Formalism and Interpretation in Quantum Theory[1]

Alexander Wilce
Department of Mathematics
Susquehanna University
Selinsgrove, PA 17870

Abstract

Quantum Mechanics can be viewed as a linear dynamical theory having a familiar mathematical framework but a mysterious probabilistic interpretation, *or* as a probabilistic theory having a familiar interpretation but a mysterious formal framework. These points of view are usually taken to be somewhat in tension with one another. The first has generated a vast literature aiming at a "realistic" and "collapse-free" interpretation of quantum mechanics that will account for its statistical predictions. The second has generated an at least equally large literature aiming to derive, or at any rate motivate, the formal structure of quantum theory in probabilistically intelligible terms. In this paper I explore, in a preliminary way, the possibility that these two programmes have something to offer one another. In particular, I show that a version of the measurement problem occurs in essentially *any* non-classical probabilistic theory, and ask to what extent various interpretations of quantum mechanics continue to make sense in such a general setting. I make a start on answering this question in the case of a simplified version of the Everett interpretation.

1 Two Views of Quantum Mechanics

Like any physical theory, Quantum Mechanics has both kinematical and dynamical aspects. The former delineate *what* changes, and the later delineate *how* it changes. In the particular case of quantum mechanics, this picture is obscured by the fact that the things that change – quantum states and observables – are related to one another probabilistically. To the extent that we view probabilities as attaching themselves to *events* – that is, to things that *happen* – and to the extent we think of these happenings as involving a *change of state*, we seem to be importing a secondary dynamics. To the extent that we think of probabilities rather as averages over static *states of affairs*, we seem to be committed to hidden variables – which must be both contextual and non-local (the former by Gleason's Theorem, the latter, by Bell's). This dilemma frames the so-called *measurement problem*: to give an account of quantum mechanics that embraces neither hidden variables nor any secondary dynamics, but still preserves probabilistic appearances.

There is, however, another way to look at quantum mechanics. It is a remarkable mathematical fact that, given only the barest essentials of its probabilistic apparatus, the rest of the structure of quantum mechanics, *including*

[1] For Jeffrey Bub on his 65th Birthday

its dynamics, is fixed, up to the choice of a Hamiltonian. [2] In view of this, one is tempted to regard quantum mechanics as *first* a probability calculus, and only secondarily a dynamical theory. Indeed, one might go further and regard quantum theory as *just being* a non-classical probability calculus – that is, not a physical theory at all, but only a stage on which to enact physics (as represented by the various possible Hamiltonians). This point of view has a lot to recommend it, in terms of conceptual and mathematical economy; and it more or less bypasses the measurement problem [35]. But it presents us with its own vexatious foundational problem, namely, that the mathematical infrastructure of quantum mechanics has little obvious motivation *as* a pure probability theory: rather, that infrastructure is most naturally seen as arising exactly from a dynamical theory, in which a system's states are complex-valued functions evolving under the action of a linear partial differential operator – but then we should expect a dynamical account of the probabilistic apparatus, and we are back where we started.

Thus, we have two problems, somewhat in tension with one another. If we view quantum mechanics as a linear dynamical theory, in which physical states are wave functions, evolving according to the Schrödinger equation, then the theory's analytical apparatus is not especially problematic (Hilbert spaces were, after all, *invented* to describe just this sort of thing); but its probabilistic content seems mysterious and ad hoc. If, on the other hand, we accept the theory's minimal probabilistic interpretation as unproblematic, then it is the theory's formal apparatus that seems mysterious and ad hoc.

I'll refer to the former problem as the *problem of interpretation*, and to the latter as the *problem of the formalism*. Both have proved remarkably refractory, withstanding decades of sustained, and often brilliant, effort by physicists, mathematicians, and philosophers of science, and, in the process, accreting substantial technical literatures. It is quite remarkable, therefore, that these two obviously related problems have for the most part been considered in isolation from one another. Superficially, perhaps, this is understandable, as each problem begins where the other wishes to end; nevertheless, when a tunnel is being dug through a mountain, it is usual for those working from opposite sides to coordinate their efforts.

In this paper, I want to urge that each of these two foundational projects has something to contribute to the other. It has become increasingly clear in recent years [4, 6, 29, 30, 41] that many of the most puzzling "quantum" phenomena – in particular, phenomena associated with entanglement, and including, as I'll show, a version of the measurement problem – are in fact quite generic features of essentially *all* non-classical probabilistic theories, quantum or otherwise. This suggests that many of the interpretive ideas that have been advanced in connection with quantum mechanics can be carried over to a much more general setting. This exercise has something to offer to both foundational projects. On the one hand, an interpretation of quantum mechanics that *can't* be made

[2] This is a one-sentence summation of a vast and intricate story, bringing together among other things the Spectral Theorem and the Theorems of Gleason, Stone and Wigner. See Varadarajan [44] for the uncondensed version.

sense of absent certain special structural features of quantum mechanics, is potentially a source of fruitful ideas with which to approach the problem of the formalism. On the other hand, if an interpretation *can* be kept aloft even in the thin atmosphere of a completely general non-classical probabilistic theory, then perhaps it has little to tell us about the physical content of quantum theory. To compress this idea into a slogan: a completely satisfactory interpretation of a physical theory should be capable of yielding (or at least, constraining!) its own formalism.

The balance of this paper should be regarded as a preface, and an invitation, to the programme just outlined. In section 2, I sketch a general (and more or less canonical) mathematical framework within which one can study and compare various kinds of classical and non-classical probabilistic models. This part of the paper is tutorial, and includes more than the minimum of detail, since the framework discussed here may be unfamiliar to many readers. In section 3, I consider coupled systems and entanglement in this setting. The main point I wish to make here is that, as mentioned above, the existence and the basic properties of entangled states are in no way specifically or characteristically quantum-mechanical phenomena. Most of the serious interpretational issues confronting quantum theory arise precisely in connection with entangled states; thus, the stage is set to rehearse familiar arguments concerning the interpretation of quantum mechanics without the usual Hilbert-space props. As I show in section 4, a version of the measurement problem can be posed for *any* non-classical probabilistic models. One can then ask to what extent familiar no-collapse interpretations of quantum mechanics – modal, many-worlds, consistent-histories, *etc.* – can be made to work in this general setting, and to what extent they depend on special structural features of quantum mechanics.

I make only the barest start on answering this question (though I hope to establish at least that it is an interesting one), focussing on a simplified version of the Everett interpretation. This depends on two features of Hilbert-space quantum mechanics that are *not* entirely generic: the fact that the conditional states of a pure bipartite state are pure, and the fact that pure bipartite quantum states always correlate at least one pair of observables. These features can be abstracted, and their consequences studied. One such consequence turns out to be a weak form of the spectral theorem.

2 General Probabilistic Models

Perhaps the most basic difference between classical and quantum probability theory is that the latter gives up the assumption, tacit in the former, that any two random quantities can jointly be measured to arbitrary accuracy. Once we admit the possibility of incommensurable random quantities, however, we open the door on a vast and rather wild landscape of possible non-classical probabilistic theories, most of them far removed from quantum mechanics. One would like to characterize the class of quantum-probabilistic models cleanly, in probabilistic (or what some would call *operational*) terms, and in such a way as sheds light on *why* this particular class of models should figure so prominently in physics.

This is a longstanding problem, already strongly foreshadowed in von Neumann's work, [45], and articulated in clear and programmatic terms by Mackey [33]. To approach it, one needs to survey the field of possible alternatives to classical and quantum probability theory, and this requires some altitude. Fortunately, this isn't hard to achieve. Indeed, there is an essentially canonical formalism of generalized probability theory, developed more or less independently by many people, beginning perhaps with Mackey himself [33], and including Ludwig [32], Davies and Lewis [16], Edwards [18], Randall and Foulis [20], Holevo [28], and, more recently, Hardy [27], D'Ariano [15] and Barrett [6], among others. The next two subsections outline a version of this canonical formalism. In the interest of simplicity (and of brevity), I will, in the main, restrict my attention to finite dimensional probabilistic models. This introduces certain distortions, but preserves the essential contours of the subject.

2.1 Convex Sets as Abstract State Spaces

One approach to a generalized probability theory begins with an abstract convex set Ω of "states". In practice, this will be a convex subset of a real vector space V, though the particular ambient space is largely irrelevant here. Unless otherwise indicated, I'll assume that V is finite-dimensional and that Ω is compact, that is, closed and bounded (as remarked below, this allows us a canonical choice for V). The idea is that, given a finite sequence of states $\alpha_1, ..., \alpha_n \in \Omega$ and a sequence of non-negative coefficients $t_1, ..., t_n$ summing to unity, the convex combination $\alpha = \sum_i t_i \alpha_i \in \Omega$ represents a statistical mixture in which state α_i occurs with probability t_i. A state is said to be *mixed* iff it can be represented in this way as a non-trivial convex combination of other states. States not so representable – that is, the extreme points of Ω – are termed *pure states*. In the present finite-dimensional setting, every state can be represented as a convex combination of pure states. [3]

Physical events (e.g., measurement outcomes) can now be defined in terms of affine – that is, convex combination preserving – functionals from Ω into the real unit interval. More exactly, let $A(\Omega)$ denote the real vector space of all real-valued affine functionals $a : \Omega \to \mathbb{R}$; call such a functional an *effect* iff it takes values in $[0, 1]$. We may interpret an effect a as representing a possible event or occurrence, with $a(\alpha)$ giving the probability of that event in state $\alpha \in \Omega$.

Let u denote the *unit effect*, that is, the constant functional with value 1; then for any effect a, the functional $u - a$ is again an effect, representing the *non*-occurrence of a. We can regard the pair $(a, u - a)$ as representing a binary *observable* associated with the system. More generally, a discrete *observable* taking values in a set E is defined to be a mapping $f : E \to A(\Omega)$ such that, for

[3] If Ω is an infinite-dimensional compact convex set, this remains largely true, provided we replace convex combinations with so-called boundary integrals. See [1] for details. On the other hand, a non-compact convex set may lack pure states entirely.

every $x \in E$, $f(x)$ is an effect, and $\sum_{x \in E} f(x) = u$. This definition allows us to pull each state $\alpha \in \Omega$ back to a classical probability weight $f^*(\alpha)(x) = f(x)(\alpha)$ on E, which we interpret as giving the statistical distribution of values of the observable when the state is α. The simplest case is that in which E is itself just a set of effects summing to u, with f the inclusion mapping. In this case, we may speak of the set E itself as an observable.

By way of illustration, the set $\Delta(E)$ of probability weights on a finite set E is a compact convex subset of $\mathbb{R}^E$. The extreme points of $\Delta(E)$ are the *point-masses* δ_x associated with points $x \in E$, defined by the condition $\delta_x(x) = 1$. Geometrically, $\Delta(E)$ is a *simplex*: every element $\alpha \in \Delta(E)$ has a *unique* expression as a mixture of extreme points, i.e, point-masses, namely $\alpha = \sum_{x \in E} \alpha(x)\delta_x$. An affine functional on $\Delta(E)$ is determined by its values on the point-masses: if $a \in A(\Delta(E))$, let $\phi(x) = a(\delta_x)$; then ϕ is a random variable on E, and, for all $\alpha \in \Delta$, $a(\alpha)$ is simply the expected value of ϕ in state α. In this way, elements of $A(\Delta(E))$ represent random variables on E. Note that any finite-dimensional simplex Δ has the form $\Delta(E)$: simply let E be extreme points of Δ.

For another example, If $\mathbf{H}$ is a finite-dimensional Hilbert space, the collection $\Omega(\mathbf{H})$ of density operators ρ on $\mathbf{H}$ is a compact convex set. The extreme points of $\Omega(\mathbf{H})$ are precisely usual quantum-mechanical pure states, that is, rank-one projection operators. The space $A(\Omega(\mathbf{H}))$ is canonically identifiable with the set of self-adjoint operators on $\mathbf{H}$: if $a \in A(\Omega(\mathbf{H}))$, there exists a unique self-adjoint operator A on $\mathbf{H}$ with $a(\rho) = \mathrm{Tr}(A\rho)$ for all density operators $\rho \in \Omega(\mathbf{H})$. Note that $\Omega(\mathbf{H})$ is not a simplex, since a density operator can typically be written as a convex combination of rank-one projections in many ways.

Returning now to an abstract convex set Ω, notice that each state $\omega \in \Omega$ induces an evaluation functional $\hat{\omega} \in A(\Omega)^*$, given by

$$\hat{\omega}(f) = f(\omega)$$

for all $f \in A(\Omega)$. It can be shown [1] that the mapping $\omega \mapsto \hat{\omega}$ is injective; clearly, it is affine. If we don't mind being a little sloppy, we can identify ω with $\hat{\omega}$, so that Ω becomes a subset the dual space, $A(\Omega)^*$, of $A(\Omega)$. The span of Ω in $A(\Omega)^*$ is denoted by $V(\Omega)$. In the context of discrete classical probability theory, where Ω is the set $\Delta(E)$ of probability weights on a set E, $V(\Omega)$ is simply the space $\mathbb{R}^E$ of all bounded real-valued functions on E. In the context of quantum probabilistic models, where Ω is the set of density operators on a Hilbert space $\mathbf{H}$, $V(\Omega)$ amounts to the space of bounded Hermitian operators on $\mathbf{H}$.

2.2 Test Spaces and Probabilistic Models

Elegant though it is, this convex-sets framework suffers both from a certain lack of concreteness, which can make it difficult to apply, and from a certain lack of flexibility: one may not want to allow all effects, or all observables, to count as "physical"; one may also want to allow measurement-outcomes to encode more than just probabilistic information – phase information, say [53, 14]. For these reasons, it turns out to be more expedient to begin with an abstract model of a set of measurements. One way to do this is in terms of so-called *test spaces* [20, 51].

Definition 1 A *test space* is is simply pair $(X, \mathfrak{A})$, where X is a set and $\mathfrak{A}$ is a covering of X by pairwise-incomparable, non-empty sets, called *tests*. The intended interpretation is that each test $E \in \mathfrak{A}$ is an exhaustive and mutually exclusive set of possible *outcomes*, as of some measurement or experiment. Accordingly, we call the set X the *outcome space* of $(X, \mathfrak{A})$. A *state* on $(X, \mathfrak{A})$ is a function $\alpha : X \to [0, 1]$ such that $\sum_{x \in E} \alpha(x) = 1$ for every test $E \in \mathfrak{A}$. We understand $\alpha(x)$ as giving the *probability* of the outcome x occurring in state α.

It is significant that we allow tests to overlap, that is, to share outcomes. We want to be as un-dogmatic as possible at this point as to what kind of thing an outcome *is*: that decision is an interpretive one, about which we may want to preserve a maximum of flexibility. For the same reason, we make no commitments as to when and why outcomes of distinct measurements should be identified: there are many different reasons why one might make such identifications, and these will vary from model to model.

The set of all states on a test space $(X, \mathfrak{A})$, denoted by $\Omega(X, \mathfrak{A})$, is of course a convex subset of $\mathbb{R}^X$; where $\mathfrak{A}$ consists of finite sets, Ω is compact in the topology of pointwise convergence. Each outcome $x \in X$ defines an affine functional $e_x : \Omega \to \mathbb{R}$ by evaluation, that is, $e_x(\omega) = \omega(x)$ for every state $\omega \in \Omega$. Evidently, if $E \in \mathfrak{A}$, then $\sum_{x \in E} e_x = u$, where u is the unit functional on Ω given by $u(\alpha) \equiv 1$ for all $\alpha \in \Omega$; thus, the mapping $x \mapsto e_x$ defines a discrete observable on Ω, in the sense of section 2.1, with values in the set E. To a certain extent, then, one can view a test space as a privileged family of observables on Ω. However, the mapping $x \mapsto e_x$ is generally non-injective: distinct outcomes may have the same probability in every state. In elementary quantum mechanics, this occurs where the outcomes in question differ by a phase. Of course, we might choose simply to *identify* x with e_x; but this presents problems when we need to consider iterated measurements. I refer the reader to the paper of Wright [53] for a thorough discussion of this point.

In practice, one almost always deals with state spaces that arise as convex subsets of the state space of a particular test space. We lose no important generality, then, in making the following

Definition 2 A *probabilistic model* is a triple $(X, \mathfrak{A}, \Gamma)$ where $(X, \mathfrak{A})$ is a test space and Γ is a convex set of states on $(X, \mathfrak{A})$.[4]

In cases where the extra structure provided by the test space is irrelevant to the discussion, I'll sometimes identify such a model with its state space Γ, and proceed as in the convex-sets approach. Later in this paper, I'll begin to speak, not of individual probabilistic *models*, but of probabilistic *theories*. For

[4]A theorem of Shultz [40] tells us that any *compact* convex set can be represented as the *full* state space of a test space – indeed, of the test space of finite partitions of unity in an orthomodular lattice.

purposes of this paper, the term can remain an informal one: roughly, a theory is simply a class of models, closed under the formation of a "tensor product", as spelled out in Section 3.

2.3 Three Examples

To help fix the foregoing ideas, and to serve as running illustrations in the balance of this paper, here are three examples of probabilistic models.

Example 1: Classical Models. Discrete classical probability theory is the theory of test spaces of the form $(E, \{E\})$, having just a single test. In this case, the state space of $\mathfrak{A}$ is simply the set of all probability weights on the set E, which we denote by $\Delta(E)$. More generally, let S be a set and Σ, a field of subsets of S. Let $\mathcal{B} = \mathcal{B}(S, \Sigma)$ be the collection of (say, countable) partitions of S into non-empty Σ-measurable sets. We can regard each partition $E \in \mathcal{B}$ as the outcome-set for a "coarse-grained" measurement of a value in S. Accordingly, we have a test space $(\Sigma^*, \mathcal{B})$, where Σ^* is the set of non-empty elements of Σ, called the *Borel test space* associated with (S, Σ). States on $(\Sigma^*, \mathcal{B})$ correspond in an obvious way to σ-additive probability measures on (S, Σ). Thus, full-dress measure-theoretic classical probability theory is also subsumed by probability theory based on test spaces.

Example 2: Quantum test spaces. Let $\mathbf{H}$ be a Hilbert space. Let $X(\mathbf{H})$ denote $\mathbf{H}$'s unit sphere and $\mathfrak{F}(\mathbf{H})$, the set of *frames*, or maximal orthonormal subsets, of $\mathbf{H}$. Then $(X(\mathbf{H}), \mathfrak{F}(\mathbf{H}))$ is a test space, called the *frame manual* of $\mathbf{H}$, representing the collection of (maximal) discrete quantum-mechanical experiments. Gleason's theorem [23] lets us represent probability weights by density operators in the usual way; that is, for every $\omega \in \Omega(\mathbf{H}) := \Omega(\mathfrak{F}(\mathbf{H}))$, there exists a unique density operator ρ on $\mathbf{H}$ with $\omega(x) = \langle \rho x, x\rangle$. In particular, every *pure* state $\alpha \in \Omega(\mathbf{H})$ is associated with a unit vector x, unique up to phase, with $\alpha(y) = |\langle x, y\rangle|^2$. Thus, elementary quantum probability theory is essentially the theory of quantum test spaces. (Much more generally, the collection $\mathfrak{A}$ of maximal pairwise-orthogonal sets of projections in a von Neumann algebra $\mathbf{A}$ is a test space. The extensions of Gleason's Theorem due to Christensen and Yeadon [13, 54] show that, where $\mathbf{A}$ contains no direct summand of type I_2, every state on $\mathfrak{A}$ extends uniquely to a state on $\mathbf{A}$.)

Example 3: The Firefly Box. One of the virtues of test spaces is the ease with which one can manufacture simple and instructive ad hoc examples. As an illustration of this, consider the following test space, known as the *Firefly box* or the *Wright Triangle*[5]. A sealed triangular box has opaque top and bottom and translucent walls. The interior is divided into three chambers, each chamber occupying one corner, as in Figure 1 below. Inside the box is a firefly, which is visible when viewed through a given wall if, and only if, the firefly occupies one

[5]The example is due originally to D. J. Foulis, but was heavily promoted by Ron Wright

of the two chambers behind that wall, and is flashing.

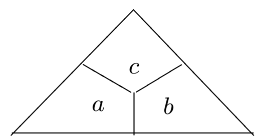

Figure 1

Each wall corresponds to an experiment: looking through the south-facing wall, for instance, we may see a light in chamber a, a light in chamber b, or we may see no light at all. Denoting this latter outcome by x, we may represent the experiment of looking through the south wall by $\{a, x, b\}$. Representing the experiments associated with the other two walls similarly by $\{b, y, c\}$ and $\{c, z, a\}$ (where y and z denote outcomes in which no light is seen), we have a test space $\{\{a, x, b\}, \{b, y, c\}, \{c, z, a\}\}$. This can conveniently be represented by a graph, as in figure 2. Here each node represents an outcome, as indicated, with the outcomes corresponding to each experiment lying along a common line.[6]

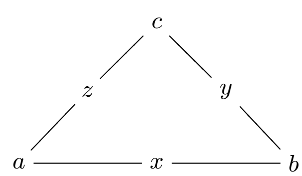

Figure 2

2.4 Classical Representations

A state ω on a test space $\mathfrak{A}$ is *dispersion-free* if it takes only the values 0 and 1, and thus predicts the outcome of each test with certainty. Equivalently, we may think of a dispersion-free state as a *transversal* of the set of tests, that is, a subset of $X = \bigcup \mathfrak{A}$ meeting each test exactly once. Evidently, every state on a classical test space $\mathfrak{A} = \{E\}$ is uniquely representable as a mixture of dispersion-free states (i.e., point masses); by Gleason's Theorem, the quantum test space $(X(\mathbf{H}), \mathfrak{F}(\mathbf{H}))$ has no dispersion-free states at all.

The Firefly Box presents an interesting intermediate case. Its dispersion-free states correspond to the four transversals pictured below. The first three of these represent the situations in which the firefly is flashing in one of the three

[6]The graphical convention we use here is an example of a useful device called a *Greechie diagram* (after R. J. Greechie). The idea is to represent each outcome of a small, finite test space by a node in a graph, connecting the outcomes belonging to a given test along a smooth arc (e.g., a straight line or, if necessary, some other smooth curve), arranging matters in such a way that the arcs corresponding to distinct but overlapping tests intersect one another transversally, so that they can readily be distinguished from one another by eye.

chambers; the fourth describes the situation in which the firefly is not flashing at all (in which case its location is unknown).

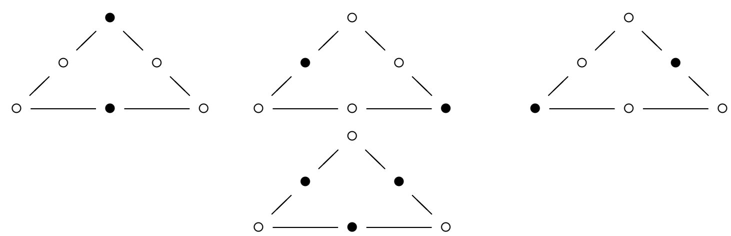

Figure 5

These four states determine the structure of the firefly-box. For each outcome $x \in X$, let $[x]$ denote the set of dispersion-free states making x certain: the mapping $x \mapsto [x]$ is injective, and takes each test in $\mathfrak{A}$ to a partition of the set of dispersion-free states. In some sense, this provides a perfectly classical model of the state space. However, not every state on the Firefly Box is explained by this model: there is a fifth pure state, given by

$$\omega(a) = \omega(b) = \omega(c) = 1/2 \text{ and } \omega(x) = \omega(y) = \omega(z) = 0, \tag{1}$$

which is obviously not an average of the four dispersion-free states pictured above. Rather, it seems to describe a "gregarious" firefly that presents itself at whichever window it is through which the observer is peering, choosing the left or right-hand chamber at random. It seems we can understand the "gregarious" state (1) very easily, but *only* if we are willing to allow that our choice of experiment (here, the act of looking through one of the three windows) perturbs the (deterministic) state of the firefly.

This illustrates a trivial but very important point: there is nothing very mysterious about non-classical models, provided we are willing to allow for contextuality. We can make this precise as follows. Call a test space $\mathfrak{A}$ *semi-classical* iff distinct tests are disjoint; that is, if one can read off from each measurement outcome, the measurement by which it was secured. The pure states of such a test space are exactly the dispersion-free states (these corresponding to arbitrary selections of one outcome from each test), and, subject to very weak analytic conditions, every state is an average of these [49]. Now given an arbitrary test space $(X, \mathfrak{A})$, let $\widetilde{X} = \{(x, E) | x \in E \in \mathfrak{A}\}$. For each test $E \in \mathfrak{A}$, let $\widetilde{E} := \{(x, E) | x \in E\}$, and let $\widetilde{\mathfrak{A}} = \{\widetilde{E} | E \in \mathfrak{A}\}$. Then $(\widetilde{X}, \widetilde{\mathfrak{A}})$ is a semi-classical test space, which I like to call the *semi-classical cover* of $(X, \mathfrak{A})$. Every state ω on $\mathfrak{A}$ defines a state $\widetilde{\omega}$ on $\widetilde{\mathfrak{A}}$ (given by $\widetilde{\omega}(x, E) = \omega(x)$ for every $(x, E) \in \widetilde{X}$), and hence, can be represented as an average over dispersion-free states on $(\widetilde{X}, \widetilde{\mathfrak{A}})$. In this sense, the latter provides a perfectly classical explanation for all of the states on $(X, \mathfrak{A})$. Alas, nothing comes for free. The cost of contextuality is non-locality – a topic to which I'll briefly return in section 3.

2.5 Characteristics of Quantum Models

Having attained a sufficient altitude, we can now revisit the problem of characterizing quantum mechanics as a probability theory. What is it that distinguishes quantum probabilistic models from non-classical probabilistic models generally? It is important to stress here that not just any kind of characterization will do: we want to distinguish the quantum-probabilistic models in a way that sheds some light on *why* we might expect these models to play an especially prominent role in physics. It is unlikely that this question will have a best, let alone a unique, answer. There are, after all, many different ways to characterize classical probability theory, and the same is doubtless true of quantum probability theory. And, indeed, such answers as we have, for example, the axiomatic reconstructions of elementary quantum probability due to Hardy [27], Goyal [24], and D'Ariano [15] are based on very diverse considerations. Nor is the answer going to be such as to single out quantum models as *uniquely reasonable*: in the face of toy models like the Firefly Box, it seems extremely unlikely that the framework of Quantum Mechanics is in any sense a "law of thought".

On the other hand, the quantum models do have many features that, from our current height, look rather special – and, indeed, rather *classical* – relative to generic probabilistic models. For one thing, in both discrete classical and quantum models, every outcome x corresponds to a unique pure state ϵ_x such that $\epsilon_x(x) = 1$: in a classical model, ϵ_x is simply the point-mass δ_x at x; in a quantum model, where x is a unit vector in a Hilbert space $\mathbf{H}$, ϵ_x is the pure state defined by the same vector, that is, $\epsilon_x(y) = |\langle x, y\rangle|^2$ for all outcomes (i.e, unit vectors) $y \in \mathbf{H}$.

Definition 3 A model $(X, \mathfrak{A}, \Omega)$ is *sharp* iff for every outcome $x \in X$, there exists a unique state $\epsilon_x \in \Omega$ with $\epsilon_x(x) = 1$. Note that the set of states making a given outcome certain is a *face* of the state space[7]; hence, the state ϵ_x must be pure.

The condition that a model be sharp very attractive mathematically, and not extraordinarily restrictive. For classical and quantum models, the association runs the other way as well: every pure state makes some outcome certain – a unique outcome, classically, and an outcome unique up to phase in the quantum case. On the other hand, as illustrated by the Firefly box, a perfectly sensible model need have neither property: the outcome x (of seeing no light in the south-facing window) is certain in either of two pure states, while the strange pure state (1) makes no outcome certain.

Another respect in which quantum and classical models are similar is this: while a quantum state has no *unique* representation as a mixture of pure states, it often has a *preferred* decomposition into a mixture of orthogonal pure states. Indeed, applied to the density matrix representing a quantum state ω, the Spectral Theorem tells us that there exists an orthonormal basis E such that $\omega = \sum_{x \in E} \omega(x)\epsilon_x$, where ϵ_x is the unique pure state corresponding to $x \in E$.

[7]A *face* of a convex set Ω is a convex set $\Gamma \subseteq \Omega$ such that if $t\alpha + (1-t)\beta \in \Gamma$, then either $\alpha \in \Gamma$ or $\beta \in \Gamma$

Definition 4 Call a family of states $\{\alpha_i\}$ *sharply distinguishable* iff there exists a test $E \in \mathfrak{A}$ such that $\forall i$, $\exists x_i \in E$ with $\alpha_i(x_i) = 1$ and $\alpha_j(x_i) = 0$ for all $j \neq i$. Call a state $\omega \in \Omega$ *spectral* iff it is a convex combination of sharply distinguishable states. Equivalently, ω is spectral iff, for some test $E \in \mathfrak{A}$, one has

$$\omega = \sum_{x \in E} \omega(x)\alpha_x,$$

where, for all $x \in E$ with $\alpha_x \in \Omega$, α_x is a state satisfying $\alpha_x(x) = 1$.

Pure *quantum* states are sharply distinguishable iff their corresponding state vectors are orthogonal; thus, the spectral theorem, as applied to density matrices, tells us that every quantum-mechanical state is spectral. This is hardly the case more generally, however. Note that a pure state is spectral iff it makes some outcome certain; thus, the *outré* pure state (1) on the Firefly box, which makes no outcome certain, isn't spectral.

One should resist the temptation to take it as an *axiom* of generalized probability theory that all states be spectral, since to do so would be to rule out examples like the Firefly Box by fiat. As it happens, one can anyway do a bit better: as we'll see in section 4, the spectrality of states follows easily from a condition with a more obvious motivation in terms of correlation between the parts of composite systems. Before we can discuss this, of course, we need to say something about how general probabilistic models can be composed. This is the subject of the next section.

3 Coupled Systems and Entanglement

As we've seen, there is nothing terribly mysterious about quantum or other non-classical probabilistic models, taken one at a time: they can readily be understood in terms of contextual hidden variables. The really interesting features of such models arise when they are combined: as is well known, composite quantum systems can be correlated in distinctly non-classical ways, via so-called entangled states, and these, by Bell's theorem, are not explicable in terms of *local* hidden variables, contextual or otherwise. It is exactly the existence of such highly correlated states that leads to the most counter-intuitive aspects of quantum mechanics. However, as we'll see in this section, entanglement is not a specifically quantum phenomenon; rather, it arises generically when one combines two non-classical probabilistic models.

3.1 Products of Probabilistic Models

Suppose we want to construct a model of a system comprising two separate subsystems. We have in mind here a rather special situation, in which the two component systems are not interacting in any obvious, causal sense (e.g., systems occupying space-like separated regions of space-time). I make two preliminary assumptions: first, that it is possible to make measurements independently on the two sub-systems, and secondly, that states of the composite system are completely determined by the probabilities they assign to the results of such measurements.[8]

To make this a bit more precise, suppose that our two sub-systems are associated with test spaces $\mathfrak{A}$ and $\mathfrak{B}$, with outcome-spaces X and Y, respectively. Following the notation of [20], let us write xy for an ordered pair (x, y) of outcomes $x \in X$ and $y \in Y$, AB for the cartesian product $A \times B$ of sets $A \subseteq X$ and $B \subseteq Y$, and so on. Our first assumption requires that the test space corresponding to a coupled system include, at a minimum, all *product tests* $EF = \{xy|x \in E, y \in F\}$, where $E \in \mathfrak{A}$ and $F \in \mathfrak{B}$. The collection of such product tests, which, in a shameless abuse of notation, we write as $\mathfrak{A} \times \mathfrak{B}$, is itself a test space with outcome set XY. We are requiring, then, that any "joint test space" $\mathfrak{C}$ for the combined system contain $\mathfrak{A} \times \mathfrak{B}$.

Our second assumption requires that distinct states of the coupled system be distinguishable by outcomes in XY. Thus, if $\Theta \subseteq \Omega(\mathfrak{C})$ is the state space of the coupled system, the natural restriction mapping $\Theta \to \Omega(\mathfrak{A} \times \mathfrak{B})$ be injective. This allows us to treat Θ, for most purposes, as a subset of $\Omega(\mathfrak{A} \times \mathfrak{B})$.

Definition 5 A state $\omega \in \mathfrak{A} \times \mathfrak{B}$ is *influence-free* [20, 29] iff the *marginal states* $\omega_2 := \omega_{2|E}$ and $\omega_1 := \omega_{1|F}$ are well-defined, i.e, independent of $E \in \mathfrak{A}$ and $F \in \mathfrak{B}$.

The idea is that a state is influence-free if the mere choice of measurement on the $\mathfrak{A}$-system has no effect on the probabilities of $\mathfrak{B}$-outcomes, and vice-versa. (Such states are also sometimes called *no-signalling* states in the literature, but I dislike the term, as it can be shown [20, 5] that influence-free states are precisely those that *allow* classical signalling between the sub-systems.)

If $\alpha \in \Omega(\mathfrak{A})$ and $\beta \in \Omega(\mathfrak{B})$, then the *product state* $\alpha \otimes \beta \in \Omega(\mathfrak{A} \times \mathfrak{B})$, defined for all $xy \in XY$ by

$$(\alpha \otimes \beta)(xy) = \alpha(x)\beta(y),$$

is clearly influence-free; hence, so is any convex combination $\omega = \sum_i t_i \alpha_i \otimes \beta_i$ of product states. Borrowing language from quantum theory, we call states of this form *separable*; influence-free states that are not separable, we call *entangled*. As is well-known, quantum mechanics allows for the existence of entangled states. As will become clear presently, this is equally true of (almost) any non-classical probabilistic theory, quantum or otherwise.

If ω is influence-free, we have, for every outcome x of $\mathfrak{A}$ and every outcome y of $\mathfrak{B}$, *conditional states* $\omega_{2|x} \in \Omega(\mathfrak{B})$ and $\omega_{1|x} \in \Omega(\mathfrak{A})$, given respectively by

$$\omega_{2|x}(y) = \frac{\omega(xy)}{\omega_1(x)} \text{ and } \omega_{1|y}(x) = \frac{\omega(xy)}{\omega_2(y)},$$

[8]We are adopting here what Barrett [6] calls the "global state" assumption. As has been noticed by many authors (see, e.g., [30]), this is sufficient to bar real and quaternionic Hilbert space models for quantum mechanics. I'll return to this point below.

with the usual proviso that $\omega_{2|x} \equiv 1$ if $\omega_1(x) = 0$, and similarly for $\omega_{1|y}$. We have the expected identities $\omega(xy) = \omega_1(x)\omega_{2|x}(y) = \omega_{1|y}(x)\omega_2(y)$ for all $x \in X, y \in Y$. Notice that if we select any test $E \in \mathfrak{A}$, we can relate the marginal states ω_1 and ω_2 by an analogue of the law of total probability, namely

$$\omega_2 \;=\; \sum_{x \in E} \omega_1(x)\omega_{2|x}. \tag{2}$$

We are now in a position to define what counts as a model of a composite system. For our purposes, the following will suffice:

Definition 6 A *separated product* of two probabilistic models $(X, \mathfrak{A}, \Omega)$ and $(Y, \mathfrak{B}, \Gamma)$ is a model $(Z, \mathfrak{C}, \Theta)$ where $Z \supseteq XY$, $\mathfrak{C} \supseteq \mathfrak{A} \times \mathfrak{B}$, and Θ is a convex set of states on $(Z, \mathfrak{A})$ such that

(a) The set XY is separating for states in Θ, so that we can identify Θ with a subset of $\Omega(\mathfrak{A} \times \mathfrak{B})$;

(b) Every state in Θ is influence-free on $\mathfrak{A} \times \mathfrak{B}$;

(c) For every state $\omega \in \Theta$ and all outcomes $x \in X$ and $y \in Y$, the conditional states $\omega_{1|y}$ and $\omega_{2|x}$ belong to Ω and Γ, respectively;

(d) Θ contains all product states.

Example: Recall that the *frame manual* of a Hilbert space $\mathbf{H}$ is the test space $\mathfrak{F}(\mathbf{H})$ consisting of all orthonormal bases for $\mathbf{H}$. The state space of $\mathfrak{F}(\mathbf{H})$, that is, the set of density operators on $\mathbf{H}$, we denote by $\Omega(\mathbf{H})$. Let X denote the unit sphere of $\mathbf{H}$. Suppose now that $\mathbf{H}$ is the tensor product, $\mathbf{H}_1 \otimes \mathbf{H}_2$, of two *complex* Hilbert spaces $\mathbf{H}_1$ and $\mathbf{H}_2$, with unit spheres X_1 and X_2, frame manuals $\mathfrak{F}_1$ and $\mathfrak{F}_2$, and state spaces Ω_1 and Ω_2, respectively. Identifying $xy \in X_1X_2$ with $x \otimes y \in X$, we may regard any product test EF in $\mathfrak{F}_1 \times \mathfrak{F}_2$ as an orthonormal basis for $\mathbf{H}$. Modulo this slight sloppiness, $\mathfrak{F} \supseteq \mathfrak{F}_1 \times \mathfrak{F}_2$. I claim that $(X, \mathfrak{F}, \Omega(\mathbf{H}))$ is a separated product of $(X_1, \mathfrak{F}_1, \Omega(\mathbf{H}_1))$ and $(X_2, \mathfrak{F}_2, \Omega(\mathbf{H}_2))$. If $\omega \in \Omega(\mathbf{H})$ corresponds to the density operator W on $\mathbf{H}_1 \otimes \mathbf{H}_2$, then the Polarization identity, applied twice, shows that W – hence, ω – is uniquely determined by the biquadratic form $\langle Wx \otimes y, x \otimes y\rangle = \omega(xy)$. Thus, condition (a) is satisfied. To see that (b) is satisfied, simply note that the marginals of $\omega \in \Omega(\mathbf{H})$ are given by $\omega_1(x) = \mathrm{Tr}(W(P_x \otimes \mathbf{1}_1))$ and $\omega_2(y) = \mathrm{Tr}(W(\mathbf{1}_2 \otimes P_x))$, where P_x and P_y are the rank-one projections on $\mathbf{H}_1$ and $\mathbf{H}_2$, respectively, corresponding to the unit vectors x and y, and where $\mathbf{1}$ represents the identity operator on $\mathbf{H}_i$, $i = 1, 2$. Conditional states are essentially just the *relative* states considered by Everett (a point to which I'll return in section 4.1), so condition (c) is satisfied. Condition (d) is trivial, so we see that $(X, \mathfrak{F}, \Omega(\mathbf{H}))$ is indeed a separated product, in the sense of Definition, of $(X_1, \mathfrak{F}_1, \Omega(\mathbf{H}_2))$ and $(X_2, \mathfrak{F}_2, \Omega(\mathbf{H}_2))$.

Remark: The use of the polarization identity above is crucial to secure condition (a) in the definition of a separated product. In fact, the models associated with tensor products of real or quaternionic Hilbert spaces are *not* separated products in the above sense. This can be seen either as a justification for the use of complex scalars, or as a defect in the definition.

3.2 Tensor products of state spaces

It is possible to give an abstract characterization of the possible state spaces of separated products in terms of the state spaces of the component systems, in a way that makes no reference to test spaces. For what follows, recall that if Ω is any compact convex set, we write $V(\Omega)$ for the span of Ω in $A(\Omega)^*$. In our finite-dimensional setting, $V(\Omega) = A(\Omega)^*$ and $V^*(\Omega) = A(\Omega)$. Recall, too, that if Ω is a set of states on $(X, \mathfrak{A})$, then every outcome $x \in X$ induces an effect $e_x \in A(\Omega)$ by $e_x(\alpha) = \alpha(x)$ for all $\alpha \in \Omega$.

Suppose now that Ω and Γ are convex sets of states on test spaces $\mathfrak{A}$ and $\mathfrak{B}$, respectively. Let ϕ be a bilinear form $\phi : A(\Omega) \times A(\Gamma) \to \mathbb{R}$ that is *positive*, in the sense that $\phi(a, b) \geq 0$ for all positive $a \in A(\Omega)$ and $b \in B(\Omega)$, and *normalized*, in the sense that $\phi(u_\Omega, u_\Gamma) = 1$. It is not difficult to see that ϕ yields an influence-free state ω on $\mathfrak{A} \times \mathfrak{B}$, given by $\omega(xy) = \phi(e_x, e_y)$, with conditional states in Ω and Γ. It is not difficult to prove the following converse (for details, as well as a discussion of the analogous result for infinite-dimensional state spaces, see [48]; see also [46, 21] for similar results in a Hilbert space context):

Proposition 1 *Let ω be an influence-free state on $\mathfrak{A} \times \mathfrak{B}$ having conditional states in Ω and Γ. Then there exists a unique positive, normalized bilinear form $\hat{\omega}$ on $A(\Omega) \times A(\Gamma)$ such that $\hat{\omega}(e_x, e_y) = \omega(xy)$ for all x, y.*

We now have a perfectly abstract characterization of compound state spaces:

Definition 7 A *tensor product* of two state spaces Ω and Γ is a convex set $\Omega \otimes \Gamma$ of positive, normalized bilinear functionals on $A(\Omega) \times A(\Gamma)$, containing all product states.

Evidently, the largest tensor product, which we call the *maximal tensor product* and denote by $\Omega \otimes_{\text{max}} \Gamma$, consists of all normalized positive bilinear functionals, while the smallest, the *minimal tensor product* $\Omega \otimes_{\text{min}} \Gamma$, consists of convex combinations of product states, that is, separable states. Thus, the minimal tensor product allows for no entanglement, and the maximal tensor product, for as much entanglement as possible (subject to states being influence-free and having conditional states in the correct state-spaces).

If α and β are pure states in Ω and Γ, respectively, then it is easy to show that the product state $\alpha \otimes \beta$ is a pure state of $\Omega \otimes \Gamma$. Unless one factor is a simplex, however, the latter will contain many extreme states that are not products. Indeed, we have the

Proposition 2 ([34]) *A compact convex set Ω satisfies $\Omega \otimes_{max} \Gamma = \Omega \otimes_{min} \Gamma$ for all compact convex sets Γ iff Ω is a simplex.*

It is worth remarking that a positive, normalized bilinear form on $A(\Omega) \times A(\Gamma)$ is effectively the same thing, in finite dimensions, as a positive linear mapping $\phi : A(\Omega) \to V(\Gamma)$ satisfying the normalization condition $\hat{\phi}(u_\Omega) \in \Gamma$. It follows [30, 48] that, if Ω and Γ are finite-dimensional convex sets, then for any choice of tensor products, we have

$$V(\Omega \otimes \Gamma) = V(\Omega) \otimes V(\Gamma) \simeq \mathcal{L}(A(\Omega), V(\Gamma))$$

where $\mathcal{L}(A(\Omega), V(\Gamma))$ is the vector space of linear mappings from $A(\Omega)$ to $V(\Gamma)$, ordered by the cone of positive linear mappings. Similarly

$$A(\Omega \otimes \Gamma) = A(\Omega) \otimes A(\Gamma) \simeq \mathcal{L}(V(\Omega), A(\Gamma))$$

for any choice of tensor product. In particular, every state a tensor product is an *affine* combination of pure tensors – that is, a linear combination of pure tensors, the coefficients of which sum to unity, but need not all be positive.

Examples: (a) If Ω and Γ are both classical state spaces, say $\Omega = \Delta(E)$ and $\Gamma = \Delta(E')$, then $\Omega \otimes_{\text{max}} \Gamma = \Omega \otimes_{\text{min}} \Gamma$, both being isomorphic to $\Delta(E \times E')$.

(b) If $\Omega(\mathbf{H}_1)$ and $\Omega(\mathbf{H}_2)$ are the state spaces associated with *complex* Hilbert spaces $\mathbf{H}_1$ and $\mathbf{H}_2$, respectively, then $\Omega(\mathbf{H}_1) \otimes_{\text{min}} \Omega(\mathbf{H}_2)$ is properly smaller than the the quantum state space $\Omega(\mathbf{H}_1 \otimes \mathbf{H}_2)$ associated with $\mathbf{H}_1 \otimes \mathbf{H}_2$; the latter, in turn, is properly smaller than $\Omega(\mathbf{H}_1) \otimes_{\text{max}} \Omega(\mathbf{H}_2)$. Indeed, any positive operator $\phi : \mathcal{B}(\mathbf{H}_1) \to \mathcal{B}(\mathbf{H}_2)$ satisfying $\text{Tr}(\phi(\mathbf{1})) = 1$, where $\mathbf{1}$ is the identity operator on $\mathbf{H}_1$, defines an element of $\Omega \otimes_{\text{max}} \Gamma$ (and conversely); but such a map corresponds to a bipartite *quantum* state iff it is *completely* positive.

What Barrett [6] calls a (probabilistic) *theory* is essentially a class of models closed under a separated product construction. I'll adopt this terminology. Thus, if one begins with complex quantum state spaces and couples these using the minimal tensor product, one obtains a theory in Barrett's sense (what Halvorson [25] calls the "Shr*dinger theory") that is quite different from ordinary quantum mechanics. Similarly, one might begin with quantum state spaces, and proceed to couple these using the maximal tensor product [47, 5]. For another example, if one begins with semi-classical test spaces consisting of dichotomies (two-outcome tests), and couples these using the maximal tensor product, one obtains essentially the class of models considered by Popescu and Rohrlich [36]. These have come to be called *PR boxes*; accordingly, one might call this the *Box Theory*. Proposition 2 tells us that if a probabilistic theory allows for non-classical (that is, non-simplex) state spaces, it must include composite systems having entangled states – unless, of course, the theory (like Halvorson's "Shr*dinger" theory) makes use exclusively of the minimal tensor product, which we may regard as a degenerate case. In this sense, entangled states are a generic feature of non-classical probabilistic theories.

3.3 Proper and Improper Mixtures

A consequence of the Law of Total Probability – Equation (2) above – is the fact, well-known in the context of quantum mechanics but in fact entirely general, that the reduced states of an entangled bipartite state are always mixed. Note that this is true *regardless* of what tensor product we use.

Lemma 1 *If either marginal, ω_1 or ω_2, of a bipartite state ω in $\Omega \otimes \Gamma$ is pure, then $\omega = \omega_1 \otimes \omega_2$.*

Proof: Suppose ω_2 is pure. We wish to show that $\omega(xy) = \omega_1(x)\omega_2(y)$ for all $x, y \in XY$. Fix x, and let E be an observable including x. By equation (2) we have, for every y,

$$\omega_2(y) = \sum_{x \in E} \omega_{2|x}(y)\omega_1(x).$$

This gives us ω_2 as a convex combination of the conditional states $\omega_{2|x}$. As ω_2 is pure, we have, for each $x \in E$, either $\omega_1(x) = 0$ or $\omega_{2|x} = \omega_2$; in either case we have $\omega(xy) = \omega_1(x)\omega_2(y)$ for all $y \in Y$. $\square$.

An immediate corollary of Lemma 1 is that if either marginal of a *pure* bipartite state is pure, then so it the other, and the state is a product state. Hence, as advertised, the marginals of a pure *entangled* state *must* be proper mixtures.

It follows that the familiar antinomy concerning the "ignorance interpretation of mixtures" in quantum mechanics is entirely generic. Indeed, suppose that a bipartite system with state space $\Omega \otimes \Gamma$ is in a pure entangled state ω. Then the reduced state ω_1 must be mixed – say, $\omega_1 = \sum_i t_i \alpha_i$. It is tempting to regard ω_1 as representing a statistical ensemble, in which system 1 is in state α_i with probability t_i. But (so runs the usual argument), if system 1 were really in pure state α_i, then the real pure state of the composite would have to be a product state, which, by assumption, it isn't.

This argument is usually glossed by saying that the marginals of pure entangled states don't admit an ignorance interpretation. Such marginals are therefore often referred to as *improper* mixtures. While the argument is certainly not water-tight from a mathematical point of view (in particular, it relies on a tacit assumption that the "true" state of system 1 must be the *marginal* of the bipartite state), it is widely accepted. What I hope to have established above is that it holds as much water as applied to general probabilistic models as it does in quantum theory.

3.4 Generic Information Theory

In recent years, it has become clear that one can use entanglement of bipartite quantum states as a *resource* with which to perform information-processing tasks. A programme associated with Brassard [8] and C. Fuchs [21] asks whether quantum mechanics might be the unique probabilistic theory having specific *information*-theoretic properties. A partial result in this direction, due to Clifton, Bub and Halvorsen [11], establishes that, indeed, quantum mechanics (with super-selection rules) is picked out uniquely from among theories having a C^*-algebraic state space by three information-theoretic constraints: the impossibility of super-luminal signalling, of bit-commitment, and of universal cloning.

Having just seen that entangled states arise generically in coupled non-classical models, it is natural to ask how far the known results of quantum information theory extend to this general setting. Barret [6] has shown the non-availability of universal cloning is generic in this way. More recently, it has been shown [4] that powerful versions of the no-cloning and no-broadcasting theorems hold for any finite-dimensional probabilistic model with a compact state space. In more detail, let Ω be any finite-dimensional compact convex set. Say that an affine mapping $B : \Omega \to \Omega \otimes \Omega$ *clones* $\alpha \in \Omega$ iff $B(\alpha) = \alpha \otimes \alpha$, and *broadcasts* α iff $B(\alpha)_1 = B(\alpha)_2 = \alpha$. We say that a finite set $\alpha_1, ..., \alpha_n$ of states is *jointly clonable* or *jointly broacastable* iff there exists a single affine mapping B that clones, or broadcasts, them all. For a proof of the following, see [4]:

Proposition 3 *States $\alpha_1, ..., \alpha_n$ are jointly clonable iff sharply distinguishable, and jointly broadcastable iff all α_i are mixtures of a single sharply distinguishable family of states.*

This has the usual no-cloning and no-broadcasting theorems as corollaries; thus, these results are not special to quantum mechanics, but simply reflect general principles governing all non-classical probabilistic theories.

On the other hand, not every theory supports a teleportation protocol [4]. It's also well known that generic models can violate Bell inequalities more strongly than quantum models do [29, 36]. Thus, the possibility remains open that *some* combination of constraints motivated by quantum information theory will single out, or nearly single out, quantum mechanics, not only among theories within a C^*-algebraic framework, but among probabilistic theories generally. Rather than pursue this possibility, however, we return in the next section to the question raised in section 1, of the extent to which familiar interpretive issues and strategies concerning quantum mechanics can be regarded as generic.

4 Correlation and Measurement

By a *realist* interpretation of quantum mechanics is usually meant one that does not take the concept of "measurement" or "outcome" as primitive, but rather, gives a principled account of which physical interactions count as measurements, and of how these come to have definite outcomes. A *no-collapse* interpretation attempts to do this within the unitary dynamical framework of standard quantum mechanics – without, that is, invoking any actual dynamical "collapse" of the quantum state. Most such accounts also adhere to what we might call *pure-statism*, i.e, the doctrine that the true state of a quantum system is, at all times, a pure state. Actually, there are two versions of this doctrine: a strict version, according to which the state of *any* quantum system is always a pure state, and a weaker but perhaps more plausible version, according to which the state of a *closed* system (one not in interaction with other systems, e.g., the universe) is at all times a pure state.

To provide such an interpretation is a non-trivial task in view of the *measurement problem*, which purports to show that, in general, unitary dynamics actually *precludes* measurements having determinate outcomes. There are three well-known strategies for overcoming this obstacle: dynamical hidden-variables theories, of which Bohmian mechanics is the best known example; Everettian relative state or "many-worlds" interpretations; and so-called *modal* interpretations. My aim in this section is, first, to show that the measurement problem, or anyway one version of it, can be formulated generically in *any* nonclassical probabilistic theory admitting entangled states. Secondly, I wish to consider – albeit only in the most preliminary way – to what extent *one* of the interpretive strategies mentioned above, that of Everett, is viable in a general setting.

4.1 The Quantum Measurement Problem

A standard, if highly idealized, account of measurement as a quantum-mechanical process runs something like this. An object system S, prepared in a state α, is coupled with another system, regarded as a measuring apparatus, initially in a "ready state" β_o. The object and apparatus are both understood to be *quantum* systems, represented by a Hilbert spaces $\mathbf{H}_O$ and $\mathbf{H}_A$, respectively. The coupled object-plus-apparatus system is represented, as usual, by the tensor product $\mathbf{H}_O \otimes \mathbf{H}_A$. During measurement, the coupled system undergoes a unitary evolution that takes the initial state $\alpha \otimes \beta_o$ of the combined system to a pure final state ω. The observable to be measured may, for our purposes, be identified with an orthonormal basis $E = \{x_i\}$ of $\mathbf{H}_O$; for each unit vector $x_i \in E$ let $\alpha_i = \epsilon_{x_i}$ denote the corresponding pure state (so that $\alpha_i(x_i) = 1$). It is required that *if* the initial state α is an eigenstate of the observable to be measured – that is, if $\alpha = \alpha_i$ for some $x_i \in E$ – then the final joint state should have the form $\alpha'_i \otimes \beta_i$, where β_i is an eigenstate of a "pointer" observable $F = \{y_i\}$ of the measuring apparatus, in which the apparatus has recorded a definite outcome x_i, and where α'_i is the corresponding post-measurement state of the system (possibly, but not necessarily, equal to α_i). Notice that there is no difficulty in constructing the desired unitary: it is uniquely defined on $\mathbf{H}_O \otimes (y_o)$ (where (y_o) is the one-dimensional subspace spanned by $y_o \in \mathbf{H}_A$) by the recipe

$$U : x_i \otimes y_o \ \mapsto x_i \otimes y_i \ \forall i, \tag{3}$$

and can be extended arbitrarily to the rest of $\mathbf{H}_O \otimes \mathbf{H}_A$.

The problem now arises that, if $\alpha = \epsilon_v$ is the pure state corresponding to a proper superposition $v = \sum_{x \in E} c_i x_i$ of the elements of E, then the linearity of the evolution requires that the final state correspond to the unit vector

$$U(v \otimes y_o) \ = \ \sum_i c_i x_i \otimes y_i. \tag{4}$$

As this state is entangled, it assigns *no* definite pure state to the apparatus system: we have only the mixed marginal state $\sum_i |c_i|^2\beta_i$. To be sure, this is statistically indistinguishable from a situation in which we end up with one of the states β_i with probability $|c_i|^2$, but it is an *improper* mixture, as discussed in section 3.4, and hence, the apparatus system isn't actually *in* any of the states β_i corresponding to the apparatus' having recorded a definite value.

Tacit here is the assumption that measurement *outcomes* are to be identified with final *states* of the measuring apparatus. This reflects the notion that, for a measurement to have taken place, some *record* of its result must come to exist in the apparatus, and that this must mean the apparatus is left in some state corresponding to that record. But there is another point of view we can take, namely, that measurement outcomes are *something other than* states – say, for instance, *events*.[9] This would seem to dissolve the measurement problem, as it is hardly surprising that a two-sorted ontology should support two distinct dynamics. On the other hand, it carries a commitment to an ontology that is not as well explored as one would wish, and leaves the formal structure of quantum theory, if anything, more mysterious than ever: why, after all, should we expect that every measurement-outcome *qua* event should correspond to some unique pure state, and vice versa? It also leaves the *fact* that some measurements are repeatable a bit of a mystery: to put it differently, such proposals face the problem of accounting for stable records.

Similar remarks apply to other proposals that have been made for the *something other* that might be the correlates of measurement outcomes. Prominent here is the suggestion of van Fraassen, Healy, Dieks and others that a quantum system has, at any time, a set of privileged observables ("beables", in John Bell's famous phrase [7]) that have definite values. In "modal" accounts of this sort, a quantum system has two distinct states: a *dynamical state*, which assigns probabilities to values of observables, and a *value state*, which determines which value of each definite-valued observable is actual.[10]

A more radical, but also far more popular, response to the measurement problem is the "many-worlds" interpretation, according to which each "branch" $x_i \otimes y_i$ of the final object-plus-apparatus system state (4) represents an equally real state of affairs, in which object and apparatus are correlated. More vividly, regarding the apparatus system as consisting of the object system's entire environment, we may wish to speak of these branches as "worlds".

I don't propose to offer here any detailed review, much less any detailed critique, of either modal or many-worlds interpretations. But I do want to make two observations. First, the problem that these interpretations purport to solve – the quantum measurement problem – depending as it does only on

[9] Other possibilities for the *something other* are mental states [26], the "value states" featuring in modal interpretations [43, 31, 17] – discussed below – and perhaps even classes of "worlds" [3] or "histories". If these other things turn out to be classical in some way, then so much the better.

[10] In order to avoid Gleason-Kochen-Specker problems, of course, the set of definite-valued observables can not be too large. Various modal interpretations differ from one another as to how they pick out this privileged set of variables [9, 10].

very general features of entanglement, arises in essentially *any* non-classical probabilistic theory. Secondly, certain versions of both modal and many-worlds interpretations depend, for their cogency, on structural features of Hilbert space quantum mechanics that are *not* completely generic – but which can fruitfully be abstracted. These points are fleshed out in the following sections.

4.2 A Generic Measurement Problem

The idealized model of measurement sketched in the previous section can be adapted to the setting of a generic probabilistic model, as follows.

Definition 8 A *measurement* of a discrete observable E on a system with state space Ω, by a second system with state space Γ, consists of (i) and affine mapping $\mu : \Omega \otimes \Gamma \to \Omega \otimes \Gamma$ (where $\Omega \otimes \Gamma$ is *some* tensor product of Ω and Γ), called the *measurement dynamics*, (ii) an *initial apparatus state* β_o, and (iii) a discrete set $\{\beta_x | x \in E\}$ of *final* apparatus states, indexed by $x \in E$, such that, for every state $\alpha \in \Omega$, the reduced apparatus state μ_2 is given by

$$\mu_2(\alpha \otimes \beta_o) = \sum_{x \in E} \alpha(x)\beta_x.$$

To simplify notation, I'll frequently refer to the mapping μ alone as a measurement, leaving tacit the initial apparatus state. As a further simplification, I shall sometimes write $\mu(\alpha)$ for $\mu(\alpha \otimes \beta_o)$, conflating μ with the corresponding mapping $\Omega \to \Omega \otimes \Gamma$. Context should make this usage unambiguous.

I should stress that this definition is intended to supply only the broadest sort of constraint on what kind of physical process *could* count as a measurement of a discrete observable. In particular, no assumption is made about the final apparatus states β_x, other than that they be distinct: they need not be pure states, they need not equal the conditional states $\mu(\alpha \otimes \beta_o)_{2|x}$, nor need they correspond to the outcomes of some "pointer observable" on the apparatus system. [11]

We can always define a measurement for an any observable E as follows: let β_x be any (distinct) states you like, indexed by $x \in E$, and let α_o be any state in Ω; the affine mapping

$$\mu : \omega \mapsto \sum_{x \in E} \omega_1(x)\alpha_o \otimes \beta_x$$

is measurement of E with final states β_x (one for which the final joint state $\mu(\alpha)$ is separable at that). However, this measurement is quite brutal, in that the post-measurement state of the object system is the constant state α_o, regardless of the original state.

[11] Ruetsche suggests [38] that an affine mapping $\alpha \otimes \beta \mapsto \beta \otimes \alpha$ can count as a measurement, as it perfectly correlates the final apparatus state with the initial system state. But this does not supply a dynamical model of any particular *observable*, which is the game here.

In contrast, the unitary operator U of Equation (3) yields a measurement, defined by

$$\mu_U(\omega)(x \otimes y) = \omega(U^{-1}(x \otimes y))$$

for all unit vectors $x \in \mathbf{H}_O$ and $y \in \mathbf{H}_A$, that is very gentle – indeed, μ_U is an affine automorphism of the joint state space $\Omega(\mathbf{H}_O) \otimes \Omega(\mathbf{H}_A) := \Omega(\mathbf{H}_O \otimes \mathbf{H}_A)$.

Definition 9 Call a measurement *purity-preserving* iff (i) the initial apparatus state β_o is pure, and (ii) the measurement dynamics $\mu : \Omega \otimes \Gamma \to \Omega \otimes \Gamma$ takes pure states to pure states. Call the measurement *reversible* iff μ is an affine automorphism of $\Omega \otimes \Gamma$.

If we are committed to pure-statism, both for the object and apparatus systems prior to interaction, and for the composite system during interaction, then we should certainly require that measurements be purity-preserving. Note that, subject to β_o being pure, reversible measurements must be purity-preserving.

Definition 10 A state α is an *eigenstate* for an observable E if and only if there exists some $x \in E$ with $\alpha(x) = 1$. If *every* pure state $\alpha \in \Omega$ is an eigenstate of E, then E is *classical*.

Note that E is classical iff Ω decomposes as a direct convex sum of the faces $F_x := \{\alpha \in \Omega | \alpha(x) = 1\}$, $x \in E$. In quantum-theoretic terms, this means that E is a (discrete) superselection rule. Note, too, that if Ω is the state space of a test space $\mathfrak{A}$ and every $E \in \mathfrak{A}$ is classical, then every pure state in Ω is dispersion-free.

Lemma 2 *Let μ be a purity-preserving measurement. Then for any pure state $\alpha \in \Omega$ not an eigenstate of E, the state $\mu(\alpha)$ is necessarily entangled.*

Proof: Since α isn't an eigenstate of E, the marginal state $\mu_2(\alpha) = \sum_{x \in E} \alpha(x)\beta_x$ is a mixed state. Thus, the pure state $\mu(\alpha)$ is not a product state. $\square$

We can now argue that $\mu(\alpha)$ is not a *proper* mixture of the pointer states β_x, $x \in E$, exactly as in the version of the quantum measurement problem glossed earlier. In particular – according to this logic – the apparatus system is not actually *in* any of the pointer states β_x. Thus, this version of the quantum-mechanical measurement problem is in fact completely generic, arising whenever we wish to model the measurement of a non-classical observable in terms of the dynamical interaction of object and apparatus systems.

It is important to note the following

Corollary: *If Γ is classical (a simplex), then there exists no purity-preserving measurement of E, unless E is classical.*[12]

[12] It is unfortunate that, in the language adopted here, classical probabilistic models support non-classical observables, namely, noisy, or "fuzzy", versions of random variables. A better term would be welcome.

Proof: As Γ is a simplex, $\Omega \otimes \Gamma = \Omega \otimes_{\max} \Gamma = \Omega \otimes_{\min} \Gamma$, by Proposition 1. Let $\mu : \Omega \to \Omega \otimes \Gamma$ be a purity-preserving measurement. Then for every pure state $\alpha \in \Omega$, $\mu(\alpha)$ is a pure state in $\Omega \otimes_{\min} \Gamma$, hence, a product of pure states – say $\mu(\alpha) = \gamma \otimes \beta$. But then $\mu(\alpha)_1 = \sum_{x \in E} \alpha(x)\beta_x = \beta$. Since β is pure and the states β_x are distinct, there must be a unique $x \in E$ with $\beta_x = \beta$ and $\alpha(x) = 1$. In particular, α is an eigenstate of E. $\Box$

Thus, to resolve our generic measurement problem, it is not sufficient to invoke the existence of classical systems that can serve as measuring devices. In the next two sections, I'll investigate whether a particular strategy for resolving the quantum measurement problem – a very simple form of Everettian relative-state interpretation – can be made to work for a general probabilistic theory.

4.3 Relative and Conditional States

According to Everett [19], if a bipartite quantum system, is in a pure entangled state represented by a vector $v \in \mathbf{H}_1 \otimes \mathbf{H}_2$, where $\mathbf{H}_1$ and $\mathbf{H}_2$ are the Hilbert spaces representing the component systems, then these component systems do not have states of their own in any absolute sense, but only *relative* states. More exactly, Everett notes that v defines, in a perfectly standard way, an operator $\hat{v}$ from $\mathbf{H}_1$ to $\mathbf{H}_2$, defined, for all vectors $x \in \mathbf{H}_1$, by the condition hat

$$\langle \hat{v}(x), y \rangle = \langle v, x \otimes y \rangle$$

for all $y \in \mathbf{H}_2$. Everett regards the vector $\hat{v}(x)$, suitably normalized, as representing the state of system 2 *relative* to system 1's being in the state corresponding to x. In fact, this relative state $v_{\text{rel},x} := \hat{v}(x)/\|\hat{v}(x)\|$ represents nothing other than the conditional state of system 2, given the *outcome* corresponding to $x \in \mathbf{H}_1$. Although Everett mentions this in passing, it is worth spelling out, since the point seems to get lost in many discussions of the Everett interpretation.

Lemma 3 *Let Ω be the state space of a bipartite quantum system with Hilbert space $\mathbf{H} = \mathbf{H}_1 \otimes \mathbf{H}_2$. If $\alpha \in \Omega$ is the pure state associated with a unit vector $v \in \mathbf{H}$, then for any unit vector a in $\mathbf{H}_1$, the conditional state α_a is the pure state associated with the unit vector $v_{rel,x} = \hat{v}(a)/\|\hat{v}(a)\|$.*

Proof: Given orthonormal bases E and F for $\mathbf{H}_1$ and $\mathbf{H}_2$, respectively, we can express the vector v as

$$v = \sum_{xy \in EF} c_{xy} x \otimes y.$$

Note that this defines an operator from $\mathbf{H}_1$ to $\mathbf{H}_2$, namely $\hat{v} : a \mapsto \sum_{xy} c_{xy} \langle x|a \rangle y$. For a fixed a, we are free here to choose the basis E so that $a \in E$; then the foregoing is more simply expressed as $\hat{v}(a) = \sum_{y \in F} c_{ay} y$. Since the vectors $y \in F$ are orthonormal, we have

$$\|\hat{v}(a)\|^2 = \sum_{y \in F} c_{ay}^2 = \alpha(aF).$$

Now the *conditional* state of α, given outcome (represented by) $a \in \mathbf{H}_1$, is given by $\alpha_a(z) = \rho(az)/\alpha(aF)$ for any orthonormal basis F. Now,

$$\begin{aligned}
\alpha(az) &= |\langle v|a \otimes z\rangle|^2 \\
&= |\sum_{xy \in EF} c_{xy}\langle x|a\rangle\langle y|z\rangle|^2 \\
&= |\sum_{y \in F} c_{ay}\langle y|z\rangle|^2. \\
&= |\langle \hat{v}(a)|z\rangle|^2.
\end{aligned}$$

Since the norm squared of $\hat{v}(a)$ is exactly $\alpha(aF)$ for any orthonormal basis F of $\mathbf{H}_2$, normalization on the two sides gives us

$$\alpha_a(z) = \langle v_{\mathrm{rel},a}|z\rangle,$$

as advertised. □

It is striking here that, in quantum mechanics, conditioning a pure bipartite state on an outcome of one system gives a *pure* state of the second. This is also true of classical probability theory, of course; but it *isn't* true of all probabilistic models.

Example: Consider the maximal tensor product of two quantum state spaces. It can be shown (e.g., [5]) that a positive linear mapping $V(\Omega_{\mathbf{H}}) \to V(\Omega_{\mathbf{K}})$ that sends extremal rays to extremal rays is either of the form $\rho \mapsto A\rho A^*$ or $\rho \mapsto A\rho^t A^*$, where $\rho \mapsto \rho^t$ is transposition relative to some orthonormal basis. Mappings of the first kind are completely positive, those of the second are said to be co-completely positive. A positive linear mapping is *decomposable* iff it is a convex combination of completely positive and co-completely positive mappings. In dimension higher than 2, there always exist non-decomposable positive mappings [12]; thus, any extreme, normalized, non-decomposable positive mapping has non-pure conditional states.

This suggests the following terminology:

Definition 11 Say that a probabilistic theory satisfies the *pure conditionalization principle* iff, for every pair of models $(X, \mathfrak{A}, \Omega)$ and $(Y, \mathfrak{B}, \Gamma)$ of the theory, for any pure bipartite state $\omega \in \Omega \otimes \Gamma$, and for any outcomes $x \in X$ and $y \in Y$, the conditional states $\omega_{2|x}$ and $\omega_{1|y}$ are pure in Ω and Γ, respectively.

If we are aiming for an Everettian solution to the generalized measurement problem, and if we wish to adhere (as I take it that Everett did) to a strict pure-statism at the level of the components of a coupled object-plus-apparatus system, then this is the least we should require of a theory (though whether such a component-wise pure-statism is indispensable to a relative-state interpretation is, I think, open to question).

4.4 Correlation and Spectrality

In the standard account of measurement interactions sketched in section 4.1, the final system-plus-apparatus state, given by Equation (4), perfectly correlates the eigenstates of the object system observable E with those of a "pointer observable" of the apparatus system. In fact, *any* unit vector v in the tensor product $\mathbf{H}_1 \otimes \mathbf{H}_2$ of two Hilbert spaces expresses a perfect correlation of this sort between *some* pair of observables. Indeed, suppose that v corresponds to the pure bipartite state ω, and let $\{x_i\}$ be an orthonormal basis diagonalizing the density matrix corresponding to the marginal state Ω. By Lemma 3, the conditional states $\omega_{2|x_i}$ are pure, and hence, correspond to unit vectors y_i in $\mathbf{H}_2$ (namely, $y_i = v_{\mathrm{rel},x_i}$). It can be shown that these are orthonormal, and that in fact,

$$v \;=\; \sum_i \lambda_i x_i \otimes y_i$$

where $\lambda_i = \sqrt{\omega(x_i y_i)}$. This *biorthogonal* or *Schmidt* decomposition of v is unique when the non-zero coefficients λ_i are distinct (which is the case for almost all choices of v).

The biorthogonal decomposition plays an important role in Everett's original formulation of his relative state interpretation [19] (Everett calls it the *canonical representation*)[13], and also in the Modal interpretations of Kochen [31] and Dieks [17]. It allows the *state itself* to select a set of preferred pairs of observables – and, in non-degenerate cases, a unique preferred pair – between which the state establishes a perfect correlation.

To consider the Everett interpretation, suppose a composite object-plus-apparatus system is in a pure state represented by a unit vector $v \in \mathbf{H}_O \otimes \mathbf{H}_A$, with biorthogonal decomposition as in (4). Then we can regard each of the pure tensors $x_i \otimes y_i$ appearing in that decomposition as representing a possible state of affairs (or, more colorfully, a "possible world"); the set of pure tensors arising from the biorthogonal decomposition of v thus gives us a family of pairwise orthogonal states of affairs, in each of which object system and measuring apparatus are perfectly correlated. The state vector v gives us also a probability weight on these, namely $|\langle v, x_i \otimes y_i \rangle|^2 = c_i^2$. Note that any observables diagonal with respect to $E = \{x_i\}$ will be perfectly correlated with corresponding observables diagonal with respect to $\{y_i\}$. Finally, note that these sets of correlated observables are not put into the account by hand, but are determined (at least in the non-degenerate case) by the state ω itself.

At this point, Everettian and modal interpretations differ in their metaphysics: the former conceives that all of the states of affairs (or "worlds") represented by the correlated pairs (x_i, y_i) are equally actual; the latter assumes that exactly one of these pairs is actual. At the present rather abstract level of discussion, it is unclear what basis, other than aesthetics, we could have for preferring one of these points of view to the other. [14]

[13] This has little role in more contemporary Everettian interpretations, which rely instead on ideas from decoherence theory.

[14] It is also unclear to me just how far a many-worlds interpretation, in particular, needs the

In what follows, I'll concentrate on the kind of Everettian theory sketched above. In order to loft a similar interpretation of an otherwise arbitrary probabilistic theory, we would seem to need a stand-in for the biorthogonal decomposition. For simplicity, I'll assume, from this point on, that both object and apparatus systems are described by the same probabilistic model $(X, \mathfrak{A}, \Omega)$.

Definition 12 Call a bipartite state $\omega \in \Omega \otimes \Omega$ *correlating* iff there exist tests $E, F \in \mathfrak{A}$ and a bijection $f : E \to F$ such that, for all $xy \in EF$, $\omega(xy) = 0$ if $y \neq f(x)$. In this case, say that ω *correlates* E and F *via* f.

If $\mathfrak{A}$ is the frame manual of a Hilbert space, then the biorthogonal decomposition theorem is exactly the statement that every pure state is correlating. This is certainly *not* true of probabilistic models generally.

Lemma 4 *The marginals of a correlating bipartite state are spectral.*

Proof: As observed above (Equation (2) in section 3.2), for every $E \in \mathfrak{A}$, we have

$$\omega_2(y) = \sum_{x \in E} \omega_1(x)\omega_{2|x}.$$

If ω is correlated via $f : E \to F$, then we have, for all $x \in E$ with $\omega_1(x) > 0$, that $\omega_{2|x}(y) = 1$ if $y = f(x) \in F$; thus, the conditional states $\omega_{2|x}$, $x \in E$, are distinguishable by $\{f(x)|\omega_1(x) > 0\}$. $\Box$.

If ω correlates tests E and F via a bijection $f : E \to F$, then $\omega_{2|x}(f(x)) = \omega_{1|f(x)}(x) = 1$. If $(X, \mathfrak{A}, \Omega)$ is sharp, this tells us that $\omega_{1|f(x)} = \epsilon_x$ and $\omega_{2|x} = \epsilon_{f(x)}$, where as usual ϵ_x denotes the unique (hence, pure) state making x certain.

Definition 13 Let us say that a probabilistic theory is

(a) *correlational* (for want of a better adjective!) iff, for all state spaces Ω and Γ of the theory, every pure state of $\Omega \otimes \Gamma$ is correlating, and

(b) *strongly correlational* iff, in addition, every state in Ω is the marginal of some pure state of $\Omega \otimes \Gamma$.

Lemma 4 gives us, for any strongly correlational theory, a kind of weak spectral theorem for states: every state is a convex combination of distinguishable states. If the theory in question also satisfies the pure conditionalization principle, or if its models are sharp, then every state is a mixture of distinguishable *pure* states. This would seem to be enough to allow a rudimentary sort of relative state interpretation to fly, even in this still very rarified air. Thus, consider a composite system having two parts, with state space $\Omega \otimes \Gamma$. If the global state ω is biorthogonal with respect to a bijection $f : E \to F$ correlating observables

uniqueness of the biorthogonal decomposition, since different choices of correlated bases $\{x_i\}$ and $\{y_i\}$ would simply correspond to different *sets* of alternative possible worlds – perhaps we should call these alternative "universes"? – all of which one could, in keeping with the spirit of this approach, regard as all equally "real".

E and F (on Ω and on Γ, respectively), then for all $x \in E$, $\omega_{2|x}$ is a pure state making $f(x)$ certain; we can say that this is the state of B *relative to* A's being in a pure state corresponding to (concentrated at) $x \in E$. Similarly, $\omega_{1|y}$ is a pure state making $f^{-1}(y)$ certain. One might wish to speak of each pair (x, y) (or perhaps better, each pair $(\omega_{1|y}, \omega_{2|x})$ with $y = f(x)$) as defining a "world" in which the system and apparatus observables have, respectively, values x and y: then the state ω assigns perfectly classical probabilities to these pairs.

4.5 Measurement again

Of course, this does not, by itself, solve the measurement problem. To do *that*, we need to say something about the measurement dynamics, i.e., the mapping μ. For this to work properly, within the interpretation sketched above, it seems we must require at a minimum that, for every initial state of the system Ω, $\mu(\alpha)$ be a pure state such that $\mu(\alpha)_x = \beta_x$ for every $x \in E$; better still, we should require that $\{\beta_x | x \in E\}$ be sharply distinguishable by an observable correlated with E by $\mu(\alpha)$.

It seems a rather special feature of quantum probabilistic models that this is always possible: the measurement μ_U associated with the unitary operator U on $\mathbf{H}_O \otimes \mathbf{H}_A$ defined by equation (3) takes an initial pure product state $\alpha \otimes \beta_o$ an entangled pure state correlating the observables $E = \{x_i\}$ and $F = \{y_i\}$. Ultimately, what allows us to manufacture the unitary U, and hence the measurement μ_U, is a strong symmetry property shared by quantum and classical test spaces.

Definition 14 A *symmetry* of a model $(X, \mathfrak{A}, \Omega)$ is a bijection $g : X \to X$ such that both $\mathfrak{A}$ and Ω are invariant under g and g^{-1}. Call $(X, \mathfrak{A}, \Omega)$ *fully symmetric* [50] iff, for every pair of tests $E, F \in \mathfrak{A}$, $|E| = |F|$ and, for every bijection $f : E \to F$ there exists a symmetry g such that with $gx = f(x)$ for every $x \in E$.

Both classical and quantum test models are fully symmetric: the former trivially, and the latter because any bijection between two orthonormal bases of a Hilbert space extends to a unitary operator. On the other hand, there is no shortage of exotic non-classical fully-symmetric test spaces that are very far from being quantum [42]. For a simple example, let X denote the set of edges of a tetrahedron, and let $\mathfrak{A}$ denote the collection of triples of edges incident at a vertex.

Recall that a probabilistic model is *sharp* iff each outcome x is made certain by unique (and necessarily therefore pure) state ϵ_x.

Lemma 5 *Suppose that $(X, \mathfrak{A}, \Omega)$ is sharp, and that there exists a fully symmetric separated product $(Z, \mathfrak{C}, \Omega \otimes \Omega)$ of $(X, \mathfrak{A}, \Omega)$ with itself. Then every test in $E \in \mathfrak{A}$ admits a correlating measurement μ with final states ϵ_x for all $x \in E$.*

Proof: Let Fix any $x_o \in E$, and let ϵ_o denote ϵ_{x_o}. Let $f : EE \to EE$ be any bijection such that $g(xx_o) = xx$ for every $x \in E$: by our full symmetry assumption, this bijection extends to a symmetry g of the composite model $(Z, \mathfrak{C}, \Omega \otimes \Omega)$. For any $\omega \in \Omega \otimes \Omega$, let

$$\mu(\omega)(xy) = \omega(g^{-1}(xy)).$$

This defines an affine mapping $\mu : \Omega \otimes \Omega \to \Omega \otimes \Omega$. Now let $\eta = \mu(\alpha \otimes \epsilon_o)$. By equation (2) in section 3 (the law of total probability),

$$\eta_2(xy) = \sum_{x \in E} \eta_1(x)\eta_{2|x}(y).$$

It's enough, then, to show that for all $x \in E$,

(i) $\eta_1(x) = \alpha(x)$ and

(ii) $\eta_{2|x} = \epsilon_x$.

For the first part, note that, for each $x \in E$, $\eta_1(x) = \eta_1(xE) = (\alpha \otimes \epsilon_o)(g^{-1}(xE))$. Now $g^{-1}(xE) = xx_o \cup R$ where $R = g^{-1}(x(E \setminus x))$. For any $uv \in R$, we have $v \neq x_o$ (else $g(uv) = g(ux_o) = uu \in xE$, whence $u = x$, contradiction). So $\alpha \otimes \epsilon_o$ is zero on R. Hence, $\omega_1(x) = (\alpha \otimes \epsilon_o)(xx_o) = \alpha(x)\epsilon_o(x_o) = \alpha(x)$. For the second part, note that

$$\eta_{2|x}(y) = \frac{\eta(xy)}{\eta_1(x)} = \frac{(\alpha \otimes \epsilon_o)(g^{-1}(xy))}{\alpha(x)}.$$

This equals 1 if $y = x$; hence, by sharpness, $\eta_{2|x} = \epsilon_x$. $\square$

Lemma 5 suggests that any theory in which all models are sharp and fully symmetric supports an Everettian interpretation, at least of the simple kind discussed above. It would therefore be of great interest to construct explicit examples of such theories that are neither classical nor quantum.

5 Summary and Further Questions

It seems that many ideas that have become standard in discussions of the interpretation of quantum mechanics can be carried over, with little change, to the much more general framework of abstract probabilistic models. In particular, the phenomenon of entanglement and many of its most familiar consequences are in fact generic features of all non-classical probabilistic theories that use any reasonably general tensor product. Among these consequences is the measurement problem. These facts present a challenge for any proposed realist interpretation of quantum mechanics. If, as I have suggested in Section 1, it is a criterion of adequacy for an interpretation of a physical theory that it help to motivate that theory's formal framework, then a proposed interpretation of quantum mechanics that makes equally good sense for any, or nearly any, probabilistic theory can hardly be adequate.

That said, most realist interpretations of quantum mechanics *do* appear to rely on structural features of Hilbert space quantum mechanics that are at least somewhat special. As we've seen, three conditions underwriting a rudimentary form of the Everett interpretation – that the conditional states of pure bipartite states be pure, that pure entangled states be states of perfect correlation between some pair of discrete observables, and that every mixed state be the marginal of some pure entangled state – are already enough to secure a form of the spectral theorem for states. Moreover, in order for this rudimentary Everettian interpretation to do its intended work of solving the measurement problem, we apparently need to place what look like a strong further constraint on the theory's dynamics. A sufficient condition seems to be that all of the theory's models be both sharp and fully symmetric – but this is a non-trivial condition, to say the least.

All this likely leaves us still some way from Hilbert space, but exactly how far is unclear: the fundamental question I've raised – that of the extent to which the familiar types of no-collapse interpretations of quantum mechanics can be made to work in a setting that is neither classical nor quantum – remains wide open. An obvious next step would be to try reconstruct, in detail, as many of the prominent "realist" interpretations as possible in the context of an otherwise generic probabilistic theory, subject to perhaps to the conditions discussed in the preceding paragraph. (At the same time, of course, it would be desirable to obtain a clean characterization of those theories picked out by these postulates). An important part of such a project would be to study decoherence at this level of abstraction, as modern many-worlds interpretations lean heavily on decoherence to pick out the which branches of the universal state count as potential worlds.

An objection that might be raised against the point of view underlying this paper, and thus against the project I am suggesting here, is that the mathematical framework of a physical theory can be understood as reflecting the *ontology* that the theory embraces. Thus, for example, in the deBroglie-Bohm interpretation, the wave function is a physical field on configuration space, governed by a linear PDE (the Schrodinger equation); it is simply a fact that such objects form a Hilbert space, at least up to certain well-understood mathematical idealizations. This is fair enough, and must be regarded, from the point of view of this paper, as a virtue of the deBroglie-Bohm interpretation. On the other hand, to urge that, similarly, the very ontology of a Many-Worlds interpretation presupposes a Hilbert space structure, would beg a very large question indeed. [15]

[15] Another possible objection is that a probabilistic/operational interpretation that takes "measurements" and "outcomes" as primitive cannot, in principle, accommodate cosmological questions, since one cannot make measurements on the entirety of space-time. But this is objection is hardly compelling. Nothing prohibits one from making measurements locally that are determined by essentially global data; or, to put it more broadly, there is no reason to suppose that a model of the universe can not be made by stitching together models of localized bits thereof, in something like the manner in which a manifold is pieced together from small Euclidean patches.

Acknowledgements: My thinking on the matters discussed here has been influenced by conversations at various times with Howard Barnum, Jonathan Barrett, Harvey Brown, Jeff Bub, Hilary Greaves and Matt Leifer, none of whom bears the slightest blame the results. I also wish to thank Howard Barnum, Matt Leifer and Rob Rynasiewicz for detailed and very helpful comments on earlier drafts of this paper.

References

[1] E. M. Alfsen, *Compact Convex Sets and Boundary Integrals*, Springer, 1971.

[2] H. Araki,

[3] D. Baker, Measurement Outcomes and Probability in Everettian Quantum Mechanics, Studies in the History and Philosophy of Modern Physics, to appear (Phil-Sci archive preprint 2717, 2006)

[4] H. Barnum, J. Barrett, M. Leifer and A. Wilce, Cloning and Broadcasting in Generic Probabilistic Models, quant-ph/0611295 (2006).

[5] H. Barnum, C. Fuchs, J. Renes and A. Wilce, Influence-free states on compound quantum systems, quant-ph/0507108 (2005).

[6] J. Barrett, Information Processing in General Probabilistic Theories, Phys. Rev. A. 75 (2007), 032304-032325 (on-line as quant-ph/0508211, 2006.)

[7] J. Bell, Against 'Measurement', Physics World, Aug. 1990, 33-41.

[8] G. Brassard, Is information the key? *Nature Physics* **1** (2005), 2-4.

[9] J. Bub, *Interpreting the quantum world*, Cambridge, 1997.

[10] J. Bub and R. Clifton, A uniqueness theorem for interpretations of quantum mechanics, *Studies in the History and Philosophy of Modern Physics* **27** (1996), 181-219.

[11] J. Bub, R. Clifton, and H. Halvorson, Characterizing quantum theory in terms of information-theoretic constraints, *Foundations of Physics* **33** (2003) 1561-1591.

[12] M.-D. Choi, Completely positive linear maps on complex matrices, Lin. Alg. Appl. **10** (1975) 285-290.

[13] E. Christensen, Measures on projections and physical states, Comm. Math. Phys. **86** (1982), 529-538.

[14] R. Cooke, and J. Hilgevoord, A new approach to equivalence in quantum logic, in E. Beltrametti and B. van Fraassen (eds.), *Current Issues in Quantum Logic*, Plenum, 1981.

[15] G. D'Ariano, Operational axioms for C^*-algebraic formulation of quantum mechanics, quant-ph/0701209 (2007)

[16] E. B. Davies and J. T. Lewis, An operatinoal approach to quantum probability, *Comm. Math. Phys.* **17** (1970) 239-260.

[17] D. Dieks, Quantum mechanics without the projection postulate, *Foundations of Physics* **19** (1989), 1397-1423.

[18] C. M. Edwards, The operational approach to algebraic quantum theory I, *Comm. Math. Phys.* **16** (1970), 207-230.

[19] H. Everett III, The theory of the universal wave function, reprinted in Bryce S. DeWitt and Neill Graham, *The Many Worlds Interpretation of Quantum Mechanics*, Princeton, 1973.

[20] D. J. Foulis and C. H. Randall, Empirical Logic and Tensor Products, in *Interpretations and Foundations of Quantum Mechanics*, H. Neumann, ed., Bibliographisches Institut, Wissenschaftsverlag, Mannheim, 1981.

[21] C. Fuchs, Quantum Mechanics as quantum information, mostly, *Journal of Modern Optics*, **50** (2003), 987-1023.

[22] D. J. Foulis, C. Piron and C. H. Randall, Realism, operationalism and quantum mechanics, *Foundations of Physics* (1983).

[23] A. Gleason, Measures on the closed subspaces of a Hilbert space, *J. Math. and Mechanics* **6** (1957) 885-893.

[24] P. Goyal, Information-theoretic approach to quantum theory, I: the abstract quantum formalism, quant-ph/0702124v1 (2007).

[25] H. Halvorson, A note on information theoretic characterizations of physical theories, *Studies in History and Philosophy of Modern Physics* **35** (2004), 277293.

[26] H. Halvorson, No eliminative materialism, no Measurement Problem, preprint, 2005.

[27] L. Hardy, Quantum theory from five reasonable axioms, quant-ph/00101012.

[28] A. Holevo, *Probabilistic and statistical aspects of quantum theory*, North-Holland, 1982.

[29] M. Kläy, Einstein-Podolski-Rosen Experiments: the structure of the sample space, *Foundations of Physics Letters* **1** (1988) 205-244.

[30] M. Kläy, D. J. Foulis and C. H. Randall, Tensor products and probability weights, *Int. J. Theor. Phys.* **26** (1987) 199-219.

[31] S. Kochen, A new interpretation of quantum mechanics, in P. Lahti and P. Mittelstaedt, eds., *Symposium on the Foundations of Modern Physics*, World Scientific, 1985.

[32] G. Ludwig, *An Axiomatic Basis for Quantum Mechanics*, vols. I, II, Springer-Verlag 1985.

[33] G. Mackey, *Mathematical Foundations of Quantum Mechanics*, Addison Wesley, 1963

[34] I. Namioka and R. R. Phelps, Tensor products of compact convex sets, *Pacific J. Math.* **9** (1969) 469-480.

[35] I. Pitowski, Quantum mechanics as a theory of probability, quant-ph/0510095 (2005).

[36] S. Popescu and D. Rohrlich, Quantum nonlocality as an axiom, *Foundations of Physics* **24** (1994), pp. 379–385.

[37] C. Rovelli, Relational quantum mechanics, *Int. J. Theor. Phys.* **35** (1996) 1637-1678.

[38] L. Ruetsche, Van Fraassen on Preparation and Measurement, Philosophy of Science **63** (1996), S338-S346.

[39] E. Schrödinger, Discussion of probability relations between separated systems, *Proceedings of the Cambridge Philosophical Society* **31** (1935), 555-563

[40] F. W. Shultz, *A characterization of the state spaces of orthomodular lattices*, *J. Comb. Theory ser. A* **17** (1974) 317-328.

[41] R. Spekkens, In defense of the epistemic view of quantum states: a toy theory quant-ph/0401052, 2005.

[42] Q. Tran and A. Wilce, Covariance in quantum logic, Int. J. Theor. Phys., to appear.

[43] B. van Fraassen, *Quantum Mechanics: an Empiricist View*, Oxford, 1991.

[44] V. S. Varadarajan *The Geometry of Quantum Theory* Springer, 1989.

[45] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik*, Springer, 1932.

[46] N. Wallach, An unentangled Gleason's Theorem, preprint arXiv:quant-ph/0002058, 2000

[47] A. Wilce, The tensor product of frame manuals, *Int. J. Theor. Phys.* **29** (1990), 805-814.

[48] A. Wilce, Tensor Products in Generalized Measure Theory, *Int. J. Theor. Physics* **31** (1992) 1915-1928.

[49] A. Wilce, Quantum logic and quantum probability, *Stanford Encyclopedia of Philosophy*, February 2002; revised March 2006.

[50] A. Wilce, Symmetry and topology in quantum logic, Int. J. Theor. Phys. **44** (2005), 2303-2316.

[51] A. Wilce, Test Spaces, in D. Gabbay, K. Engesser and D. Lehman (eds.), *The Handbook of Quantum Logic*, vol. II, Elsevier, to appear.

[52] G. Wittstock, Ordered normed tensor products, in A. Hartkamper and H. Neumann (eds.), *Foundations of Quantum Mechanics and Ordered Linear Spaces*, Springer Lecture Notes in Physics **29** (1974).

[53] R. Wright, Spin Manuals, in A. R. Marlow (ed.), *Mathematical Foundations of Quantum Theory*, Academic Press, 1978.

[54] F. J. Yeadon, Finitely additive measures on projections in finite W^*-algebras, Bull. London Math. Soc. **16** (1984), 145-150.