecting on what these expressions mean physically. Doing so, as we will see, casts the standard no-go theorems of quantum theory in a new light, showing them to be the theory’s way of forbidding the interpretation of state-space decompositions as defining “objects” as they are conceived of within the classical worldview.

When the observationally-distinguishable states $|0\rangle$ and $|1\rangle$ are viewed as states of \mathbf{B}_O as a single, whole system, the superposition $|\mathbf{B}_O(t_O)\rangle = \alpha_0 e^{-i\varphi_0(t_O)}|0\rangle + \alpha_1 e^{-i\varphi_1(t_O)}|1\rangle$ (Eq. 5) has the straightforward meaning given in §5.7: the complex phases and hence the superposition itself reflect O ’s inability to observe the inner workings of \mathbf{B}_O and hence inability to predict what \mathbf{B}_O will do next. This is a standard epistemic view of the quantum state: it expresses O ’s objective ignorance of \mathbf{B}_O ’s internal state, ignorance that cannot, even in principle, be resolved by observation.

The system \mathbf{B}_O does not, however, have to be regarded as a single, whole system. Like any system, it can be arbitrarily decomposed. One natural decomposition of \mathbf{B}_O is into two systems \mathbf{S}_0 and \mathbf{S}_1 , where \mathbf{S}_0 is defined to be in state $|0\rangle$ and \mathbf{S}_1 is defined to be in state $|1\rangle$. This decomposition renders \mathbf{B}_O a composite system. It also renders the state of the composite system \mathbf{B}_O unobservable: with this decomposition, O can observe either the state of \mathbf{S}_0 or the state of \mathbf{S}_1 but can never observe their joint state. Letting $\mathbf{S}(t_O)$ be the notional single system occupying the states that O can observe at t_O and replacing states with the systems that occupy them, Eq. 5 can be re-written:

$$\mathbf{S}(t_O) = \alpha_0 e^{-i\varphi_0(t_O)}\mathbf{S}_0 + \alpha_1 e^{-i\varphi_1(t_O)}\mathbf{S}_1. \quad (6)$$

Here the *system* \mathbf{S} is a superposition of the systems \mathbf{S}_0 and \mathbf{S}_1 . This superposition can be given an epistemic interpretation as above: it expresses O ’s objective ignorance regarding the system that will be observed next.

Suppose now that O has an additional observable; for example, suppose O can re-orient the meter stick vertically and receive a “1” or “0” outcome if an encountered object does or does not have a vertical dimension of 1 m. In this case the “system” \mathbf{S}_1 can be in one of two vertical “states”: some 1 m wide objects are also 1 m high, while other are not (Fig. 2). The “system” \mathbf{S}_0 , however, can only be in one vertical state. If, on the other hand, O is supposed to possess a shape observable that distinguishes objects by the numbers of corners they contain, both \mathbf{S}_0 and \mathbf{S}_1 can be in any of the three distinct “states” defined by having zero, three or four corners. Nothing, moreover, prevents O from choosing to define “systems” by numbers of corners and to regard horizontal and vertical dimensions as “states” of those systems. Decompositional equivalence allows any subset of

the available observables to be regarded as defining systems, in which case the remaining subset of observables defines states of those systems.

If observers can obtain information only by observation – i.e., if they occupy a universe satisfying decompositional equivalence – the distinction between “systems” and “states” drawn above is the only such distinction available. Decompositional equivalence explicitly forbids system boundaries, and hence the choices of observables that define them, from having any physical significance. Human observers, for example, tend to use position, size and shape to define systems, e.g. to pick out their favorite chairs or voltmeters from among the other furniture of the laboratory. If the universe satisfies decompositional equivalence, from the perspective of fundamental physics these choices of system-defining degrees of freedom are entirely arbitrary, and there is no principled reason to assume that other observers would make them.

The lack of any physical distinction between observables that define systems and observables that characterize states has an important consequence: the claim that “system \mathbf{S} is in state $|\mathbf{S}\rangle$ ” is simply the claim that some set of outcome values have been observed together. These outcome values characterize the \mathbf{O} - $\mathbf{B}_{\mathbf{O}}$ interaction and hence the state $|\mathbf{B}_{\mathbf{O}}\rangle$, entirely independently of any interpretative choices \mathbf{O} may have made. There is, therefore, no *physical* sense in which $|\mathbf{B}_{\mathbf{O}}\rangle$ is separable; $|\mathbf{B}_{\mathbf{O}}\rangle$ remains a superposition as described by Eq. 5, or its generalization to some larger basis, at all times. Entanglement is not, therefore, an occasional characteristic of some components of $\mathbf{B}_{\mathbf{O}}$ but is rather, as Everett [30] recognized, the permanent status of $\mathbf{B}_{\mathbf{O}}$ itself.

6.2 No-go theorems

The non-objectivity of system boundaries required by decompositional equivalence corresponds, from an observer’s perspective, to an unresolvable ambiguity in the source of any detected outcome. In the world of Fig. 2, for example, \mathbf{O} cannot determine whether a change in observed outcome from “0” to “1” reflects a change in the length of some single object that is in continuous contact with the meter stick or a change in which object has been placed in contact with the meter stick. Similarly, \mathbf{O} cannot determine whether a fixed outcome of “0” or “1” indicates continuous contact with a single object or an exchange of one object for another with the same length. These ambiguities characterize the outcomes obtainable by any observer of any black box, regardless of the number of measurement devices that can be deployed or their resolution, as long as both are finite, i.e. as long as Landauer’s Principle is respected.

The well-known no-go theorems of quantum theory, including Bell’s Theorem, the Kochen-Specker Theorem and the no-cloning theorem, follow from these ambiguities about the provenance of observational outcomes (for details, see [31]). Each of these theorems can be viewed as stating that a particular inference from observations is illegitimate; in doing so, each reveals assumptions that conflict with either Landauer’s principle, decompositional equivalence, or both. The no-cloning theorem blocks the inference that the state of one

system is a “copy” of the state of another system. Two systems can be in the same state, however, only if the systems themselves are copies. Decompositional equivalence prevents the access to system boundaries that would be required to establish this by finite observations. The Kochen-Specker theorem blocks the inference that two distinct measurement procedures, both of which act on a single subset of degrees of freedom of a particular system, will produce the same sequence of outcomes. Such an inference depends on the assumption that the outcomes of the measurements result *only* from interactions with the specified degrees of freedom. As noted earlier, decompositional equivalence prevents such exclusivity from being established by finite observations. Bell’s theorem blocks the inference that two measurements performed on distant parts of a system are independent; it shows, in particular, that both local causation and counterfactually-definite states cannot be assumed. Decompositional equivalence and Landauer’s principle together block the inference that *any* two observations are causally independent. They place the “ontic” causal structure of the world beyond observation, even in principle, and thus enforce a working assumption that this causal structure, whatever it is, is arbitrarily complex. Even if the ontic causal structure of the world is classical as depicted in Fig. 2, the consequences of that structure for an observer can only be described, from the observer’s own perspective, by quantum theory.

When the ubiquitous superpositions in quantum theory are viewed as superpositions of systems, i.e. as indicating an unresolvable ambiguity in the sources of observational outcomes, they cease to be so mysterious. While natural language expresses the conceit that “systems” are fixed and enduring things, we all know that this is not actually the case. The microstructures of all systems are in constant flux. Living systems obviously gain and lose material components, and hence become different *physical systems*, on a daily if not hourly or even moment-to-moment basis; your own breathing provides an example. It is, moreover, just an item of faith, sometimes explicitly propped up by appeals to Occam’s razor, that systems do not undergo more radical alterations or even substitutions of one for another while we are not observing them. The cognitive mechanisms underlying this item of faith are becoming better understood, largely by studying their points of failure [32, 33, 34]. Quantum theory reminds us that this faith-driven inference of object identity through time is unsupportable, in principle, by finite observations.

7 Quantum theory in practice: The double-slit experiment

Quantum theory was originally developed not by considering thought experiments such as that presented here, but by searching for a mathematical formalism that could explain the results of otherwise-mystifying experiments. Once it was understood that light behaved like a particle in the quantum domain, the already century-old double-slit experiment, which when performed by Thomas Young in 1803 had demonstrated that light behaved as a wave, became a mystery. Indeed Feynman, in an often-quoted passage, describes it as “a

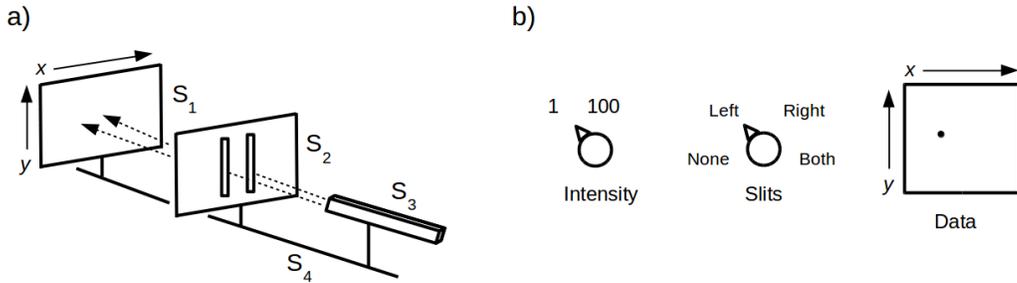


Figure 3: a) Schematic representation of the double-slit apparatus from the perspective of its designer, comprising a detection screen S_1 , slits S_2 , photon or ion source S_3 and stabilizing structure S_4 . Real instruments may be substantially more complex, particularly in the implementation of the slits. The dashed arrows show classical trajectories used to collimate the apparatus. b) The double-slit apparatus from the perspective of an observer, comprising a knob for regulating the source intensity, a knob for determining which slit(s) to open, and a display screen on which detection events are recorded. The x and y coordinates of the display correlate with the x and y coordinates of the detector as shown.

phenomenon which is impossible, *absolutely* impossible, to explain in any classical way, and which has in it the heart of quantum mechanics. In reality, it contains the only mystery” ([35] Vol. III, Sect. 1.1, emphasis in original).

The apparatus for the double-slit experiment is schematically illustrated, from the perspective of its designer, in Fig. 3a. A source emits particles or waves of some kind: photons, electrons, C_{60} “Buckyballs” [36] or even heavier molecules with masses up to 10,000 amu [37]. These pass through a pair of slits, the implementation of which depends on the type of particle/wave passing through them, and are detected at some distance beyond the slits. These components of the apparatus are connected by a rigid structure that also isolates them, in darkness and under high vacuum, from the outside world. Observers employing the double-slit apparatus to make measurements do not, therefore, have observational access to the internals of the instrument. Observers interact, instead, with a “user interface” comprising controls that set the source intensity and the states, either open or closed, of the two slits, together with a display screen that shows the positions, on the $x - y$ plane perpendicular to the instrument’s axis, of any detected particles. The state of the instrument from the observer’s perspective is defined by the states of these interface elements, as shown in Fig. 3b.

The phenomenology of the double-slit experiment is straightforward. If neither slit is open, nothing is detected on the screen. If only one slit is open, particles are detected in an approximately Gaussian distribution, elongated on the y axis and centered on the $x - y$

position of the open slit. If both slits are open, an interference pattern appears along the x axis of the display, with the interference peak and valley intensities indicated on the y axis. These patterns are maintained even if the source intensity is set to just one particle per unit time. If, however, the detection screen is moved to be immediately behind the slits (for photons) or some means of detecting which slit a particle passed through is installed (for massive particles), the interference pattern disappears. The standard interpretation of this phenomenology is that individual particles, even individual large molecules such as the porphyrins employed by [37], act as waves, passing through both slits if both are open and subsequently interfering with themselves. Both this ambiguous wave-particle behavior and the fact that the interference pattern disappears if the particle's trajectory is "observed" by monitoring the slit it passed through are canonical "quantum" phenomena.

Let us now consider this standard interpretation in somewhat more detail. If the source intensity is adjusted so that many particles are emitted per unit time, the resulting "beam" of particles illuminates the plate \mathbf{S}_2 in which the slits are cut. Some fraction of the "beam" passes through the slits and subsequently, the "beam" components that passed through the left slit interfere with the components that passed through the right slit. This situation is analogous to what happens when water waves encounter a breakwater barrier with two openings, as often demonstrated in laboratory exercises for beginning physics students. Hence the "beam" behaves like a classical wave. This classical analogy breaks down, however, when the source intensity is turned down to one particle per unit time. As the interference pattern persists, built up one detection event at a time, it must be assumed either that each particle went through both slits and subsequently interfered with itself, or else that the particles can interfere with each other across time, with the first assumption by far the most commonly chosen. In this case the trajectory of each particle can be thought of as comprising two components. First, the particle travels from the source to both slits; representing the x coordinates of the slits as l and r and ignoring the other coordinates, this first part of the trajectory can be represented as $|l + (r - l)/2\rangle$ (the x coordinate location of the source) $\rightarrow (1/\sqrt{2})(|l\rangle + |r\rangle)$. This first part of the trajectory has, in other words, a fixed starting point but a superposition of end points. The second part of the trajectory can then be represented as $(1/\sqrt{2})(|l\rangle + |r\rangle) \rightarrow |x_d\rangle$, where x_d is the x coordinate of the detection event. Each particle is, therefore, considered to be in a superposition of x coordinate states for its entire trajectory, excepting only its initial location at the source and its final location on the detector.

The x coordinate of each particle is, however, *observed* only at the detector. No aspect of its trajectory is observed other than its detection at S_1 . Indeed, that the particle even has a trajectory is an interpretative assumption based on the designer's, not the observer's, perspective on the apparatus. The discussion in the preceding sections tells us that, from the observer's perspective, the apparatus is a black box, and it warns us that assumptions made from other perspectives can rapidly lead to contradictions. Let us, therefore, set all such assumptions aside and focus exclusively on the perspective of the observer. What the observer can determine by observation is that the number of detection events correlates with the source intensity provided that at least one slit is open, and that the number of

open slits determines the pattern - either Gaussian or interference - that is detected. By appeal to the classical theory of waves, the observer can infer that the slits are far enough from the detection screen for an interference pattern to develop whenever both slits are open. Where, however, is the source? It is consistent with the observed detection patterns that each slit is also a source; if a slit is open, it emits particles at the chosen intensity, but if it is closed it does not. The question: “What is the source of a particle detected at time t ?” has, therefore, a determinate answer if only one slit is open, but if both slits are open, the answer is ambiguous. This ambiguity can be expressed by representing the source as being in a superposition $(1/\sqrt{2})(|l\rangle + |r\rangle)$ of positions along the x axis, or following the reasoning in §6, by representing the source as a superposition of systems $(1/\sqrt{2})(\mathbf{S}_{3l} + \mathbf{S}_{3r})$, where the notations “ \mathbf{S}_{3l} ” and “ \mathbf{S}_{3r} ” refer to left and right sources of particles. These alternative expressions equally and completely capture what an observer can know about the source of an observed particle. No more detail can be given. The question, “Where was the particle before it interacted with the slits” cannot be properly posed.

It is important to note that taking the apparatus apart to allow inspection of the internal components does not resolve the ambiguity. Once the instrument is disassembled, the interference pattern disappears. Indeed any attempt to determine whether the particle has a trajectory prior to encountering the slits disrupts the pattern, just as attempting to determine which slit has been traversed does. The source superpositions $(1/\sqrt{2})(|l\rangle + |r\rangle)$ or $(1/\sqrt{2})(\mathbf{S}_{3l} + \mathbf{S}_{3r})$, and the interference pattern they produce, can therefore be viewed as a consequence - perhaps even a side-effect - of the isolation and consequent unobservability of the internals of the apparatus. They are, in other words, consequences of the apparatus being a black box.

Viewing the double-slit experiment in this way enables a clear prediction: interference patterns should be observable in *any* situation in which the epistemic position of the observer is as represented in Fig. 3b. From this perspective, the steadily-increasing reports of interference phenomena in studies of human categorization that access information from multiple locations in the category network (reviewed by [38, 39]) are not surprising.

8 Conclusion

A scientific theory is a collection of intuitions and assumptions around which a mathematical formalism that enables precise and testable predictions has been constructed. Whether the formalism is fully consistent with the intuitions and assumptions on which it is based, and with the support of which it is used to generate predictions, may not be obvious. Quantum theory provides a dramatic case in point: the formal axioms of quantum theory are notoriously non-intuitive, and the extent to which they are consistent with common physical intuitions and assumptions is not obvious. The unexpected experimental observations that motivated quantum theory, the no-go theorems, and the continuing demonstrations of entanglement and other distinctly “quantum” phenomena all indicate that some, at least, of the intuitions and assumptions that constitute the “classical worldview” are wrong. It

has often been argued that it is the implicit addition of such wrong assumptions that render quantum theory paradoxical (e.g. [40]). The challenge is to determine precisely which assumptions are wrong.

What has been shown here is that two deep intuitions, that there is an observer-independent way the world works and that acquiring information requires spending energy, are consistent with quantum theory, and when formulated more precisely as the principle of decompositional equivalence and Landauer's principle, together materially imply quantum theory. Both decompositional equivalence and Landauer's principle place limits on the information that can be obtained by observation; quantum theory provides a formalism for describing observations of the physical world that enforces these limits. Empirical evidence supporting quantum theory is empirical evidence that these limits are real. It is also evidence that all intuitions or assumptions that conflict with decompositional equivalence or Landauer's principle are wrong. Combining intuitions or assumptions that conflict with either of these principles with quantum theory is guaranteed to lead to paradoxes and inconsistencies. When such intuitions or assumptions are rigorously removed from quantum theory, the paradoxes can be expected to disappear.

Several intuitions or assumptions that play central roles in the classical worldview clearly do conflict with decompositional equivalence, Landauer's principle, or both. These include:

- That observers can interact exclusively with particular, identified subsystems of the world.
- That observational outcomes can be exclusively attributed to particular, identified subsystems of the world.
- That observers can “prepare” two or more particular, identified subsystems of the world independently of each other.
- That multiple observers can reliably identify and exclusively interact with the same subsystem of the world.
- That observers can reliably identify and exclusively interact with the same subsystem of the world at multiple times.

These assumptions all reflect underlying intuitions that *apparent subsystem boundaries are objectively real* and that *upper limits can be placed on the complexity of the world*. These deep intuitions directly contradict decompositional equivalence and Landauer's principle, respectively. They therefore contradict quantum theory.

Because these classical intuitions or assumptions are taken for granted in ordinary life, they appear regularly in foundational discussions of quantum theory. As noted earlier, Zurek's “axiom o” can easily be construed as a statement that apparent subsystem boundaries are objectively real, and in this form it underlies the assumption of an objective system - environment boundary in decoherence calculations. The assumption that observers can

reliably identify and exclusively interact with the same subsystem of the world at multiple times underlies the frequentist conception of probability. The recent Pusey-Barrett-Rudolph theorem (for an extensive review, see [29]) relies on the assumption that spatially-separated systems can be prepared independently. In discussing this assumption, Leifer remarks that “we generally think that experiments on separated systems are independent of one another ... If we allow that genuinely global properties may be relevant even to an isolated system then we open up a Pandoras box” ([29] p. 104). If the global state of the world is non-separable, there are no isolated systems. Pandora’s box is open and cannot be closed. As Galileo put it so long ago, Nature is indifferent to the difficulties that this might cause us.

Both decompositional equivalence and Landauer’s principle are straightforwardly realist physical principles: they are claims about how a real, physical universe works. As shown here, however, their consequences for observers are deeply if not radically instrumentalist. It could be argued that this level of instrumentalism makes science impossible. Perhaps it does make a certain naïve-realist ideal of science impossible, but a pragmatic, everyday science is obviously not just possible but extraordinarily successful. *What makes our pragmatic science possible* becomes, in the present framework, a deep empirical question. How, for example, are we able to identify systems of interest through time, isolate them “well enough” to conduct experiments, and agree among ourselves that we have replicated both the system of interest and the experimental manipulations? How are we able to treat the world as a substantially white box instead of a completely black one? The fact that we can get away, in most ordinary circumstances, with pragmatic assumptions that violate decompositional equivalence and/or Landauer’s principle can only be an outcome of the way the world works, including the way that we happen to be embedded in it. It is an empirical challenge for physics to understand this happy outcome.

Acknowledgements

Thanks to Don Hoffman for encouraging me to think about 1-bit information transfers, and to The Federico and Elvia Faggin Foundation for financial support during the final stages of this work.

References

- [1] Landauer R. Irreversibility and heat generation in the computing process. *IBM J. Research Development* 1961; 5(3):183–195.
- [2] Landauer R. Information is a physical entity. *Physics A* 1999; 263(1-4):63–67.
- [3] Ashby WR. *An Introduction to Cybernetics*. London: Chapman and Hall, 1956.

- [4] Moore EF. Gedankenexperiments on sequential machines. In: *Automata Studies*, Shannon CW, McCarthy J (editors), Princeton, NJ: Princeton University Press, 1956, pp.129–155.
- [5] Nagel E. *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt, Brace & World, 1961.
- [6] Butterfield J. Emergence, reduction and supervenience: A varied landscape. *Found. Phys.* 2011; 41(6):920–959.
- [7] Ghirardi GC, Rimini A, Weber T. Unified dynamics for microscopic and macroscopic systems. *Phys. Rev. D* 1986; 34(2):470–491.
- [8] Penrose R. On gravity’s role in quantum state reduction. *General Relativity and Gravitation* 1996; 28(5):581–600.
- [9] Weinberg S. Collapse of the state vector. *Phys. Rev. A* 2012; 85(6):062116.
- [10] Schlosshauer M. *Decoherence and the Quantum-to-Classical Transition*, Berlin: Springer, 2007.
- [11] Ollivier H, Poulin D, Zurek WH. Objective properties from subjective quantum states: Environment as a witness. *Phys. Rev. Lett.* 2004; 93(22):220401.
- [12] Ollivier H, Poulin D, Zurek WH. Environment as a witness: Selective proliferation of information and emergence of objectivity in a quantum universe. *Phys. Rev. A* 2005; 72(4):042113.
- [13] Tegmark M. Importance of quantum decoherence in brain processes. *Phys. Rev. E* 2000; 61(4):4194–4206.
- [14] Tegmark M. How unitary cosmology generalizes thermodynamics and solves the inflationary entropy problem. *Phys. Rev. D* 2012; 85(12):123517.
- [15] Koenderink J. The all-seeing eye? *Perception* 2014; 43(1):1–6.
- [16] Shannon CE. A mathematical theory of communication. *Bell System Tech. J.* 1948; 27:379–423.
- [17] Zurek WH. Decoherence, einselection, and the quantum origins of the classical. *Rev. Mod. Phys.* 2003; 75(3):715–775.
- [18] Fields C. Classical system boundaries cannot be determined within quantum Darwinism. *Phys. Essays* 2011; 24(4):518–522.
- [19] Fields C. On the Ollivier-Poulin-Zurek definition of objectivity. *Axiomathes* 2014; 24(1):137–156.

- [20] Kastner R. ‘Einselection of pointer observables: The new H-theorem?’ *Stud. Hist. Phil. Mod. Phys.* 2014; 48(Part A):56–58.
- [21] Nielsen MA, Chuang II. *Quantum Information and Quantum Computation*. Cambridge, UK: Cambridge University Press, 2000.
- [22] Wang Q, Schoenlein RW, Peteanu LA, Mathies RA, Shank CV. Vibrationally coherent photochemistry in the femtosecond primary event of vision. *Science* 1994; 266(5184):422–424.
- [23] Jennings D, Leifer M. No return to classical reality. *Contemp. Phys.* 2015; in press (arxiv:1501.03202v2 [quant-ph]).
- [24] Rovelli C. Relational quantum mechanics. *Int. J. Theor. Phys.* 1996; 35(8):1637–1678.
- [25] Peres A., Terno DR. Quantum information and relativity theory. *Rev. Mod. Phys.* 2004; 76:93–123.
- [26] Juch JM. *Foundations of Quantum Mechanics*. Reading, MA: Addison-Wesley, 1968.
- [27] Rovelli C. Why do we remember the past and not the future? The ‘time oriented coarse graining’ hypothesis. Preprint, 2014.
- [28] Fuchs CA. QBism: The perimeter of quantum Bayesianism. Preprint arXiv:1003.5209, 2010.
- [29] Leifer MS. Is the quantum state real? An extended review of ψ -ontology theorems. *Quanta* 2014; 3(1):67–155.
- [30] Everett H III. “Relative state” formulation of quantum mechanics. *Rev. Mod. Phys.* 1957; 29(3):454–462.
- [31] Fields C. Bell’s Theorem from Moore’s Theorem. *Int. J. General Systems* 2013; 42(4):376–385.
- [32] Fields C. The very same thing: Extending the object-token concept to incorporate causal constraints on individual identity. *Adv. Cognitive Psychology* 2012; 8(3):234–247.
- [33] Fields C. The principle of persistence, Leibniz’s Law and the computational task of object identification. *Human Development* 2013; 56(3):147–166.
- [34] Fields C. Motion, identity and the bias toward agency. *Front. Human Neuroscience* 2014; 8:# 597, 2014.
- [35] Feynman RP, Leighton RB, Sands M. *Feynman Lectures on Physics*. New York: Addison-Wesley, 1964.

- [36] Arndt M, Nairz O, Vos-Andreae J, Keller C, van der Zouw G, Zeilinger A. Wave-particle duality of C₆₀ molecules. *Nature* 1999;401, 680–682.
- [37] Eibenberger S, Gerlich S, Arndt M, Mayor M, Tüxen J. Matter-wave interference of particles selected from a molecular library with masses exceeding 10,000 amu. *Phys. Chem. Chem. Phys.* 2013; 15:14696–14700.
- [38] Pothos EM, Busemeyer JM. Can quantum probability provide a new direction for cognitive modeling? *Behav. Brain Sci.* 2013; 36:255–327.
- [39] Aerts D, Gabora L, Sozzo S. Concepts and their dynamics: A quantum-theoretic modeling of human thought. *Topics Cogn. Sci.* 2013; 5:737–772.
- [40] Peres A, Fuchs C. Quantum theory needs no “interpretation.” *Physics Today* 2000; 53(3):70–73.