# A Royal Road to Quantum Theory
# (or Thereabouts)

Alexander Wilce

Department of Mathematics, Susquehanna University

July 1, 2016

## Abstract

This paper fails to derive quantum mechanics from a few simple postulates. But it gets very close — and it does so without much exertion. More exactly, I obtain a representation of finite-dimensional probabilistic systems in terms of euclidean Jordan algebras, in a strikingly easy way, from simple assumptions. This provides a framework within which real, complex and quaternionic QM can play happily together, and allows some — but not too much — room for more exotic alternatives.

## 1  Introduction and Overview

Whatever else it may be, Quantum mechanics (QM) is a machine for making *probabilistic* predictions about the results of measurements. To this extent, QM is, at least in part, about information. Recently, it has become clear that the formal apparatus of quantum theory, at least in finite dimensions, can be recovered from constraints on how physical systems store and process information. To this extent, finite-dimensional QM is *just* about information.

The broad idea of regarding QM in this way, and of attempting to derive its formal apparatus from simple operational or probabilistic axioms, is not new. Efforts in this direction go back at least to the work of Von Neumann [33], and include also attempts by Schwinger [31], Mackey [25], Ludwig [24], Piron [29], and many others. However, the consensus is that these were not entirely successful: partly because the results they achieved (e.g., Piron's well-known representation theorem) did not rule out certain rather exotic alternatives to QM, but mostly because the axioms deployed seem, in retrospect, to lack sufficient physical or operational motivation.

More recently, with inspiration from quantum information theory, attention has focussed on finite-dimensional systems, where the going is a bit easier. At the same time, and

1

perhaps more significantly, quantum information theory prompts us to treat properties of composite systems as fundamental, where earlier work focussed largely on systems in isolation.[1]These shifts of emphasis were strikingly illustrated by the work of Hardy [19], who presented five simple, broadly information-theoretic postulates governing the states and measurements associated with a physical system, and showed that these lead to a very restricted class of theories parametrized by a positive integer, with finite-dimensional quantum and classical probability theory as the first two entries. Following this lead, several recent papers, notably [10, 12, 27], have derived standard formulation of finite-dimensional QM from various packages of axioms governing the information-carrying and information-processing capacity of finite-dimensional probabilistic systems.

**Problems with existing approaches** These recent reconstructive efforts suffer from two related problems. First, they make use of assumptions that seem *too strong*; secondly, in trying to derive *exactly* complex, finite-dimensional quantum theory, they derive *too much*.

All of the cited papers assume *local tomography*. This is the doctrine that the state of a bipartite composite system is entirely determined by the joint probabilities it assigns to outcomes of measurements on the two subsystems. This rules out both real and quaternionic QM, both of which are legitimate quantum theories [3]. These papers also make some version of a *uniformity assumption*: that all systems having the same information-carrying capacity are isomorphic, or that all systems are composed, in a uniform way, from "bits" of a uniform structure.[2] This rules out systems involving superselection rules, i.e., those that admit both real *and* classical degrees of freedom. [3] More seriously, it rules out any theory that includes, e.g., real *and* complex, or real *and* quaternionic systems, as the state spaces of the bits of these theories have different dimensions. As I'll discuss below, one can indeed construct mathematically reasonable theories that embrace finite-dimensional quantum systems of all three types.

An additional shortcoming, not related to the exclusion of real and quaternionic QM, is the assumption (explicit in [27] for bits) that all positive affine functionals on the state space that take values between 0 and 1, represent physically accessible effects. From an operational point of view, this principle (called the "no-restriction hypothesis" in [18]) seems to call for further motivation.

**Another approach** In these notes, I'm going to describe an alternative approach that avoids these difficulties. This begins by isolating two striking features shared by classical and quantum probabilistic systems. The first is the possibility of finding a joint state that perfectly correlates a system $A$ with an isomorphic system $\overline{A}$ — call it a *conjugate*

---

[1]A recent exception to this trend is the paper [9] of Barnum, Müller and Ududec.

[2]Here, "information capacity" means essentially the maximum number of states that can be sharply distinguished (distinguished with probability one) from one another by a single measurement, and a bit is a system for which this number is two.

[3]for example, the quantum system corresponding to $M_2(\mathbb{C}) \oplus M_2(\mathbb{C})$, corresponding to a classical choice between one of two qubits, has the same information-carrying capacity as a single, four-level quantum system.

system — in the sense that every basic observable on $A$ is perfectly correlated with the corresponding observable on $\overline{A}$. In finite-dimensional QM, where $A$ is represented by a finite-dimensional Hilbert space $\mathcal{H}$, $\overline{A}$ corresponds to the conjugate Hilbert space $\overline{\mathcal{H}}$, and the perfectly correlating state is the maximally entangled "EPR" state on $\mathcal{H} \otimes \overline{\mathcal{H}}$.

The second feature is the existence of what I call *filters* associated with basic observables. These are processes that independently attenuate the "response" of each outcome by some specified factor. Such a process will generally not preserve the normalization of states, but up to a constant factor, in both classical and quantum theory one can prepare any desired state by applying a suitable filter to the maximally mixed state. Moreover, when the target state is not singular (that is, when it does not assign probability zero to any nonzero measurement outcome), one can reverse the filtering process, in the sense that it can be undone by another process with positive probability.

The upshot is that all probabilistic systems having conjugates and a sufficiently lavish supply of reversible filters can be represented by *formally real Jordan algebras*, a class of structures that includes real, complex and quaternionic quantum systems, and just two further (and well studied) additional possibilities, which I'll discuss below.

In addition to leaving room for real and quaternionic quantum mechanics — which I take to be a virtue — this approach has another advantage: *it is much easier!* The assumptions involved are few and easily stated, and the proof of the main technical result (Lemma 1 in Section 4) is short and straightforward. By contrast, the mathematical developments in the papers listed above are significantly more difficult, and ultimately lean on the (even more difficult) classification of compact groups acting on spheres. My approach, too, leans on a received result, but one that's relatively accessible. This is the Koecher-Vinberg theorem, which characterizes formally real, or euclidean, Jordan algebras in terms of ordered real vector spaces with homogeneous, self-dual cones. A short and non-taxing proof of this classical result can be found in [13].

These ideas were developed in [36, 37] and especially [38], of which this paper is, to an extent, a summary. However, the presentation here is slightly different, and some additional ideas are also explored. I also briefly discuss recent work with Howard Barnum and Matthew Graydon [4] on the construction of probabilistic theories in which real, complex and quaternionic quantum systems coexist.

**A bit of background** At this point, I'd better pause to explain some terms. A Jordan algebra is a real commutative algebra — a real vector space $\mathbf{E}$ with a commutative bilinear multiplication $a, b \mapsto a \bullet b$ — having a multplicative unit $u$, and satisfying the *Jordan identity*: $a^2 \bullet (a \bullet b) = a \bullet (a^2 \bullet b)$, for all $a, b, c \in \mathbf{E}$, where $a^2 = a \bullet a$. A Jordan algebra is *formally real* if sums of squares of nonzero elements are always nonzero. The basic, and motivating, example is the space $\mathcal{L}(\mathcal{H})$ of self-adjoint operators on a complex Hilbert space, with Jordan product given by $a \bullet b = \frac{1}{2}(ab + ba)$. Note that here $a \bullet a = aa$, so the notation $a^2$ is unambiguous. To see that $\mathcal{L}(\mathcal{H})$ is formally real, just note that $a^2$ is always a positive operator.

If $\mathcal{H}$ is finite dimensional, $\mathcal{L}(\mathcal{H})$ carries a natural inner product, namely $\langle a, b \rangle = \text{Tr}(ab)$. This plays well with the Jordan product: $\langle a \bullet b, c \rangle = \langle b, a \bullet c \rangle$. More generally, a finite-dimensional Jordan algebra equipped with an inner product having this property is said to be *euclidean*. For finite-dimensional Jordan algebras, being formally real and being euclidean are equivalent [13]. In what follows, I'll abbreviate "euclidean Jordan algebra" to EJA.

Jordan algebras were originally proposed, with what now looks like slightly thin motivation, by P. Jordan [21]: if $a$ and $b$ are quantum-mechanical observables, represented by $a, b \in \mathcal{L}(\mathcal{H})$, then while $a+b$ is again self-adjoint, $ab$ is not, unless $a$ and $b$ commute; however, their average, $a \bullet b$, *is* self-adjoint, and thus, represents another observable. Almost immediately, Jordan, von Neumann and Wigner showed [22] that all formally real Jordan algebras are direct sums of simple such algebras, with the latter falling into just five classes, parametrized by positive integers $n$: self-adjoint parts of matrix algebras $M_n(\mathbb{F})$, where $\mathbb{F} = \mathbb{R}, \mathbb{C}$ or $\mathbb{H}$ (the quaternions) or, for $n = 3$, over $\mathbb{O}$ (the Octonions); and also what are called *spin factors* $V_n$ (closely related to Clifford algebras). [4] In all but one case, one can show that a simple Jordan algebra is a Jordan subalgebra of $M_n(\mathbb{C})$ for suitable $n$. The *exceptional Jordan algebra*, $M_3(\mathbb{O})_{\text{sa}}$, admits no such representation.

Besides this classification theorem, there is only one other important fact about euclidean Jordan algebras that's needed for what follows. This is the Koecher-Vinberg (KV) theorem alluded to above. Any EJA is also an ordered vector space, with positive cone $\mathbf{E}_+ := \{a^2 | a \in A\}$.[5] This cone has two special features: first, it is *homogeneous*, i.e, for any points $a, b$ in the *interior* of $\mathbf{E}_+$, there exists an automorphism of the cone — a linear isomorphism $\mathbf{E} \rightarrow \mathbf{E}$, taking $\mathbf{E}_+$ onto itself — that maps $a$ onto $b$. In other words, the group of automorphisms of the cone acts transitively on the cone's interior. The other special property is that $\mathbf{E}_+$ is *self-dual*. This means that $\mathbf{E}$ carries an inner product — in fact, the given one making $\mathbf{E}$ euclidean — such that $a \in \mathbf{E}_+$ iff $\langle a, b \rangle \geq 0$ for all $b \in \mathbf{E}_+$. In the following, by a *euclidean order unit space*, I mean an ordered vector space $\mathbf{E}$ equipped with an inner product $\langle, \rangle$ with $\langle a, b \rangle \geq 0$ for all $a, b \in \mathbf{E}_+$, and a distinguished order-unit $u$. I will say that such a space $\mathbf{E}$ is *HSD* iff $\mathbf{E}_+$ is homogeneous, and also self-dual with respect to the given inner prouct.

**Theorem [Koecher 1958; Vinberg 1961]:** *Let $\mathbf{E}$ be a finite-dimensional euclidean order-unit space. If $\mathbf{E}$ is HSD, then there exists a unique product $\bullet$ with respect to which $\mathbf{E}$ (with its given inner product) is a euclidean Jordan algebra, $u$ is the Jordan unit, and $\mathbf{E}_+$ is the cone of squares.*

It seems, then, that if we can motivate a representation of physical systems in terms of HSD order-unit spaces, we will have "reconstructed" what we might call (with a little license) finite-dimensional *Jordan*-quantum mechanics. In view of the classification theorem glossed above, this gets us into the neighborhood of orthodox QM, but still leaves open

---

[4]There is some overlap: $V_2 \simeq M_2(\mathbb{R}), V_3 \simeq M_2(\mathbb{C})$ and $V_5 \simeq M_2(\mathbb{H})$.

[5]Recall here than an ordered vector space is a real vector space, call it $\mathbf{E}$, spanned by a distinguished convex cone $\mathbf{E}_+$ having its vertex at the origin. Such a cone induces a translation-invariant partial order on $\mathbf{E}$, namely $a \leq b$ iff $b - a \in \mathbf{E}_+$.

the possibilty of taking real and quaternionic quantum systems seriously. (It also leaves the door open to two possibly unwanted guests, namely spin factors and the exceptional Jordan algebra. I'll discuss below some constraints that at least bar the latter.)

**Some Notational Conventions**: My notation is *mostly* consistent with the following conventions (more standard in the mathematics than the physics literature, but in places slightly excentric relative to either). Capital Roman letters $A, B, C$ serve as labels for systems. Vectors in a Hilbert space $\mathcal{H}$ are denoted by little roman letters $x, y, z$ from the end of the alphabet. Operators on $\mathcal{H}$ will usually be denoted by little roman letters $a, b, c, ...$ from the begining of the alphabet. Roman letters $t, s$ typically stand for real numbers. The space of linear operators on $\mathcal{H}$ is denoted $\mathcal{L}(\mathcal{H})$; $\mathcal{L}_h(\mathcal{H})$ is the (real) vector space of hermitian operators on $\mathcal{H}$.

As above, the conjugate Hilbert space is denoted $\overline{\mathcal{H}}$. I'll write $\overline{x}$ for the vectors in $\overline{\mathcal{H}}$ corresponding to $x \in \mathcal{H}$. From a certain point of view, this is the same vector; the bar serves to remind us that $\overline{cx} = \overline{c}\,\overline{x}$ for scalars $c \in \mathbb{C}$. (Alternatively, one can regard $\overline{\mathcal{H}}$ as the space of "bra" vectors $\langle x|$ corresponding to the "kets" $|x\rangle$ in $\mathcal{H}$, i.e., as the dual space of $\mathcal{H}$.)

The inner product of $x, y \in \mathcal{H}$ is written as $\langle x, y \rangle$, and is linear in the *first* argument (if you like: $\langle x, y \rangle = \langle y|x \rangle$ in Dirac notation.). The inner product on $\overline{\mathcal{H}}$ is then $\langle \overline{x}, \overline{y} \rangle = \langle y, x \rangle$. The rank-one projection operator associated with a unit vector $x \in \mathcal{H}$ is $p_x$. Thus, $p_x(y) = \langle y, x \rangle x$[6]. I denote functionals on $\mathcal{L}(\mathcal{H})$ by little Greek letters, e.g., $\alpha, \beta...$, and operators on $\mathcal{L}(\mathcal{H})$ by capital Greek letters, e.g., $\Phi$. Two exceptions to this scheme: a generic density operator on $\mathcal{H}$ is denoted by the capital Roman letter $W$, and a certain unit vector in $\mathcal{H} \otimes \overline{\mathcal{H}}$ is denoted by the capital Greek letter $\Psi$. With luck, context will help keep things straight.

# 2   Homogeneity and self-duality in quantum theory

Why *should* a probabilistic physical system be represented by a euclidean order-unit space that's either homogeneous or self-dual? One place to start hunting for an answer might be to look a standard quantum probability theory, to see if we can isolate, *in operational terms*, what makes *this* self-dual and homogeneous.

**Correlation and self-duality** Let $\mathcal{H}$ be a finite-dimensional complex Hilbert space, representing some finite-dimensional quantum system. As above, $\mathcal{L}_h(\mathcal{H})$ is the space of self-adjoint operators on $\mathcal{H}$. The system's states are represented by density operators, i.e., positive trace-one operators $W \in \mathcal{L}_h(\mathcal{H})$; possible measurement-outcomes are represented by effects, i.e., positive operators $a \in \mathcal{L}_h(\mathcal{H})$ with $a \leq \mathbf{1}$. The probability of observing effect $a$ in state $W$ is $\text{Tr}(Wa)$. If $W$ is a pure state, i.e., $W = p_v$ where $v$ is a unit vector

---

[6]This is my main reason for abjuring Dirac notation: $\langle y, x \rangle x$ is simply nicer than $\langle x|y\rangle x$, and I'm too wedded to having scalars on the left to write $x\langle x|y\rangle$.

in $\mathcal{H}$, then $\mathrm{Tr}(Wa) = \langle av, v \rangle$; by the same token, if $a = p_x$, then $\mathrm{Tr}(Wa) = \langle Wx, x \rangle$.

For $a, b \in \mathcal{L}_h(\mathcal{H})$, let $\langle a, b \rangle = \mathrm{Tr}(ab)$. This is an inner product. By the spectral theorem, $\mathrm{Tr}(ab) \geq 0$ for all $b \in \mathcal{L}_h(\mathcal{H})_+$ iff $\mathrm{Tr}(ap_x) \geq 0$ for all unit vectors $x$. But $\mathrm{Tr}(ap_x) = \langle ax, x \rangle$. So $\mathrm{Tr}(ab) \geq 0$ for all $b \in \mathcal{L}_h(\mathcal{H})_+$ iff $a \in \mathcal{L}_h(\mathcal{H})_+$, i.e., the trace inner product is self-dualizing. But this now leaves us with the

**Question:** *What does the trace inner product represent, probabilistically?*

Let $\overline{\mathcal{H}}$ be the conjugate Hilbert space to $\mathcal{H}$. Suppose $\mathcal{H}$ has dimension $n$. Any unit vector $\Psi$ in $\mathcal{H} \otimes \overline{\mathcal{H}}$ gives rise to a joint probability assignment to effects $a$ on $\mathcal{H}$ and $\bar{b}$ on $\overline{\mathcal{H}}$, namely $\langle (a \otimes \bar{b})\Psi, \Psi \rangle$. Consider the *EPR state* for $\mathcal{H} \otimes \overline{\mathcal{H}}$ defined by the unit vector

$$\Psi = \frac{1}{\sqrt{n}} \sum_{x \in E} x \otimes \overline{x} \in \mathcal{H} \otimes \overline{\mathcal{H}},$$

where $E$ is any orthonormal basis for $\mathcal{H}$. A straightforward computation shows that

$$\langle (a \otimes \bar{b})\Psi, \Psi \rangle = \tfrac{1}{n} \mathrm{Tr}(ab).$$

In other words, *the normalized trace inner product* just is *the joint probability function $\eta$ determined by the pure state vector $\Psi$*!

As a consequence, the state represented by $\Psi$ has a very strong correlational property: if $x, y$ are two orthogonal unit vectors with corresponding rank-one projections $p_x$ and $p_y$, we have $p_x p_y = 0$, so $\langle (p_x \otimes \overline{p_y})\Psi, \Psi \rangle = 0$. On the other hand, $\langle (p_x \otimes \overline{p_x})\Psi, \Psi \rangle = \tfrac{1}{n} \mathrm{Tr}(p_x) = \tfrac{1}{n}$. Hence, $\eta$ *perfectly, and uniformly, correlates* every basic measurement (orthonormal basis) of $\mathcal{H}$ with its counterpart in $\overline{\mathcal{H}}$.

**Filters and homogeneity** Next, let's see why the cone $\mathcal{L}_h(\mathcal{H})_+$ is homogeneous. Recall that this means that any state in the interior of the cone — here, any non-singular density operator — can be obtained from any other by an order-automorphism of the cone. But in fact, something better is true: this order-automorphism can be chosen to represent a probabilistically reversible physical process, i.e., an invertible CP mapping with CP inverse.

To see how this works, suppose $W$ is a positive operator on $\mathcal{H}$. Consider the pure CP mapping $\Phi_W : \mathcal{L}_h(\mathcal{H}) \to \mathcal{L}_h(\mathcal{H})$ given by

$$\Phi_W(a) = W^{1/2} a W^{1/2}.$$

Then $\Phi_W(\mathbf{1}) = W$. If $W$ is nonsingular, so is $W^{1/2}$, so $\Phi_W$ is invertible, with inverse $\Phi_W^{-1} = \Phi_{W^{-1}}$, again a pure CP mapping. Now given another nonsingular density operator $M$, we can get from $W$ to $M$ by applying $\Phi_M \circ \Phi_{W^{-1}}$.

All well and good, but we are still left with the

**Question:** *What does the mapping $\Phi_W$ represent, physically?*

To answer this, suppose $W$ is a density operator, with spectral expansion $W = \sum_{x \in E} t_x p_x$. Here, $E$ is an orthonormal basis for $\mathcal{H}$ diagonalizing $W$, and $t_x$ is the eigenvalue of $W$ corresponding to $x \in E$. Then, for each vector $x \in E$,

$$\Phi_W(p_x) = t_x p_x$$

where $p_x$ is the projection operator associated with $x$. We can understand this to mean that $\Phi_W$ acts as a *filter* on the test $E$: the *response* of each outcome $x \in E$ is attenuated by a factor $0 \le t_x \le 1$.[7] Thus, if $M$ is another density operator on $\mathcal{H}$, representing some state of the corresponding system, then the probability of obtaining outcome $x$ after preparing the system in state $M$ and applying the process $\Phi$, is $t_x$ times the probability of $x$ in state $M$. (In detail: suppose $p_x$ is the rank-one projection operator associated with $x$, and note that $W^{1/2}p_x = p_x W^{1/2} = t_x^{1/2}p_x$. Thus,

$$\begin{aligned}\Phi_W(M)(x) = \mathrm{Tr}(W^{1/2}MW^{1/2}p_x) &= \mathrm{Tr}(W^{1/2}Mt_x^{1/2}p_x) = \mathrm{Tr}(t_x^{1/2}p_x W^{1/2}M) \\ &= \mathrm{Tr}(t_x p_x M) = t_x \mathrm{Tr}(Mp_x).\end{aligned}$$

If we think of the basis $E$ as representing a set of alternative *channels* plus *detectors*, as in the figures below, we can add a classical filter attenuating the response of one of the detectors — say, $x$ — by a fraction $t_x$ What the computation above tells us is that we can achieve the same result by applying a suitable CP map to the system's state. Moreover, this can be done independently for each outcome of $E$. In Figure 1 below, this is illustrated for 3-level quantum system: $E = \{x, y, z\}$ is an orthonormal basis, representing three possible outcomes of a Stern-Gerlach-like experiment; the filter $\Phi$ acts on the system's state in such a way that the probability of outcome $x$ is attenuated by a factor of $t_x = 1/2$, while outcomes $y$ and $z$ are unaffected.
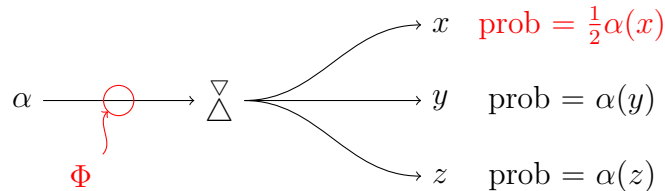


Figure 1: $\Phi$ attenuates $x$'s sensitivity by $1/2$.

If we apply the filter $\Phi_W$ to the maximally mixed state $\frac{1}{n}\mathbf{1}$, we obtain $\frac{1}{n}W$. Thus, we can *prepare* $W$, up to normalization, by applying the filter $\Phi_W$ to the maximally mixed state.

**Filters are Symmetric** Here is a final observation, linking these last two: The filter $\Phi_W$ is *symmetric* with respect to the uniformly correlating state $\eta$, in the sense that

$$\langle (\Phi_W(a) \otimes \bar{b})\Psi, \Psi \rangle = \langle (a \otimes \overline{\Phi}_W(b))\Psi, \Psi \rangle$$

---

[7]My usage here is slightly non-standard, in that I allow filters that "pass" the system with a probability strictly between 0 and 1.

for all effects $a, b \in \mathcal{L}_h(\mathcal{H})_+$. Remarkably, *this is all that's needed* to recover the Jordan structure of finite-dimensional quantum theory: the existence of a conjugate system, with a uniformly correlating joint state, plus the possibility of preparing non-singular states by means of filters that are symmetric with respect to this state, and doing so reversibly when the state is nonsingular.

In very rough outline, the argument is that states preparable (up to normalization) by symmetric filters have spectral decompositions, and the existence of spectral decompositions makes the uniformly correlating joint state a self-dualizing inner product. But to spell this out in any precise way, I need a general mathematical framework for discussing states, effects and processes in abstraction from quantum theory. The next section reviews the necessary apparatus.

# 3  General probabilistic theories

A characteristic feature of quantum mechanics is the existence of incompatible, or non-comeasurable, observables. This suggests the following simple, but very fruitful, notion:

**Definition:** A *test space* is a collection $\mathcal{M}$ of non-empty sets $E, F, ....$, each representing the *outcome-set* of some measurement, experiment, or *test*. At the outset, one makes no special assumptions about the combinatorial structure of $\mathcal{M}$. In particular, distinct tests are permitted to overlap. Let $X := \bigcup \mathcal{M}$ denote the set of all outcomes of all tests in $\mathcal{M}$: a *probability weight* on $\mathcal{M}$ is a function $\alpha : X \to [0,1]$ such that $\sum_{x \in E} \alpha(x) = 1$ for every $E \in \mathcal{M}$.[8]

It can happen that a test space admits no probability weights at all. However, to serve as a model of a real family of experiments associated with an actual physical system, a test space should obviously carry a lavish supply of such weights. One might want to single out some of these as describing physically (or otherwise) possible states of the system. This suggests the following

**Definition:** A *probabilistic model* is a pair $A = (\mathcal{M}, \Omega)$, where $\mathcal{M}$ is a test space and $\Omega$ is some designated convex set of probability weights, called the *states* of the model.

The definition is deliberately spare. Nothing prohibits us from adding further structure (a group of symmetries, say, or a toplogy on the space of outcomes). However, no such additional structure is needed for the results I'll discuss below. I'll write $\mathcal{M}(A), X(A)$ and $\Omega(A)$ for the tet space, associated outcome space, and state space of a model $A$. The

---

[8]Test spaces were introduced and studied by D. J. Foulis and C. H. Randall and their students (of whom I'm one) in a long series of papers beginning around 1970. The original term for a test was an *operation*, which has the advantage of signaling that the concept has wider applicability than simply reading a number off a meter: anything an agent can *do* that leads to a well-defined, exhaustive set of mutually exclusive outcomes, defines an operation. Accordingly, test spaces were originally called "manuals of operations".

convexity asumption on $\Omega(A)$ is intended to capture the possibility of forming mixtures of states. To allow the modest idealization of taking outcome-wise limits of states to be states, I will also assume that $\Omega(A)$ is closed as a subset of $[0,1]^{X(A)}$ (in its product topology). Note that this makes $\Omega(A)$ compact, and so, guarantees the existence of pure states, that is, extreme points of $\Omega(A)$. If $\Omega(A)$ is the set of *all* probability weights on $\mathcal{M}(A)$, I'll say that $A$ has a *full state space*.

**Two Bits** Here is a simple but instructive illustration of these notions. Consider a test space $\mathcal{M} = \{\{x, x'\}, \{y, y'\}\}$. Here we have two tests, each with two outcomes. We are permitted to perform either test, but not both at once. A probability weight is determined by the values it assigns to $x$ and to $y$, and — since the sets $\{x, x'\}$ and $\{y, y'\}$ are disjoint — these values are independent. Thus, geometrically, the space of all probability weights is the unit square in $\mathbb{R}^2$ (Figure 2(a), below). To construct a probabilistic model, we can choose any closed, convex subset of the square for $\Omega$. For instance, we might let $\Omega$ be the convex hull of the four probability weights $\delta_x, \delta_{x'}, \delta_y$ and $\delta_{y'}$ corresponding to the midpoints of the four sides of the square, as in Figure 2(b) — that is,

$$\delta_x(x) = 1, \ \delta_x(x') = 0, \ \delta_x(y) = \delta_x(y') = 1/2,$$

$$\delta_{x'}(x) = 0, \ \delta_{x'}(x') = 1, \ \delta_{x'}(y) = \delta_{x'}(y') = 1/2,$$

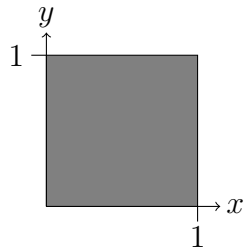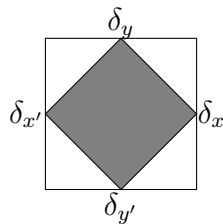and similarly for $\delta_y$ and $\delta_{y'}$.



Figure 2(a)          Figure 2(b)

The model of Figure 2(a), in which we take $\Omega$ to the be entire set of probability weights on $\mathcal{M} = \{\{x, x'\}, \{y, y'\}\}$ is sometimes called the *square bit*. I'll call the model of Figure 2(b) the *diamond bit*.

**Classical, Quantum and Jordan Models** If $E$ is a finite set, the corresponding *classical model* is $A(E) = (\{E\}, \Delta(E))$ where $\Delta(E)$ is the simplex of probability weights on $E$. If $\mathcal{H}$ is a finite-dimensional complex Hilbert space, let $\mathcal{M}(\mathcal{H})$ denote the set of orthonormal bases of $\mathcal{H}$: then $X = \bigcup \mathcal{M}(\mathcal{H})$ is the unit sphere of $\mathcal{H}$, and any density operator $W$ on $\mathcal{H}$ defines a probability weight $\alpha_W$, given by $\alpha_W(x) = \langle Wx, x \rangle$ for all $x \in X$. Letting $\Omega(\mathcal{H})$ denote the set of states of this form, we obtain the *quantum model*, $A(\mathcal{H}) = (\mathcal{M}(\mathcal{H}), \Omega(\mathcal{H}))$, associated with $\mathcal{H}$.[9]

More generally, every euclidean Jordan algebra $\mathbf{E}$ gives rise to a probabilistic model as follows. A minimal or *primitive* idempotent of $\mathbf{E}$ is an element $p \in \mathbf{E}$ with $p^2 = p$ and,

---

[9]The content of Gleason's Theorem is that $A(\mathcal{H})$ has a full state space for $\dim(\mathcal{H}) > 2$. We will not need this fact.

for $q = q^2 < p$, $q = 0$. A *Jordan frame* is a maximal pairwise orthogonal set of primitive idempotents. Let $X(\mathbf{E})$ be the set of primitive idempotents, let $\boldsymbol{\mathcal{M}}(\mathbf{E})$ be the set of Jordan frames, and let $\Omega(\mathbf{E})$ be the set of probability weights of the form $\alpha(p) = \langle a, p \rangle$ where $a \in \mathbf{E}_+$ with $\langle a, u \rangle = 1$. This data defines the *Jordan model* $A(\mathbf{E})$ associated with $\mathbf{E}$. In the case where $\mathbf{E} = \mathcal{L}_h(\boldsymbol{\mathcal{H}})$ for a finite-dimensional Hilbert space $\boldsymbol{\mathcal{H}}$, this *almost* gives us back the quantum model $A(\boldsymbol{\mathcal{H}})$: the difference is that we replace unit vectors by their associated projection operators, thus conflating outcomes that differ only by a phase.

**Sharp Models** Jordan models enjoy many special features that the generic probabilistic model lacks. I want to take a moment to disuss one such feature, which will be important below.

**Definition:** A model $A$ is *unital* iff, for every outcome $x \in X(A)$, there exists a state $\alpha \in \Omega(A)$ with $\alpha(x) = 1$, and *sharp* if this state is unique (from which it follows easily that it must be pure). If $A$ is sharp, I'll write $\delta_x$ for the unique state making $x \in X(A)$ certain.

If $A$ is sharp, then there is a sense in which each test $E \in \boldsymbol{\mathcal{M}}(A)$ is *maximally informative:* if we are certain which outcome $x \in E$ will occur, then we know the system's state exaclty, as there is only one state in which $x$ has probability 1.

Classical and quantum models are obviously sharp. More generally, every Jordan model is sharp. To see this, note first that every state $\alpha$ on a euclidean Jordan algebra $\mathbf{E}$ has the form $\alpha(x) = \langle a, x \rangle$ where $a \in \mathbf{E}_+$ with $\langle a, u \rangle = 1$, and where $\langle \, , \, \rangle$ is the given inner product on $\mathbf{E}$, normalized so that $\|x\| = 1$ for all primitive idempotents (equivalently, so that $\|u\| = n$, the rank of $\mathbf{E}$). The spectral theorem for EJAs shows that $a = \sum_{p \in E} t_p p$ where $E$ is a Jordan frame and the coefficients $t_p$ are non-negative and sum to 1 (since $\langle a, u \rangle = 1$) . If $\langle a, x \rangle = 1$, then $\sum_{p \in E} t_p \langle p, x \rangle = 1$ implies that, for every $p \in E$ with $t_p > 0$, $\langle p, x \rangle = 1$. But $\|p\| = \|x\| = 1$, so this implies that $\langle p, x \rangle = \|p\| \|x\|$ which in turn implies that $p = x$.

In general, a probabilistic model need not even be unital, much less sharp. On the other hand, given a unital model $A$, it is often possible to construct a sharp model by suitably restricting the state space. This is illustrated in Figure 2(b) above: the full state space of the square bit is unital, but far from sharp; however, by restricting the state space to the convex hull of the barycenters of the faces, we obtain a sharp model. This is possible whenever $A$ is unital and carries a group of symmetries acting transitively on the outcome-set $X(A)$. For details, see Appendix A. The point here is that sharpness is not, by itself, a very stringent condition: since we should expect to encounter highly symmetric, unital models abundantly "in nature", we can also expect to encounter an abundance of sharp models.

**The spaces $\mathbf{V}(A)$, $\mathbf{V}^*(A)$** Any probabilistic model gives rise to a pair of ordered vector spaces in a canonical way. These will be essential in the development below, so I'm going to go into a bit of detail here.

**Definition:** Let $A$ be any probabilistic model. Let $\mathbf{V}(A)$ be the span of the state space $\Omega(A)$ in $\mathbb{R}^{X(A)}$, ordered by the cone $\mathbf{V}(A)_+$ consisting of non-negative multiples of states, i.e.,

$$\mathbf{V}(A)_+ = \{t\alpha | \alpha \in \Omega(A), \ t \geq 0\}.$$

Call the model $A$ *finite-dimensional* iff $\mathbf{V}(A)$ is finite-dimensional. *From now on, I assume that all models are finite-dimensional.*

Let $\mathbf{V}^*(A)$ denote the dual space of $\mathbf{V}(A)$, ordered by the *dual cone* of positive linear functionals, i.e., functionals $f$ with $f(\alpha) \geq 0$ for all $\alpha \in \mathbf{V}(A)_+$. Any measurement-outcome $x \in X(A)$ yields an evaluation functional $\widehat{x} \in \mathbf{V}^*(A)$, given by $\widehat{x}(\alpha) = \alpha(x)$ for all $\alpha \in \mathbf{V}(A)$. More generally, an *effect* is a positive linear functional $f \in \mathbf{V}^*(A)$ with $0 \leq f(\alpha) \leq 1$ for every state $\alpha \in \Omega(A)$. The functionals $\widehat{x}$ are effects. One can understand an arbitrary effect $a$ to represent a *mathematically* posssible measurement outcome, having probability $a(\alpha)$ in state $\alpha$. I stress the adjective *mathematically* because, *a priori*, there is no guarantee that every effect will correspond to a *physically* realizable measurement outcome. In fact, at this stage, I make no assumption at all about what, apart from the tests $E \in \boldsymbol{\mathcal{M}}(A)$, is or is not physically realizable. (Later, it will follow from further assumptions that every element of $\mathbf{V}^*(A)$ represents a random variable associated with some $E \in \boldsymbol{\mathcal{M}}(A)$, and is, therefore, operationally meaningful. But this will be a theorem, not an assumption.)

The *unit effect* is the functional $u_A := \sum_{x \in E} \widehat{x}$, where $E$ is *any* element of $\boldsymbol{\mathcal{M}}(A)$. This takes the constant value 1 on $\Omega(A)$, and thus, represents a trivial measurement outcome that occurs with probability one in every state. This is an *order unit* for $\mathbf{V}^*(A)$: given any $a \in \mathbf{V}^*(A)$, one can find some constant $N > 0$ such that $a \leq Nu_A$. (To see this, just let $N$ be the maximum value of $|a(\alpha)|$ for $\alpha \in \Omega(A)$, remembering that the latter is compact.)

For both classical and quantum models, the ordered vector spaces $\mathbf{V}^*(A)$ and $\mathbf{V}(A)$ are naturally isomorphic. If $A(E)$ is the classical model associated with a finite set $E$, both are isomorphic to the space $\mathbb{R}^E$ of all real-valued functions on $E$, ordered pointwise. If $A = A(\boldsymbol{\mathcal{H}})$ is the quantum model associated with a finite-dimensional Hilbert space $\boldsymbol{\mathcal{H}}$, $\mathbf{V}(A)$ and $\mathbf{V}^*(A)$ are both naturally isomorphic to the space $\mathcal{L}_h(\boldsymbol{\mathcal{H}})$ of hermitian operators on $\boldsymbol{\mathcal{H}}$, ordered by its usual cone of positive semi-definite operators. More generally, if $\mathbf{E}$ is a euclidean Jordan algebra and $A = A(\mathbf{E})$ is the corresponding Jordan model, then $\mathbf{V}(A) \simeq \mathbf{V}^*(A) \simeq \mathbf{E}$, the latter ordered by its cone of squares.

**The space $\mathbf{E}(A)$** It's going to be technically useful to introduce a third ordered vector space, which I will denote by $\mathbf{E}(A)$. This is the span of the evaluation-effects $\widehat{x}$, associated with measurement outcomes $x \in X(A)$, in $\mathbf{V}^*(A)$, ordered by the cone

$$\mathbf{E}(A)_+ := \left\{ \sum_i t_i \widehat{x}_i | t_i \geq 0 \right\}.$$

That is, $\mathbf{E}(A)_+$ is the set of linear combinations of effects $\widehat{x}$ having non-negative coefficients. It is important to note that this is, in general, a *proper* sub-cone of $\mathbf{V}(A)_+$. To

see this, revisit again example 1: Here, $x$ is the outcome corresponding to the right face of the larger (full) state space pictured on the left, while $y$ is the outcome corresponding to the top face. Now consider the functional $f := \widehat{x} + \widehat{y} - \frac{1}{2}u$. This takes positive values on the smaller state space $\Omega$, but is negative on, for example, the state $\gamma$ corresponding to the lower-left corner of the full state space. Thus, $f \in \mathbf{V}(A)_+$, but $f \notin \mathbf{E}(A)$.
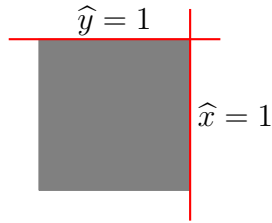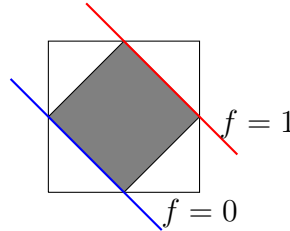


<div align="center">Figure 3(a)        Figure 3(b)</div>

Since we are working in finite dimensions, the outcome-effects $\widehat{x}$ span $\mathbf{V}^*(A)$. Thus, *as vector spaces*, $\mathbf{E}(A)$ and $\mathbf{V}^*(A)$ are the same. However, as the diamond bit illustrates, they can have quite different positive cones, and thus, need not be isomorphc as *ordered vector spaces*.

**Processes and subnormalized states** A *subnormalized state* of a model $A$ is an element $\alpha$ of $\mathbf{V}(A)_+$ with $u(\alpha) < 1$. These can be understood as states that allow a nonzero probability $1 - u(\alpha)$ of some generic "failure" event, (e.g., the destruction of the system), represented by the 0 functional in $\mathbf{V}^*(A)$.

More generally, we may wish to regard two systems, represented by models $A$ and $B$, as the input to and output from some *process*, whether dynamical or purely information-theoretic, that has some probability to destroy the system or otherwise "fail". In general, such a process should be represented mathematically by an affine mapping $\Omega(A) \to \mathbf{V}(B)_+$, taking each normalized state $\alpha$ of $A$ to a possibly *sub-normalized* state $T(\alpha)$ of $B$. One can show that such a mapping extends uniquely to a positive linear mapping $T : \mathbf{V}(A) \to \mathbf{V}(B)$, so from now on, this is how I represent processes.

Even if a process $T$ has a nonzero probability of failure, it may be possible to reverse its effect with nonzero probability.

**Definition:** A process $T : A \to B$ is *probabilistically reversible* iff there exists a process $S$ such that, for all $\alpha \in \Omega(A)$, $(S \circ T)(\alpha) = p\alpha$, where $p \in (0, 1]$.

This means that there is a probability $1 - p$ of the composite process $S \circ T$ failing, but a probability $p$ that it will leave the system in its initial state. (Note that, since $S \circ T$ is linear, $p$ must be constant.) Where $T$ preserves normalization, so that $T(\Omega(A)) \subseteq \Omega(B)$, $S$ can also be taken to be normalization-preserving, and will undo the result of $T$ with probability 1. This is the more usual meaning of "reversible" in the literature.

Given a process $T : \mathbf{V}(A) \to \mathbf{V}(B)$, there is a dual mapping $T^* : \mathbf{V}^*(B) \to \mathbf{V}^*(A)$, also positive, given by $T^*(b)(\alpha) = b(T(\alpha))$ for all $b \in \mathbf{V}^*(B)$ and $\alpha \in \mathbf{V}(A)$. The assumption

<div align="center">12</div>

that $T$ takes normalized states to subnormalized states is equivalent to the requirement that $T^*(u_B) \leq u_A$, that is, that $T^*$ maps effects to effects.

*Remark:* Since we are attaching no special physical interpretation to the cone $\mathbf{E}_+$, it is important that we do *not* require physical processes $T : \mathbf{V}(A) \to \mathbf{V}(B)$ to have dual processes $T^*$ that map $\mathbf{E}_+(B)$ to $\mathbf{E}_+(A)$. That is, we do *not* require $T^*$ to be positive as a mapping $\mathbf{E}(B) \to \mathbf{E}(A)$.

**Joint probabilities and joint states** If $\boldsymbol{\mathcal{M}}_1$ and $\boldsymbol{\mathcal{M}}_2$ are two test spaces, with outcome-spaces $X_1$ and $X_2$, we can construct a space of *product tests*

$$\boldsymbol{\mathcal{M}}_1 \times \boldsymbol{\mathcal{M}}_2 = \{E \times F | E \in \boldsymbol{\mathcal{M}}_1, F \in \boldsymbol{\mathcal{M}}_2\}.$$

This models a situation in which tests from $\boldsymbol{\mathcal{M}}_1$ and from $\boldsymbol{\mathcal{M}}_2$ can be performed separately, and the results colated. Note that the outcome-space for $\boldsymbol{\mathcal{M}}_1 \times \boldsymbol{\mathcal{M}}_2$ is $X_1 \times X_2$. A *joint probability weight* on $\boldsymbol{\mathcal{M}}_1$ and $\boldsymbol{\mathcal{M}}_2$ is just a probability weight on $\boldsymbol{\mathcal{M}}_1 \times \boldsymbol{\mathcal{M}}_2$, that is, a function $\omega : X_1 \times X_2 \to [0, 1]$ such that $\sum_{(x,y) \in E \times F} \omega(x, y) = 1$ for all tests $E \in \boldsymbol{\mathcal{M}}_1$ and $F \in \boldsymbol{\mathcal{M}}_2$. One says that $\omega$ is *non-signaling* iff the *marginal* (or *reduced*) probability weights $\omega_1$ and $\omega_2$, given by

$$\omega_1(x) = \sum_{y \in F} \omega(x, y) \ \text{ and } \ \omega_2(y) = \sum_{x \in E} \omega(x, y)$$

are well-defined, i.e., independent of the choice of the tests $E$ and $F$, respectively. One can understand this to mean that the choice of which test to measure on $\boldsymbol{\mathcal{M}}_1$ has no observable, i.e., no statistical, influence on the outcome of tests made of $\boldsymbol{\mathcal{M}}_2$, and vice versa. In this case, one also has well-defined *conditional* probability weights

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

(with, say, $\omega_{2|x} = 0$ if $\omega_1(x) = 0$, and similarly for $\omega_{1|y}$). This gives us the following bipartite version of the law of total probability [15]: for any choice of $E \in \boldsymbol{\mathcal{M}}_1$ or $F \in \boldsymbol{\mathcal{M}}_2$,

$$\omega_2 = \sum_{x \in E} \omega_1(x) \omega_{2|x} \text{ and } \ \omega_1 = \sum_{y \in F} \omega_2(y) \omega_{1|y}. \tag{1}$$

**Definition:** A *joint state* on a pair of probabilistic models $A$ and $B$ is a non-signaling joint probability weight $\omega$ on $\boldsymbol{\mathcal{M}}(A) \times \boldsymbol{\mathcal{M}}(B)$ such that, for every $x \in X(A)$ and every $y \in X(B)$, the conditional probability weights $\omega_{2|x}$ and $\omega_{1|y}$ are valid states in $\Omega(A)$ and $\Omega(B)$, respectively. It follows from (1) that the marginal weights $\omega_1$ and $\omega_2$ are also states of $A$ and $B$, respectively.[10]

This naturally suggests that one should define, for models $A$ and $B$, a *composite model* $AB$, the states of which would be precisely the joint states on $A$ and $B$. If one takes

---

[10]Notice that I do not, at this point, offer any notion of a composite model for a joint state to be a state *of*. The major results of this section do not depend on such a notion. Composite models will be discussed later, in Section 6, where, among other things, I will assume that states of such models induce (but may not be determined by) joint states in the sense defined here.

$\mathcal{M}(AB) = \mathcal{M}(A) \times \mathcal{M}(B)$, this is essentially the "maximal tensor product" of $A$ and $B$ [8]. However, this does not coincide with the usual composite of quantum-mechanical systems. In section 6, I will discuss composite systems in more detail. Meanwhile, for the main results of this paper, the idea of a joint state is sufficient.

For a simple example of a bipartite state that is neither classical nor quantum, let $B$ denote the "square bit" model discussed above. That is, $B = (\mathcal{B}, \Omega)$ where e $\mathcal{B} = \{\{x, y\}, \{a, b\}\}$ is a test space with two non-overlapping, two-outcome tests, and $\Omega$ is the set of all probability weights thereon, amounting to the unit square in $\mathbb{R}^2$. The bipartite state on $B \times B$ given by the table below (a variant of the "non-signaling box" of Popescu and Rohrlich [30]) is clearly non-signaling. Notice that it also establishes a perfect, uniform correlation between the outcomes of any test on the first system and its counterpart on the second.

|   | $x$ | $y$ | $a$ | $b$ |
|---|-----|-----|-----|-----|
| $x$ | 1/2 | 0 | 1/2 | 0 |
| $y$ | 0 | 1/2 | 0 | 1/2 |
| $a$ | 0 | 1/2 | 1/2 | 0 |
| $b$ | 1/2 | 0 | 0 | 1/2 |

**The conditioning map** If $\omega$ is a bipartite state on $A$ and $B$, define the associated *conditioning maps* $\widehat{\omega} : X(A) \to \mathbf{V}(B)$ and $\widehat{\omega}^* : X(B) \to \mathbf{V}(A)$ by

$$\widehat{\omega}(x)(y) = \omega(x, y) = \widehat{\omega}^*(y)(x)$$

for all $x \in X(A)$ and $y \in X(B)$. Note that $\widehat{\omega}(x) = \omega_1(x)\omega_{2|x}$ for every $x \in X(A)$, i.e., $\widehat{\omega}(x)$ can be understood as the *un-normalized* conditional state of $B$ given the outcome $x$ on $A$. Similarly, $\widehat{\omega}^*(y)$ is the unnormalized conditional state of $A$ given outcome $y$ on $B$.

**The conditioning map linearized** The conditioning map $\widehat{\omega}$ extends uniquely to a positive linear mapping $\mathbf{E}(A) \to \mathbf{V}(B)$, which I also denote by $\widehat{\omega}$, such that $\widehat{\omega}(\widehat{x}) = \widehat{\omega}(x)$ for all outcomes $x \in X(A)$. To see this, consider the linear mapping $T : \mathbf{V}^*(A) \to \mathbb{R}^{X(B)}$ defined, for $f \in \mathbf{V}^*(A)$, by $T(f)(y) = f(\widehat{\omega}^*(y))$ for all $y \in X(B)$. If $f = \widehat{x}$, we have $T(\widehat{x}) = \omega_1(x)\omega_{2|x} \in \mathbf{V}(B)_+$, whence, for all $y \in X(B)$, $T(\widehat{x})(y) = \omega(x, y) = \widehat{\omega}(x)(y)$. Since the evaluation functionals $\widehat{x}$ span $\mathbf{E}(A)$, the range of $T$ lies in $\mathbf{V}(B)$, and, moreover, $T$ is *positive* on the cone $\mathbf{E}(A)_+$. Hence, as advertised, $T$ defines a positive linear mapping $\mathbf{E}(B) \to \mathbf{V}(A)$, extending $\widehat{\omega}$. In the same way, $\widehat{\omega}^*$ defines a positive linear mapping $\widehat{\omega}^* : \mathbf{E}(B) \to \mathbf{V}(A)$.

In general, $\widehat{\omega}$ need *not* take $\mathbf{V}^*(A)_+$ into $\mathbf{V}(B)_+$. This is the principal reason for working with $\mathbf{E}(A)$ rather than $\mathbf{V}^*(A)$.

# 4   Conjugates and Filters

We are now in a position to abstract the two features of QM discussed earlier. Call a test space $(X, \mathcal{M})$ *uniform* iff all tests $E \in \mathcal{M}$ have the same size. The test spaces

associated with quantum models have this feature, and it is quite easy to generate many other examples (cf Appendix A). A uniform test space, say with tests of size $n$, always admits at least one probability weight, namely, the *maximally mixed* probability weight $\rho(x) = 1/n$ for all $x \in X$. I will say that a probabilistic *model $A$* is uniform if the test space $\mathbf{\mathcal{M}}(A)$ is uniform *and* the maximally mixed state $\rho$ belongs to $\Omega(A)$.

**Definition:** Let $A$ be uniform probabilistic model with tests of size $n$. A *conjugate* for $A$ is a model $\overline{A}$, plus a chosen isomorphism[11] $\gamma_A : A \simeq \overline{A}$ and a joint state $\eta_A$ on $A$ and $\overline{A}$ such that for all $x, y \in X(A)$,

(a) $\eta_A(x, \overline{x}) = 1/n$

(b) $\eta_A(x, \overline{y}) = \eta(y, \overline{x})$

where $\overline{x} := \gamma_A(x)$.

Note that if $E \in \mathbf{\mathcal{M}}(A)$, we have $\sum_{x,y \in E \times E} \eta_A(x, \overline{y}) = 1$ and $|E| = n$. Hence, $\eta_A(x, \overline{y}) = 0$ for $x, y \in E$ with $x \neq y$. Thus, $\eta_A$ establishes a perfect, uniform correlation between *any* test $E \in \mathbf{\mathcal{M}}(A)$ and its counterpart, $\overline{E} := \{\overline{x} | x \in E\}$, in $\mathbf{\mathcal{M}}(\overline{A})$.

The symmetry condition (b) is pretty harmless. If $\eta$ is a bipartite state on $A$ and $\overline{A}$ satisfying (a), then so is $\eta^t(x, \overline{y}) := \eta(y, \overline{x})$; thus, $\frac{1}{2}(\eta + \eta^t)$ satisfies both (a) and (b). In fact, if $A$ is sharp, (b) is automatic: if $\eta$ satisfies (a), then the conditional state $(\eta_A)_{1|\overline{x}}$ assigns probability one to the outcome $x$. If $A$ is sharp, this implies that $\eta_{1|\overline{x}} = \delta_x$ is uniquely defined, whence, $\eta(x, \overline{y}) = n\delta_y(x)$ is also uniquely defined. In other words, for a sharp model $A$ and a given isomorphism $\gamma : A \simeq \overline{A}$, there exists *at most one* bipartite state $\eta$ satisfying (a) — whence, in particular, $\eta = \eta^t$.

If $A = A(\mathbf{\mathcal{H}})$ is the quantum-mechanical model associated with an $n$-dimensional Hilbert space $\mathbf{\mathcal{H}}$, then we can take $\overline{A} = A(\overline{\mathbf{\mathcal{H}}})$ and define $\eta_A(x, \overline{y}) = |\langle \Psi, x \otimes \overline{y} \rangle|^2$, where $\Psi$ is the EPR state, as discussed in Section 3.

So much for conjugates. We generalize the filters associated with pure CP mappings as follows:

**Definition:** A *filter* associated with a test $E \in \mathbf{\mathcal{M}}(A)$ is a positive linear mapping $\Phi : \mathbf{V}(A) \to \mathbf{V}(A)$ such that for every outcome $x \in E$, there is some coefficient $t_x \in [0, 1]$ with $\Phi(\alpha)(x) = t_x \alpha(x)$ for every state $\alpha \in \Omega(A)$.

Equivalently, $\Phi$ is a filter iff the dual process $\Phi^* : \mathbf{V}^*(A) \to \mathbf{V}^*(A)$ satisfies $\Phi^*(\widehat{x}) = t_x \widehat{x}$ for each $x \in E$. Just as in the quantum-mechanical case, a filter independently attenuates the "sensitivity" of the outcomes $x \in E$.[12]

---

[11]By an *isomorphism* from a model $A$ to a model $B$, I mean the obvious thing: a bijection $\gamma : X(A) \to X(B)$ such taking $\mathbf{\mathcal{M}}(A)$ onto $\mathbf{\mathcal{M}}(B)$, and such that $\beta \mapsto \beta \circ \gamma$ maps $\Omega(B)$ onto $\Omega(A)$.

[12]The extreme case is one in which the coefficient $t_x$ corresponding to a particular outcome is 1, and the other coefficients are all zero. In that case, all outcomes other than $x$ are so to say, blocked by the

Call a filter $\Phi$ *reversible* iff $\Phi$ is an order-automorphism of $\mathbf{V}(A)$; that is, iff it is probabilistically reversible as a process. Evidently, this requires that all the coefficients $t_x$ be nonzero. We'll eventually see that the existence of a conjugate, plus the *preparability* of arbitrary nonsingular states by *symmetric* reversible filters, will be enough to force $A$ to be a Jordan model. Most of the work is done by the easy Lemma 1, below. First, some terminology.

**Definition:** Suppose $\Delta = \{\delta_x | x \in X(A)\}$ is a family of states indexed by outcomes $x \in X(A)$, and such that $\delta_x(x) = 1$. Say that a state $\alpha$ is *spectral* with respect to $\Delta$ iff there exists a test $E \in \mathcal{M}(A)$ such that $\alpha = \sum_{x \in E} \alpha(x)\delta_x$. Say that the model $A$ itself is spectral with respect to $\Delta$ if every state of $A$ is spectral with respect to $\Delta$.

**Lemma 1:** *Let $A$ have a conjugate $(\overline{A}, \eta_A)$. Suppose $A$ is spectral with respect to the states $\delta_x := \eta_{1|\overline{x}}$, $x \in X(A)$. Then*

$$\langle a, b \rangle := \eta_A(a, \overline{b})$$

*defines a self-dualizing inner product on $\mathbf{E}(A)$, with respect to which $\mathbf{V}(A)_+ \simeq \mathbf{E}(A)_+$. Moreover, $A$ is sharp, and $\mathbf{E}(A)_+ = \mathbf{V}^*(A)_+$.*

*Proof:* That $\langle \ , \ \rangle$ is symmetric and bilinear follows from $\eta$'s being symmetric and non-signaling. We need to show that $\langle \ , \ \rangle$ is positive-definite. Since $\widehat{A} \simeq A$, and the latter is spectral, so is the former. It follows that $\widehat{\eta}$ takes $\mathbf{E}(A)_+$ *onto* $\mathbf{V}(A)_+$, and hence, is an order-isomorphism. From this, it follows that every $a \in \mathbf{E}(A)_+$ has a "spectral" decomposition of the form $\sum_{x \in E} t_x x$ for some coefficients $t_x \geq 0$ and some test $E \in \mathcal{M}(A)$. In fact, *any* $a \in \mathbf{E}(A)$, positive or otherwise, has such a decomposition (albeit with possibly negative coefficients). For if $a \in \mathbf{E}(A)$ is arbitrary, with $a = a_1 - a_2$ for some $a_1, a_2 \in \mathbf{E}(A)_+$, we can find $N \geq 0$ with $a_2 \leq Nu$. Thus, $b := a + Nu = a_1 + (Nu - a_2) \geq 0$, and so $b := \sum_{x \in E} t_x x$ for some $E \in \mathbf{A}$, and hence $a = b - Nu = \sum_{x \in E} t_x x - N(\sum_{x \in E} x) = \sum_{x \in E}(t_x - N)x$.

Now let $a \in \mathbf{E}(A)$. Decomposing $a = \sum_{x \in E} t_x x$ for some test $E$ and some coeffcients $t_x$, we have
$$\langle a, a \rangle = \sum_{x,y \in E \times E} t_x t_y \eta_A(x, \overline{y}) = \frac{1}{n} \sum_{x \in E} t_x^2 \geq 0.$$

This is zero only where all coefficients $t_x$ are zero, i.e., only for $a = 0$. So $\langle \ , \ \rangle$ is an inner product, as claimed.

We need to show that $\langle \ , \ \rangle$ is self-dualizing. Clearly $\langle a, b \rangle = \eta(a, \overline{b}) \geq 0$ for all $a, b \in \mathbf{E}(A)_+$. Suppose $a \in \mathbf{E}(A)$ is such that $\langle a, b \rangle \geq 0$ for all $b \in \mathbf{E}(A)_+$. Then $\langle a, y \rangle \geq 0$ for all $y \in X$. Now, $a = \sum_{x \in E} t_x x$ for some test $E$; thus, for all $y \in E$ we have $\langle a, y \rangle = t_y \geq 0$, whence, $a \in \mathbf{E}(A)_+$.

---

filter. Conversely, given such an "all or nothing" filter $\Phi_x$ for each $x \in E$, we can construct an arbitrary filter with coefficients $t_x$ by setting $\Phi = \sum_{x \in E} t_x \Phi_x$.

Since $\widehat{\eta} : \mathbf{E}(A) \to \mathbf{V}(\overline{A})$ is an order-isomorphism, for every $\alpha \in \mathbf{V}(A)$ there exists a unique $a \in \mathbf{E}(A)$ with $\widehat{\eta}(a) = \overline{\alpha}$. In particular,

$$\langle a, x \rangle = \eta(a, \overline{x}) = \overline{\alpha}(\overline{x}) = \alpha(x).$$

It follows that if $b \in \mathbf{E}(A) = \mathbf{V}^*(A)$,

$$b(\alpha) = \overline{b}(\overline{\alpha}) = \widehat{\eta}(a)(\overline{b}) = \eta(a, \overline{b}) = \langle a, b \rangle.$$

Since every $a \in \mathbf{E}(A)_+$ has the form $a = \widehat{\eta}^{-1}(\overline{\alpha})$ for some $\alpha \in \mathbf{V}(A)_+$, if $b \in \mathbf{V}^*(A)_+$, we have $\langle a, b \rangle \geq 0$ for all $a \in \mathbf{E}(A)_+$, whence, by the self-duality of the latter cone, $b \in \mathbf{E}(A)_+$. Thus, $\mathbf{V}^*(A) = \mathbf{E}(A)_+$.

Finally, let's see that $A$ is sharp. If $\alpha \in \mathbf{V}(A)$, let $a$ be the unique element of $\mathbf{E}(A)_+$ with $\langle a, x \rangle = \alpha(x)$, as discussed above. If $a$ has spectral decomposition $a = \sum_{x \in E} t_x x$, where $E \in \mathbf{\mathcal{M}}(A)$, then $\sum_{x \in E} \langle a, x \rangle = \sum_{x \in E} t_x = 1$. Thus, $\|a\|^2 = \sum_{x \in E} t_x^2 \leq 1$, whence, $\|a\| \leq 1$. Now suppose $\alpha(x) = 1$ for some $x \in X(A)$: then $1 = \langle a, x \rangle \leq \|a\| \|x\|$; as $\|x\| = 1$, we have $\|a\| = 1$. But now $\langle a, x \rangle = \|a\| \|x\|$, whence, $a = x$. Hence, there is only one weight $\alpha$ with $\alpha(x) = 1$, namely, $\alpha = \langle x, \cdot \rangle$, so $A$ is sharp. $\square$

If $A$ is sharp, then we say that $A$ is *spectral* iff it is spectral with respect to the pure states $\delta_x$ defined by $\delta_x(x) = 1$. If $A$ is sharp and has a conjugate $\overline{A}$, then, as noted earlier, the state $\eta_{1|\overline{x}}$ is exactly $\delta_x$, so the spectrality assumption in Lemma 1 is fulfilled if we simply say that $A$ is spectral. Hence, *a sharp, spectral model with a conjugate is self-dual*.

For the simplest systems, this is already enough to secure the desired representation in terms of a euclidean Jordan algebra.

**Definition:** Call $A$ a *bit* iff it has rank 2 (that is, all tests have two outcomes), and if every state $\alpha \in \Omega(A)$ can be expressed as a mixture of two sharply distinguishable states; that is, $\alpha = t\delta_x + (1 - t)\delta_y$ for some $t \in [0, 1]$ and states $\delta_x$ and $\delta_y$ with $\delta_x(x) = 1$ and $\delta_y(y) = 1$ for some test $\{x, y\}$.

**Corollary 1:** *If $A$ is a sharp bit, then $\Omega(A)$ is a ball of some finite dimension $d$.*

The proof is given in Appendix C. If $d$ is $2, 3$ or $5$, we have a real, complex or quaternionic bit. For $d = 4$ or $d \geq 6$, we have a non-quantum spin factor.

For systems of higher rank (higher "information capacity"), we need to assume a bit more. Suppose $A$ satisfies the hypotheses of Lemma 1. Appealing to the Koecher-Vinberg Theorem, we see that if $\mathbf{V}(A)$ and, hence, $\mathbf{V}^*(A)$ are also homogeneous, then $\mathbf{V}^*(A)$ carries a canonical Jordan structure. In fact, we can say something a little stronger.

**Theorem 1:** *Let $A$ be spectral with respect to a conjugate system $\overline{A}$. If $\mathbf{V}(A)$ is homogeneous, then there exists a canonical Jordan product on $\mathbf{E}(A)$ with respect to which $u$ is the Jordan unit. Moreover with respect to this product $X(A)$ is exactly the set of primitive idempotents, and $\mathbf{\mathcal{M}}(A)$ is exactly the set of Jordan frames.*

The first part is almost immediate from the Koecher-Vinberg Theorem, together with Lemma 1. The KV Theorem gives us an isomorphism between the ordered vector spaces $\mathbf{V}(A)$ and $\mathbf{E}(A)$, so if one is homogeneous, so is the other. Since $\mathbf{E}(A)$ is also self-dual by Lemma 1, the KV theorem yields the requisite unique euclidean Jordan structure having $u$ as the Jordan unit. One can then show without much trouble that every outcome $x \in X(A)$ is a primitive idempotent of $\mathbf{E}(A)$ with respect to this Jordan structure, and that every test is a Jordan frame. The remaining claims — that every minimal idempotent belongs to $X(A)$ and every Jordan frame, to $\mathcal{M}(A)$ — take a little bit more work. I won't reproduce the proof here; the details (which are not difficult, but depend on some facts concerning euclidean Jordan algebras) can be found in [38].

The homogeneity of $\mathbf{V}(A)$ can be understood as a *preparability assumption:* it is equivalent to saying that every state in the interior of $\Omega(A)$ can be obtained, *up to normalization,* from the maximally mixed state by a reversible process. That is, if $\alpha \in \Omega(A)$, there is some such process $\phi$ such that $\phi(\rho) = p\alpha$ where $0 < p \leq 1$. One can think of the coefficient $p$ as the probability that the process $\phi$ will yield a nonzero result (more dramatically: will not destroy the system). Thus, if we prepare an ensemble of identical copies of the system in the maximally mixed state $\rho$ and subject them all to the process $\phi$, the fraction that survive will be about $p$, and these will all be in state $\alpha$.

In fact, if the hypotheses of Lemma 1 hold, the homogeneity of $\mathbf{E}(A)$ follows directly from the mere existence of reversible filters with arbitrary non-zero coefficients. To see this, suppose $a \in \mathbf{E}(A)_+$ has a spectral decomposition $\sum_{x \in E} t_x x$ for some $E \in \mathcal{M}(A)$, with $t_x > 0$ for all $x$ when $a$ belongs to the interior of $\mathbf{E}(A)_+$. Now if we can find a reversible filter for $E$ with $\Phi(x) = t_x x$ for all $x \in E$, then applying this to the order-unit $u = \sum_{x \in E} x$ yields $a$. Thus, $\mathbf{V}^*(A)$ is homogeneous.

**Two paths to spectrality** Some axiomatic treatments of quantum theory have taken one or another form of spectrality as an axiom [16, 9]. If one is content to do this, then Lemma 1 above provides a very direct route to the Jordan structure of quantum theory. However, spectrality can actually be derived from assumptions that, on their face, seem a good deal weaker, or anyway more transparent.[13]

I will call a joint state on models $A$ and $B$ *correlating* iff it sets up a perfect correlation between *some* pair of tests $E \in \mathcal{M}(A)$ and $F \in \mathcal{M}(B)$. More exactly:

**Definition:** A joint state $\omega$ on probabilistic models $A$ and $B$ *correlates* a test $E \in \mathcal{M}(A)$ with a test $F \in \mathcal{M}(B)$ iff there exist subsets $E_0 \subseteq E$ and $F_0 \subseteq F$, and a bijection $f : E_0 \to F_0$ such that $\omega(x, y) = 0$ for $(x, y) \in E \times F$ unless $y = f(x)$. In this case, say that $\omega$ correlates $E$ with $F$ *along $f$.* A joint state on $A$ and $B$ is *correlating* iff it correlates some pair of tests $E \in \mathcal{M}(A), F \in \mathcal{M}(B)$.

Note that $\omega$ correlates $E$ with $F$ along $f$ iff $\omega(x, f(x)) = \omega_1(x) = \omega_2(f(x))$, which, in turn, is equivalent to saying that $\omega_{2|x}(f(x)) = 1$ for $\omega_1(x) \neq 0$.

---

[13]A different path to spectrality is charted in a recent paper [11] by G. Chiribella and C. M. Scandolo.

**Lemma 2:** *Suppose $A$ is sharp, and that every state $\alpha$ of $A$ arises as the marginal of a joint state between $A$ and some model $B$. Then $A$ is spectral.*

*Proof:* Suppose $\alpha = \omega_1$, where $\omega$ is a bipartite state correlating a test $E \in \mathcal{M}(A)$ with a test $F \in \mathcal{M}(B)$, say along an bijection $f : E_0 \to F_0$, $E_o \subseteq E$ and $F_0 \subseteq F$. Then for any $x \in E$ with $\alpha(x) \neq 0$, $\omega_{1|f(x)}(x) = 1$, whence, as $A$ is sharp, $\omega_{1|f(x)} = \delta_x$, the unique state making $x$ certain. It follows from the Law of Total Probability that $\alpha = \sum_{x \in E} \alpha(x)\delta_x$. $\square$

In principle, the model $B$ can vary with the state $\alpha$. Lemma 2 suggests the following language:

**Definition:** A model $A$ satisfies the *correlation condition* iff every state $\alpha \in \Omega(A)$ is the marginal of some correlating joint state of $A$ and some model $B$.

This has something of the same flavor as the *purification postulate* of [10], which requires that all states of a given system arise as marginals of a *pure* state on a larger, composite system, unique up to symmetries on the purifying system. However, note that we do not require the correlating joint state to be either pure (which, in classical probability theory, it will not be), or unique.

If $A$ is sharp and satisfies the correlation condition, then every state of $A$ is spectral. If, in addition, $A$ has a conjugate, then for every $x \in X(A)$, we have $\eta_{1|\overline{x}} = \delta_x$. In this case, $A$'s states are spectral with respect to the family of states $\eta_{1|\overline{x}}$, and the hypotheses of Lemma 1 are satisfied. So we have

Here is another, superficially quite different, way of arriving at spectrality. Call a transformation $\Phi$ *symmetric* with respect to $\eta_A$ iff, for all $x, y \in X(A)$,

$$\eta_A(\Phi^* x, \overline{y}) = \eta_A(x, \overline{\Phi}^* y).$$

**Lemma 2:** *Let $A$ have a conjugate, $\overline{A}$, and suppose every state of $A$ is preparable by a symmetric filter. Then $A$ is spectral.*

*Proof:* Let $\alpha = \Phi(\rho)$ where $\Phi$ is a filter on a test $E \in \mathcal{M}(A)$, say $\Phi(x) = t_x x$ for all $x \in E$. Then

$$\alpha = \Phi(\widehat{\eta}^*(\overline{u})) = \eta(\Phi^*(\cdot), \overline{u}) = \eta(\,\cdot\,, \overline{\Phi}^*(\overline{u})) = \sum_{x \in E} \eta(\,\cdot\,, t_x \overline{x}) = \sum_{x \in E} t_x \tfrac{1}{n} \delta_x. \;\; \square$$

Thus, the hypotheses of either Corollary 2 or Lemma 3 will supply the needed spectral assumption that makes Lemma 1 work. (In fact, it is not hard to see that these hypotheses are actually equivalent — an exercise I leave for the reader.)

To obtain a Jordan model, we still need homogeneity. This is obviously implied by the preparability condition in Lemma 3, provided the preparing filters $\Phi$ can be taken to be reversible whenever the state to be prepared is non-singular. On the other hand, as noted above, in the presence of spectrality, it's enough to *have* arbitrary reversible filters, as

these allow one to prepare the spectral decompositions of arbitrary non-singular states. Thus, conditions (a) and (b), below, both imply that $A$ is a Jordan model. Conversely, one can show that any Jordan model satisfies both (a) and (b), closing the loop [38]:

**Theorem 2:** *The following are equivalent:*
  (a) *$A$ has a conjugate, and every non-singular state can be prepared by a reversible symmetric filter;*
  (b) *$A$ is sharp, has a conjugate, satisfies the correlation condition, and has arbitary reversible filters;*
  (c) *$A$ is a Jordan model.*

## 5 Measurement, Memory and Correlation

Of the spectrality-underwriting conditions given in Lemmas 2 and 3, the one that seems less transparent (to me, anyway) is that every state arise as the marginal of a correlating biparite state. While surely less *ad hoc* than spectrality, this still calls for further explanation. Suppose we hope to implement a measurement of a test $E \in \mathcal{M}(A)$ *dynamically*. This would involve bringing up an ancilla system $B$ — also uniform, suppose; and which we can suppose (by suitable coarse-graining, if necessary) to have tests of the same cardinality as $A$'s — in some "ready" state $\beta_o$. We would then subject the combined system $AB$ to some physical process, at the end of which $AB$ is in some final joint state $\omega$, and $B$ is (somehow!) in one of a set of *record states*, $\beta_x$, each corresponding to an outcome $x \in Ea$.[14] We would like to insist that
  (a) The states $\beta_x$ are *distinguishable*, or *readable*, by some test $F \in \mathcal{M}(B)$. This means that for each $x \in E$, there is a unique $y \in F$ such that $\beta_x(y) = 1$. Note that this sets up an injection $f : E \to F$.
  (b) The record states must be *accurate*, in the sense that if we *were* to measure $E$ on $A$, and secure $x \in E$, the record state $\beta_x$ should coincide with the conditional state $\omega_{2|x}$. (If this is not the case, then a measurement of $A$ cannot correctly calibrate $B$ as a measuring device for $E$.)

It follows from (a) and (b) that, for $x \in E$ and $y \neq f(x) \in F$,

$$\omega(x, y) = \omega_1(x)\omega_{2|x}(y) = \omega_1(x)\beta_x(y) = 0.$$

In other words, $\omega$ must correlate $E$ with $F$, along the bijection $f$. If the measurement process leaves $\alpha$ undisturbed, in the sense that $\omega_1 = \alpha$, then $\alpha$ dilates to a correlating state. This suggests the following *Non-disturbance principle:* Every state can be measured, by some test $E \in \mathcal{M}(A)$, without disturbance. Lemma 2 then tells us that if $A$ is sharp and satifies the non-disturbance principle, every state of $A$ is spectral.

---

[14]This way of putting things takes us close to the usual formulation of the quantum-mechanical "measurement problem". which I certainly don't propose to discuss here. The point is only that, if *any* dynamical process, describable within the theory, can account for measurement results, it should be consistent with this description.

Here is a slightly different, but possibly more compelling, version of this story. Suppose we can make a test $E$ on $A$ directly (setting aside, that is, any issue of whether or not this can be achieved through some dynamical process): this will result in an outcome $x$ *occuring*. To *do* anything with this, we need to *record* its having occurred. This means we need a storage medium, $B$, and a family of states $\beta_x$, one for each $x \in E$, such that *if* we get $x$, then $B$ *will* be in state $\beta_x$. Moreover, these record states need to be *readable* at a later time, i.e., distinguishable by a later measurement on $B$. To arrange this, we need $AB$ to be in a joint state, associated with a joint probability weight $\omega$, such that $\omega_1 = \alpha$ (because we want to have prepared $A$ in the state $\alpha$) and $\beta_x = \omega_{2|x}$ for every $x \in E$. We then measure $E$ on $A$; upon our obtaining outcome $x \in E$, $B$ is in the state $\beta_x$. Since the ensemble of states $\beta_x$ is readable (by some $F \in \mathcal{M}(B)$ with $|F| \geq |E|$), we have correlation, and $\alpha$ must also be spectral.

Of course, these desiderata can not always be satisfied. What *is* true, in QM, is that for every choice of state $\alpha$, there will exist *some* test that is recordable that state, in the foregoing sense. If we promote this to general principle, we again see that every state is the marginal of a correlating state, and hence spectral, if $A$ is sharp.

# 6    Composites and Categories

Thus far, we've been referring to the correlator $\eta_A$ as a *bipartite state*, but dodging the question: *state of what?* Mathematically, of course, nothing much hangs on this question: it is sufficient to regard $\eta_A$ as a bipartite probability assignment on $A$ and $\overline{A}$. But it would surely be more satisfactory to be able to treat it as an actual physical state of some composite system $A\overline{A}$. How should this be chosen?

One possibility is to take $A\overline{A}$ to be the *maximal* tensor product of the models $A$ and $\overline{A}$ [8]. By definition, this has for its states all nonsignaling probability assignments with conditional states belonging to $A$ and $\overline{A}$. However, we might want composite systems, in particular $A\overline{A}$, to satisfy the same conditions we are imposing on $A$ and $\overline{A}$, i.e., to be a Jordan model. If so, we need to work somewhat harder: the maximal tensor product will be self-dual only if $A$ is classical.

In order to be more precise about all this, the first step is to decide what ought to count as a composite of two probabilistic models. If mean to capture the idea of two physical systems that can be acted upon separately, but which cannot influence one another in any observable way (e.g., two spacelike-separated systems), the following seems to me to capture the minimal requirements:

**Definition:** A non-signaling composite of models $A$ and $B$ is a model $AB$, together with a mapping $\pi : X(A) \times X(B) \to \mathbf{V}^*(AB)$ such that

$$\sum_{x \in E, y \in F} \pi(x, y) = u_{AB}$$

21

and, for $\omega \in \Omega(AB)$, $\omega \circ \pi$ is a joint state on $A$ and $B$, as defined in Section 2.

So the question becomes: can one construct, for Jordan models $A$ and $B$, a nonsignaling composite $AB$ that is also a Jordan model? At present, and in this generality, this question seems to be open, but some progress is made in [5]. I will have more to say about this below. As shown in [5], neither $A$ nor $B$ contains the exceptional Jordan algebra as a summand, such a composite can indeed be constructed, and in multiple ways. Under a considerably more restrictive definition of "Jordan composite", it is also shown in [5] that no Jordan composite $AB$ can exist if either factor has an exceptional summand.

**Categories of Self-Dual Probabilistic Models** It's natural to interpret a physical theory as a *category*, in which objects represent physical systems and morphisms represent physical processes having these systems (or there states) as inputs and outputs. In order to discuss composite systems, this should be a *monoidal* category. That is, for every pair of objects $A, B$, there should be an object $A \otimes B$, and for every pair of morphisms $f : A \to A'$ and $g : B \to B'$, there should be a morphism $f \otimes g : A \otimes B \to A' \otimes B'$, representing the two processes $f$ and $g$ occuring "in parallel". One requires that $\otimes$ be associative and commutative, and have a unit object $I$, in the sense that there exist canonical isomorphisms $\alpha_{A,B;C} : A \otimes (B \otimes C) \simeq (A \otimes B) \otimes C$, $\sigma_{A,B} : A \otimes B \simeq B \otimes A$, $\lambda_A : I \otimes A \simeq A$ and $\rho_A : A \otimes I \to A/$ These must atisfy various "naturality conditions", guaranteeing that they interact correctly; see [26] for details. One also requires that $\otimes$ be bifunctorial, meaning that $\mathrm{id}_A \otimes \mathrm{id}_B = \mathrm{id}_{A \otimes B}$, and if $f : A \to A'$, $f' : A' \to A''$, $g : B \to B'$ and $g' : B' \to B''$, then

$$(f' \otimes g') \circ (f \otimes g) = (f' \circ f) \otimes (g' \circ g).$$

Following this lead, by a *probabilistic theory*, I mean a category of probabilistic models and processes — that is, objects of $\mathcal{C}$ are models, and a morphism $A \to B$, where $A, B \in \mathcal{C}$, is a process $\mathbf{V}(A) \to \mathbf{V}(B)$ A *monoidal* probabilistic theory is such a category, $\mathcal{C}$, carrying a symmetric monoidal structure $A, B \mapsto AB$, where $AB$ is a non-signaling composite in the sense of the Definition above. I also assume that the monoidal unit, $I$, is the trivial model 1 with $\mathbf{V}(1) = \mathbb{R}$, and that, for all $A \in \mathcal{C}$,

(b) $\alpha \in \Omega(A)$ iff the mapping $\alpha : \mathbb{R} \to \mathbf{V}(A)$ given by $\alpha(1) = \alpha$ belongs to $\mathcal{C}(I, A)$;

(b) The evaluation functional $\widehat{x}$ belongs to $\mathcal{C}(A, I)$ for all outcomes $x \in X(A)$.

Call $\mathcal{C}$ *locally tomographic* iff $AB$ is a locally-tomographic composite for all $A, B \in \mathcal{C}$;

Much of the qualitative content of (finite-dimensional) quantum information theory can be formulated in purely categorical terms [1, 3, 32]. In particular, in the work of Abramsky and Coecke [1], it is shown that a range of quantum phenomena, notably gate teleportation, are available in any *dagger-compact* category. For a review of this notion, as well as a proof of the following result, see Appendix D:

**Theorem 3:** *Let $\mathcal{C}$ be a locally tomographic monoidal probabilistic theory, in which every object $A \in \mathcal{C}$ is sharp, spectral, and has a conjugate $\overline{A} \in \mathcal{C}$, with $\eta_A \in \Omega(A\overline{A})$. Assume also that, for all $A, B \in \mathcal{C}$,*

*(i)* $\overline{\overline{A}} = A$, *with* $\eta_{\overline{A}}(\overline{a}, b) = \eta_A(a, \overline{b})$;
*(ii)* *If* $\phi \in \mathcal{C}(A, B)$, *then* $\overline{\phi} \in \mathcal{C}(\overline{A}, \overline{B})$.

*Then* $\mathcal{C}$ *has a canonical dagger-compact structure, in which* $\overline{A}$ *is the dual of* $A$ *with* $\eta_A :$ $\mathbb{R} \to \mathbf{V}(A\overline{A})$ *as the co-unit.*

**Jordan Composites** The local tomography assumption in Theorem 3 is a strong constraint. As is well known, the standard composite of two real quantum systems is not locally tomographic, yet the category of finite-dimensional real mixed-state quantum systems is certainly dagger-compact, and satisfies the other assumptions of Theorem 3, so local tomography is definitely not a necessary condition for dagger-compactness.

This raises some questions. One is whether local tomography can simply be dropped in the statement of Theorem 3. At any rate, at present I don't know of any non-dagger-compact monoidal probabilistic theory satisfying the other assumptions.

Another question is whether there exist examples *other* than real QM of non-locally tomographic, but still dagger-compact, monoidal probabilistic theories satisfying the assumptions of Theorem 1. The answer to this is *yes*. Without going into detail, the main result of [5] is that one can construct a dagger-compact category in which the objects are hermitian parts of finite-dimensional real, complex and quaternionic matrix algebras — that is, the euclidean Jordan algebras corresponding to finite-dimensional real, complex or quaternionic quantum-mechanical systems — and morphisms are certain completely positive mappings between enveloping complex $*$-algebras for these Jordan algebras. The monoidal structure gives *almost* the expected results: the composite of two real quantum systems is the real system corresponding to the usual (real) quantum-mechanical composite of the two components (and, in particular, is not locally tomographic). The composite of two quaternionic systems is a real system (see [3] for a account of why this is just what one wants). The composite of a real and a complex, or a quaternionic and a complex, system is again complex. The one surprise is that the composite of two standard *complex* quantum systems, in this category, is *not* the usual thing, but rather, comes with an extra superselection rule. This functions to make time-reversal a legitimate physical operation on complex systems, as it is for real and quaternionic systems. This is part of the price one pays for the dagger-compactness of this category.

# 7   Conclusion

As promised, we have here an easy derivation of something *close* to orthodox, finite-dimensional QM, from operationally or probabilistically transparent assumptions. As discussed earlier, this approach offers — in addition to its relative simplicity — greater lattitude than the locally tomographic axiomatic reconstructions of [19, 12, 27, 10], putting us in the slightly less constrained realm of formally real Jordan algebras. This allows for real and quaternionic quantum systems, superselection rules, and even theories, such as the ones discussed in section 6, in which real, complex and quaternionic quantum systems

23

coexist and interact.

There remains some mystery as to the proper interpretation of the conjugate system $\overline{A}$. Operationally, the situation is clear enough: if we understand $A$ as controlled by Alice and $\overline{A}$, by Bob, then if Alice and Bob share the state $\eta_A$, then they will always obtain the same result, as long as they perform the same test. But what does it mean *physically* that this should be possible (in a situation in which Alice and Bob are still able to chose their tests indpendently)? In fact, there is little consensus (that I can find, anyway) among physicists as to the proper interpretation of the conjugate of the Hilbert space representing a given quantum-mechanical system. One popular idea is that the conjugate is a *time-reversed* version of the given system — but why, then, should we expect to find a state that perfectly correlates the two? At any rate, finding a clear physical intepretation of conjugate systems, even — or especially! — in orthodox quantum mechanics, seems to me an important problem.

I'd like to close with another problem, this one of mainly mathematical interest. The hypotheses of Theorem 1 yield a good deal more structure than just an homogeneous, self-dual cone. In particular, we have a distinguished set $\boldsymbol{\mathcal{M}}(A)$ of orthonormal observables in $\mathbf{V}^*(A)$, with respect to which every effect has a spectral decomposition. Moreover, with a bit of work one can show that this decomposition is essentially unique. More exactly, if $a = \sum_i t_i p_i$ where the coefficients $t_i$ are all distinct and the effects $p_1, ..., p_k$ are associated with a coarse-graining of a test $E \in \boldsymbol{\mathcal{M}}(A)$, then both the coefficients and the effects are uniquely determined. The details are in Appendix B. Using this, we have a functional calculus on $\mathbf{V}^*(A)$, i.e., for any real-valued function $f$ of a real variable, and any effect $a$ with spectral decomposition $\sum_i t_i p_i$ as above, we can define $f(a) = \sum_i f(t_i) p_i$. This gives us a unique candidate for the Jordan product of effects $a$ and $b$, namely,

$$a \bullet b = \tfrac{1}{2}((a+b)^2 - a^2 - b^2)).$$

We know from Theorem 1 — and thus, ultimately, from the KV theorem — that this is bilinear.The challenge is to show this *without* appealing to the KV theorem. (The fact that the state spaces of "bits" are always balls, as shown in Appendix C, is perhaps relevant here.)

# References

[1] Abramsky and Coecke, Abstract Physical Traces, Theory and Applications of Categories **14** (2005) 111–124 (`arXiv:   arXiv:0910.3144`, 2009)

[2] E. M. Alfsen and F. W. Shultz, *Geometry of state spaces of operator algebras.* Birkhäuser (2003).

[3] J. Baez, Division algebras and quantum theory, Found. Phys. **42** (2012), 819-855; `arXiv:1101.5690`

[4] H. Barnum, M. Graydon and A. Wilce, Some nearly quantum theories, EPTCS **195** (2015), 59-705; `arXiv:1507.06278`

[5] H. Barnum, M. Graydon and A. Wilce, Composites and categories of euclidean Jordan algebras, to appear.

[6] H. Barnum and A. Wilce, Information processing in convex operational theories, *Electronic Notes in Theoretical Computer Science* **270** (2011), 3-15; `arXiv:0908.2352`.

[7] H. Barnum and A. Wilce, Local tomography and the Jordan structure of quantum theory, *Found. Phys.* **44** (2014) 192-212 (2014); `arXiv:1202.4513`.

[8] H. Barnum and A. Wilce, Post-classical probability theory, to appear in G. Chiribella and R. Spekkens (eds.), *Quantum Theory: Informational Foundations and Foils*, Springer, 2016; `arXiv:   1205.3833`

[9] H. Barnum, M. Müller and C. Ududec, Higher-order interference and single-system postulates characterizing quantum theory, *New J. Phys.* **16** (2014) ; `arXiv:1403.4147`.

[10] G. Chiribella, M. D'Ariano and P. Perinotti, Informational derivation of quantum theory, Phys. Rev. A **84** (2011); `arXiv:1011:6451`

[11] G. Chribella, C. M. Scandolo, Operational axioms for state diagonalization, `arXiv: 1506:00380`

[12] B. Dakic and Č. Brukner, Quantum theory and beyond: is entanglement special? `arXiv:0911.0695`

[13] J. Faraut and A. Koranyi, *Analysis on Symmetric Cones.* Oxford University Press (1994).

[14] D. J. Foulis and C. H. Randall, An approach to empirical logic, Amer. Math. Montly **77** (1970), 363-374.

[15] D. J. Foulis and C. H. Randall, Empirical logic and tensor products, in H. Neumann (ed.), *Interpretations and Foundations of Quantum Theory*, Bibliographisches Institut, 1981

[16] J. Gunson, On the algebraic structure of quantum mechanics, Comm. Math. Phys. **6** (1967), 262-285.

[17] H. Hanche-Olsen, On the structure and tensor products of JC-algebras, Can. J. Math. **35** (1983) 1059-1074.

[18] P. Janotta and R. Lal, Generalized probabilistic theories without the no-restriction hypothesis, Phys. Rev. A. **87** (2013)

[19] L. Hardy, Quantm theory from five reasonable axioms, `arXiv:quant-ph/0101012` (2001)

[20] A. Holevo, *Probabilistic and Statistical Aspects of Quantum Mechanics*, North-Holland, 1982

[21] P. Jordan, Über ein Klasse nichtassoziativer hypercomplexe algebren, *Nachr. Akad. Wiss. Göttingen Math. Phys. Kl. I.* **33** 569-575 (1933).

[22] P. Jordan, J. von Neumann and E. Wigner, On an algebraic generalization of the quantum mechanical formalism, Annals of Math. **35** (1934)

[23] M. Koecher, Positivitatsbereiche im $\mathbb{R}^n$, Amer. J. Math. **97** (1957), 575596.

[24] G. Ludwig, *Foundations of Quantum Mechanics I* Springer, 1983

[25] G. Mackey, *Mathematical Foundations of Quantum Mechanics*, Addison-Wesley, 1962

[26] S. MacLane, *Categories for the Working Mathematician*, Springer, 1978

[27] Ll. Masanes and M. Müller, A derivation of quantum theory from physical requirements, New J. Phys. **13** (2011); `arXiv:1004.1483`

[28] M. Müller and C. Ududec, The structure of reversible computation determines the self-duality of quantum theory, Phys. Rev. Lett. **108** (2012); `arXiv:1110.3516`

[29] C. Piron, *Mathematical Foundations of Quantum Mechanics*, Academic Press, 1978

[30] S. Popescu and D. Rohrlich, Nonlocality as an axiom, Foundations of Physics **24** (1994), 379385.

[31] J. Schwinger, The algebra of microscopic measurement, Proc. Nat. Acad. USA **45** (1959), 1542-1553

[32] P. Selinger, Dagger compact closed categories and completely positive maps, Electronic Notes in Theoretical Computer Science **170** (2007), 139163.

[33] J. von Neumann, *Mathematical Foundations of Quantum Mechanics*, Princeton, 1955

[34] E. Vinberg, Homogeneous cones, Soviet Math Dokl. **1** (1961), 787-790

[35] A. Wilce, Tensor products in generalized measure theory, Int. J. Theor. Phys. **31** (1992), 1915-1928

[36] A. Wilce, 4.5 axioms for finite-dimensional quantum probability, in Y. ben-Menahem and M. Hemmo, *Probability in Physics*, Springer, 2012; `arXiv:0912.5530`

[37] A. Wilce, Symmetry, self-duality and the Jordan structure of finite-dimensional quantum mechanics, `arxiv:1110.6607` (2011)

[38] A. Wilce, Conjugates, Filters and Quantum Mechanics, `arxiv.org/pdf/1206.2897` (2012)

[39] A. Wilce, Symmetry and composition in probabilistic theories `arXiv:0910.1527` (2009)

# A  Models with symmetry

Recall that a probabilistic model $A$ is *sharp* iff, for every measurement outcome $x \in X(A)$, there exists a *unique* state $\delta_x \in \Omega(A)$ with $\delta_x(x) = 1$. While this is clearly a very strong condition, it is not an unreasonable one. In fact, given the test space $\mathcal{M}(A)$, we can often *choose* the state space $\Omega(A)$ in such a way as to guarantee that $A$ is sharp. In particular, this is the case when $\mathcal{M}(A)$ enjoys enough symmetry.

**Definition:** Let $G$ be a group. A *G-test space* is a test space $(X, \mathcal{M})$ where $X$ is a $G$-space — that is, where $X$ comes equipped with a preferred $G$-action $G \times X \to X$, $(g, x) \mapsto gx$ — such that $gE \in \mathcal{M}$ for all $E \in \mathcal{M}$. A *G-model* is a probabilistic model $A$ such that (i) $\mathcal{M}(A)$ is a $G$-test space, and (ii) $\Omega(A)$ is invariant under the action of $G$ on probability weights given by $\alpha \mapsto g\alpha := \alpha \circ g^{-1}$ for $g \in G$.

**Lemma A1:** *Let $A$ be a finite-dimensional $G$-model and suppose $G$ acts transitively on the outcome space $X(A)$. Suppose also that $A$ is unital, i.e., for every $x \in X(A)$, there exists at least one state $\alpha$ with $\alpha(x) = 1$. Then there exists a $G$-invariant convex subset $\Delta \subseteq \Omega(A)$ such that $A' = (\mathcal{M}(A), \Delta)$ is a sharp $G$-model.*

*Proof:* For each $x \in X(A)$, let $F_x$ denote the face of $\Omega(A)$ consisting of states $\alpha$ with $\alpha(x) = 1$. Let $\beta_x$ be the barycenter of $F_x$. It is easy to check that $F_{gx} = gF_x$ for every $g \in G$. Thus, $g\beta_x = \beta_{gx}$, i.e., the set of barycenters $\beta_x$ is an orbit. Let $\Delta$ be the convex hull of these barycenters. Then $\Delta$ is invariant under $G$. If $\alpha \in \Delta$ with $\alpha(x) = 1$, then $\alpha \in F_x \cap \Delta = \{\beta_x\}$, so $(\mathcal{M}(A), \Delta)$ is sharp. $\square$

# B  Uniqueness of Spectral Decompositions

Let $A$ be a model satisfying the conditions of Lemma 1. In particular, then, every $a \in \mathbf{V}^*(A)$ has a spectral representation $a = \sum_{x \in E} t_x x$ for some test $E \in \mathcal{M}(A)$. In general, this expansion is highly non-unique. For instance, the unit $u$ can be expanded as $\sum_{x \in E} \widehat{x}$ for any test $E \in \mathcal{M}(A)$. The aim in this Appendix is to obtain a form of spectral expansion for effects that *is* unique.

Call a subset of a test an *event* of $A$. That is, $D \subseteq X(A)$ is an event iff there exists a test $E \in \mathcal{M}(A)$ with $D \subseteq E$. Any event gives rise to an effect

$$p(D) := \sum_{x \in D} \widehat{x}.$$

Thus, a test is a maximal event, and for any test $E \in \mathcal{M}(A)$, $p(E) = u$.

Let's say that an effect $a \in \mathbf{V}^*(A)$ is *sharp* iff there exists a state $\alpha \in \Omega(A)$ with $\alpha(a) = 1$. Clearly, every effect of the form $p(D)$ is sharp. The converse is also true:

**Lemma B1:** *If $a \in \mathbf{V}^*(A)_+$ is sharp, then it has the form $p(D)$ for some event $D$.*

*Proof:* Let $a = \sum_{x \in E} t_x x$. Since $a \in \mathbf{V}^*(A)_+$, each $t_x$ is non-negative. Since $a$ is an effect, $t_x = \delta_x(a) \leq \delta_x(u) = 1$. Now let $\alpha$ be a state with $\alpha(a) = 1$: we have $\sum_{x \in E} t_x \alpha(x) = 1$. Since the terms $\alpha(x)$ are convex coefficients, the values $t_x$ all lie in $[0, 1]$, and and 1 is an extreme point of $[0, 1]$, it follows that, for every $x \in E$, $t_x$ is either 1 or 0. Letting $D = \{x \in E | t_x = 1\}$ gives the result. $\square$

**Definition:** A set of sharp effects $p_1, ..., p_n \in \mathbf{V}^*(A)$ is *jointly orthogonal* with respect to $\mathcal{M}(A)$ iff there exists a test $E \in \mathcal{M}(A)$ and pairwise disjoint events $D_1, ..., D_n \subseteq E$ with $p_i = p(D_i)$ for $i = 1, ..., n$. Given an arbitrary element $a \in \mathbf{V}^*(A)$ with spectral decomposition $a = \sum_{x \in E} t_x x$, we can isolate distinct values $t_o > t_1 > ... > t_k$ of the coefficients $t_x$. Letting $E_i = \{x \in E | t_x = t_i\}$ and setting $p_i = p(E_i) = \sum_{x \in E_i} x$, we have $a = \sum_i t_i p_i$, with $p_1, ..., p_n$ jointly orthogonal. Suppose there is another such decomposition, say $a = \sum_j s_j q_j$, with $q_j = p(F_j) = \sum_{y \in F_j} \widehat{y}$, where $F_1, ..., F_l \subseteq F \in \mathcal{M}(A)$ are pairwise disjoint, and again with the coefficients in descending order, say $s_o > s_1 > \cdots > s_l$.

**Lemma B2:** *In the situation described above, $t_o = s_o$ and $p_o = q_o$.*

*Proof:* First, notice that for any choice of $x_o \in E_o$, and with $\alpha = |x_o\rangle$, we have $\alpha(p_o) = 1$ and $\alpha(p_i) = 0$ for $i > o$. Thus,

$$t_o = \alpha(a) = \sum_j s_j \alpha(q_j).$$

Since the coefficients $\alpha(q_j)$ are convex, the right-hand side lies in the convex hull of the values $s_j$ and, in particular, is no larger than the largest of these values, $s_o$. Thus, $t_o \leq s_o$. A similar argument shows that $s_o \leq t_o$. Thus, $s_o = t_o$.

Now let $x \in E_o$: then
$$\langle \widehat{x}, p_o \rangle = \sum_{y \in E_o} \langle \widehat{x}, \widehat{y} \rangle = 1,$$
whence, $\langle \widehat{x}, a \rangle = t_o$. But we then have (using the fact that $s_o = t_o$)

$$t_o = \langle \widehat{x}, a \rangle = \left\langle \widehat{x} \ , \ t_o q_o + \sum_{j=1}^{l} s_j q_j \right\rangle = t_o \langle \widehat{x}, q_o \rangle + \sum_{j=1}^{l} s_j \langle x, \widehat{q}_j \rangle.$$

Since $\sum_{j=0}^{l} \langle x, q_j \rangle \leq 1$, the last expression above is a sub-convex combination of the values $s_o, ...., s_m$. This can equal $t_o = s_o$, the maximum of these values, only if $\langle \widehat{x}, q_o \rangle = 1$ and $\langle \widehat{x}, q_j \rangle = 0$ for the remaining $q_j$. It follows that $\langle p_o, q_o \rangle = \sum_{x \in E_o} \langle \widehat{x}, q_o \rangle = |E_o| = \|p_o\|^2$. The same argument, with $p$'s and $q$'s interchanged, shows that $\langle p_o, q_o \rangle = \|q_o\|^2$. Hence, $\|p_o\| = \|q_o\|$, and $\langle p_o, q_o \rangle = \|p_o\|^2 = \|p_o\| \|q_o\|$, whence, $p_o = q_o$ $\square$

**Proposition B1:** *Every $a \in \mathbf{V}^*(A)$ has a unique expansion of the form $a = \sum_{i=0}^{k} t_i p_i$ where $t_0 > t_1 > ... > t_k$ are non-zero coefficients and $p_1, ..., p_n$ are jointly orthogonal sharp effects.*

*Proof:* Suppose $a = \sum_{i=1}^{k} t_i p_i$, as above, and also $a = \sum_{j=1}^{l} s_j q_j$, $s_0 > \cdots > s_l > 0$, and $q_j$ pairwise orthogonal extremal effects. We shall show that $k = l$, and that $t_i = s_i$ and $p_i = q_i$ for each $i = 1, ..., k$. The Lemma above tells us that $t_o = s_o$ and $p_o = s_o$. Hence,

$$\sum_{i=1}^{k} t_i p_i = a - t_o p_o = a - s_o q_o = \sum_{j=1}^{l} s_j q_j.$$

Applying the Lemma recursively, we find that $t_i = s_i$ and $p_i = q_i$ for $i = 1, ..., \min(k, l)$. If $k \neq l$, say $k < l$, we then have

$$t_k p_k = s_k q_k + \sum_{j=k+1}^{l} s_j q_j = t_k p_k + \sum_{j=k+1}^{l} s_j q_j$$

whence, $\sum_{j=k+1}^{l} s_j q_j = 0$, which is impossible since all $q_j$ are sharp and the coefficients $s_j$ are strictly positive. Hence, $l = k$ and the proof is complete. $\square$

# C   Bits are balls

In most other reconstructions of QM, the first step is to show that the state space of a "bit" — that is, a system in which every state is the mixture of two sharply-distinguishable pure states — is a ball. In our approach, this fact is an easy consequence of Lemma 1. In our framework, we will define a bit to be a sharp model $A$ with uniform rank 2, in which every state has the form $t\delta_x + (1-t)\delta_y$, where $\{x, y\} \in \mathcal{M}(A)$. Note that this implies $A$ is spectral.

29

**Lemma C1:** *Let $A$ be a bit with conjugate $\overline{A}$. Then $\Omega(A)$ is a ball, with extreme points $\delta_x$.*

*Proof:* By Lemma 1, $\mathbf{E}(A)$ carries a self-dualizing inner product such that $\|x\| = 1/n$ for each outcome $x \in X(A)$ and $\|u\| = 1$. Also, if $\{x, y\} \in \mathcal{M}(A)$, then $\langle x, y \rangle = 0$, whence, $\langle u, x \rangle = \langle x, x \rangle = 1/n$. It will be convenient to adjust the normalization so that $\langle x, x \rangle = 1$ for outcomes $x \in X(A)$, whence, $\langle u, u \rangle = n$. We can now represent states by vectors $a \in \mathbf{E}(A)_+$ with $\langle a, u \rangle = 1$. In particular, the maximally mixed state corresponds to the vector $\frac{1}{n}u$. To simplify the notation, let us agree for the moment to write $\rho$ for this vector. Then

$$\langle \rho, x \rangle = \frac{1}{n} \text{ and } \langle \rho, \rho \rangle = \frac{1}{n^2}\langle u, u \rangle = \frac{1}{n}.$$

Hence,

$$\|\rho - x\|^2 = \|\rho\|^2 - 2\langle \rho, x \rangle + \|x\|^2 = \frac{1}{n} - 2\frac{1}{n} + 1 = 1 - \frac{1}{n}.$$

If $n = 2$, we see that $\|\rho - x\| = 1/\sqrt{2}$. Thus, $X(A)$ lies on the sphere of radius $1/\sqrt{2}$ about the unit $\rho$. I now claim that any $a \in \mathbf{E}(A)$ with $\langle a, u \rangle = 1$ — that is, any state — wth $\|u - a\| \leq 1/\sqrt{2}$, belongs to the positive cone $\mathbf{E}(A)_+$. To see this, use spectrality to decompose $a$ as $sx + ty$ where $\{x, y\} \in \mathcal{M}(A)$. Consider now the two-dimensional subspace $\mathbf{E}_{x,y}$ spanned by $x$ and $y$. With respect to the inner product inherited from $\mathbf{E}$, we can regard this as a 2-dimensional euclidean space, in which $a$ is represented by the Cartesian coordinate pair $(s, t)$. Expanding $\rho$ as $\rho = \frac{1}{2}(x + y)$, we see that $\rho \in \mathbf{E}_{x,y}$ with coordinates $(1/2, 1/2)$. The point $(t, s)$ lies, therefore, in the disk of radius $1/\sqrt{2}$ centered at $(1/2, 1/2)$ in $\mathbf{E}_{x,y}$. Moreover, as $\langle u, a \rangle = 1$, we see that $s + t = 1$, i.e., $(s, t)$ lies on the line of slope $-1$ through $(1/2, 1/2)$. This puts $(s, t)$ in the positive quadrant of this plane, i.e., $s \geq 0$ and $t \geq 0$. But then $a \in \mathbf{E}_+$, as claimed. $\square$

It follows that, for rank-two models, we do not even need to invoke homogeneity: they all correspond to spin factors. Letting $d$ denote the dimension of the state space (that is, $d = \dim(\mathbf{E}) - 1$), we see that if $d = 1$, we have the clasical bit; $d = 2$ gives the real quantum-mechanical bit, $d = 3$ gives the familiar Bloch sphere, i.e., the usual qubit of complex QM, while $d = 5$ corresponds to the quaternionic unit sphere, giving us the quaternionic bit. The generalized bits with $d = 4$ and $d \geq 6$ are more exotic "post-quantum" possibilities.

# D  Locally tomography and dagger-compactness

A *dagger* on a category $\mathcal{C}$ is a contravariant functor $\dagger : \mathcal{C} \to \mathcal{C}$ that is the identity on objects, and satisfies $\dagger \circ \dagger = \mathrm{id}_{\mathcal{C}}$. That is, if $A \xrightarrow{f} B$ is a morphism in $\mathcal{C}$, then $A \xleftarrow{f^\dagger} B$, with $f^{\dagger\dagger} = f$ and $(f \circ g)^\dagger = g^\dagger \circ f^\dagger$ whenever $f \circ g$ is defined. An isomorphism $f : A \simeq B$ in $\mathcal{C}$ is then said to be *unitary* iff $f^\dagger = f^{-1}$. One says that $\mathcal{C}$ is $\dagger$-*monoidal* iff $\mathcal{C}$ is equipped with a symmetric monoidal structure $\otimes$ such that $(f \otimes g)^\dagger = f^\dagger \otimes g^\dagger$, and such that the canonical isomorphisms $\alpha_{A,B,C}$, $\sigma_{A,B}$, $\lambda_A$ and $\rho_A$ are all unitary.

Recall that a *dual* for an object $A$ in a symmetric-monoidal category $\mathcal{C}$ is a structure $(A', \eta, \epsilon)$ where $A' \in \mathcal{C}$ and $\eta : I \to A \otimes A'$ and $\epsilon : A' \otimes A \to I$, such that

$$(\mathrm{id}_A \otimes \epsilon) \circ (\eta \otimes \mathrm{id}_A) = \mathrm{id}_A \quad \text{and} \quad (\epsilon \otimes \mathrm{id}_{A'}) \circ (\mathrm{id}_{A'} \otimes \eta) = \mathrm{id}_{A'}$$

up to the natural associator and unit isomorphisms. If $\mathcal{C}$ is †-monoidal and $\epsilon = \sigma_{A,A'} \circ \eta_A^\dagger$, then $(A', \eta, \epsilon)$ is a *dagger-dual*. A category in which every object has a specified dual is *compact closed*, and a dagger-monoidal category in which every object has a given dagger-dual is *dagger-compact* [1, 32]

An important example of all this is the category — I'll denote it by $\mathbf{FdHilb}_{\mathbb{R}}$ — of finite-dimensional real Hilbert spaces and linear mappings. If $A$ and $B$ are two such spaces and $\phi : A \to B$, let $\phi^\dagger$ be the usual adjoint of $\phi$ with respect to the given inner products. Letting $A \otimes B$ be their usual tensor product (in particular, with $\langle x \otimes y, u \otimes v \rangle = \langle x, u \rangle \langle y, v \rangle$ for $x, u \in A$ and $y, v \in B$) . Then $\mathbf{FdHilb}_{\mathbb{R}}$ is a dagger-monoidal category with $\mathbb{R}$ as the monoidal unit.

Since any $A \in \mathbf{FdHilb}_{\mathbb{R}}$ is canonically isomorphic to its dual space, we have also a canonical isomorphism $A \otimes A \simeq A^* \otimes A = \mathcal{L}(A, A)$, and a canonical trace functional $\mathrm{Tr} : A \otimes A \to \mathbb{R}$. Taking $A = A'$, let $\eta_A \in A \otimes A$ be given by $\eta_A = \frac{1}{n} \sum_i x_i \otimes x_i$, where the sum is taken over any orthonormal basis $\{x_i\}$ for $A$; then for any $a \in A \otimes A$, $\langle \eta_A, a \rangle = \mathrm{Tr}(a)$. It is routine to show that $\epsilon_A = \sigma_{A,A} \circ \eta_A^\dagger$, so that $\eta_A$ and $\epsilon_A$ make $A$ its own dagger-dual.

In any compact closed symmetric monoidal category $\mathbb{C}$, every morphism $\phi : A \to B$ yields a *dual* morphism $\phi' : B' \to A'$ by setting

$$\phi' = (\mathrm{id}_{A'} \otimes \epsilon_B) \circ (\mathrm{id}_{A'} \otimes f \otimes \mathrm{id}_{B'}) \circ (\eta_A \otimes \mathrm{id}_{B'}).$$

(again, suppressing associators and left and right units). For $\phi : A \to B$ in $\mathbf{FdHilb}_{\mathbb{R}}$, one has, for any $v \in A$,

$$\phi'(v) = \sum_{x \in M} \langle v, f(x) \rangle x = \sum_{x \in M} \langle f^\dagger(v), x \rangle x = f^\dagger(v),$$

i.e., $\phi' = \phi^\dagger$.

Now let $\mathcal{C}$ be a monoidal probabilistic theory — that is, a category of probabilistic models and processes, with a symmetric monoidal structure $A, B \mapsto AB$, where $AB$ is a (non-signaling) composite in our sense, and with tensor unit $I = \mathbb{R}$. I will further assume that

(a) Every $A \in \mathcal{C}$ has a conjugate, $\overline{A} \in \mathbb{C}$, with $\overline{\overline{A}} = A$;

(b) For all $A, B \in \mathcal{C}$ and $\phi \in \mathbb{C}(A, B)$, $\overline{\phi} \in \mathbb{C}(\overline{A}, \overline{B})$;

(c) $\overline{\overline{A}} = A$, with $\eta_{\overline{A}}(\overline{a}, b) := \eta_A(a, \overline{b})$.

*Remarks:* (1) The chosen conjugate $\overline{A}$ for $A \in \mathcal{C}$ required by condition (a) is equipped with a canonical isomorphism $\gamma_A : A \simeq \overline{A}$, with $\overline{x} = \gamma(x)$ for every $x \in X(A)$. As discussed in Section ..., this extends to an order-isomorphism $\mathbf{E}(A) \simeq \mathbf{E}(\overline{A})$, which we again write as $\gamma_A(a) = \overline{a}$ for $a \in \mathbf{E}(A)$. Notice, however, that $\gamma_A$ is *not* assumed to be a morphism in $\mathcal{C}$.

(2) In spite of this, condition (b) requires that $\overline{\phi} = \gamma_B \circ \phi \circ \gamma_A^{-1}$ *does* belong to $\mathcal{C}(\overline{A}, \overline{B})$ for $\phi \in \mathcal{C}(A, B)$. Notice here that $\phi \mapsto \overline{\phi}$ is functorial.

(3) The second part of condition (c) is redundant if every model $A$ in $\mathbb{C}$ is sharp (since in this case there is at most one correlator between $\overline{A}$ and $A$). Notice, too, that condition (c) implies that

$$\langle \overline{x}, \overline{y} \rangle = \eta_{\overline{A}}(\overline{x}, y) = \eta_A(x, \overline{y}) = \langle x, y \rangle$$

for all $x, y \in \mathbf{E}(A)$.

We are now ready to prove Theorem 3. We wish to show that if every $A \in \mathcal{C}$ is spectral, then $\mathcal{C}$ has a canonical dagger, with respect to which it is dagger compact.

Before proceeding, it will be convenient to dualize our representation of morphisms, so that $\phi \in \mathcal{C}(A, B)$ means that $\phi$ is a positive linear mapping $\mathbf{E}(B) \to \mathbf{E}(A)$.[15] By Lemma 1, for every $A \in \mathcal{C}$, the space $\mathbf{E}(A)$ carries a canonical self-dualizing inner product $\langle \, , \, \rangle_A$, with respect to which $\mathbf{E}(A) \simeq \mathbf{V}(A)$.

**Lemma D1:** *For all models $A, B \in \mathcal{C}$, the inner product on $\mathbf{E}(AB)$ factors, in the sense that if $a, x \in \mathbf{E}(A)$ and $b, y \in \mathbf{E}(B)$, then $\langle a \otimes b, x \otimes y \rangle = \langle a, x \rangle \langle b, y \rangle$.*

*Proof:* This follows from the sharpness of $A, B$ and $AB$. For $a \in X(A)$, $b \in X(B)$, let $\delta_a, \delta_b$ and $\delta_{a,b}$ denote the unique states of $A$, $B$ and $AB$ such that $\delta_a(a) = \delta_b(b) = \delta_{a,b}(a, b) = 1$. Since $(\delta_a \otimes \delta_b)(a \otimes b)$ is also 1, we conclude that $\delta_{a,b} = \delta_a \otimes \delta_b$. But we also have $\delta_a(x) = n\langle a, x \rangle$, $\delta_b(y) = m\langle b, y \rangle$ and $\delta_{a \otimes b}(x \otimes y) = nm\langle a \otimes b, x \otimes y \rangle$, which establishes the claim. $\square$

It follows that $\mathcal{C}$ is a monoidal subcategory of $\mathbf{FdHilb}_{\mathbb{R}}$. In effect, we are going to show that $\mathcal{C}$ inherits a dagger-compact structure from $\mathbf{FdHilb}_{\mathbb{R}}$, with the minor twist that we will take $\overline{A}$, rather than $A$, as the dual for $A \in \mathcal{C}$. We define the dagger of $\phi \in \mathcal{C}(A, B)$ to be the hermitian adjoint of $\phi : \mathbf{E}(A) \to \mathbf{E}(B)$ with respect to the canonical inner products on $\mathbf{E}(A)$ and $\mathbf{E}(B)$. At this point, it is not obvious that $\phi^\dagger$ belongs to $\mathcal{C}$. In order to show that it does, we first need to show that $\mathcal{C}$ is compact closed. To define the unit, let $e_A \in \mathbf{E}(\overline{A}) \otimes \mathbf{E}(\overline{A}) = \mathbf{E}(\overline{A}A)$ (note the use of local tomography here) to be the vector with $\langle e_A, \cdot \rangle = \eta_A$, i.e, for all $a, b \in \mathbf{E}(A)$,

$$\langle e_A, \overline{a} \otimes b \rangle = \eta_A(a \otimes \overline{b}) = \langle a, b \rangle.$$

Since $\mathbf{E}(A\overline{A})$ is self-dual, $e_A \in \mathbf{E}(A\overline{A})_+$.

---

[15]Thus, our co-unit $\eta \in \mathcal{C}(I, A \otimes A')$ should becomes a positive linear mapping $\eta_A : \mathbf{E}(A \otimes A') \to \mathbb{R}$, and similarly, a unit $\epsilon_A \in \mathcal{C}(A' \otimes A, I)$ should be a positive linear mapping $\mathbb{R} \to \mathbf{E}(A' \otimes A)$, i.e, an element of $\mathbf{E}(A \otimes A')$.

**Lemma D2:** *With $\eta_A$ and $e_A$ defined as above, $\overline{A}$ is a dual for $A$ for every $A \in \mathcal{C}$. In particular, $\mathcal{C}$ is compact closed.*

*Proof:* Choose an orthonormal basis $M \subseteq \mathbf{E}(A)$. Local tomography and Lemma D1 tell us that $\overline{M} \otimes M = \{\overline{x} \otimes x | x \in M\}$ is then an orthonormal basis for $\mathbf{E}(\overline{A}A)$. (Note here that $x, y \in M$ are not necessarily even positive, let alone in $X(A)$.) If we expand $e_A$ with respect to this basis, we have

$$e_A = \sum_{x,y \in M} c_{x,y} \overline{x} \otimes y.$$

Then

$$\langle e_A, \overline{x} \otimes x \rangle = \langle x, x \rangle = \|x\|^2 = 1$$

since the basis is orthogonal, and for $x \neq y$, both in $M$,

$$\langle e_A, \overline{x} \otimes y \rangle = \langle x, y \rangle = 0$$

Hence, $e_A = \sum_{x \in M} \overline{x} \otimes x$. We now have

$$
\begin{aligned}
(\eta_A \otimes \mathrm{id}_A) \circ (\mathrm{id}_A \otimes e_A)(v) &= (\eta_A \otimes id_A)\left(\sum_{x \in M} v \otimes \overline{x} \otimes x\right) \\
&= \sum_{x \in M} \eta_A(v \otimes \overline{x})x \\
&= \sum_{x \in M} s\langle v, x \rangle x = v.
\end{aligned}
$$

Similarly, for $\overline{v} \in \overline{A}$,

$$
\begin{aligned}
(\mathrm{id}_{\overline{A}} \otimes \eta_{\overline{A}}) \circ (e_{\overline{A}} \otimes \mathrm{id}_{\overline{A}})(\overline{v}) &= (\mathrm{id}_{\overline{A}} \otimes \eta_{\overline{A}})\left(\sum_{x \in M} \overline{x} \otimes x \otimes \overline{v}\right) \\
&= \sum_{x \in M} \overline{x}\eta_A(x, \overline{v}) = \sum_{x \in E} \langle x, v \rangle \overline{x} \\
&= \sum_{x \in E} \langle \overline{v}, \overline{x} \rangle \overline{x} = \overline{v}. \quad \square
\end{aligned}
$$

**Lemma D3:** *If $\phi : \mathbf{E}(A) \to \mathbf{E}(B)$ belongs to $\mathcal{C}$, then so does $\phi^\dagger : \mathbf{E}(B) \to \mathbf{E}(A)$.*

*Proof:* Using the compact structure on $\mathcal{C}$ defined above, if $\phi : A \to B$, we construct the dual of $\overline{\phi}$:

$$\overline{\phi}' := (\eta_B \otimes \mathrm{id}_A) \circ (\mathrm{id}_B \otimes \overline{\phi} \otimes \mathrm{id}_A) \circ (\mathrm{id}_B \otimes e_A) : \mathbf{E}(B) \to \mathbf{E}(A)$$

a morphism in $\mathcal{C}$. Now, if $b \in \mathbf{E}(B)$, applying this last mapping gives us

$$b \mapsto (\eta_B \otimes \mathrm{id}_A)\left(\sum_{x \in M} b \otimes \overline{\phi}(\overline{x}) \otimes x\right) = \sum_{x \in M} \eta_B(b, \overline{\phi}(\overline{x}))x.$$

33

This last expression can be rewritten:

$$\sum_{x \in M} \eta_B(b, \overline{\phi}(\overline{x}))x = \sum_{x \in M} \langle b, \phi(x) \rangle x = \sum_{x \in M} \langle \phi^\dagger(b), x \rangle x = \phi^\dagger(b).$$

Thus, the mapping $\overline{\phi}'$ above is $\phi^\dagger$, which, therefore, belongs to $\mathcal{C}$.

Thus, $\mathcal{C}$ is a dagger, as well as a monoidal, sub category of $\mathbf{FdHilb}_\mathbb{R}$. Hence, the associator, swap, and left- and right-unit morphisms associated with an object $A \in \mathcal{C}$ are all unitary (since they are unitary in $\mathbf{FdHilb}_\mathbb{R}$), whence, $\mathcal{C}$ is dagger-monoidal. To complete the proof of Theorem 3, we need to check that $\eta_A = e_A^\dagger \circ \sigma_{A,\overline{A}} : \mathbf{E}(A\overline{A}) \to \mathbb{R}$. In view of our local tomography assumption, it is enough to check this on pure tensors, where a routine computation gives us $e_A^\dagger(\sigma_{A,\overline{A}}(a \otimes \overline{b})) = \langle e_A^\dagger(\overline{b} \otimes a), 1 \rangle_1 = \langle \overline{b} \otimes a, e_A \rangle_{\overline{A}A} = \langle a, b \rangle = \eta_A(a \otimes \overline{b})$. $\square$

*Remark:* Given that $\mathcal{C}$ is compact closed, with $\overline{A}$ the dual of $A$, the functoriality of $\phi \mapsto \overline{\phi}$ makes $\mathcal{C}$ *strongly* compact closed, in the sense of [1]. This is equivalent to dagger-compactness.