Unitary quantum theory as a formal framework for a theory of consciousness

Chris Fields

243 West Spain Street, Sonoma, CA 95476, USA

fieldsres@gmail.com

Abstract

If the extraneous and mutually contradictory assumptions of its various "interpretations" are uniformly rejected, unitary quantum theory provides a clear and self-consistent description of the physical world. The continuing empirical success of unitary quantum theory requires, moreover, that we take this description seriously. The physical world as described by unitary quantum theory is globally deterministic, but is provably locally non-deterministic at every scale and from every perspective definable within the theory. It contains no physically-efficacious boundaries, either in "physical" coordinate space or in Hilbert space; hence it contains bounded "objects" or "systems" only as semantic stipulations that may or may not be useful as physical approximations. Measurable space and time are strictly derivative in this world, not fundamental. Observers - generators of classical information, the kind of information that can be recorded using finite, time-persistent symbols - are ubiquitous at all scales, but like all bounded, time-persistent systems exist only as fiat semantic entities. The physical states of the collections of degrees of freedom constituting observers are mutually correlated by quantum entanglement, but the classical information that they encode is context-sensitive in principle; hence quantum "intersubjectivity" is guaranteed, while classical intersubjectivity is at best an approximation. In particular, no classically-characterized observation is strictly repeatable. It is argued that the characteristics of classical information predicted by unitary quantum theory in the absence of extraneous interpretative assumptions are sufficiently similar, *prima facie*, to the characteristics of human phenomenal experience of external objects through either perception or memory to regard quantum theory as a candidate formal theory of phenomenological object-directed awareness. If viewed in this way, quantum theory immediately entails that structural or functional models of the encoding of phenomenal experience by bounded macromolecules, cells, or anatomical systems are at best observer- and context-relative approximations.

Introduction

Since its inception in the 1920s, quantum theory has been regarded by many prominent physicists, and by the vast majority of philosophical commentators on physics, as a "bare" mathematical framework that does not directly describe the physical world. Bohr, for example, is widely quoted as saying:

> "There is no quantum world. There is only an abstract physical description. It is wrong to think that the task of physics is to find out how nature *is*. Physics concerns what we can *say* about nature ..."
>
> (e.g. in Petersen, 1963)

While the first, third and fourth sentences of this statement express an instrumentalist or "non-ontic" view of quantum theory shared by Bohr, Heisenberg, and many current researchers (e.g. Fuchs, 2010; Chiribella, D'Ariano & Perinotti, 2011; Healey, 2012; Friederich, 2013), the second sentence expresses the much more broadly held opinion that unlike classical physics, quantum theory is *only* an abstract mathematical construct without direct, intuitively-clear physical content. As such, it is fundamentally unsatisfactory as a *physical* theory regardless of its predictive success. The motivation behind this broadly held opinion that quantum theory is unsatisfactory as a physical theory is clear. We do not directly experience the world as having the properties, such as superposition and entanglement, that quantum theory ascribes to it; we do not experience cats as being both dead and alive, for example, or chairs as being both here and there. Indeed our ordinary, shared, pre-theoretical human experience of the world is not even "classical" in the sense of being described by classical physics; it rather conforms to an essentially Aristotelian "folk physics" that we develop as young children (McCloskey, 1983; Karmiloff-Smith, 1995). The world as classical physics describes it is, however, at least close to this intuitive world; the world as quantum theory describes it is not close at all.

Both physicists and philosophers have responded to this misalignment between quantum theory and our pre-theoretical physical intuitions by adding assumptions to the theory that make its mathematical predictions more consonant with our everyday experience. The most well-known added assumption is the "collapse postulate" of von Neumann (1932), which interprets measurement as an irreversible, non-unitary physical process that converts a quantum state into a classical state. Since the starting point of this non-unitary process cannot be measured even in principle, von Neumann's postulate is widely regarded as a philosophical or interpretative assumption instead of a physical assumption; indeed it has recently become commonplace to reformulate it as the claim that measurement converts quantum information into classical information, thus dropping the idea of a *physical* "collapse" altogether (e.g. Nielsen & Chaung, 2000). Many physicists, however, still regard "collapse" as a physical process and have developed an enormous diversity of theories to explain how it is implemented (e.g. Ghirardi, Rimini & Weber, 1986; Bohm, Hiley & Kaloyerou, 1987, Hameroff & Penrose, 1996; Stapp, 2001; Weinberg, 2012; Kastner, 2013). Theorists who do not regard "collapse" as a physical process nonetheless tend to regard it as objective, i.e. as fully independent of any characteristics of the observer (e.g. Griffiths, 2002; Zurek, 2009). Even the "many worlds" interpretation of quantum theory generally invokes an objective, classical past shared by multiple observers, and hence implicitly incorporates an objective "collapse" to classicality (e.g. Zeh, 2000; Tegmark, 2010).

The present chapter explores the consequences of an alternative approach to quantum theory: that of taking unitary quantum theory *without any added assumptions* as a literal description of physical reality. There are at least three reasons to contemplate doing this. First, unitary quantum theory is both enormously successful and apparently unlimited in its applicability (e.g. Schlosshauer, 2006); increasingly challenging experiments designed to reveal entanglement over macroscopic spatial and temporal scales invariably have done so (e.g. Bennet *et al*., 2012; Kaiser *et al*., 2012; Peruzzo *et al*., 2012; Christensen *et al*., 2013; Inagaki *et al*., 2013; Saeedi *et al*., 2013). It seems, therefore, at least worth considering as a possibility that quantum theory might be *literally correct* in what it is telling us about nature. Second, the assumptions that are added to quantum theory in order to explain the "emergence of classicality" are invariably classical assumptions. Any "explanation" of classicality derived from quantum theory supplemented by such classical assumptions is clearly circular, and hence is not an explanation at all. To take just one example, all models of physical interactions that take the classical boundaries of ordinary objects to be physically significant violate a deep mathematical symmetry of quantum theory and are therefore either strictly approximations or are inconsistent (e.g. Fields, 2012a; 2012b; 2014). Finally, what "observers" and "observations" are is still not understood in physical terms. Some interpreters of quantum theory – most notably von Neumann – give observers an explicit role in the theory; most do not. However, the alternative of asking quantum theory itself to characterize observation in physical terms has not been rigorously pursued. Perhaps it should be.

In what follows, therefore, I have two main aims. The first is to briefly describe the world in which quantum theory says we live. Bas van Fraassen (1991) famously asked "how could the world possibly be the way a quantum theory says it is?" Perhaps we should examine what quantum theory says, and then ask how it could be otherwise. My second aim is to sketch an account of observation within unitary quantum theory, one that draws on the work of Zanardi (2001) and others on quantum reference frames and on the concept of quantum holography developed by 't Hooft (1993) and Susskind (1994). If observation is modeled in this purely quantum-theoretic way, "observers" are ubiquitous at all scales and "observation" is a way of describing what the universe is doing at all times. It is natural, in this picture, to simply identify "consciousness" or "awareness" with physical dynamics, and in particular with quantum entanglement as a resource driving physical dynamics. I conclude by suggesting that this identification of physicality with awareness might be useful, both for understanding what kinds of systems we human beings are and for appreciating the kind of universe that we live in.

The world according to quantum theory

Let us begin by dropping our preconceived notions of what the world is like, and ask what the world would be like if quantum theory – in particular, unitary quantum theory with no "collapse" – was literally correct. This is the question that Everett (1957) asked, but the "relative state" picture that he proposed as an answer did not have the benefit of the last 50 years of work on quantum information, and has been replaced, both in physics and in the popular imagination, by the "many worlds" view of a universe that "branches" outward from a single past into multiple futures. The many worlds view requires for its coherent formulation not just the classical idea of a bounded object that is persistent through time, but the idea that the "same" bounded, persistent object – e.g. the same individual observer – can exist in and have different outcomes "happen to" it in different "branches" of the universe. It is, therefore, a semi-classical interpretative addition to quantum theory, not a simple statement of quantum theory as is often advertised (e.g. by Tegmark, 2010).

When not burdened with extraneous classical or semi-classical assumptions, quantum theory provides a remarkably clear picture of the physical world. The key components of this picture are briefly discussed below.

Physical dynamics are unitary

As Everett (1957) noted, quantum theory describes the evolution of the universe as unitary, i.e. as a process U such that $U^{\dagger}U = UU^{\dagger} = Id$, where Id is the identity operator on the Hilbert space on which U is defined and $U^{\dagger}$ is the complex conjugate of U. That the dynamic evolution of any closed quantum system is unitary is generally considered to be the second axiom of quantum theory, complementing the first axiom, that the state of any closed quantum system can be represented by a unit vector in a Hilbert space (e.g. Nielsen & Chaung, 2000); unitary evolution of the universe as a whole then follows from the fact that the "universe" is closed by definition. The Schrödinger equation describing the evolution of the physical state $|\mathbf{U}>$ of the universe $\mathbf{U}$, $(\partial/\partial t)\,|\mathbf{U}> = (-i/\hbar)\,H_{\mathbf{U}}\,|\mathbf{U}>$, expresses unitarity. Here $H_{\mathbf{U}}$ is the universal Hamiltonian operator (*cf*. Schlosshauer, 2006); the universal unitary time-propagator is then $U(t) = exp((-i/\hbar)H_{\mathbf{U}}t)$. The "time" parameter t in this expression is purely formal; as discussed below, its identification with the recordable time of human experience is an interpretative assumption.

As noted by physicists from Schrödinger onward, unitary time evolution turns any separable physical state $|\mathbf{U}> = |\mathbf{A}> \otimes |\mathbf{B}>$ into an entangled state $|\mathbf{U}> = |\mathbf{A} \otimes \mathbf{B}>$, where "entangled" simply means "not separable," i.e. not expressible as a mathematical combination of the states of two independent entities **A** and **B**. Since the work of Bell (1964) and Kochen and Specker (1967), entanglement has become increasingly recognized as *the* fundamental characteristic of the world described by quantum theory. A fully-entangled universal state is a quantum superposition of all physical possibilities: all possible combinations of the values of all physical degrees of freedom. A fully-entangled universe contains no separable systems, and hence contains no systems that evolve independently of any other systems: a fully-entangled universe is a universe that evolves through time as a *single entity*. Indeed since the state of a fully-entangled universe is a superposition of all physical possibilities, its "evolution" only alters the relative amplitudes of these possibilities. No "new" possibilities are introduced, nor or any "old" possibilities removed from the mix.

An immediate consequence of unitarity is that any closed quantum system – including **U** as a whole – satisfies the fundamental symmetry of *decompositional equivalence*; the dynamical evolution of any such system is strictly independent of any decomposition of the system into "subsystems" or "parts" (Fields, 2012a, 2012b). Formally, decompositional equivalence requires that all tensor-product structures (TPSs) of a closed quantum system are strictly dynamically equivalent; that is, if $\mathbf{A} \otimes \mathbf{B}$ and $\mathbf{A'} \otimes \mathbf{B'}$ are both TPSs of **U**, then $\mathbf{A} \otimes \mathbf{B} = \mathbf{A'} \otimes \mathbf{B'} = \mathbf{U}$. The equality of alternative TPSs is an axiom of quantum theory, e.g. axiom 4 in the presentation of Nielsen and Chaung (2000). Alternatively, decompositional equivalence is satisfied by any system with a linear Hamiltonian, i.e. any system **U** for which $H_{\mathbf{U}} = \Sigma_{ij} H_{ij}$, where the indices i and j range over all degrees of freedom of **U**. The linearity of the Hamiltonian of any closed quantum system is similarly axiomatic within quantum theory.

Decompositional equivalence imposes a powerful constraint on the physics of any system that satisfies this symmetry: that physics cannot depend in any way on subsystem boundaries. Thus if quantum theory is true of our universe, the boundaries that separate "systems" within our universe – including the boundaries around observers – are *physically irrelevant*. Such boundaries are obviously relevant to any *description* of what is going on, but they are irrelevant to what is going on. This situation can be summarized in a slogan: system boundaries are *semantic not dynamic*. The semantic nature of system boundaries has significant consequences for both the theory of decoherence and the physical description of observers; these are discussed below.

Time is symmetric

Unitary evolution is symmetric in the "time" parameter t; indeed the conjugate operator $U^{\dagger}$ can be viewed as U acting backwards in t. As pointed out by Rovelli (1996) and others, t is not an observable in quantum theory. Recordable "laboratory time" is measured using some other observable, typically position; for example, one measures the position of the hands of a clock, or the position of the illuminated segments of an LED display. Measuring time in this way requires that the "clock" be singled out as an entity, one that, *by assumption*, maintains its identity through time and hence can be repeatedly accessed to make sequential time measurements. This assumption of a time-persistent clock, measurements of which can be recorded in a time-persistent memory, breaks the symmetry of t that is imposed by unitarity.

From a logical perspective, the assumption that a clock that *defines* time maintains its identity over time is clearly circular. More importantly for the present purposes, this assumption violates decompositional equivalence by defining a boundary – the boundary of the clock – that remains inviolate under unitary transformations. To see this clearly, consider the process by which an observer re-identifies the clock measured previously in order to make a second measurement of time. The second measurement only counts as a second measurement if the *very same physical degrees of freedom* are measured the second time around. Measuring the very same physical degrees of freedom that were measured previously requires picking them out from among all the degrees of freedom of the universe. No experimental procedure allowed by quantum theory permits doing this (Fields, 2010; 2011); indeed, any such procedure violates unitarity. As discussed in the more general case below, re-identifications of systems, including clocks, as the *same things* over time is always an approximation. As with system boundaries themselves, such re-identifications are items of classical semantics that are added to quantum theory, not components of quantum theory itself.

Once the boundary in **U** that picks out the degrees of freedom of the designated clock is recognized as a semantic boundary and not a dynamic, and hence not a physical boundary, it becomes clear that observed time is itself an item of semantics that only appears, within the interpretation that defines it, to break the physical symmetry of t. The semantic assumption that defines measured or experienced time is the assumption of *identity*. Without measured or experienced time, the notion that any system maintains its identity over time cannot be defined, while without the assumption that at least one system, the experiencer or the designated clock, maintains its identity over time, experienced or measured time cannot be defined. The assumption of measured or experienced time and the assumption of identity over time are, therefore, effectively a single assumption. This assumption is strictly a semantic addition to quantum theory, and can have no physical consequences in a world correctly described by quantum theory.

Global determinism implies local autonomy

All known microscopic physical interactions – in particular, the Standard Model interactions and gravity – are in isolation fully deterministic. As the Hamiltonian $H_{\mathbf{U}}$ characterizing a quantum theoretic universe is a linear sum of component interactions, it is fully deterministic if the component interactions are. Explicitly taking decompositional equivalence into account reinforces this conclusion, by ruling out any “higher-order interactions” that are non-linear in the fundamental component interactions, and that therefore privilege some TPSs of **U** over others. All evidence supporting unitary quantum theory is, therefore, evidence that the evolution of our universe is fully deterministic.

When combined with universal entanglement, universal determinism yields an interesting and *prima facie* surprising consequence: the behavior of any given degree of freedom is determined by, but is determined *only* by, the behavior of all other degrees of freedom in the universe. No *local* collection of degrees of freedom is sufficient to determine any behavior. A deterministic, quantum theoretic universe thus satisfies a particularly strong form of the Conway-Kochen “free will theorem” (2006): the behavior of a given degree of freedom is not only not determined, even stochastically, by the information available in its past lightcone, it is not determined by the information available in *any* local neighborhood, even any local spacelike neighborhood (Fields, 2013). This result is consistent with a recent demonstration that predicting the results of one's own decision-making process under given circumstances has greater computational complexity than simply making the decision, a result that holds whether the decision making process is deterministic or stochastic, and whether the computational resources available are limited or not (Lloyd, 2012).

In a quantum-theoretic universe, therefore, global determinism implies local autonomy, not only from the perspective of a particular decision-making agent, but from every other local perspective as well. As in the case of system re-identification, local determinism in such a universe is strictly an approximation. These two approximations are strongly coupled. We intuitively expect what we re-identify as “the same system” to exhibit “the same behavior” in response to “the same causes.” This intuitive inference cannot even be formulated within unitary quantum theory without imposing extraneous classical assumptions.

Decoherence is a semantic interpretation, not a physical process

Since its development by Zeh (1970) and Zurek (1982), the theory of decoherence has become the standard approach to explaining the “emergence of classicality” within quantum theory (e.g. Schlosshauer, 2007). Briefly put, decoherence theory provides methods to calculate the observable quantum states of any physical system that interacts with, and is observed from a perspective within, a surrounding environment. As a mathematical formalism, decoherence theory is simply an application of quantum theory to the system-environment interaction. Because decoherence theory predicts that the observable states of any system in interaction with a surrounding environment will appear classical to observers embedded in that environment, it is widely regarded as explaining the apparent classicality of observable states, and hence as solving the notorious “measurement problem” of quantum theory (e.g. Landsman, 2007; Wallace, 2008).

When one examines decoherence calculations closely, however, it becomes immediately clear that the mathematical methods provided by decoherence theory are only applicable in practice when supplemented by classical assumptions. In particular, the interaction between a system and its environment can only be characterized within a TPS that treats the “system” and “environment” as components of **U**. This interaction can only be characterized as a function of time if the physical degrees of freedom composing the system and those composing the environment are re-identifiable as such over time. As noted earlier, any such assumption violates decompositional equivalence and is, therefore, an approximation at best. Decoherence calculations make, in addition, the assumption that the observer obtains no information from the environment; this assumption is often rendered mathematically by assuming that the dynamics of the "environment" can be treated using classical statistical mechanics, i.e. can be treated as classically random. Again as noted earlier, there is no randomness in quantum theory; the assumption of randomness – and the use of classical statistics to characterize it – is an explicitly classical assumption, as is the underlying assumption that the observer is informationally decoupled from the environment. Making either of these assumptions introduces classicality into decoherence theory, and hence renders any "explanation" of classicality by decoherence circular (Fields, 2014).

Stripped of the classical assumptions of a time-persistent system-environment boundary and an informationally-decoupled observer, the mathematics of decoherence describes system-environment entanglement, a non-random process that effectively erases any stipulated system-environment boundary. Entanglement with a surrounding environment does not remove quantum coherence from a system; it rather couples the system and the environment in a way that makes their joint state exhibit quantum coherence. What "removes coherence" in decoherence calculations is the assumption that the observer obtains no information from the environment and hence can treat its state using classical statistics: a classically random state is precisely an incoherent state. What decoherence calculations describe, therefore, is not a *physical* process of coherence removal, but simply the logical consequence of the use of classical statistics to characterize the environmental state. As noted earlier, assumptions of system boundaries are *semantic* assumptions that are overlaid on the dynamics described by quantum theory; the use of classical statistics to describe the environment depends on the assumption of a system-environment boundary and is therefore itself a semantic assumption. The "process" of decoherence is, therefore, a *semantic* process in which a classical system state – and its classical "encoding" by the environment – are inferred from *a priori* classical assumptions.

Summary: The quantum world

The quantum-theoretic picture of the world briefly outlined here is, as noted earlier, utterly unlike the essentially Aristotelian picture of the world that we develop by observing and manipulating objects as infants and young children. Indeed it is difficult to imagine a picture of the physical world that is less like our intuitive picture than is the quantum world. Some 600 years of increasingly precise experiments tell us, however, that while our intuitive picture of the physical world is accurate enough for hunting, agriculture, social life and even substantial feats of architecture, it fails utterly when pushed into new domains. The mechanical devices ushered into human culture by the industrial revolution required the classical physics of Galileo and Newton – as radical in its day as quantum theory is now – for their comprehension; the electronic devices of the mid-to-late 20^{th} century require quantum theory for their comprehension. Every time we send an email or take a digital photograph, we affirm – Bohr's quip not withstanding – that the world we live in is a quantum world. It is time that we understood this world on its own terms.

The fundamental concept that quantum theory introduces is entanglement. The quantum world is an entangled world. It is *not separable* into bounded, time-persistent, re-identifiable entities that can be understood, *ceteris paribus*, independently of one another. Our idea of independent bounded systems that persist through time is, from a physical point of view, strictly an approximation. It is often a good approximation, good enough to enable experiments that employ bounded apparatus that can be re-identified easily from day to day in the laboratory, experiments that have uncertainties better than one part in a billion. What this means, however, is that it is a good approximation for *us*: we *define* the systems that we consider to be bounded in space and time, and these are the systems for which the approximation is good. How does this work: what enables us to define systems that appear, to us, to be bounded? Understanding how this works requires understanding how observers interact with the "systems" that they have defined. This, in turn, requires understanding what "observation" and "observers" are.

This question of what counts, physically, as an observation or an observer is almost always avoided. Zurek (2003), for example, explicitly places this question outside of science, stating that "the observer's mind (that verifies, finds out, etc.) constitutes a primitive notion which is prior to that of scientific reality" (p. 363-364) . Fuchs (2010) asks, somewhat more pointedly, "would one ever imagine that the notion of an agent, the user of the theory, could be derived out of its conceptual apparatus?" (p. 8). Stapp (2001) follows von Neumann (1932) in treating observers as amalgams of physical and "psychical" components, with the latter inexplicable in physical terms. Hameroff and Penrose (1996) develop a theory of observation centered on neuronal microtubules localized within the brains of observers, but offer no physical account of why these particular macroscopic structures and not others should be privileged bearers of quantum coherence in an otherwise-classical world. Most physicists avoid the question altogether, remaining content to regard an "observer" as just another quantum system (e.g. Schlosshauer, 2007) or even, à la Einstein, as a moving coordinate system (e.g. Rovelli, 1996) and an "observation" as any event that transfers classical information to an observer.

It is my contention that understanding the quantum world requires understanding observation as a physical process, and hence understanding *how* semantics are overlaid on dynamics. From a quantum-theoretic perspective, this is the question of how observer-system entanglement implements a mathematical operator – a von Neumann projection or, in more modern terms, a positive operator-valued measure (POVM) – representing a quantum-theoretic observable. From an information-theoretic perspective, it is the question of how classical, symbol-based information processing – classical computation – can be implemented by unitary dynamics. The latter question is central to quantum computing: it is the question of how to define a denotational semantics that allows a quantum *process* to be interpreted as a quantum *computation*, in particular, a quantum computation that accepts a classical question as input and yields a classical answer as output. A number of formal answers to this question now exist, in the form of well-defined semantics for quantum programming languages (e.g. Gay & Mackie, 2010). These formal answers, however, assume an observer. The challenge is to replace this assumed observer with a physical model of observation. The next section further characterizes this challenge and outlines a solution.

Boundaries implement observers

Treating an observer as simply a quantum system with associated coordinates unfortunately provides no insight into how an observer accomplishes the two key tasks that compose observation: the *acquisition* of classical information and its *storage* in a stable, long-term memory. It is successful execution of the second of these tasks that renders an observation *reportable*, and hence usable by either the original or some other observer as a basis for comparison with future observations. As noted earlier in the special case of systems employed as clocks, this ability to compare current and previous observations is essential to the definitions of both measured time – as opposed to symmetric t – and time-persistent individuality. Zurek (2003) takes this memory encoding and re-accessing ability as criterial, remarking that what distinguishes observers from mere apparatus is the ability of the former to "readily consult the content of their memory" (p. 759) . The common picture of an observer as a system – or even as a conscious agent – that deploys a measurement operator to interact with a quantum system (e.g. Fuchs, 2010, Fig. 1) is therefore incomplete. Any observer capable of reporting comparative observations must deploy *two* measurement operators that acquire classical information from two separate quantum systems: the "system of interest" and the observer's own physically-implemented memory. These two systems must be separate – not entangled – if the memory of a previous event is to be independent of the event currently being observed. Again as noted earlier, such separability is always an assumption and always an approximation. It is relatively straightforward to arrange circumstances in which this separability fails, sometimes spectacularly, in the case of human memory (e.g. in change-blindness paradigms; see Simons & Levin (1998) for a striking example).

How can this assumption-laden semi-classical view of observation be reformulated in a way that makes it fully consistent with quantum theory? Two components of the standard view clearly must be dropped: the idea that observer, observed and memory are separable, and the idea that current and previous memory contents can be compared side-by-side. It should be emphasized that these ideas are not incidental, but rather are central to the traditional conception of observation and indeed of experimental science. David Bohm, one of the foremost advocates of "holism" in physics, still insisted that "the very idea of making an observation implies that what is observed is totally distinct from the person observing it" (1989, p. 585). The ability to compare current and previous observations side-by-side, with a non-problematic presumption of independence, is essential to the notion of experimental replication. These ideas are, however, inconsistent with unitary quantum theory and must be rejected – or treated as "for all practical purposes" approximations – if unitary quantum theory is to be taken seriously as a correct description of the world.

Quantum theory replaces the classical idea that observer and observed are separate with its opposite: the idea that they are entangled. Consider the Bell state $\psi = (1/\sqrt{2})(|\downarrow\rangle \otimes |\uparrow\rangle - |\uparrow\rangle \otimes |\downarrow\rangle)$. If an action on this state with a "spin" observable $(\hat{s}_z \otimes Id)$ produces an outcome value of 1, an action with $(Id \otimes \hat{s}_z)$ will, if performed, produce an outcome value of 0; this expectation has now been confirmed with measurements made 300 km apart (Inagaki *et al*., 2013). What entanglement does, in this or any case, is classically correlate conditional information: the classical correlation only appears *if* a particular measurement is made, but if the measurement is made, the correlation is exact. Hence quantum theory replaces the traditional notion of absolute, objective classical correlation with a notion of conditional classical correlation. It also replaces the idea that classical correlation can be viewed as classical information *transfer* from the observed system to the observer with a notion of correlation full stop. Entanglement is not a classical causal process; observer and observed are equally entangled both before and after the observer records the measured outcome value.

The above discussion treats the observer and the observed as "monogamously" entangled, i.e. entangled with each other and with nothing else. It is only under this condition that classical correlation can be exact. Monogamous entanglement, however, requires isolation; in particular, it requires isolation from any surrounding environment. Any *physically efficacious* isolation, however, violates decompositional equivalence. No system is actually isolated. So what can it mean to say that entanglement provides a mechanism that correlates an observer with an observed system?

As shown in detail by Zanardi (2001), whether two components of an entangled state are monogamously entangled – or measurably entangled at all – is not an objective fact, but is rather dependent on both the chosen TPS and the availability of resources, such as time or energy, that enable particular measurements with respect to that TPS. These dependencies can be combined into the single notion of a quantum reference frame, a bounded quantum – and hence physical – system that two observers must jointly access in order to exchange classical information, and hence to communicate observational outcomes (e.g. Bartlett, Rudolph & Spekkens, 2007). Because a quantum reference frame is not an abstraction but a physical system, accessing it requires observation; hence what observers can be thought of as sharing is an observable, a POVM (Fields, 2012b). An "observer" in this context is, moreover, any physical system that is classically correlated with and hence shares classical information with another system. The idea that entanglement creates a classical correlation between an observer and a system being observed depends, therefore, on the assumption that observer and observed share a POVM. As noted above in the case of systems defined to be clocks, however, such an assumption is always a piece of semantics: it is a specific carving-up and naming of the world that is taken to be privileged. It is not something that the *physics* of the situation provides.

"Observers" and "observed systems" having well-defined boundaries in ordinary 3-dimensional space provide an interesting case in point. Here the notion of holographic encoding ('t Hooft, 1993; Susskind, 1994) allows one to think of the surface bounding the observer as encoding a record of every quantum bit ("qubit") received by the observer. A classical black hole can be considered an optimal observer in this framework; the surface "covering" the event horizon encodes one classical bit for every qubit that has fallen into the black hole, and does so at the maximal possible encoding density (Bekenstein, 1981). A system with a larger boundary than a black hole of the same mass encodes information on its boundary at a lower density; such a system is less efficient than a black hole as an accumulator of qubits and encoder of classical information.

An observer external to a black hole cannot share a quantum reference frame with the black hole: any quantum system that falls into the black hole becomes inaccessible to the outside observer, while any system that has not fallen in remains inaccessible to the black hole. Hence an external observer only has access to the *number of bits* encoded by a black hole; such an observer has no access to the POVM employed by the black hole to convert qubits to bits, and hence no access to the semantics assigned by the black hole to the bits it encodes. This situation generalizes to any three-way division of the universe into a spatially-bounded observer, a spatially-bounded system that serves as a quantum reference frame, and a second spatially-bounded observer or system being observed. Neither observer has observational access to the POVM being used by the other observer to determine the state of the shared reference frame; each can access, at most, the classical information encoded by the boundary of the other observer. As above, the observers must *assume* that they share a POVM, and hence assume that they assign the same semantics to the shared quantum reference frame.

The resource that quantum theory provides for establishing classical correlations and hence classical information sharing between systems – the resource of entanglement – thus comes with important strings attached. Entanglement only provides for the sharing of conditional classical information: information of the form "*if* classical information is recorded, it will be correlated across systems." Entanglement provides even this conditional sharing, moreover, only in the context of an assumed quantum reference frame or POVM. This latter assumption is a *semantic* assumption that any two parties sharing classical information must make about each other.

The conditionality and contextuality that quantum theory imposes on classical information, and hence on the concept of observation, makes these a far cry from their traditional usages in classical physics, in which information was assumed to be objectively available for the taking in the environment. They are not, however, surprising when viewed from the perspective of 20th century systems theory, philosophy of language, or even perceptual psychology. The lesson of the 20th century in all of these domains is that shared meaning is an assumption, one that can easily turn out to be wrong. Abandoning the still-classical assumptions about observation embedded in standard interpretations of quantum theory for a "bare" account in terms of entanglement and quantum reference frames catches the physics of measurement up to the realizations made by logicians (e.g. Tarski, 1944), computer scientists (e.g. Turing, 1950), systems theorists (e.g. Ashby, 1956; Moore, 1956), philosophers (e.g. Quine, 1969; Wettstein, 1999), and cognitive scientists (e.g. Newell, 1980) about the relationship between semantics and its physical encodings.

Toward a quantum theory of reportable experience

All reportable contents of experience – all sensory perceptions, memories, imaginations, emotional feelings, and epistemic, aesthetic or moral feelings – are remembered and reported as classical information. The theory sketched above therefore applies to any physical encodings of such reportable experiential contents, including not just encodings in artificial media but also encodings produced by voices or bodily motions, or encodings in biological media such as electrochemical activity patterns in brains. Taking quantum theory seriously as a correct description of the physical world involves taking seriously what it has to say about the origins and dynamics of such encodings.

Quantum theory first specifies the conditions under which encodings of reportable experiences are generated: reportable experiences are encoded whenever a spatially-bounded collection of physical degrees of freedom composing an observer is entangled with a spatially-bounded collection of degrees of freedom composing an observed system. It further specifies that while such entanglement characterizes all spatially-bounded collections of degrees of freedom at all times, it does not have an objective description. How a system is entangled with its environment, and hence what classical information it encodes as reportable experiences, is dependent on the quantum reference frame from which the system-environment interaction is viewed. Such reference frames are not unique. Hence as Dugić and Jeknić-Dugić (2012) put it, "the emergent classical world is not unique"; indeed every observer can be regarded as recording a different "emergent" world. The common assumption that what other observers experience can be assumed to be "like" what we experience is, therefore, in general mistaken. The experienced world of the vast majority of other observers can, instead, be safely assumed to be *nothing like* the world as we experience it. The vast majority of other observers may, for example, experience nothing like what we call "space" or "time."

Quantum theory also specifies how the classical information encoding reportable experiences is encoded: it is encoded on the observer-environment boundary. Different ways of drawing the observer-environment boundary correspond to different quantum reference frames and hence to different encoded experiential contents. How these reference frames are chosen or boundaries are drawn, however, can have no physical consequences; hence what contents are encoded can have no physical consequences. The classical encodings that render experiences reportable are, therefore, physical phenomena but not causal phenomena. Their lack of causality, either within or outside of the boundary defining the observer, is assured by the quantum-theoretic prohibition of local determinism. Indeed the Conway-Kochen (2006) theorem assures that such encodings are not even stochastically deterministic.

The picture that emerges when the implications of quantum theory for encodings of experience are taken seriously is thus both expansive and bracingly counterintuitive. Quantum theory makes observers ubiquitous in all locations and at all scales; it allows all interactions between spatially-bounded systems to be considered “observations.” As Fuchs puts it, quoting William James, an entangled quantum universe “consists of an all-pervasive `pure experience'” (2010; p. 27). This “pure experience” of entanglement is encoded as reportable, classical information at all boundaries between distinct collections of physical degrees of freedom. The universe is full of such boundaries, but from the perspective of physical dynamics they are all both arbitrary and notional. Hence all reportable experience, from a fleeting sensory impression of redness to the memory of an experienced past to the most startling and profound insights, are from this perspective created equal: all are encodings on boundaries drawn between degrees of freedom that are not, in fact, separable. Such boundaries are, moreover, only drawable with respect to a quantum reference frame, the choice of which is always pragmatic and never objectively determined.

The intimate relationship between reportable experience and a choice of quantum reference frame suggests that experimental approaches that characterize the implementation of reportable experience in human beings or other animals also implicitly characterize quantum reference frames. If this is the case, suitable experimental manipulations would be expected to reveal measurable quantum effects at all biologically-relevant scales, from biochemical and macromolecular processes in neurons to cognitive processes in individuals or groups. The growing evidence for quantum effects in human categorization and conceptual reasoning (e.g. Pothos & Busemeyer, 2013) suggests, from this perspective, that activation processes in large-scale neural networks will soon be found to exhibit measurable quantum effects as well.

Conclusion

Quantum theory has traditionally been regarded as unsatisfactory as an explanation of what is going on in the physical world. It has, therefore, been supplemented with a wide variety of interpretative assumptions, all of which introduce aspects of classicality and hence reduce the “quantum-ness” of the theory. While such interpretative assumptions bring the theory closer to the familiar realm of human experience, they render its explanations circular and systematically hide what quantum theory is attempting to tell us about our world.

When quantum theory is stripped of interpretative assumptions, it tells us two things of great significance. First, it tells us that semantics is not given by dynamics. Semantics is, instead, given by *boundaries*, and boundaries have no physical consequences. Aspects of the world that have no physical consequences are often thought of as epiphenomena. Boundaries are not epiphenomena; indeed boundaries are not even phenomena. Boundaries *enable* phenomena by creating and encoding classical information. The arbitrariness and ephemerality of boundaries renders phenomena arbitrary and ephemeral; selecting a different boundary yields a different set of phenomena, one that may be utterly different from the original. Moreover, *any* boundary may be selected, and yield a semantics and a set of phenomena as a result.

The second lesson of quantum theory is that our position as experiencers is not privileged. We are experiencers, but so are all other systems, seen and unseen, that surround us. We are not even particularly complex experiencers; huge numbers of distinct systems contain each one of us, and huge numbers more contain all of us. Moreover, none of our experiences are privileged with respect to any others. All reportable experiences are equally encodings on ephemeral, consequence-free boundaries traversing entangled states. The challenge for an emerging science of consciousness is to develop experimental techniques that reveal how entanglement between definable collections of degrees of freedom produces the experiences encoded on the boundaries that we call “human beings.”

Acknowledgment

Thanks to Revered Prof. P. S. Satsangi for inviting me to give the talk on which this chapter is largely based, and to the HESA Institute for financial support.

References

Ashby, W. R. (1956). *An Introduction to Cybernetics*. London: Chapman & Hall.

Bartlett, S. D., Rudolph, T. & Spekkens, R. W. (2007). Reference frames, superselection rules, and quantum information. *Reviews of Modern Physics* 79, 555-609.

Bekenstein, J. D. (1981). Universal upper bound on the entropy-to-energy ratio for bounded systems. *Physical Review D* 23, 287-298.

Bell, J. S. (1964). On the Einstein-Podolsky-Rosen paradox. *Physics* 1, 195-200.

Bennet, A. J., Evans, D. A., Saunders, D. J. *et al*. (2012). Arbitrary loss-tolerant Einstein-Podolsky-Rosen steering allowing a demonstration over 1 km of optical fiber with no detection loophole. *Physical Review X* 2, 031003.

Bohm, D., Hiley, B. J. & Kaloyerou, P. N. (1987). An ontological basis for the quantum theory. *Physics Reports* 144, 321-375.

Bohm, D. (1989). *Quantum theory*. New York: Dover.

Chiribella, G., D'Ariano, G. M., & Perinotti, P. (2011). Informational derivation of quantum theory. *Physical review A* 84, 012311.

Christensen, B. G., McCusker, K. T., Altepeter, J. *et al*. (2013). Detection-loophole-free test of quantum nonlocality, and applications. Preprint arxiv:1306.5772.

Conway, J., & Kochen, S. (2006). The free will theorem. *Foundations of Physics* 36, 1441-1473.

Dugić, M. & Jeknić-Dugić, J. (2012). Parallel decoherence in composite quantum systems. *Pramana* 79, 199-209.

Everett, H. III (1957). ‘Relative state’ formulation of quantum mechanics. *Review of Modern Physics* 29, 454-462 .

Fields, C. (2010). Quantum Darwinism requires an extra-theoretical assumption of encoding redundancy. *International Journal of Theoretical Physics* 49, 2523-2527.

Fields, C. (2011). Classical system boundaries cannot be determined within quantum Darwinism . *Physics Essays* 24, 518-522.

Fields, C. (2012a). A model-theoretic interpretation of environment-induced superselection. *International Journal of General Systems* 41, 847-859.

Fields, C. (2012b). Implementation of classical communication in a quantum world. *Information* 3, 809-831 .

Fields, C. (2013). A whole box of Pandoras: Systems, boundaries and free will in quantum theory. *Journal of Experimental and Theoretical Artificial Intelligence* 25, 291-302.

Fields, C. (2014). On the Ollivier-Poulin-Zurek definition of objectivity. *Axiomathes* 24, 137-156.

Friederich, S. (2013). In defense of non-ontic accounts of quantum states. *Studies in the History and Philosophy of Modern Physics* 44, 77-92.

Fuchs, C. (2010). QBism: The Perimeter of quantum Bayesianism. Preprint arxiv:1003.5209 [quant-ph].

Gay, S. & Mackie, I. (2010). *Semantic Techniques in Quantum Computation*. Cambridge: Cambridge University Press.

Ghirardi, G. C., Rimini, A. & Weber, T. (1986). Unified dynamics for microscopic and macroscopic systems. *Physical Review D* 34, 470-491.

Griffiths, R. B. (2002). *Consistent Quantum Theory*. New York: Cambridge University Press.

Hameroff, S & Penrose, R (1996). Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness. *Mathematics and Computation in Simulation* 40, 453-480.

Healey, R. (2012). Quantum theory: A pragmatist approach. *British Journal for the Philosophy of Science* 63, 729-771.

Inagaki, T., Matsuda, N., Tadanaga, O., Asobe, M. & Takesue, H. (2013). Entanglement distribution over 300 km of fiber. *Optics Express* 21, 23241-23249.

Kaiser, F., Coudreau, T., Milman, P., Ostrowsky, D. B. & Tanzilli, S. (2012). Entanglement-enabled delayed choice experiment. *Science* 338, 637-640.

Karmiloff-Smith, A. (1995). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, MA: MIT Press.

Kastner, R. (2013). *The Transactional Interpretation of Quantum Mechanics.* Cambridge: Cambridge University Press.

Kochen, S., & Specker, E. P. (1967). The problem of hidden variables in quantum mechanics. *Journal of Mathematics and Mechanics* 17, 59-87.

Landsman, N. P. (2007). Between classical and quantum. In Butterfield, J. & Earman, J. (Eds) *Handbook of the Philosophy of Science: Philosophy of Physics*. Amsterdam: Elsevier (pp. 417-553).

Lloyd, S. (2012). A Turing test for free will. *Philosophical Transactions of the Royal Society A* 370, 3597-3610.

McCloskey, M. (1983). Naïve theories of motion. In Gentner, D. & Stevens, A. L. (Eds) *Mental Models*. Hillsdale, NJ: Erlbaum (pp. 269-305).

Moore, E. F. (1956). Gedankenexperiments on sequential machines. In Shannon, C. W. & McCarthy, J. (Eds) *Automata Studies*. Princeton, NJ: Princeton University Press (pp. 129-155).

Newell, A. (1980). Physical symbol systems. *Cognitive Science* 4, 135-183.

Nielsen, M. A. & Chaung, I. L. (2000). *Quantum Information and Quantum Computation*. Cambridge: Cambridge University Press.

Petersen, A. (1963). The philosophy of Niels Bohr. *Bulletin of the Atomic Scientists* 19(7), 8-14.

Peruzzo, A., Shadbolt, P., Brunner, N., Popescu, S. & O'Brien, J. L. (2012). A quantum delayed-choice experiment. *Science* 338, 634-637.

Pothos, E. M. & Busemeyer, J. R. (2013). Can quantum probability provide a new direction for cognitive modeling? *Behavioral and Brain Sciences* 36, 225-327.

Quine, W. V. O. (1969). Ontological relativity. In *Ontological Relativity and Other Essays*. New York: Columbia University Press (pp. 26-68).

Rovelli, C. (1996). Relational quantum mechanics. *International Journal of Theoretical Physics* 35, 1637-1678.

Saeedi, K., Simmons, S., Salvail, J. Z. *et al*. (2013). Room-temperature quantum bit storage exceeding 39 minutes using ionized donors in Silicon-28. *Science* 342, 830-833.

Schlosshauer, M. (2006). Experimental motivation and empirical consistency of minimal no-collapse quantum mechanics. *Annals of Physics* 321, 112-149.

Schlosshauer, M. (2007). *Decoherence and the Quantum to Classical Transition*. Berlin: Springer.

Simons, D. & Levin, D. (1998). Failure to detect changes to people in a real-world interaction. *Psychonomic Bulletin and Review* 5 , 644-649.

Stapp, H. P (2001). Quantum theory and the role of mind in nature. *Foundations of Physics* 31, 1465-1499.

Susskind, L. (1994). The world as a hologram. *Journal of Mathematical Physics* 36, 6377-6396.

't Hooft, G. (1993). Dimensional reduction in quantum gravity. In Ali, A., Ellis, J. & Randjbar-Daemi, S. (Eds) *Salamfestschrift: A Collection of Talks*. Singapore: World Scientific (arxiv:gr-qc/9310026).

Tarski, A. (1944). The semantic conception of truth and the foundations of semantics. *Philosophy and Phenomenological Research* 4, 341-376.

Tegmark, M. (2010). Many worlds in context. In Saunders, S. Barrett, J. Kent, A. & Wallace, D (Eds) *Many Worlds? Everett, Quantum Theory and Reality*. Oxford: Oxford University Press (pp. 553-581).

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind* 59, 433-460.

van Fraassen, B. (1991). *Quantum Mechanics: An Empiricist View*. Oxford: Clarendon.

von Neumann, J. (1932). *Mathematical Foundations of Quantum Mechanics*. Berlin: Springer.

Wallace, D. (2008). Philosophy of quantum mechanics. In Rickles, D. (Ed) *The Ashgate Companion to Contemporary Philosophy of Physics*. Aldershot: Ashgate (pp. 16-98).

Weinberg, S. (2012) Collapse of the state vector. *Physical Review A* 85, 062116.

Wettstein, H. (1999). A father of the revolution. *Philosophical Perspectives* 13, 443-457.

Zanardi P. (2001). Virtual quantum subsystems. *Physical Review Letters* 87, 077901.

Zeh, D. (1970). On the interpretation of measurement in quantum theory. *Foundations of Physics* 1, 69-76.

Zeh, D. (2000). The problem of conscious observation in quantum mechanical description. *Foundations of Physics Letters* 13, 221-233 .

Zurek, W. H. (1982). Environment-induced superselection rules. *Physical Review D* 26, 1862-1880.

Zurek, W. H. (2003). Decoherence, einselection, and the quantum origins of the classical. *Reviews of Modern Physics* 75, 715-775.

Zurek, W. H. (2009). Quantum Darwinism. *Nature Physics* 5, 181-188.