A whole box of Pandoras: Systems, boundaries and free will in quantum theory[*]

Chris Fields

21 Rue des Lavandières
Caunes Minervois, 11160 France

fieldsres@gmail.com

October 7, 2012

Abstract

It is shown that the paradoxes of quantum theory can be traced to an assumption that observers can know which physical degrees of freedom are causally responsible for each of their experiences. The untenability of this assumption is demonstrated, and a quantum theory without it is proposed. Removing this assumption from quantum theory sheds new light on the nature of semantics, and perhaps also on the nature of freedom.

Keywords: Determinism; Explanation; Causation; Probability; Bayesianism; System identification

1 Introduction

Were one to examine what physicists actually do with quantum theory, one would conclude with considerable justification that quantum theory applies across the board to the stuff that physics is about: matter and energy, including even "dark" matter and "dark" energy. One would, moreover, conclude that quantum theory is about *all of it*: that the domain of the theory is the entire universe (or "multiverse" as some would have it). If one then

---

[*]The delightful phrase "a whole box of Pandoras" is the coinage of Bruce King (1924 - 2009), former governor of New Mexico.

looked at the equations of motion employed in quantum theory - in the non-relativistic case, the familiar Schrödinger equation $(\partial/\partial t)\psi = -(\imath/\hbar)H\psi$, where H is the Hamiltonian operator acting on the quantum state ψ - one would also conclude that quantum theory is fully *deterministic*. The Hamiltonian H represents the instantaneous interaction between physical degrees of freedom - for example, the Coulomb interaction between electric charges or the gravitational interaction between masses - and these interactions are themselves described by equations that are fully deterministic. The Schrödinger equation is, therefore, a deterministic equation; the unitary time-propagator $e^{-(\imath/\hbar)Ht}$ derived from the Schrödinger equation maps an initial state $\psi(t_i)$ to a final state $\psi(t_f)$ deterministically. Many physicists explicitly treat quantum theory as both universal and fully deterministic, and all available experimental evidence is consistent with doing so (e.g. Schlosshauer, 2006).

It is easy, however, to find physicists who assert that quantum theory is not deterministic but rather *probabilistic*. The source of this probabilistic view is the Born rule, the formula for calculating the probabilities of observational outcomes. It states that if a quantum state ψ can be represented as a weighted sum $\sum_i \lambda_i |x_i\rangle$ of mutually-orthogonal, normalized Hilbert-space basis vectors $|x_i\rangle$ that collectively represent all possible outcomes of experiments that measure some observable x, then the probability $P(|x_k\rangle)$ of observing the kth possible outcome in an experiment that assays the value of x is given by $P(|x_k\rangle) = |\lambda_k|^2\langle x_k|x_k\rangle = |\lambda_k|^2$. The Born rule separates the world described by quantum theory into two parts: *what is happening* - the deterministic, time-symmetric dynamics described by the Schrödinger equation and its associated time-propagator - and *what can be observed* - discrete experimental outcomes characterized by Born-rule probabilities. These two parts of the theory describe different things: the deterministic *what is happening* part describes the time-evolution of quantum states, while the probabilistic *what can be observed* part describes experimental outcomes. The Born rule thus encapsulates the ancient divide between realism and phenomenalism - between Platonists and Aristotelians, or in more popular parlance, between advocates of "ontic" versus "non-ontic" views of quantum states - in a single, simple equation.

It is perhaps not surprising that physicists, being mindful of the practical matter of conducting experiments and recording their outcomes, should reject realism about quantum states in favor of an essentially phenomenal, probabilistic outlook on the world. What surely *is* surprising, however, is that many physicists also assert that quantum theory is not only consistent with, but requires a concept of *free will*, specifically a concept of human free will. Quantum information theorist Christopher Fuchs, for example, quotes Hans Primus as claiming that "we assume that the experimenter has a certain freedom of action which is not accounted for by first principles of physics" and that this freedom is a prerequisite of science (2010, p. 18). Nicolas Gisin, one of the first to demonstrate quantum entanglement over kilometer distances, similarly claims that "if humans are not assumed to have the possibility to decide which experiment to perform, and when to perform them, then there is no way to test any scientific theory: this would be the end of Science" (2012, p. 84). Even Niels Bohr, one of the founders of quantum theory, held that "the freedom of experimentation ... is fully retained and corresponds to the free choice of experimental

arrangement for which the mathematical structure of the quantum mechanical formalism offers the appropriate latitude" (1958, p. 71). How can this be? If quantum theory is universal - if it applies to everything across the board - then surely it applies to us. If one opts for phenomenalism, this application of quantum theory describes us probabilistically. Why add free will?

A formal result, the "free will theorem" of Conway and Kochen (2006, 2009), puts the tension between the Schrödinger equation and the assumption that human experimenters have free will into sharp relief, and in the process reveals the deep connection between probability and free will in quantum theory. The free will theorem states that if an experimenter's choice of which observable to deploy in an experiment is not determined by the information available to the observer - in relativistic terms, by the events in the observer's backwards-facing light cone - then the response of the observed system is not determined by the information available to the system. The sense of "not determined" here is local: information is assumed to travel no faster than light. It is also strong: "not determined" means not determined even stochastically, and hence not determined even as a probability distribution. Simply put, if experimenters make *free* choices, not determined even stochastically by the local information available to them, the free will theorem requires that the systems they observe make free choices too, even if these systems are elementary particles. Conway and Kochen are themselves believers in free will, and they state the implication of their theorem explicitly: "fundamental particles are continually making their own decisions" (2006, p. 1463). Hence if free will exists for anyone, "chance" or "probability" are not just phenomenal results of lacking information that could, in principle, be obtained. They are rather objective consequences of an objective, dynamical freedom from any local determination - an objective local *autonomy* - that characterizes all, without exception, of the various constituents of the world.

The free will theorem faces physicists committed to some degree of conceptual coherence with an even starker choice than that forced on them by the Born rule: if they assume free will even just for themselves, they must reject the idea that anything - at least anything for which the relativistic notion of locality makes sense - "obeys" the Schrödinger equation or any other equation. They must reject the idea that even probability distributions are predictable as a matter of fact. What then is science? The "quantum Bayesianism" (or "QBism") of Fuchs (2010, 2012) and colleagues is one answer, and is perhaps the only one. QBism interprets probabilities computed using the Born rule in personalist Bayesian terms and treats all physical systems as *autonomous agents* that employ quantum theory - *universal* quantum theory - as a "users' manual" to support rational decision making. Science, on this view, is basically a coping strategy, one available to all systems across the board. Fuchs asks, "Does this mean even 'elementary' physical events just after the big bang must make use of *concepts* that, to the reductionist mind, must be 15 billion years removed down the evolutionary chain?" and answers immediately, "You bet it does" (2010, p. 21, fn. 37; emphasis in original). QBism treats the Schrödinger equation not as a factual description of reality but as simply a tool, something used to make practical calculations to predict, as well as possible, the likely outcomes of freely-chosen actions.

The entire theory-driven ontology of quantum states is rejected in favor of an apparently nominalist ontology of "this and this and this ... every particular that is and every way of carving up every particular that is" (2010, p. 22) in which "the real world ... is taken for granted" (2010, p. 7). The key feature of this ontology is its intentional naïveté: whatever "particulars" appear phenomenally are taken to be real, full stop. No attempt is made to "explain" the structure of this reality; indeed the idea of "explaining" either the structure or the dynamical behavior of the various objects in the world disappears entirely in a theory that treats autonomy - i.e. freedom - as a fundamental and defining characteristic of such objects.

Fuchs presents QBism as an approach to quantum theory that avoids the classic paradoxes of quantum states: a QBist who describes Schrödinger's cat as $(1/\sqrt{2})(|alive\rangle + |dead\rangle)$, for example, is just saying that the chances, as far as she knows, are 50:50, i.e. the Born-rule probabilities are $(1/\sqrt{2})^2\langle alive|alive\rangle = (1/\sqrt{2})^2\langle dead|dead\rangle = 0.50$, regarding the cat's health. As Fuchs puts it, "quantum states do not exist" (2010, p. 2), so cannot be paradoxical. I have argued elsewhere, however, that QBism becomes paradoxical as soon as one asks how an observer knows that an object observed at time t_2 is the same thing that was observed at some previous time t_1 (Fields, 2012a). The mathematical formalism employed by QBism requires its "particulars" - its objects - to be identifiable as such by observers, but the QBist ontology provides nothing to enable such identification. Fuchs offers no conception of a physical state besides a quantum state, and a quantum state for Fuchs is not an objective thing but a personalist-Bayesian "degree of belief." If observers are unable to determine the physical state of an object - how big it is, for example, how much it weighs or where it is located - how can they establish its continuing identity? How can they recognize it as a persistent thing in a persistent world, something that can be "taken for granted" as a thing to which observational outcomes refer, and to which the quantum formalism can be applied? Removing this paradox without introducing ontic quantum states appears to require dropping the assumption that observations refer to individual, re-identifiable objects in favor of an alternative formalism in which observations refer to the world as a whole (Fields, 2012a). Dropping the assumption that observations refer to individual re-identifiable objects, however, requires dropping the goal of taking the real world for granted. The resulting theory can be overlaid with a Bayesian view of probability, but it is ontologically nothing like QBism.

The present paper takes this argument one step further. It examines the ubiquitous assumption that observers can re-identify physical systems more broadly, and shows that it leads to paradox in realist as well as phenomenalist quantum theory. It then argues that this troublesome assumption does not, in fact, make sense: quantum theory does not allow observers to precisely and unambiguously determine which systems they are interacting with when they make observations and record their outcomes, and real life - specifically, real life in the laboratory - does not allow this either. A reformulation of quantum theory that drops the assumption that observers can precisely identify the systems they observe is outlined. In this reformulated theory, physical states are real, but all observed "systems" are purely semantic: in the language of computer science, they are virtual machines. The dynamics

of the reformulated theory are fully deterministic, but without ontic boundaries separating systems from their environments, these deterministic dynamics are definable only at the scale of the universe as a whole. The semantics of the theory depend on the dynamics, but do so non-locally and hence unpredictably from any perspective definable within the theory. Observers within the theory are therefore, from every definable perspective, effectively free.

2 The assumption of re-identification as a source of paradox

That quantum theory is paradoxical has been recognized since the theory was founded; what has remained unclear for more than eighty years is the *source* of the quantum paradoxes. From a phenomenal perspective, the problem is the quantum state: one only observes cats that are either alive or dead, so the idea of a cat being in a *physical state* $(1/\sqrt{2})(|alive\rangle + |dead\rangle)$ is paradoxical. From the perspective of the Schrödinger equation, on the other hand, the problem is the discreteness and definiteness of experimental observations: if the dynamics by which quantum states evolve is unitary, the probability of any observation yielding a definite outcome in some preselected basis is vanishingly small. The traditional Copenhagen interpretation attempts to split the difference, by proposing both that quantum states undergo a non-unitary "collapse" into the observer's chosen measurement basis when observed, and that quantum theory is not applicable - simply by *fiat* - to the macroscopic "ordinary" world. The steady march of observable quantum phenomena into the macroscopic domain, the increasing application of quantum theory to cosmology, and above all the non-observation of direct evidence for non-unitary dynamics render the Copenhagen view increasingly untenable; Tegmark (1998), for example, estimated that only about a quarter of working physicists - as against more than a third for "none of the above" - still accept Copenhagen. Those of a phenomenalist persuasion who reject Copenhagen in favor of universality have little option but to side with Fuchs in rejecting the ontic reality of quantum states. Realists who reject Copenhagen are faced with a theoretical problem: the problem of explaining how phenomenal classicality can "emerge" from unitary dynamics.

Since its introduction by Dieter Zeh (1970), the primary realist approach to explaining the emergence of classicality has been decoherence theory (Joos et al., 2003; Zurek, 2003; Schlosshauer, 2007). The basic idea of decoherence is that the interactions between a system and its surrounding environment dissipate quantum coherence - the delicate balance between competing possibilities that allows particles to be in two places at once or cats to be both alive and dead - from the system into the environment. Such dissipation mimics collapse; a system in interaction with even as ethereal an environment as the cosmic microwave background radiation is forced into an eigenstate of the system-environment interaction Hamiltonian - a state with a particular value of the system-environment interaction energy - and will remain in such an eigenstate unless something else happens. Physical interactions tend to depend on the spatial separation between the interacting systems; hence energy eigenstates tend to also be at least approximate eigenstates of (relative)

position. This forcing of interacting systems into eigenstates of their interaction - termed environment-induced superselection or "einselection" by Wojciech Zurek (1982) - is the source of classicality in decoherence theory. Zurek and colleagues have shown that multiple observers interacting with a shared environment can, under appropriate observational conditions, extract classical information about the systems embedded in that environment (Ollivier, Poulin & Zurek, 2005; Blume-Kohout & Zurek, 2006; Zurek, 2009). With this model of "quantum Darwinism" acting on environmentally-encoded information, an explanation of the emergence of phenomenal classicality from unitary dynamics appears to be complete. Realist quantum theory with decoherence is, therefore, regarded at least by some as free of paradox.

Hidden beneath the elegant formalism of decoherence theory, however, is an unanswered question: with which quantum systems does "the environment" interact? A quantum system is just a collection of quantum degrees of freedom - a collection of the positions, momenta, masses, spins, isospins, and so on of elementary particles, for example. Any arbitrary slicing up of the quantum degrees of freedom of "everything" into subsets creates quantum systems. Every quantum system is embedded in an environment: the environment consisting of the quantum degrees of freedom of everything else. Any finite collection of quantum systems in which no system is a proper subsystem of any of the others share an environment: the environment consisting of everything other than those systems. If all that is required to render the states of a quantum system effectively classical is interaction with an environment, then the states of every quantum system would be effectively classical. This is not what we observe: what we observe is that *how one looks* at the environment surrounding a collection of quantum systems determines not just whether the systems appear classical, but what systems one sees (Fields, 2012b). An observer who has looked at the environment surrounding a collection of quantum systems and recorded an experimental outcome cannot, moreover, determine by further observations *which* system produced the observed outcome (Fields, 2010), nor can two observers determine by further observations that they have recorded outcomes from the *same* system (Fields, 2011). These are not mere failures of decoherence theory as an explanation: uniquely identifying the system that produced the outcome by further observations would, in either case, require the use of procedures that are explicitly forbidden by quantum theory.

Realist quantum theory supplemented by the assumption of decoherence due to system-environment interactions is, therefore, deeply paradoxical after all: it employs a formalism that implicitly assumes that systems can be fully and explicitly defined by specifying sets of Hilbert-space basis vectors - e.g. by writing expressions such as "let $|\mathbf{S}\rangle = \sum_i \lambda_i |s_i\rangle$" - but forbids observers from uniquely identifying the systems that they have observed. In effect, realist quantum theory provides observers with a formalism, but denies them the observational tools needed to use it. As noted above, phenomenalist quantum theory plays the same shell game. It too employs formal expressions such as "let $|\mathbf{S}\rangle = \sum_i \lambda_i |s_i\rangle$" - such expressions are required to make sense of "$P(|s_k\rangle) = |\lambda_k|^2$" - in its "users' manual," but allows its users no observational means to associate terms such as "$|\mathbf{S}\rangle$" or "$|s_k\rangle$" with something they observe, and no observational means for multiple observers to assure each

other that they are using such terms in the same way (Fields, 2012a). Both formulations of quantum theory, in other words, assume that observers have information about *which* degrees of freedom compose the quantum systems that they observe, but both formulations prevent them from obtaining it. This paradox cannot be escaped by changing one's perspective: escaping it requires changing an assumption within quantum theory itself.

3 Interlude: Does the assumption of known system boundaries make sense?

The most immediate and common-sensible response to the paradox outlined above is to look for a defect or a loophole in the quantum-theoretic restrictions on observation. It seems just obvious, after all, that observers can know what they are looking at, and if they can know, there must be some way for them to find out. If there were any doubts about this, would the repeated experimental successes of quantum theory, to say nothing of the experimental successes of more ordinary sciences like biology or geology, not assuage them? How could observers generate and test predictions if they were unable, for example, to repeatably identify their experimental apparatus, and repeatably assure that their apparatus was interacting only with the quantum system of interest? Is this not what "controlled experiments" are all about?

Suppose we define a system $\mathbf{S}$ by explicitly specifying some finite collection of degrees of freedom $\{\xi_i\}$; the specified ξ_i then compose the Hilbert space $\mathcal{H}_{\mathbf{S}}$ of $\mathbf{S}$. An observation of $\mathbf{S}$ is represented, within quantum theory, by the action of a positive operator-valued measure (POVM; Nielsen & Chaung, 2000, Ch. 2), a set $\{E_i\}$ of positive semi-definite operators defined over $\mathcal{H}_{\mathbf{S}}$ that sum to the Identity; such an action is a traditional projective measurement as defined by von Neumann (1932) in the special case in which the component operators E_i are all mutually orthogonal. An observational outcome is a POVM component E_k that is associated with a real outcome value α_k; to say that an observation yields an outcome α_k is to say that the action of $\{E_i\}$ on $\mathbf{S}$ is followed by the selection by $\mathbf{S}$ of E_k and hence of α_k. Fuchs (2010, Fig. 1) emphasizes that measurements require active interventions into the world by representing a POVM as an experimenter's prosthetic hand; he emphasizes that the selection of a POVM component E_k is the outcome of such an action by identifying the selected E_k as the "experience" of the observer. Let us provisionally adopt this terminology: the "experience" of the observer is just *what happens to* the observer following the deployment of a POVM.

The representation of an observation by the action of a POVM is not the only representation available within quantum theory. If we assume that observers are implemented by collections of physical degrees of freedom - or alternatively and more generally, assume that all information must be physically encoded (Landauer, 1999) - then the observer-system interaction can also be represented as a straightforward physical interaction, i.e. by a Hamiltonian $H_{\mathbf{OS}}$, where $\mathbf{O}$ represents the degrees of freedom of the observer. If the selection of a particular POVM element E_k as the outcome of an observation at time t can be

regarded as the "experience" of the observer at t, then the action of the Hamiltonian $H_{\mathbf{OS}}$ at t can also be regarded as the "experience" of the observer at t. To say this is to say that the experience of the observer, however we choose to define it, is caused *only* by the observed system $\mathbf{S}$, i.e. only by the behavior of the degrees of freedom ξ_i. This amounts to the claim that the combined system $\mathbf{S} \otimes \mathbf{O}$ is isolated, i.e. free of interactions with any other degrees of freedom that might contribute causally to the observer's experience. *Isolated* composite systems, however, are by definition free from environmental decoherence; they obey the Schrödinger equation, their components are mutually entangled, and their joint states $|(\mathbf{S} \otimes \mathbf{O})\rangle$ can at any time be arbitrary superpositions in any basis. Hence once again we have a paradox: the paradox of the von Neumann chain that never terminates (von Neumann, 1932; Schlosshauer, 2007), the observational outcome that is never definite, and hence never recordable as classical information. If both system and observer are exposed to environmental decoherence, quantum information (i.e. coherence) cannot flow between them and no observation is made; if they are both isolated from the environment, quantum information does flow between them, but neither observer nor observed system can appear effectively classical - or even *visible* - to any third-party observer.

The Copenhagen interpretation avoids the paradox of a continuing von Neumann chain by appealing to irreversible physical collapse - and hence violation of the Schrödinger equation - somewhere within or near the observer; this preserves the physicality of information at the expense of the universality of quantum theory. Believers in free will might be expected to avoid the paradox by rejecting the physicality of information, and hence the representation of observation by a Hamiltonian. If such a view is followed through, however, it soon becomes clear that the action of a POVM is not a physical process either, and the question of why "observation" should require any physical events at all becomes unanswerable. Fuchs, at least, does not embrace this option, but rather states explicitly that "For a Quantum Bayesian, the only physical process in a quantum measurement is what was previously seen as 'the selection step' - i.e. the agent's action on the external world and its unpredictable consequence for her" (Fuchs & Schack, 2011, p. 12). It is this insistence on the physicality of information exchange that, when combined with Fuchs' identification of quantum states with information, renders his thesis that quantum states are not physical states of physical systems essentially unintelligible (Fields, 2012a).

Fortunately, another option for escaping the paradox of a never-ending von Neumann chain is available: one can reject the claim that the observer's experience is exclusively caused by the observed system. Before dwelling on the formal manipulations required to implement this option, it is worth briefly examining it in practice. Why is it that experiments are conducted in multiple replicates? If experimenters were assured that only the system of interest was causally responsible for the observed outcome, would not one instance of each experiment be enough? The fact that one instance is *never* considered enough suggests a hypothesis: that any given outcome is caused in part by the system of interest, and in part by unknown other stuff. In actual practice, no observer can know exactly which degrees of freedom, out of all the degrees of freedom in the universe, will cause her next experience - no observer can know exactly what her next experience will be, let alone how it will be

caused. The hope in any repeated experiment is that over sufficiently many trials, the causal contributions of the "other stuff" will average to zero. This hope remains a *hope* because whether they in fact average to zero cannot be tested without assuming some behavior on the part of the system, i.e. without assuming the very outcome that the observations are designed to obtain. It also remains a hope for a more subtle reason. Each instance of observation, once performed, lies in the past, forever inaccessible to further characterization. To call any later instance of observation a "replicate" of the first is, therefore, simply to express the hope that the degrees of freedom that in fact contributed to the outcome of the first instance - *whatever they were* - contributed to the outcome again.

4   Towards a solution: Universal POVMs and 'whatever'

Suppose an observer is provided with a POVM $\{E_i^{\mathbf{S}}\}$ defined over $\mathcal{H}_{\mathbf{S}}$ and another POVM $\{E_j^{\mathbf{E}}\}$ defined over $\mathcal{H}_{\mathbf{E}}$, where "$\mathbf{E}$" for *environment* represents any "other stuff" that might contribute to an experienced observational outcome. The sum $\{E_i^{\mathbf{S}}\}+\{E_j^{\mathbf{E}}\}$ is then a POVM defined over all degrees of freedom that could, in principle, contribute to the outcome. Conventional quantum theory enforces the assumption that systems have well-defined boundaries by setting each of the $E_j^{\mathbf{E}}$ equal to the Identity; suppose, however, that this simplifying assumption is dropped. Suppose the observer deploys the combined POVM $\{E_i^{\mathbf{S}}\} + \{E_j^{\mathbf{E}}\}$ and regards as "$\mathbf{S}$" any collection of degrees of freedom for which a detectable outcome is produced. Such an observer clearly does not know *which* degrees of freedom within the composite system $\mathbf{S} \otimes \mathbf{E}$ caused any particular outcome, and hence any particular "experience" of $\mathbf{S}$. However, if outcomes are observed relatively rarely, and if the information contained in any particular outcome is rich and detailed, the observer might suppose that there was at least some overlap between the collections of degrees of freedom causing different experiences of $\mathbf{S}$. The observer might, in particular, be able to use the information obtained by deploying $\{E_i^{\mathbf{S}}\} + \{E_j^{\mathbf{E}}\}$ on multiple occasions to construct a mathematical model that enabled accurate predictions of the outcomes of deploying $\{E_i^{\mathbf{S}}\} + \{E_j^{\mathbf{E}}\}$ in the future. Were this the case, the observer might treat the $\mathbf{S}$ described by her predictive model as a *thing*, even if its degrees of freedom could not be circumscribed within a well-defined boundary. This assumption of "thing-ness" would be hypothetical, as it would be entirely dependent on the continued predictive success of the observer's constructed model. It would, in other words, be an assumption of thing-ness of the sort that we make when we regard model-derived entities such as electrons or quarks as "things" deserving a place in our ontology.

Adding such an $\mathbf{E}$ comprising "other stuff" that might affect observations begins to address the question of why observers run replicate experiments, but unless $\mathbf{E}$ can itself be identified and shown to actually contain *all* of the relevant degrees of freedom it accomplishes nothing. What assures, in any particular case, that a proposed $\mathbf{E}$ is large enough? If $\mathbf{E}$ comprises events in the backward light cone, for example, how far back in time must this cone extend

to capture all potentially-relevant causal influences? How, moreover, does the observer re-identify $\mathbf{E}$ from replicate to replicate? If quantum theory is universal, these questions are unanswerable: $\mathbf{E}$ must include *all* other stuff, the entire rest of the universe. If $\mathbf{E}$ includes the entire rest of the universe, the identification problem does not arise: in this case the observer just has to deploy $\{E_i^{\mathbf{S}}\} + \{E_j^{\mathbf{E}}\}$, and it automatically samples $\mathbf{S} \otimes \mathbf{E}$. As pointed out by Everett (1957) but still often unappreciated, such an extension of $\mathbf{E}$ to include the entire rest of the universe in fact follows from the assumption that quantum theory is universal and hence that the evolution of the universe itself is unitary. The Bell (1964) and Kochen-Specker (1967) theorems both formally demonstrate that the information necessary to predict outcomes of experiments conducted on a local collection of degrees of freedom cannot be extracted from local observations of those degrees of freedom. It is, therefore, not possible to establish by observations that any local $\mathbf{E}$ is sufficient to include all degrees of freedom relevant to the outcome of observations of any system $\mathbf{S}$ for which $\mathbf{E}$ is the surrounding environment. In particular, any degree of freedom of $\mathbf{S}$ might be entangled with degrees of freedom outside of the backward light cone of $\mathbf{S}$ and hence outside any local $\mathbf{E}$; hence events that are unobservable in principle from $\mathbf{E}$ might influence the outcomes of observations of $\mathbf{S}$.

I have argued previously that all finite observers are in fact equipped with POVMs defined over the entire universe, that they are equipped only with such POVMs, and that they employ them to identify as particular "systems" - for example, particular items of laboratory apparatus - *whatever* collections of degrees of freedom in fact produce detectable outcomes (Fields, 2012b). For such an observer, a "quantum state" $|\mathbf{S}\rangle$ at some time t is not an element of some well-defined $\mathcal{H}_{\mathbf{S}}$, but rather the inverse image, at t, of whatever components of $\{E_i^{\mathbf{S}}\} + \{E_j^{\mathbf{E}}\}$ have measures that are detectable at the observer's observational resolution. Quantum states defined in this way have the properties of conventional quantum states; they exhibit unclonability, enforce non-commutativity of operators with overlapping inverse images, and behave as would be expected on the basis of Bell's theorem and the Kochen-Specker theorem. Indeed, quantum states defined this way *are* conventional quantum states of the universe $\mathbf{U}$ as a whole. The difference between this formulation of quantum theory and conventional formulations is, however, profound: in this formulation, "systems" have no boundaries. Systems smaller than the universe as a whole - including observers - are in this formulation instantaneous grouping of degrees of freedom that happen, at t, to be mapped to a detectable outcome by a POVM. Outcomes of POVMs are classical information: hence a POVM can be viewed as a semantic mapping from a Hilbert space - i.e. from physics - to a collection of classical symbols of the form $(\mathbf{S}, \alpha_k)$, where '$\mathbf{S}$' names a "system" and 'α_k' names an outcome value. A sequence of POVM outcomes ordered by discrete observation times can be thought of as a sequence of states of a classical finite state machine. Hence in the current formulation, "systems" are virtual machines implemented by $\mathbf{U}$ (Fields, 2012b, 2012c). Observers are systems, so observers are also virtual machines implemented by $\mathbf{U}$. Observers do not *have* information in this formulation: observers *are* information, information encoded by an unknown, and possibly ever-changing, collection of physical degrees of freedom.

Viewing systems and observers in these information-theoretic terms removes the standard paradoxes of quantum theory, not by resolving them but by preventing their formulation. The notion of "collapse" as a non-unitary physical process, for example, evaporates; an observation is a semantic relationship between physical dynamics and classical information, not a particular kind of physical event. The von Neumann chain terminates not because entanglement is somehow reconciled with decoherence, but because POVMs map physical dynamics not to more physical dynamics, but to items of classical information that those dynamics encode. Probabilities, in this formulation, reflect the inescapable ignorance of observers regarding system boundaries. To see that these probabilities satisfy the Born rule, consider an observer who deploys a POVM $\{E_i^{\mathbf{S}}\} + \{E_j^{\mathbf{E}}\}$ to observe a "system" $\mathbf{S}$. Because $\sum_{ij}(\{E_i^{\mathbf{S}}\} + \{E_j^{\mathbf{E}}\}) = 1$, the probabilities $P(\alpha_k)$ of the possible outcome values α_k sum to one and are hence well defined. In each instance of observation, the apparent "system" state $|\mathbf{S}\rangle$ is just some physical state $|\mathbf{U}\rangle$ of $\mathbf{U}$; a notional expansion $\sum_k \lambda_k |s_k\rangle$ of $|\mathbf{S}\rangle$ thus corresponds to a real expansion $\sum_k \lambda_k |u_k\rangle$ of $|\mathbf{U}\rangle$. The Born rule statement $P(\alpha_k) = |\lambda_k|^2$ for such a notional representation of $|\mathbf{S}\rangle$ is, therefore, just the statement that the Born rule applies in $\mathbf{U}$.

In place of the standard paradoxes, a boundary-free formulation of quantum theory substitutes a fundamental question: the question of the origin of semantics. In a boundary-free formulation of quantum theory, observation is a mapping from physical dynamics to classical information. If the assumption that all information is physically encoded is to be maintained, the classical information produced by this mapping must also be encoded by physical dynamics. Hence physical dynamics must encode physical dynamics: it must stand in a semantic relationship to itself. What is it about physical dynamics - about the interactions between physical degrees of freedom - that not only enables but produces this semantic relationship, and hence not only enables but produces observers? How is it that physical systems themselves overlay classical information onto the world? This question is not new: it has been with us, in this form, since the birth of computer science, and with us in the form of philosophical questions about the origin of "meaning" for millennia. It has, however, routinely been dodged by characterizations of observers as operating from outside of the theory. Fuchs exemplifies this external perspective when he asks, somewhat derisively, "Would one ever imagine that the notion of an agent, the user of the theory, could be derived out of its conceptual apparatus?" (2010, p. 8). David Bohm takes the external perspective to be a logical requirement: "the very idea of making an observation implies that what is observed is totally distinct from the person observing it" (1989, p. 585). Treating observers as virtual machines that are implemented, as all "systems" are implemented, by collections of physical degrees of freedom brings them *inside* the theory. To insist that quantum theory be rendered conceptually coherent is to insist that this question of how physics produces semantics be confronted for what it is: a demand to explain from the inside how physics produces observations, how Hamiltonian dynamics and entangled states implement POVMs and produce outcomes stored in classical memories. To ask this question is to ask how "what happens" to an observer can plausibly be viewed within the theory itself as an "experience" in the normal meaning of that word.

5   Semantic freedom

I cannot answer the question of why there is experience - real experience - in the physical world, and do not know if the question is answerable. The formulation of quantum theory discussed here does, however, suggest that even the most deterministic physics allows a particular kind of freedom, freedom of semantic interpretation. An observer considered to be outside a theory has complete semantic freedom: an observer considered to be outside a theory can choose any interpretation of the theory's domain, and hence any model theory desired. If information must be physically encoded, however, the observer's representations of the chosen model theory are physical encodings of classical information. If the observer is considered within the theory, these physical encodings are physical states that are produced by the dynamics specified by the theory; if the specified dynamics are deterministic, these encodings have a deterministic source. Hence semantic freedom might be expected to disappear altogether in a deterministic theory. Our question is: does it disappear in a deterministic but boundaryless theory?

In the current formulation of quantum theory, an observer is a collection of physical degrees of freedom that can be regarded as encoding time-dependent classical information, the observer's "experiences" $e(t)$. The behavior of these physical degrees of freedom is fixed by the Schrödinger equation as it applies to the universe $\mathbf{U}$ as a whole, $(\partial/\partial t)|\mathbf{U}\rangle = -(\imath/\hbar)H_\mathbf{U}|\mathbf{U}\rangle$. From this alone, it is clear that the observer's experience at some time $t + \Delta t$ is not predictable from $e(t)$, or indeed from any sequence of previous experiences, even if the mapping from the degrees of freedom composing the observer to the observer's experience at t or at any sequence of previous times is known: the observer's encodings of classical information are local, but the determination of the observer's state at $t + \Delta t$, and hence the determination of *what* information is locally encoded at $t + \Delta t$, is not. In a deterministic universe without boundaries, semantic mappings from subsystems of $\mathbf{U}$ to classical information are all nonlocal: they all depend via unitary dynamics on the current state $|\mathbf{U}\rangle$, and $|\mathbf{U}\rangle$ cannot be determined even from complete knowledge of the state of one of its subsystems. Hence if the $e(t)$ correspond to observations of some external system $\mathbf{S}$, i.e. if they are outcomes of some POVM that recognizes $\mathbf{S}$, $e(t + \Delta t)$ cannot be predicted from the physical state $|\mathbf{S}\rangle$ at t, or indeed from any sequence of previous physical states recognized as "$\mathbf{S}$" by the relevant POVM. Just as they are in Fuchs' world of bounded but autonomous agents, all observational outcomes in a boundaryless deterministic world are indeterminate in principle from any within-theory perspective.

A POVM deployed by an observer is a semantic mapping from physics to experienced outcomes; from the above, it follows that an observer cannot know at t what POVM she will deploy at $t+\Delta t$, nor can any other observer determine what POVM she will deploy, even in principle. From the observer's own perspective, in other words, she is fully autonomous - not even *stochastically* determined - with regard to the choice of which POVM to deploy at $t + \Delta t$, and she is autonomous in this choice from every other perspective as well. It is worth noting, however, that such strong autonomy is the very antithesis of "being in control" in its usual, informal sense. If the behavior of an observer is not determined by

any local degrees of freedom, it is clearly not determined by just the local degrees of freedom composing the observer. An observer may be able to *predict* what POVM she will deploy next on the basis of a model devised by observing her own behavior, but she cannot even stochastically determine what she will do next; what she will do next is only determined by her *plus the entire rest of the universe.* Whether the observer *feels* autonomous, or *feels* in control at $t + \Delta t$ is similarly unpredictable from every definable perspective; feeling autonomous is a (possible) part of the experience at $t + \Delta t$, so its occurrence is dependent on the unknowable, nonlocal semantic mapping that will be instantiated by $|\mathbf{U}\rangle$ at $t + \Delta t$. It is, indeed, consistent with the current theory that an observer could *feel* completely under the control of some other local system - i.e. feel utterly *non*autonomous - when performing an observation, even though her performance of the observation would in fact be fully autonomous from every definable perspective, including her own.

Nothing special has been assumed about the nature or complexity of the "observer" in this discussion; the observer is a virtual machine implemented by a collection of physical degrees of freedom. If any one collection of degrees of freedom implements a "real" observer - i.e. if outcomes are "really" experienced - nothing prevents all collections of degrees of freedom from implementing "real" observers. The present theory is, therefore, strongly panpsychist, or at least "panobserverist." All collections of degrees of freedom smaller than $\mathbf{U}$ as a whole are, moreover, autonomous in the same way and to the same extent: the behavior of any collection of degrees of freedom is determined by the unitary dynamics of $\mathbf{U}$ as a whole, and hence is not determined, even stochastically, by the dynamics of anything smaller than $\mathbf{U}$. Hence in the present boundaryless theory, semantic freedom not only exists, it satisfies the free will theorem.

6   Conclusion

In conventional quantum theory, the relationship between determinism and chance - between the Schrödinger equation and the Born rule - is a well-known generator of paradoxes. Adding free will to the mix makes these paradoxes more intense by emphasizing that "chance" in quantum theory involves a deeper kind of indeterminacy that it did in classical physics, an indeterminacy so deep that it is fairly called autonomy. It has been argued here that the quantum paradoxes can be traced to an unspoken but ubiquitous assumption: the assumption of well-defined systems, i.e. the assumption of knowable boundaries in Hilbert space. This assumption can be removed by treating quantum states as time-dependent inverse images of detectable outcomes produced by POVMs acting on the universe as a whole; if this is done, a "system" is not a fixed thing, but is rather whatever a given POVM detects at a given time. Removing system boundaries makes the "emergence of classicality" entirely a semantic affair; all systems, including observers, become semantic entities, virtual machines. The determinism of the Schrödinger equation assures that these virtual machines exhibit a particular kind of freedom: semantic freedom, the freedom of an open experiential future. Such a future may or may not feel free, and it is decidedly not under one's own control.

Is semantic freedom real freedom, the freedom to actually change the world in accord with one's desires? From the present perspective, even to ask this question is to presume knowledge - knowledge of one's own boundaries, and of one's own past - that it is impossible, in principle, to obtain.

Acknowledgment

References

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

Blume-Kohout, R. & Zurek, W. H. (2006). Quantum Darwinism: Entanglement, branches, and the emergent classicality of redundantly stored quantum information. *Physical Review A* 71, 052105 [22 pp]. arXiv:quant-ph/0405161v2

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

Bohr, N. (1958). *Atomic Physics and Human Knowledge.* New York: Wiley.

Conway, J. & Kochen, S. (2006). The free will theorem. *Foundations of Physics* 36, 1441-1473. arXiv:quant-ph/0604079v1

Conway, J. & Kochen, S. (2009). The strong free will theorem. *Notices of the AMS* 56, 226-232. arXiv:0807.3286v1 [quant-ph]

Fields, C. (2010). Quantum Darwinism requires an extra-theoretical assumption of encoding redundancy. *International Journal of Theoretical Physics* 49, 2523-2527. arXiv:1003.5136v2 [quant-ph]

Fields, C. (2011). Classical system boundaries cannot be determined within quantum Darwinism. *Physics Essays* 24, 518-522. arXiv:1008.0283v4 [quant-ph]

Fields, C. (2012a). Autonomy all the way down: systems and dynamics in quantum Bayesianism. *Physics and Philosophy* 2012, 018 [27 pp]. arXiv:1108.2024 [quant-ph]

Fields, C. (2012b). If physics is an information science, what is an observer? *Information* 3, 92-123. arxiv:1108.4865v2 [quant-ph]

Fields, C. (2012c). A model-theoretic interpretation of environment-induced superselection. *International Journal of General Systems* in press (DOI:10.1080/03081079.2012.707197). arXiv:1202.1019v2 [quant-ph]

Fuchs, C. (2010). QBism: The perimeter of quantum Bayesianism. Preprint arXiv:1003.5209v1 [quant-ph].

Fuchs, C. (2012). Interview with a quantum Bayesian. Preprint arXiv:1207.2141v1 [quant-ph] (a compilation of Fuchs' contributions to M. Schlosshauer (Ed.) *Elegance and Enigma: The Quantum Interviews*. Berlin: Springer, 2011).

Fuchs, C. & Schack, R. (2011). Bayesian conditioning, the reflection principle and quantum decoherence, in Y. Ben-Menahem & M. Hemmo (Eds.) *Probability in Physics*. Berlin: Springer (pp. 233-247). arXiv:1103.5950v1 [quant-ph]

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

Gisin, N. (2012). Non-realism: Deep thought or a soft option? *Foundations of Physics* 42, 80-85.

Joos, E., Zeh, D., Kiefer, C., Giulini, D., Kupsch, J. & Stamatescu, I.-O. (2003). *Decoherence and the Appearance of a Classical World in Quantum Theory* (2nd Ed.) Berlin: Springer.

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

Landauer, R. (1999). Information is a physical entity. *Physica A* 263, 63-67.

Ollivier, H., Poulin, D. & Zurek, W. H. (2005). Environment as a witness: Selective proliferation of information and emergence of objectivity in a quantum universe. *Physical Review A* 72, 042113 [21 pp]. arXiv:quant-ph/0408125v3

Schlosshauer, M. (2006). Experimental motivation and empirical consistency of minimal no-collapse quantum mechanics. *Annals of Physics* 321, 112-149. arXiv:quant- ph/0506.199v3

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

Tegmark, M. (1998). The interpretation of quantum mechanics: Many worlds or many words? *Fortschritte der Physik* 46, 855-862.

von Neumann, J. (1932). *Mathematische Grundlagen der Quantenmechanik*. Berlin: Springer.

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

Zurek, W. H. (2003). Decoherence, einselection, and the quantum origins of the classical. *Reviews of Modern Physics* 75, 715-775. arXiv:quant-ph/0105127v3

Zurek, W. H. (2009). Quantum Darwinism. *Nature Physics* 5, 181-188. arXiv:0903.5082v1 [quant-ph]