A Primer on Probabilistic Models

John Harding and Alex Wilce

Abstract We give a basic introduction to a generalized setting for probability theory, known as probabilistic models, that evolved over the past half century from attempts to provide an axiomatic foundation for quantum mechanics. Classical probability theory and finite-dimensional quantum mechanics are just two examples from a spectrum of possibilities.

Keywords: Base-normed space. Convex. Effect. Order-unit space. Orthoalgebra. Probabilistic model. Quantum mechanics. Simplex. Test space.

1 Introduction

The simplest situation in probability is that of a finite probability space. Here, one begins with a finite set $X = \{x_1, \ldots, x_n\}$ of mutually exclusive possible *outcomes*. A *probability weight* on X is a mapping $\alpha : X \to [0,1]$ that sums to unity. Probability weights are also called *states* since they describe the way a system modeled by X can be. For instance, a generic 2-sided coin can be modeled by $X = \{h,t\}$ and the various probability weights describe the inherent propensity for the coin to provide a head or tail when flipped.

The pair $A = (X, \Omega)$ consisting of a finite probability space together with its set Ω of states is an example of a *probabilistic model* (PM). The collection of all PMs arising from finite probability spaces is an example of a *probabilistic theory*. Here we presume that this theory includes a description of what will be admissible morphisms between the PMs that it contains, and also a means to combine two PMs

---

John Harding
New Mexico State University (USA)
e-mail: jharding@nmsu.edu

Alex Wilce
Susquehanna University (USA)
e-mail: wilce@susqu.edu

A and B into a single model $A \otimes B$ in the theory, representing the compound system. For instance, if A is the PM for a generic 2-sided coin, then $A \otimes A$ would be the PM for a pair of such coins.

The situation described above can be generalized in many ways. For instance the notion of a finite probability space can be generalized to that of a measurable space, and that of a probability weight to a finitely or countably additive probability measure. This will also provide an example of a probabilistic theory. But there is another direction of generalization that is of quite a different nature: we allow the possibility that some outcomes cannot be simultaneously tested. Roughly speaking, we begin with a *test space*, a set X of outcomes together with a family of subsets $E, F, \ldots \subseteq X$ called *tests*, representing classical discrete outcome-sets, subject to some modest conditions. A *probability weight* on this is a mapping $\alpha : X \to [0,1]$ that sums to unity over each test.

A probabilistic model (PM) is a pair $A = (\mathscr{A}, \Omega)$ consisting of a test space and a chosen collection of its probability weights, which we call the *states* of the model. A *probabilistic theory* is a collection of such PMs, together with suitable morphisms, and a means to form compound systems.

The origins of probabilistic models in this broad sense lie in studies of quantum mechanics, where a fundamental property is the incompatibility of certain quantities such as position and momentum that cannot be precisely measured together in a single experiment. The aim was to consider features that are common to a range of probabilistic situations, whatever their origin, and then to find constraints on probabilistic theories that lead to classical or quantum mechanics. These probabilistic models provide a language to speak precisely about probabilistic concepts such as incompatibility, contextuality, and entanglement.

The study of probabilistic models, though not by that name, began with work of Mackey [13] in the 1950's, Ludwig [11] in the 50's and 60's, Foulis and Randall [15] and Piron [8] beginning in the 1970's, and Wilce [16, 17] from the 1990's to the present. In the past decade Hardy [9], Masanes and Müller [14], and many others [6] have incorporated aspects of probabilistic models in developing axiomatizations of finite-dimensional quantum theory.

A feature of the study of probabilistic models is the wide range of tools, drawn from diverse areas of mathematics, that are used to gain insight into probabilistic phenomena. This is true even in the setting of a finite probability space $A = (X, \Omega)$: a subset of the outcome-set X is an *event*, and the collection of all events is a Boolean algebra called the *logic* of the probability space. The collection of all probability weights on X is a compact convex subset of the vector space $\mathbb{R}^X$, in this case an $n-1$-dimensional simplex. Its extreme points are the point masses on X. The state space Ω is the base of a cone in the vector space $\mathbb{R}^X$ and this cone provides a partial ordering on V, in this case yielding a vector lattice.

Thus, one can view a finite probability space through various lenses, ranging from logical to linear algebraic to geometric. As we will see, the same is true for a general PM. However, in general the "logics" will not be Boolean, the state-spaces will not be simplices, and the ordered vector spaces will not be vector lattices.

This note is arranged in the following way. The second section provides basic definitions and examples. The third section describes how to associate "logics" to certain test spaces. The fourth section the discusses convex geometrical view of state spaces. The fifth section describes the ordered vector space view of PMs. The sixth section discusses the relation between PMs and classical probability spaces. The seventh section treats compound systems and the eighth treats the non-classical phenomenon of entanglement. The final section briefly discusses probabilistic theories. This is a wide range of topics, and our aim here, given space constraints, is simply to provide an outline of the theory.

2 Basic setup and examples

Definition 1. A *test space* $\mathscr{A}$ is a set of non-empty sets, none of which properly contains another.

We call the sets that comprise a test space its *tests* and the elements of a test its *outcomes*. The union of all of the tests of the test space is the set of all possible outcomes of its tests, and is called the *outcome-set* of the test space. Tests may be infinite, and an outcome can belong to any number of tests except zero.

Definition 2. Let $\mathscr{A}$ be a test space with outcome-set X. A *probability weight* on $\mathscr{A}$ is a map $\alpha : X \to [0,1]$ that sums to unity over each test in the sense that the associated net of finite partial sums converges to 1.

Definition 3. A *probabilistic model* (PM) is a pair $A = (\mathscr{A}, \Omega)$ consisting of a test space $\mathscr{A}$ and a family Ω of probability weights on $\mathscr{A}$. We refer to members of Ω as *states*.

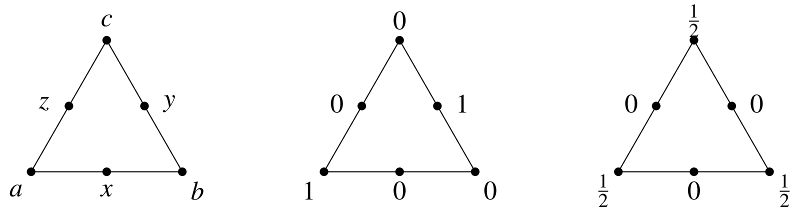
The idea is that we have a situation where we can perform tests, and each test has its set of possible outcomes. One may be able to test for a given outcome in different ways, and it may be impossible to test for certain pairs of outcomes simultaneously. In defining states as maps $\alpha : X \to [0,1]$, we make the assumption that when the system is in a given state, the probability of obtaining an outcome x is independent of how x is tested. This is called *non-contextuality*.

Example 1. A finite probability space X is a PM $A = (\mathscr{A}, \Omega)$ where the test space $\mathscr{A} = \{X\}$ consists of the single set X, and the state space Ω consists of the set of all probability weights on X.

Example 2. For a measurable space (S, Σ), we can construct a test space $\mathscr{A}$ whose outcome-set is the collection of all non-empty measurable sets, and whose tests are the *finite partitions of unity*, that is, the collections of finitely many pairwise disjoint measurable sets whose union covers S. The probability weights on $\mathscr{A}$ are the finitely additive probability measures on the given measurable space. Alternately, we can form a test space $\mathscr{A}'$ from the finite or countable measurable partitions of unity, and the probability weights on this test space are the probability measures in the usual sense.

Example 3. For a separable Hilbert space $\mathscr{H}$, form a test space $\mathscr{A}$ by taking its outcomes to be unit vectors and its tests to be orthonormal bases. If $\dim \mathscr{H} > 2$, Gleason's theorem [7] gives that probability weights on $\mathscr{A}$ correspond to density operators W via $\alpha_W(v) = \langle Wv \mid v \rangle$.

Example 4 (Firefly). A firefly is in a triangular box divided into three chambers, a, b and c. Each side of the box gives a view of two of the chambers. An experiment consists of viewing the box from one side and noting whether a light appears on one side or the other, or not at all. We represent this with a test space $\mathscr{A} = \{\{a,x,b\},\{b,y,c\},\{c,z,a\}\}$ where x,y and z represent seeing no light in the appropriate window. The diagrams below depict this test space in Figure (a), and two probability weights on it in Figures (b) and (c).



(a) (b) (c)

3 The logic of a test space

For a finite probability space on a set X, its *events* are the subsets of X. These form a Boolean algebra, a powerset algebra, and provide a *logic* of events. We have notions of the conjunction, disjunction, and negation of events, and can reason with these using the connections between classical logic and Boolean algebras. The same holds true of probability models based on a measurable space, where the events are the measurable subsets and the collection of all events is the Boolean σ-algebra of measurable sets. This connection to classical logic is a fundamental aspect of probability.

Definition 4. For a test space $\mathscr{A}$ with outcome-set X, a subset $a \subseteq X$ is an *event* if a is a subset of a test of $\mathscr{A}$. Two events a,c are *complementary* if they are disjoint and their union is an event. Two events a,b are perspective, written $a \sim b$, if they have a common complement.

In classical models, each event has a unique complement, and perspectivity is trivial. Not so in our quantum models, or in the firefly model. For instance, in the latter, the events $\{a,x\}$ and $\{y,c\}$ are perspective, both being complements for the event $\{b\}$.

Our aim is to construct a logic for a test space through equivalence classes of events. In general, this is not possible since test spaces are very general structures, and perspectivity can be poorly behaved. However, when a test space is *algebraic*, meaning that perspective elements have exactly the same complements, we can do much better. In particular, for an algebraic test space, perspectivity is an equivalence relation.

Definition 5. For an algebraic test space $\mathscr{A}$, its *logic* $\Pi\mathscr{A}$ is the set of equivalence classes of events under perspectivity.

The logic of an algebraic test space carries natural structure: a partial ordering, a complementation, and a partially defined sum operation. Writing the equivalence class of an event a as $[a]$, the partial ordering puts $[a] \leq [b]$ if $a \subseteq b$. That this is well-defined follows from algebraicity. Similarly, the complementation is given by $[a]' = [c]$ where c is any complement of a. The partial sum operation is defined for pairs $[a]$, $[b]$ where the events a, b are disjoint and contained in a common test and then their sum is given by $[a] \oplus [b] = [a \cup b]$. These are well-defined thanks to algebraicity. The resulting structure $(\Pi\mathscr{A}, \oplus)$ is an *orthoalgebra*:

Definition 6. An *effect algebra* is a set Z with a partially defined binary operation $\oplus$ that is commutative and associative in the natural sense and constants $0, 1$ that satisfies for all $z \in Z$

1. $z \oplus 0 = z$;
2. if $z \oplus 1$ is defined, then $z = 0$;
3. there is a unique element z' with $z \oplus z' = 1$.

This structure is an *orthoalgebra* if $z \oplus z$ defined implies $z = 0$.

Proposition 1 (Foulis). *The logic of an algebraic test space is an orthoalgebra, and all orthoalgebras arise as such logics.*

Example 5. Each Boolean algebra is an orthoalgebra. The test spaces arising from finite probability spaces and from measurable spaces are algebraic and their logics are Boolean algebras. The test space constructed from a separable Hilbert space $\mathscr{H}$ in Example 3 is algebraic. Its logic is an orthoalgebra, in fact it is the orthomodular lattice of closed subspaces of $\mathscr{H}$. The test space called the Firefly in Example 4 is algebraic. Its logic consists of three 8-element Boolean algebras glued together in pairs at an atom and coatom.

Example 6. The real unit interval $[0, 1]$ is an effect algebra where $z' = 1 - z$ and $x \oplus y = x + y$ when this sum is at most 1. This extends in a natural way to produce an effect algebra from an interval $[0, u]$ of any partially ordered abelian group, and in particular from any partially ordered vector space. A primary example of this is its application to the real vector space of bounded self-adjoint operators on a Hilbert space $\mathscr{H}$. Its unit interval $[0, I]$ is an effect algebra and its elements are called the *standard effects* of the Hilbert space.

An element z of an effect algebra is called *sharp* if 0 is the only lower bound of z, z'. Under the mild conditions, the sharp elements of an effect algebra form an orthoalgebra [10].

Example 7. Consider the ordered vector space V of continuous real-valued functions on a space X. Its interval effect algebra $[0,1]_V$ consists of the continuous functions taking values in the real interval $[0,1]$ and its sharp elements are the Boolean algebra of continuous functions $f : X \to \{0,1\}$. The sharp elements of the effect algebra of standard effects of a Hilbert space is the orthomodular lattice of projection operators.

4 The geometry of the state space

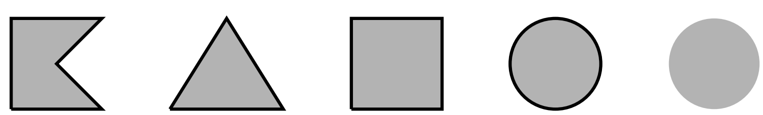
As we discussed, a PM $A = (\mathscr{A}, \Omega)$ consists of a test space $\mathscr{A}$ and a collection Ω of probability weights on $\mathscr{A}$ called states. We are free to choose whatever collection of probability weights we like. One frequent choice is to take Ω to be all probability weights on $\mathscr{A}$. If $\mathscr{A}$ has outcome-set X, the probability weights on $\mathscr{A}$ are a subset of the vector space $\mathbb{R}^X$. This subset is *convex* meaning that if α, β are probability weights and $s, t \geq 0$ with $s + t = 1$, then $s\alpha + t\beta$ is also a probability weight. Further, the probability weights are compact as a subset of $\mathbb{R}^X$ with the product topology.

Definition 7. A PM $A = (\mathscr{A}, \Omega)$ with outcome-set X is called *standard* if Ω is a closed, hence compact, convex subset of $\mathbb{R}^X$.

One can motivate the requirement of convexity as follows: if α, β are allowed states of the system, then by preparing a large number of system with the proportion s in state α and t in state β, we effectively have a system in state $s\alpha + t\beta$. Compactness can be motivated by viewing the pointwise limit of a family of states as a sort of ideal state.

Definition 8. Suppose $A = (\mathscr{A}, \Omega)$ is a PM with Ω is convex. A state α is a *pure state* of A if it is an extreme point of the state space Ω, that is, if it cannot be expressed as a non-trivial convex combination of other states.

Example 8. In the illustration, the figure at left is not convex. The triangle is compact and convex, as is the square and the disc with boundary. The disc without boundary is convex but not compact.



The extreme points of the triangle are its vertices, those of the square are its corners. Each point on the boundary of the disc with boundary is an extreme point, while the disc without boundary has no extreme points.

A result of Caratheodory provides that if Ω is a compact convex subset of $\mathbb{R}^n$, then each point in Ω is a convex combination of its extreme points. A subset $\Omega \subseteq \mathbb{R}^n$ is a k-dimensional *simplex* if it has $k+1$ extreme points and each point in Ω can be uniquely expressed as a convex combination of its extreme points. So for example, a 2-dimensional simplex is a triangle and a 3-dimensional simplex is a tetrahedron.

Example 9. For a finite probability space $X = \{x_1, \ldots, x_n\}$, its probability weights are an $(n-1)$-dimensional simplex. The extreme points are the point masses δ_i taking value 1 at x_i and 0 otherwise, and each probability weight α can be uniquely expressed as a convex combination of these, $\alpha = s_1\delta_1 + \cdots + s_n\delta_n$ where $s_i = \alpha(x_i)$.

Example 10. Consider the PM $A = (\mathscr{A}, \Omega)$ where $\mathscr{A}$ has two tests $\{a,x\}$ and $\{b,y\}$ and Ω is all of its probability weights. This can be viewed as a system where there are 4 outcomes a,x,b,y, but one of a,x must occur if tested, and one of b,y must occur if tested. Probability weights are given by specifying the probability that x occurs if tested, and the probability that y occurs if tested, since the probabilities of a,b are then determined. Thus, Ω is given by the unit square in the $x-y$ plane.

Example 11. Let $A = (\mathscr{A}, \Omega)$ be the PM obtained from the 2-dimensional Hilbert space $\mathbb{C}^2$ whose outcomes are unit vectors and whose states are those given by density operators. It is well known that this state space is convexly and topologically isomorphic to a closed ball in $\mathbb{R}^3$ known as the *Bloch sphere*. State spaces of higher-dimensional quantum models are also compact convex sets, but with much richer geometry [4].

The basis of the convexity approach to PMs is that geometric properties of the state space provide some (but not all) information about the system. For instance, with a finite probability space $X = \{x_1, \ldots, x_n\}$ there is no uncertainty associated with the point weights δ_i, when the system is in such a state, there is no doubt about the outcome of a measurement. The uncertainty associated with a general state $\alpha = s_1\delta_1 + \cdots + s_n\delta_n$ is solely in the proportions s_i used to form the mixture representing α. If we have sufficiently many copies of the identically prepared systems, we can estimate these proportions through repeated experiments.

For general PMs, there are other sources of uncertainty. Pure states might not be $\{0,1\}$-valued, i.e. be inherently probabilistic. This is the case for the quantum system associated to a Hilbert space. When the state space is not a simplex, i.e. when the representation of a state as a convex combination of pure states is not unique, there is uncertainty in how a mixture was prepared, and this uncertainty cannot be removed. For instance, the state that is the center of a square or of a closed unit ball can be prepared in infinitely many ways as a mixture of pure states.

Remark 1. Due to space constraints, and the intent to provide a first introduction to the subject, we will not focus on more mathematically complex issues related to infinite-dimensional convex sets. We only mention that Carathéodory's result for compact convex subsets of $\mathbb{R}^n$ has extensions. The Krein-Milman theorem [1] says that a compact convex subset Ω of $\mathbb{R}^X$, or more generally of any locally convex space, is the closure of the convex hull of its extreme points. The theorem of Bishop and De Leeuw [1] then says that each point in Ω is the barycenter of a probability measure on the extreme points, essentially an integral convex combination of them by some probability measure. The notion of a simplex then extends to the infinite-dimensional setting by requiring this barycenter representation to be unique.

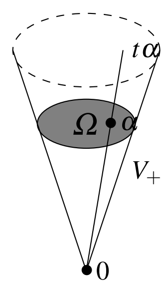
5 PMs and linearity

One we have a state space, which is by definition, a subset of $\mathbb{R}^X$, we have one foot in the subject of linear algebra. We outline how this can be moved forward in a more refined manner by associating to each PM certain kinds of ordered normed vector spaces, and how to in turn construct PMs from such ordered normed vector spaces.

Definition 9. A *cone* of a real vector space V is a non-empty set $C \subseteq V$ that is closed under vector addition, multiplication by positive scalars, and satisfies $C \cap -C = \{0\}$.

Given a cone C there is a partial ordering on V given by $a \leq b$ iff $b - a \in C$. This ordering makes addition and scalar multiplication by positive scalars monotone. Conversely, for an ordering on V satisfying these conditions, the its set of positive elements is a cone, written V_+. Thus, cones correspond to orderings. A cone is called *generating* it spans V, and an ordered space is *archimedean* if $b \in V_+$ and $na \leq b$ for all natural numbers n implies $a \leq 0$.

Definition 10. For an ordered vector space V with non-trivial cone, a convex subset Ω of the cone V_+ is called a *base* for the cone if each $0 \neq \beta$ in the cone can be uniquely expressed as $\beta = t\alpha$ for some $t > 0$ and $\alpha \in K$.



Proposition 2. *If $A = (\mathscr{A}, \Omega)$ is a standard* PM *with V_A the span in $\mathbb{R}^X$ of its state space, then Ω is the base for a generating cone on V_A.*

For an archimedean ordered space V with Ω as the base of a generating cone, let B be the convex hull of $\Omega \cup -\Omega$. For $x \in V$ set $\|x\|_\Omega \ = \ \inf\{t \geq 0 \mid x \in tB\}$. This is always a semi-norm, when it is a norm we call (V, Ω) a *base-normed space* (BNS).

Theorem 1. *If $A = (\mathscr{A}, \Omega)$ is a standard* PM *then (V_A, Ω) is a complete* BNS.

Example 12. For a measurable space (S,Σ), its signed measures form a vector space V and the set Ω of probability measures are the base for a cone on V making it a complete BNS. The base-norm of this space is the variation norm on V.

Example 13. For a separable Hilbert space $\mathscr{H}$, its trace-class operators form a vector space and the density operators are a cone base for this vector space that make it a complete BNS. The base-norm of this space is given by the trace.

There is more to this story, which we will discuss only briefly, as the technical details become involved; see [1, 2, 3]. There is another type of ordered normed space, called an *order unit space* (OUS), that comes in pairs with BNSs: the dual of a BNS is an OUS and conversely. A pair consisting of a BNS and an OUS is in *separating order duality* if each space effectively sits in the dual of the other in a natural way. A primary example is the BNS of trace-class operators on a separable Hilbert space $\mathscr{H}$ and the OUS of bounded self-adjoint operators on $\mathscr{H}$. Here, an element of each space effects a functional on the other in a natural way. Each standard PM $A=(\mathscr{A},\Omega)$ produces such a pair with the BNS created through the states of A, and the OUS through evaluation of the events of A on the states.

6 Classicality

It is natural to wonder whether PMs really take us beyond the scope of classical probability theory, or whether they represent classical situations with constraints or limitations placed on what can be done, (e.g, what states can be prepared, or which classical measurements can be made). This was essentially Einstein's question when asking if quantum theory could be explained by a so-called "hidden variables" theory. In this section, we will see that, subject to certain hypotheses, any probabilistic model can be explained in this way, but generally at a cost in terms of one's overall outlook.

Example 14. Consider the test space of Example 10 with tests $\{a,x\}$ and $\{b,y\}$ and whose state space is all probability weights, which we noted was affinely isomorphic to the unit square. We can view this as representing a system with 4 outcomes: $p=a,b$ occur, $q=a,y$ occur, $r=x,b$ occur, $d=x,y$ occur. However, we have limited possibility to conduct tests — we can test which of the events $\{p,q\},\{r,s\}$ occurs and we can test which of the events $\{p,r\},\{q,s\}$ occur, and that is all.

A PM is *semi-classical* if each event is contained in exactly one test. If we take a test space $A=(\mathscr{A},\Omega)$ and form a new test space by duplicating each outcome once for each test that contains it, we obtain a new test space with tests $\{(x,E)\mid x\in E\}$ for each test E of $\mathscr{A}$. Extending each state α of A to $\alpha'(x,E)=\alpha(x)$ obtain

Proposition 3. *For each* PM *A there is a semi-classical* PM *S and an surjection π from the outcomes of S to those of A mapping onto the tests of A such that states of S are obtained as $\alpha\circ\pi$ for α a state of A.*

A PM A is *locally finite* if each of its tests is finite. This includes the test spaces obtained from unit vectors in a finite-dimensional Hilbert space, the setting of quantum information. The following is essentially due to Cook [5].

Theorem 2. *If S is a locally finite semi-classical* PM *there is a measurable space and an embedding from the outcomes of S to those of the* PM *C of its measurable sets, taking tests to tests, so that every state on S is the restriction of a state of C.*

Thus, if A is locally finite, the semi-classical PM S produced in Proposition 3 is locally finite, hence is embedded into a classical test space C. So A is a quotient of a submodel of C, and hence the statistics of A can be reproduced from those of a classical model. This can be applied to finite-dimensional quantum mechanics!

However, as mentioned above, there is a cost. There may be states of C that are not obtained from ones of A, and that behave in undesirable ways. In particular, there may be a state β of C and an outcome x of A that is contained in two different tests E and F of A and such that $\beta(x,E) \neq \beta(x,F)$. So β provides different probabilities for an outcome of A depending on how we test for it. This is *contextuality*! Of course, from a mathematical viewpoint we can simply "ignore" such states. But if C is being used to "explain" the behavior of A, then simply ignoring problematic features is itself problematic. The issue is further complicated when considering compound systems.

7 Compound systems

A PM reflects a system under study. If two systems are modeled by PMs A and B, we would like a single PM C that reflects the two systems considered as a whole. There may be various ways to produce such a composite system from those of A and B.

Definition 11. Let A and B be PMs with outcome-sets X and Y. Their *product* $A \times B$ has outcome-set $X \times Y$, tests $E \times F$ where E and F are tests of A and B, and its states are *product states* $\alpha \times \beta$ of states α of A and β of B where

$$(\alpha \times \beta)(x,y) \; = \; \alpha(x)\beta(y).$$

If A and B arise from finite probability spaces, then one might guess that $A \times B$ would be the natural model for a composite. However, even for classical probability, this notion is too restrictive.

Example 15. Let A and B be the PMs for the measurable spaces (S,Σ_S) and (T,Σ_T). Their outcome-sets are the measurable subsets of S and T. The composite system is usually formed by taking the measurable space structure on $S \times T$ induced by those on its components, this is the smallest σ-algebra obtaining all products $U \times V$ of measurable sets of S and T. Thus, the PM for this product measurable space contains all products of outcomes of A and B, but contains many more as well.

Definition 12. A composite of PMs A and B is a PM C that contains the outcomes and tests of $A \times B$, and is such that each product state $\alpha \times \beta$ is the restriction to $A \times B$ of a state on C.

The product $A \times B$ of two PMs built from finite probability spaces is a composite as is the model obtained from two measurable spaces in Example 15. The following standard tool from quantum mechanics is another example of a composite.

Example 16. For Hilbert spaces $\mathscr{H}$ and $\mathscr{K}$, their tensor product $\mathscr{H} \otimes \mathscr{K}$ is the metric space completion of their algebraic tensor product. This has the property that for any orthonormal bases (u_i) of $\mathscr{H}$ and (v_j) of $\mathscr{K}$, that $(u_i \otimes v_j)$ is an orthonormal basis of $\mathscr{H} \otimes \mathscr{K}$. Further, for any density operators W and W' on $\mathscr{H}$ and $\mathscr{K}$, their tensor product $W \otimes W'$ is a density operator on the tensor product. This shows that product states extend to the tensor product. So if A, B, C are the PMs constructed from unit vectors of $\mathscr{H}$, $\mathscr{K}$ and $\mathscr{H} \otimes \mathscr{K}$, then C is a composite of A and B.

When dealing with composites of PMs, we encounter two phenomena that are not usually seen in classical probability theory — the lack of local tomography and the possibility of signaling. We discuss the former here, and the latter in the next sub-section.

Definition 13. A composite C of PMs A and B is *locally tomographic* if each state of C is determined by its restriction to the outcomes in $A \times B$.

Clearly the composite of PMs given by finite probability spaces is locally tomographic since it is their product. We consider several other examples.

Example 17. Suppose that C is a composite of PMs A and B that are formed from measurable spaces (S, Σ_S) and (T, Σ_T) as in Example 15. Each probability measure on the product of these measurable spaces is determined by its values on measurable sets of the form $U \times V$ where U, V are measurable subsets of S and T. Thus C is a locally tomographic composite.

Example 18. Suppose A and B are PMs obtained from Hilbert spaces $\mathscr{H}$ and $\mathscr{K}$ and that C is their composite obtained from $\mathscr{H} \otimes \mathscr{K}$ as in Example 16. If these Hilbert spaces are taken in the usual sense of spaces over the scalar field of complex numbers, then the composite is locally tomographic; if they are Hilbert spaces over the field of real numbers, then the composite is not locally tomographic. This is related to the fact that a self-adjoint operator A on a complex Hilbert space is determined by the values of the inner products $\langle Av, v \rangle$ where v ranges over all unit vectors, and that this is not the case for a self-adjoint operator on a real Hilbert space.

8 Signaling and entanglement

We consider states of a compound system in more detail. Assume until further notice that A, B are PMs with outcome-sets X and Y and test-spaces $\mathscr{A}$ and $\mathscr{B}$, and let ω be a probability weight on $\mathscr{A} \times \mathscr{B}$.

Definition 14. For each test E of A, define the *marginal* probability weight $\omega_{2,E}$ on B by setting for all $y \in Y$

$$\omega_{2,E}(y) \;=\; \sum\{\omega(x,y) \mid x \in E\}$$

Given a test F of B, define the marginal $\omega_{1,F}$ on A symmetrically.

In general, marginal probability weights will depend on the choice of a test E of A or F of B. The dependence on the choice of a test poses a problem when we consider a compound system consisting of two systems separated by a great distance and with a large supply of pairs of systems prepared in state ω. The experimenter with the second system can estimate the state $\omega_{2,E}$ through repeated testing, and thereby potentially gain information about the test E performed on the first system. This could constitute transmission of information faster than light.

Definition 15. A probability weight ω on $A \times B$ is *non-signaling* from A to B if the marginal weight $\omega_{2,E}$ is independent of the test E.

The definition of ω being non-signaling from B to A is symmetric, and ω is called *non-signaling* if it is non-signaling in both directions. For a non-signaling When ω is non-signaling, we can define its conditionals in the usual way.

Definition 16. A joint state of $A \times B$ is a non-signaling probability weight whose conditionals and marginals are states of A and B.

We say a state on a composite C of A and B is non-signaling if its restriction to $A \times B$ is a joint state and that C is a *non-signaling composite* if all of its states are non-signaling states.

Example 19. Let A, B be PMs given by measurable spaces S, T and μ be a probability measure on their product. A test E of A is a countable measurable partition U_n of S and an outcome of B is a measurable set $V \subseteq Y$. Then $\sum_n \mu(U_n \times V) = \mu(S \times V)$ is independent of the test E. So μ is a non-signaling state on this composite.

Example 20. Let A, B be PMs associated to Hilbert spaces $\mathscr{H}, \mathscr{K}$ and W be a density operator on their tensor product. A test E of A is an orthonormal basis u_n of $\mathscr{H}$ and an outcome v of B is a unit vector of $\mathscr{K}$. The density operator W gives a state ω where $\omega(u,v) = \mathrm{Tr}(W(P_u \otimes P_v))$ using the trace Tr and projections P_u and P_v. A calculation gives $\sum_n \omega(u_n, v) = \mathrm{Tr}(1 \otimes p_v)$ which is independent of the test E. Thus, density operators on the tensor product yield non-signaling states.

We next consider a different issue with states of composites, one that is also unfamiliar to our experience with classical systems. To simplify our discussion, we restrict attention to finite-dimensional (f.d.) standard PMs, those models A whose states span a finite-dimensional space V_A. These include models from f.d. quantum mechanics since density operators on $\mathbb{C}^n$ span the f.d. space of self-adjoint $n \times n$ matrices.

Definition 17. A joint state of $A \times B$ is *separable* if it is in the closed convex hull of the products states and is *entangled* otherwise.

A convex combination of product states need not be a product state, but physically it can be thought of as representing a statistical mixture of product states. Entangled states can not be represented in this way. Entangled states are ubiquitous in quantum mechanics, and indeed are essential for nearly all quantum information applications.

Example 21. Probability weights on the product of two finite probability spaces are separable. More generally, probability measures on the product of two finite measure spaces are separable.

Example 22. Let A, B be PMs associated to two complex Hilbert spaces and let v be a unit vector in the tensor product of this Hilbert spaces that cannot be expressed as a pure tensor $a \otimes b$. Then the projection onto v is a density operator on the tensor product, so yields a state on the tensor product, hence its restriction to $A \times B$ is a joint state. This state is entangled since it is an extreme point of the state space of the tensor product and is not itself a product state.

We come now to a well-known result with origins in the work of Bell. A “qubit” is treated using a PM A constructed from the Hilbert space $\mathbb{C}^2$. Among its tests are ones $E = \{x, y\}$ and $F = \{x', y'\}$. We take two qubits separated at a distance and perform tests on the pair where one of E, F is performed on the first qubit and one of E, F is performed on the second. We do this for a large number of identically prepared couplets in state ω, and the collect statistics on the results. In particular, we compute the statistic

$$S(\omega) \;=\; C(E,E) + C(E,F) + C(F,E) - C(F,F)$$

where $C(E,E)$ is the correlation $\omega(x,x) + \omega(y,y) - \omega(x,y) - \omega(y,x)$, and so forth. A routine, but tedious, calculation shows that for any product state we have $|S(\omega)| \leq 2$, and hence the same holds for any separable state. However, if we take i, j to be the standard basis of $\mathbb{C}^2$ and form the the entangled state ω given by the normalized form of the vector $i \otimes j - j \otimes i$, a computation yields $S(\omega) = 2\sqrt{2}$.

In the early 1980’s an experiment was performed by A. Aspect in which a stream of identically prepared qubit couplets were produced; the two qubits in each couplet sent rapidly in different directions, and tests E or F randomly performed on each half of a couplet so close to simultaneous that these tests occurred outside of the light cone of the other. The statistics matched those of an entangled state. We note that this does not imply signaling, since as we noted in Example 20, quantum states are non-signaling. Aspect won a Nobel prize in physics, Bell did not.

The situation can be treated formally in terms of composites. While the details of a complete treatment are somewhat technically involved, but we can make the point with an approximate version. We say composite has a *local classical explanation* if A and B have classical explanations by measurable spaces and the product of these measurable spaces provides a classical explanation of the composite.

Theorem 3. *If C is a non-signaling composite with an entangled state, then C does not have a local classical explanation.*

9 Generalized probabilistic theories

A *probabilistic theory*, or simply a *theory*, consists of a collection of probabilistic models as well as a specification of the morphisms between models satisfying natural conditions that amount to it forming a category. Roughly, this means that each member of the theory has an identity morphism and that morphisms are closed under composition provided the codomian of one is the domain of the next. PMs are quite general structures, and the definition of morphisms between them that we provide is also very general, allowing for a wide range of applications. We note that there are still more general definitions of morphisms that are considered for some purposes.

Definition 18. A *morphism* ϕ from a PM $A = (\mathscr{A}, \Omega_A)$ to a PM $B = (\mathscr{B}, \Omega_B)$ is a mapping ϕ from the events of A to the events of B taking tests to tests that preserves orthogonality, perpectivity, and compatible unions.

If $\phi : A \to B$ is a morphism of PMs, then for each probability weight β on B we have $\alpha = \beta \circ \phi$ where $\alpha(x) = \sum\{\beta(y) \mid y \in \phi(\{x\})\}$ is a probability weight on A. It is common to consider morphisms that have $\beta \circ \phi$ a state of A for each state β of B. Clearly the identity map provides a such a PM morphism from a model A to itself. Several other examples are given below.

Example 23. Suppose A and B are PMs associated with finite probability spaces X and Y. Then each function $f : Y \to X$ provides a morphism $\phi : A \to B$ through inverse image $\phi(e) = f^{-1}(e)$ for each event $e \subseteq X$. This extends to the setting where A, B arise from measurable spaces (X, Σ_X) and (Y, Σ_Y) and $f : Y \to X$ is a measurable function. Thus, we have means to treat random variables through morphisms of appropriate models.

Example 24. If A and B are PMs arising from Hilbert spaces $\mathscr{H}$ and $\mathscr{K}$, each unitary operator $U : \mathscr{H} \to \mathscr{K}$ gives rise to a morphism $\phi : A \to B$ taking an orthonormal set in $\mathscr{H}$ to its image in $\mathscr{K}$.

Example 25. By the spectral theorem, self-adjoint operators on a Hilbert space $\mathscr{H}$ correspond to spectral measures on the projection lattice $\mathscr{P}(\mathscr{H})$. These spectral measures are σ-additive mappings $E : \mathrm{Bor}(\mathbb{R}) \to \mathscr{P}(\mathscr{H})$ from the Borel subsets of the reals to $\mathscr{P}(\mathscr{H})$. If A is the PM associated to the measurable space given by the Borel subsets of the reals and B is the PM whose tests are countable pairwise orthogonal families of projections that sum to unity, i.e. decompositions of unity, then E yields a morphism $\phi : A \to B$ of PMs and for each state β of B we have $\beta \circ \phi$ is a Borel probability measure on the reals.

There are many examples of probabilistic theories. One might take the collection of all PMs and the morphisms between them, or the f.d. ones, or the locally finite ones, or those arising from finite probability spaces, or from measurable spaces, or from separable Hilbert spaces. Usually we impose additional structure on the category to reflect the ability to form compound systems. While the following is quite involved to state precisely [12], we give an informal account.

Definition 19. A *symmetric monoidal category* is a category C equipped with a means to produce, from objects A and B, an object $A \otimes B$, in such a way that $\otimes$ is symmetric and associative and has a unit object I satisfying $A \otimes I \simeq A$.

In slightly more detail, $\otimes$ is a bifunctor and symmetry, associativity, and the existence of a unit are all only up to isomorphism; these isomorphisms, moreover, must be natural in a precise technical sense [12].

When forming a probabilistic theory, one often additionally requires that $A \otimes B$ be a specified non-signaling composite. We may assume additional properties such as local tomography. The setting of probabilistic theories considered as symmetric monoidal categories provides a general and flexible setting to speak precisely about probabilistic situations beyond those encountered in classical probability theory.

Acknowledgements The first listed author was partially supported by the National Science Foundation under Grant No. DMS-2231414.

References

1. Erik M. Alfsen. *Compact convex sets and boundary integrals*, volume Band 57 of *Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas]*. Springer-Verlag, New York-Heidelberg, 1971.
2. Erik M. Alfsen and Frederic W. Shultz. *State spaces of Operator Algebras*. Mathematics: Theory & Applications. Birkhäuser Boston, Inc., Boston, MA, 2001. Basic theory, orientations, and C^*-products.
3. Erik M. Alfsen and Frederic W. Shultz. *Geometry of State Spaces of Operator Algebras*. Mathematics: Theory & Applications. Birkhäuser Boston, Inc., Boston, MA, 2003. Basic theory, orientations, and C^*-products.
4. Ingemar Bengtsson and Karol Życzkowski. *Geometry of Quantum States: An Introduction to Quantum Entanglement*. Cambridge University Press, 2006.
5. T. Cook. Banach spaces of weights on quasimanuals. *International Journal of Theoretical Physics*, 2024:1113–1131, 1985.
6. G. M. D'Ariano, G. Chiribella, and P. Perinotti. *Quantum from Principles*. Cambridge University Press, 2017.
7. Anatolij Dvurečenskij. *Gleason's theorem and its applications*, volume 60 of *Mathematics and its Applications (East European Series)*. Kluwer Academic Publishers Group, Dordrecht; Ister Science Press, Bratislava, 1993.
8. D. Foulis, C. Piron, and C. Randall. Realism, operationalism, and quantum mechanics. *Found. Phys.*, 13(8):813–841, 1983.
9. L. Hardy. Quantum theory from five reasonable axioms. 2000.
10. G. Jenca. Blocks of homogeneous effect algebras. *Bulletin of the Australian Mathematical Society*, 64:81–98, 2001.

11. Günther Ludwig. *Foundations of quantum mechanics. I*. Texts and Monographs in Physics. Springer-Verlag, New York, 1983. Translated from the German by Carl A. Hein.
12. S. Mac Lane. *Categories for the Working Mathematician*, volume 5 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1998.
13. George W. Mackey. *The Mathematical Foundations of Quantum Mechanics*. W.A. Benjamin, New York, 1963. A lecture-note volume.
14. Ll. Masanes and M. Müller. A derivation of quantum theory from physical requirements. *New J. Phys.*, 13, 2011.
15. C. H. Randall and D. J. Foulis. The operational approach to quantum mechanics. In C. A. Hooker, editor, *Physical Theory as Logico-Operational Structure*, volume 7 of *University of Western Ontario Series in Philosophy of Science*. D. Reidel Publishing Company, 1979.
16. A. Wilce. Test spaces. In *Handbook of quantum logic and quantum structures—quantum logic*, pages 443–549. Elsevier/North-Holland, Amsterdam, 2009.
17. Alex Wilce. General probabilistic theories: notes for a short course, 2025.