Orthodox Quantum Theory from Contextual Probability Models

Jacob A. Barandes*

April 20, 2026

Abstract

In this paper, I introduce a new kind of mathematical structure, called a contextual probability model, that is fully grounded in conventional definitions in metaphysics and ordinary notions of probability. I then argue that these contextual probability models are broad enough to encompass all quantum systems based on the orthodox Dirac-von Neumann axioms, at least for the case of finite-dimensional Hilbert spaces. I also use contextual probability models as a platform for re-examining various exotic features of quantum theory, including the physical status of the wave function and the necessity of the complex numbers.

1 Introduction

My primary claim in this paper is that all quantum systems based on the 'orthodox' Dirac-von Neumann axioms can be identified as specific kinds of *contextual probabilities models*, which are a new class of mathematical structures based on conventional tools from metaphysics and ordinary notions of probability theory.[1] Other kinds of contextual probability models may then provide alternative interpretations of quantum theory, including those based explicitly on hidden variables. (In particular, the construction to follow might vaguely resemble some versions of the *modal interpretations*.)

A running theme throughout this paper will be the idea of *representation*, and the importance of not confusing matters of representation with matters of substance. I will argue in this paper that various debates in the philosophy of quantum theory, including the physical status of the wave function, turn on this representation-substance dichotomy, and that one can deflate these debates, partially at least, by getting clearer on which things are matters of representation and which things are matters of substance.

*Departments of Philosophy and Physics, Harvard University, Cambridge, MA 02138; jacob_barandes@harvard.edu; ORCID: 0000-0002-3740-4418

[1]Due to the ubiquity of the term 'context' in quantum theory, the term 'contextual probability' has shown up many times in the research literature before. The notion of a contextual probability model in this paper is new.

2 States of Affairs

In metaphysics, a *state of affairs* is an aspect of the world that serves as a truth-maker for propositions. There can be many different *possible* states of affairs, and a state of affairs that obtains is an *actual* state of affairs. States of affairs can change with time. I will not make any other assumptions or assertions about states of affairs, their interpretation, or their ontological significance in this paper. (For a review, see Textor 2025.)

A *possible world* is then, on some views, a maximal state of affairs that does not change with time. The *actual world* is then whichever of these possible worlds obtains. (Again, for a review, see Menzel 2025.)

Let the label i index a collection of N mutually exclusive and exhaustive possible states of affairs, such as the $N = 6$ possible showings on a die. I will assume that N is a finite integer, perhaps on some notion of coarse-graining over observationally indistinguishable possible states of affairs. The core mysteries in the philosophy and foundations of quantum theory arguably do not hinge on whether N is finite or infinite, or on whether one makes appeals to infinities, especially in light of the questionable status of true or literal infinities in physics.

3 Representations of Propositional Logic

As a first example of a representation, one can introduce a representation of our N possible states of affairs by introducing the formalism of *propositional logic*. Let P_i be a proposition that evaluates to 'true' if and only if the specific state of affairs i obtains, and otherwise evaluates to 'false.' Then with the usual meanings of the logical connectives $\neg$ (not/negation), $\wedge$ (and/conjunction), $\vee$ (or/disjunction), letting $\mathbb{T}$ denote the tautology, letting $\mathbb{F}$ denote the anti-tautology, and letting $\iff$ denote logical equivalence, the following formulas hold:

$$P_i \wedge P_i \iff P_i \quad \text{[idempotence]}, \tag{1}$$

$$P_i \wedge P_{j \neq i} \iff \mathbb{F} \quad \text{[mutual exclusivity]}, \tag{2}$$

$$P_1 \vee \cdots \vee P_N \iff \mathbb{T} \quad \text{[completeness]}. \tag{3}$$

Rather than juggle logical connectives and truth tables, a convenient move is to introduce a further layer of representation, this one from mathematics: *Boolean algebra*. To this end, I will introduce the following table of formal replacements:

Propositional Logic	Boolean Algebra
‘true’	1
‘false’	0
P_i (‘true’ or ‘false’)	P_i (1 or 0)
$\mathbb{T}$	$\mathbb{1}$ (constant always equal to 1)
$\mathbb{F}$	$\mathbb{0}$ (constant always equal to 0)
$\wedge$	$\times$ (multiplication)
$\neg$	$1-$ (difference from 1)

These replacements then fix all other logical connectives. It follows that each of the mathematical symbols $P_1, \ldots, P_N$ takes the value 1 or 0, where P_i takes the value 1 if and only if the state of affairs i obtains, and otherwise takes the value 0. The logical formulas (1)–(3) are then collectively equivalent to the following arithmetic equations:

$$P_i^2 = P_i \quad \text{[idempotence]}, \tag{4}$$

$$P_i P_j = \delta_{ij} P_i \quad \text{[mutual exclusivity]}, \tag{5}$$

$$\sum_{i=1}^{N} P_i = \mathbb{1} \quad \text{[completeness/resolution of the identity]}. \tag{6}$$

Here δ_{ij} is the usual Kronecker delta,

$$\delta_{ij} \equiv \begin{cases} 1 & \text{for } i = j, \\ 0 & \text{for } i \neq j, \end{cases} \tag{7}$$

where the symbol $\equiv$ denotes a definition. The idempotence property (4) identifies each of the mathematical symbols P_i as a *projector*, and the conditions of mutual exclusivity (5) and completeness (6) identify the collection $P_1, \ldots, P_N$ as a *projection-valued measure*, or *PVM* (Mackey 1952, 1957).

4 Commutative Algebras

One can extend this Boolean algebra to a more expansive kind of mathematical representation. Let A denote some real-valued numerical variable whose magnitude depends on which of the N possible states of affairs obtains. I will call these real-valued magnitudes the *possible outcomes* of A, and I will denote them as $a_1, \ldots, a_N$, where a_i occurs if the possible state of affairs i obtains. Altogether, these possible outcomes comprise the *spectrum* $\sigma(A)$ of A. For example, A could be the specific number that shows on a die, or the number of points awarded to a player who rolls a specific showing on the die. It is then trivial that A can be expressed as the following *spectral decomposition*, which explicitly displays its spectrum of possible outcomes $a_1, \ldots, a_N$ and their associated projectors $P_1, \ldots, P_N$:

$$A = \sum_{i=1}^{N} a_i P_i = \begin{cases} a_1 & \text{if the state of affairs 1 obtains,} \\ \vdots & \\ a_N & \text{if the state of affairs } N \text{ obtains.} \end{cases} \tag{8}$$

The mutual-exclusivity condition (5) is symmetric between i and j, so it defines a commutative rule for multiplication of these numerical variables:

$$P_i P_j = P_j P_i. \tag{9}$$

As a consequence, if B is any other such numerical variable, then multiplication with A is commutative:

$$AB = BA. \tag{10}$$

Indeed, the product AB will have its own spectral decomposition of the form

$$AB = \sum_{i=1}^{N} a_i b_i P_i, \tag{11}$$

where cross terms $a_i b_{j\neq i}$ are not present due to the mutual exclusivity $P_i P_{j\neq i} = 0$. Together with $\mathbb{1}$, which serves as the multiplicative identity or unity element, and with obvious definitions of addition, $A + B$, and multiplication by scalars, cA, these numerical variables $A, B, \ldots$ therefore constitute a *commutative algebra with unity* over the real numbers.

5 Probability Models

Consider any suitable account of probability—epistemic credence, objective or aleatory chance, evidentiary support, or something else entirely. I will then associate each possible state of affairs i with a numerical, real-valued probability p_i, where these probabilities then satisfy the standard rules:

$$p_i \geq 0 \quad \text{[non-negativity]}, \tag{12}$$

$$\sum_{i=1}^{N} p_i = 1 \quad \text{[completness/normalization]}. \tag{13}$$

It will be convenient to gather these probabilities together into a spectral decomposition that defines a new kind of formal symbol that I will call a *probability variable* ρ:

$$\rho \equiv \sum_{i=1}^{N} p_i P_i. \tag{14}$$

It is easy to check that this probability variable is idempotent if and only if the probability distribution $p_1, \ldots, p_N$ is *pure*, meaning that just one probability p_i is 1 and the rest are 0:

$$\rho^2 = \rho \quad \text{[pure]}. \tag{15}$$

Otherwise one calls the probability distribution *mixed*:

$$\rho^2 \neq \rho \quad \text{[mixed]}. \tag{16}$$

Introducing a linear *trace map* Tr on the commutative algebra according to the formal rules

$$\mathrm{Tr}(P_i) \equiv 1, \tag{17}$$

$$\mathrm{Tr}(cA + dB) \equiv c\,\mathrm{Tr}(A) + d\,\mathrm{Tr}(B), \tag{18}$$

it follows immediately from these two definitional properties, together with the spectral decomposition (14), that a probability variable ρ always has trace equal to 1:

$$\mathrm{Tr}(\rho) = 1. \tag{19}$$

Furthermore, because $P_i\rho = p_i P_i$, it follows that each probability p_i itself has an expression in terms of the trace map Tr, the projector P_i, and the probability variable ρ given by

$$p_i = \mathrm{Tr}(P_i\rho). \tag{20}$$

With the N possible states of affairs now associated with specific probabilities, each numerical variable A now becomes a *random variable*, meaning that its possible outcomes are associated with specific probabilities as well. With the *expectation value* $\langle A \rangle$ of a random variable A defined as the probability-weighted average value of A, it follows from the formula (20) for each probability $p_i = \mathrm{Tr}(P_i\rho)$, together with the spectral decomposition (8) of A and the linearity of the trace map Tr that

$$\langle A \rangle \equiv \sum_{i=1}^{N} a_i p_i = \mathrm{Tr}(A\rho). \tag{21}$$

Notice the formal replacements

$$\sum_{i=1}^{N} \mapsto \mathrm{Tr}, \quad a_i \mapsto A, \quad p_i \mapsto \rho. \tag{22}$$

If all these formulas are beginning to resemble well-known formulas from quantum theory, it is no accident, for reasons that we will see shortly.

Our commutative algebra with unity over the real numbers is now, in particular, a commutative algebra of *random variables* with unity over the real numbers. I will denote this commutative algebra by $\mathcal{A}$.

Letting the *sample space* Ω be a set that collects together the possible values of the label i for our possible states of affairs,

$$\Omega \equiv \{1, \ldots, N\}, \tag{23}$$

letting the *event space* $\mathcal{F}$ denote the power set $\mathcal{P}(\Omega) \equiv 2^{\Omega}$ of Ω,

$$\mathcal{F} \equiv \mathcal{P}(\Omega) \equiv 2^{\Omega}, \tag{24}$$

and letting μ denote a *probability measure* on the event space $\mathcal{F}$,

$$\mu : \mathcal{F} \to [0, 1], \tag{25}$$

$$\mu(\Omega) = 1 \quad \text{[normalization]}, \tag{26}$$

$$\mu(E \cup F) = \mu(E) + \mu(F) - \mu(E \cap F) \quad \text{[additivity]}, \tag{27}$$

$$\mu(\{i\}) \equiv p_i, \tag{28}$$

the three-tuple $(\Omega, \mathcal{F}, \mu)$ defines a (finite) *probability space*. Together with the commutative algebra $\mathcal{A}$ of random variables with unity over the real numbers, I will call the four-tuple

$$\mathcal{M} \equiv (\Omega, \mathcal{F}, \mu, \mathcal{A}) \tag{29}$$

a (finite) *probability model*.

6 Changes with Time

Letting t denote a real-valued, physically distinguished time parameter, our possible states of affairs, together with their associated probabilities, can presumably change with time. To be more precise, suppose that at some earlier time t, we have N possible states of affairs $1, \ldots, N$, with a corresponding probability distribution $\{p_1, \ldots, p_N\}$ and an associated PVM $\{P_1, \ldots, P_N\}$. Then either or both of two things can happen in going from that earlier time t to a later time $t' > t$:

1. The possible states of affairs can change qualitatively and/or in number.

2. The numerical values of the probabilities for the possible states of affairs can change.

As an example of the first kind of change, imagine a die that can change color from one time to another, so that the possible states of affairs at the earlier time all have one color, and the possible states of affairs at the later time all have some other color. As an example of the second kind of change, imagine a dice-rolling machine that rolls a new showing on the die according to some internal algorithm that may be probabilistic, or *stochastic*.

To capture the first kind of change, I will let N' denote the number of possible states of affairs $1, \ldots, N'$ at the later time t', and I will let $P'_1, \ldots, P'_{N'}$ denote the projectors associated with each of the possible states of affairs at the later time t'. We then have a new 'primed' PVM $\left\{P'_1, \ldots, P'_{N'}\right\}$

that then satisfies conditions analogous to (4)–(6). This new PVM underwrites a new commutative algebra $\mathcal{A}'$ of random variables $A', B', \ldots$, each with its own spectral decomposition

$$A' = \sum_{i=1}^{N'} a'_i P'_i = \begin{cases} a'_1 & \text{if the later state of affairs 1 obtains,} \\ \vdots & \\ a'_{N'} & \text{if the later state of affairs } N' \text{ obtains.} \end{cases} \tag{30}$$

The probabilities $p'_1, \ldots, p'_{N'}$ of the possible states of affairs at the later time t' then combine to form a distinguished probability variable ρ' formally defined as the spectral decomposition

$$\rho' \equiv \sum_{i=1}^{N'} p'_i P'_i. \tag{31}$$

With a linear trace map Tr defined analogously to (17)–(18), we can express the probabilities $p'_1, \ldots, p'_{N'}$ just like in (20),

$$p'_i = \mathrm{Tr}\big(P'_i \rho'\big), \tag{32}$$

and we can express expectation values $\langle A' \rangle$ of random variables A' just like in (21),

$$\big\langle A' \big\rangle \equiv \sum_{i=1}^{N'} a'_i p'_i = \mathrm{Tr}\big(A' \rho'\big). \tag{33}$$

To capture the second kind of change, it will be necessary to introduce a specific kind of conditional probability. I will define the *conditional 'transition' probability* for the 'primed' state of affairs i to obtain at the later time t', given that the 'unprimed' state of affairs j obtained at the earlier time t, to be the conditional probability

$$p(i \leftarrow j) \equiv p\big(i, t' | j, t\big) \equiv p\big(P'_i \text{ is true at } t' \,|\, P_j \text{ is true at } t\big). \tag{34}$$

These conditional transition probabilities are inherently *first-order*, in the sense that in addition to their *target argument* "P'_i is true at t'," they involve just a *single* choice of *conditioning argument* "P_j is true at t." Then from the law of total probability, it follows that the standalone probabilities p'_i at the later time t' are related to the standalone probabilities p_j at the earlier time t according to the equation

$$p'_i = \sum_{j=1}^{N} p(i \leftarrow j)\, p_j. \tag{35}$$

More explicitly,

$$p\big(P'_i \text{ is true at } t'\big) = \sum_{j=1}^{N} p\big(P'_i \text{ is true at } t' \,|\, P_j \text{ is true at } t\big)\, p(P_j \text{ is true at } t). \tag{36}$$

The collection of conditional transition probabilities $p(i \leftarrow j)$ naturally form an $N' \times N$ matrix $\mathbf{\Gamma}$ with individual entries Γ_{ij}:

$$\mathbf{\Gamma} = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} & \\ \Gamma_{21} & \ddots & \\ & & \Gamma_{N'N} \end{pmatrix}, \quad \Gamma_{ij} \equiv p(i \leftarrow j). \tag{37}$$

I will call $\mathbf{\Gamma}$ a *transition matrix*. With an $N \times 1$ column matrix $\mathbf{p}$ and an $N' \times 1$ column matrix $\mathbf{p}'$ defined in the obvious ways,

$$\mathbf{p} \equiv \begin{pmatrix} p_1 \\ \vdots \\ p_N \end{pmatrix}, \quad \mathbf{p}' = \begin{pmatrix} p'_1 \\ \vdots \\ p'_{N'} \end{pmatrix}, \tag{38}$$

the law of total probability then takes the form of matrix multiplication,

$$\mathbf{p}' = \mathbf{\Gamma}\mathbf{p}. \tag{39}$$

The transition matrix $\mathbf{\Gamma}$ consists of conditional probabilities, so its entries are all non-negative, and each of its columns sums to 1:

$$\Gamma_{ij} \geq 0 \quad \text{[non-negativity]}, \tag{40}$$

$$\sum_{i=1}^{N'} \Gamma_{ij} = 1 \quad \text{[completeness/normalization]}. \tag{41}$$

These two conditions are the defining properties of a *column-stochastic matrix*, or just a *stochastic matrix* for short.

One can capture both forms of change in one mathematical construction. I will define the *channel* $\mathcal{E}$ by

$$\mathcal{E}(P_j) \equiv \sum_{i=1}^{N'} \Gamma_{ij} P'_i = \sum_{i=1}^{N'} p(i \leftarrow j)\, P'_i, \tag{42}$$

and then extend $\mathcal{E}$ by formal linearity. When it is necessary to be explicit about the earlier time t and the later time t', one can write the channel in more detail as

$$\mathcal{E}_t^{t'}(P_j) \equiv \sum_{i=1}^{N'} p\big(P'_i \text{ is true at } t' \,|\, P_j \text{ is true at } t\big)\, P'_i. \tag{43}$$

Letting ρ be the probability variable at the earlier time t, the probability variable ρ' at the later time t' is then, invoking the law of total probability (35),

$$\mathcal{E}(\rho) = \mathcal{E}\left(\sum_{j=1}^{N} p_j P_j\right) = \sum_{i=1}^{N'} \sum_{j=1}^{N} p(i \leftarrow j) p_j P'_i = \sum_{i=1}^{N'} p'_i P'_i = \rho',$$

so the channel map $\mathcal{E}$ neatly implements the change from the earlier time t to the later time t':

$$\rho' = \mathcal{E}(\rho). \tag{44}$$

7 Contextual Probability Models

The earlier time t corresponds to a probability model $(\Omega, \mathcal{F}, \mu, \mathcal{A})$, and the later time t' corresponds to a different probability model $(\Omega', \mathcal{F}', \mu', \mathcal{A}')$. These two probability models are connected to each other by the channel $\mathcal{E}_t^{t'}$. In general, if a collection of probability models are sewn together in this manner by channels, I will refer to the pairs $(\Omega, \mathcal{A})$ as *contexts*, and I will refer to their random variables as *observables*.

Note that although each context then has its own commutative algebra of random variables, or observables, there is, as yet, no well-defined meaning for arithmetic combinations involving observables from more than one context. In particular, if $A, B \in \mathcal{A}$ and $A', B' \in \mathcal{A}'$ are observables belonging to different contexts, then arithmetic combinations *within* a single context, like

$$A + B = B + A, \qquad AB = BA \tag{45}$$

and

$$A' + B' = B' + A', \qquad A'B' = B'A' \tag{46}$$

are all well-defined and commutative. By contrast, arithmetic combinations *across* contexts, like

$$A + A', \qquad AA' \tag{47}$$

are not well-defined at this point because, the random variables A and A' *qua* maps from sample spaces to the real numbers take different respective sample spaces Ω and Ω' as their domains.

I will define a *contextual probability model* to be a collection of probability models together with all the channels needed to sew them together into a unified 'web' or whole. A contextual probability model is therefore grounded in the standard metaphysics of states of affairs and possible worlds as well as in ordinary notions of probability, but apparently cannot be captured in terms of mathematical constructs currently available in the research literature.

Indeed, a moment's reflection indicates that a contextual probability model is a *category.* The objects of this category are probability models, and the morphisms of this category are channels. A channel that merely permutes the possible states of affairs of a single probability model *qua* object can be regarded as an endomorphism of that object.

8 Matrix Representations

My starting place in arguing for the primary claim of this paper will be introducing another layer of mathematical representation, this one phrased in terms of *matrices.* We have already encountered matrices in (37)–(39). The matrices ahead will be more abstract, and will eventually lead to an overall matrix representation of the entire contextual probability model that we can regard as a *functor*.

A representation that handles all three of the matrix operations of addition, multiplication, and multiplication by scalars is to represent observables $A, B, \ldots$ in the given context as diagonal $N \times N$ matrices $\mathbf{A}, \mathbf{B}, \ldots$, with the spectrum of possible outcomes of each observable laid out along the main diagonal:

$$\mathbf{A} \equiv \begin{pmatrix} a_1 & 0 & \\ 0 & \ddots & \\ & & a_N \end{pmatrix} \equiv \operatorname{diag}(a_1, \ldots, a_N). \tag{48}$$

From its definition (48), it is obvious that we can express $\mathbf{A}$ as a spectral decomposition in parallel with (8),

$$\mathbf{A} = \sum_{i=1}^{N} a_i \mathbf{P}_i, \tag{49}$$

where $\mathbf{P}_1, \ldots, \mathbf{P}_N$ are diagonal projection matrices defined by

$$\mathbf{P}_1 \equiv \operatorname{diag}(1, 0, \ldots, 0), \quad \ldots, \quad \mathbf{P}_N \equiv \operatorname{diag}(0, \ldots, 0, 1). \tag{50}$$

Letting $\mathbf{1}$ denote the $N \times N$ identity matrix, the projection matrices $\mathbf{P}_1, \ldots, \mathbf{P}_N$ satisfy the matrix analogues of the PVM conditions (4)–(6):

$$\mathbf{P}_i^2 = \mathbf{P}_i \quad \text{[idempotence]}, \tag{51}$$

$$\mathbf{P}_i \mathbf{P}_j = \delta_{ij} \mathbf{P}_i \quad \text{[mutual exclusivity]}, \tag{52}$$

$$\sum_{i=1}^{N} \mathbf{P}_i = \mathbf{1} \quad \text{[completeness/resolution of the identity]}. \tag{53}$$

In analogy with (17), each projection matrix $\mathbf{P}_i$ has the value 1 for its *matrix trace* tr:

$$\operatorname{tr}(\mathbf{P}_i) = 1. \tag{54}$$

The spectral decomposition (14) defining the probability variable ρ then has a matrix version

$$\boldsymbol{\rho} \equiv \sum_{i=1}^{N} p_i \mathbf{P}_i, \tag{55}$$

where $\boldsymbol{\rho}$ is called a *density matrix* and automatically has matrix trace equal to 1:

$$\operatorname{tr}(\boldsymbol{\rho}) = 1. \tag{56}$$

The products $\mathbf{P}_i\boldsymbol{\rho}$ and $\mathbf{A}\boldsymbol{\rho}$ are then given respectively by

$$\mathbf{P}_i\boldsymbol{\rho} = \mathrm{diag}(0, \dots, 0, p_i, 0, \dots 0),$$

$$\mathbf{A}\boldsymbol{\rho} = \mathrm{diag}(a_1 p_1, \dots, a_N p_N),$$

so we can express each standalone probability p_i as the matrix trace of the product $\mathbf{P}_i\boldsymbol{\rho}$, just like in (20),

$$p_i = \mathrm{tr}(\mathbf{P}_i\boldsymbol{\rho}), \tag{57}$$

and we can express the expectation value $\langle A \rangle$ as the matrix trace of the product $\mathbf{A}\boldsymbol{\rho}$, just like in (21),

$$\langle A \rangle \equiv \sum_{i=1}^{N} a_i p_i = \mathrm{tr}(\mathbf{A}\boldsymbol{\rho}). \tag{58}$$

Nothing important hinged on this particular choice of *diagonal* matrix representation. *Any* matrix representation for which the PVM projectors $P_1, \dots, P_N$ are represented by projection matrices $\mathbf{P}_1, \dots, \mathbf{P}_N$, diagonal or non-diagonal, with real- or complex-valued entries, that satisfy the basic conditions (51)–(54) would have led to a commutative algebra of matrices representing observables $\mathbf{A}, \mathbf{B}, \dots$ and a density matrix $\boldsymbol{\rho}$ obeying all the same essential rules.

Indeed, there is a continuous infinity of such options. This abundance of options is a crucial resource, because we can use each distinct such matrix representation to represent a different context in the overall contextual probability model. In order to define the contextual probability model, the remaining question is then how to choose a network of channels to sew all the distinct contexts together.

9 Constructing Orthodox Quantum Theory

It is here that actual specialization will be necessary. In order to single out the class of contextual probability models that will encompass orthodox quantum theory, I will impose the following axioms:

1. All the contexts should have the same number N of possible states of affairs.

2. The matrix PVMs should be the maximal set consisting solely of positive semidefinite matrices

$$\mathbf{P}_i \geq 0 \quad \text{[positive semidefinite]}. \tag{59}$$

3. The conditional transition probabilities (34) should be defined to be

$$\Gamma_{ij} \equiv p(i \leftarrow j) \equiv \mathrm{tr}\big(\mathbf{P}'_i \mathbf{P}_j\big). \tag{60}$$

Axiom 1 is needed to ensure that all the matrices are $N \times N$, so that adding or multiplying them together, as in (60), is well-defined.[2] Axiom 2 guarantees that the conditional transition probabilities in Axiom 3 are non-negative. To see why, note that positive semidefiniteness and self-adjointness are equivalent for idempotent matrices, and then use

$$\mathrm{tr}\big(\mathbf{P}'_i\mathbf{P}_j\big) = \mathrm{tr}\Big(\big(\mathbf{P}'_i\mathbf{P}_j\big)^{\dagger}\mathbf{P}'_i\mathbf{P}_j\Big), \tag{61}$$

which, as the square of a *Frobenius norm*

$$\mathrm{tr}(\mathbf{Z}^{\dagger}\mathbf{Z}) = \sum_{ij} |z_{ij}|^2 \geq 0, \tag{62}$$

is manifestly non-negative. It follows from this non-negativity together with the completeness relation (53) for $\mathbf{P}'_i$ that the transition matrix $\mathbf{\Gamma}$ is indeed a column-stochastic matrix, as in (40)–(41).

It is now straightforward to show from the law of total probability that if the context at an earlier time t corresponds to an 'unprimed' matrix PVM $\mathbf{P}_1, \ldots, \mathbf{P}_N$, and the context at a later time $t' > t$ corresponds to a 'primed' matrix PVM $\mathbf{P}'_1, \ldots, \mathbf{P}'_N$, then, with the density matrix $\boldsymbol{\rho}$ at t given by a spectral decomposition $\sum_j p_j \mathbf{P}_j$ as in (55), and with the conditional transition probabilities given by $p(i \leftarrow j) \equiv \mathrm{tr}(\mathbf{P}'_i\mathbf{P}_j)$ as in the axiom (60), each standalone probability p'_i at t' is given by

$$p'_i = \mathrm{tr}\big(\mathbf{P}'_i\boldsymbol{\rho}\big), \tag{63}$$

which is the *Born rule* in its general form.

The commutative algebras belonging to each context now join together, via this matrix representation, to form a noncommutative algebra across the contextual probability model as a whole. The observables that make up this noncommutative algebra, being based on self-adjoint PVMs, are therefore, likewise, self-adjoint,

$$\mathbf{A}^{\dagger} = \mathbf{A} \quad \text{[self-adjoint/Hermitian]}, \tag{64}$$

as are all density matrices $\boldsymbol{\rho}$,

$$\boldsymbol{\rho}^{\dagger} = \boldsymbol{\rho} \quad \text{[self-adjoint/Hermitian]},$$

which, furthermore, due to their spectrum consisting of non-negative probabilities, are positive semidefinite,

$$\boldsymbol{\rho} \geq 0 \quad \text{[positive semidefinite]},$$

and have matrix trace equal to 1, (56). Matrix arithmetic *across* contexts is then commutative

[2]One can relax this condition somewhat by requiring only that there is some maximum number N of states of affairs that any one context can have. Then for any context with fewer than N possible states of affairs, one can formally extend it to N possible states of affairs by including 'null' states of affairs that have zero probability.

under addition but not generically under multiplication:

$$\mathbf{A} + \mathbf{A}' = \mathbf{A}' + \mathbf{A}, \qquad \mathbf{A}\mathbf{A}' \neq \mathbf{A}'\mathbf{A} \text{ (generically)}. \tag{65}$$

Altogether, we therefore arrive at an overall mapping, or *functor*, of the category encompassing our contextual probability model and a corresponding category made up of analogous self-adjoint $N \times N$ matrices.

We are now in a position to answer some very basic questions about quantum theory. Why, for instance, do the matrices that make up observables generically require the complex numbers, but not more general number systems, such as the quaternions?

As to the necessity of the complex numbers, the answer comes from looking at a feature common to *all* quantum systems: noncommutativity. Precisely in virtue of the noncommutativity of the algebra of observables for a quantum system, it is possible to define nonzero, self-adjoint matrices out of commutators of other observables:

$$\mathbf{Z} \equiv i\big[\mathbf{A}, \mathbf{A}'\big] = \mathbf{Z}^\dagger. \tag{66}$$

Without the imaginary unit $i \equiv \sqrt{-1}$, or an algebraic construct essentially isomorphic to i (such as suitable 2×2 matrices), such observables would not be available, eliminating a great many observables that arise, for example, from the infinitesimal generators of continuous symmetries.

As for number systems that are more general than the complex numbers, the loss of multiplicative commutativity (as occurs for the quaternions), the loss of associativity, or the existence of zero divisors would threaten the required arithmetic properties that should hold for observables *within* each individual context, so such number systems are not permissible. The complex numbers are special in being the largest commutative division algebra over the real numbers.

A number of important additional facts now follow from elementary theorems of linear algebra. For any given matrix PVM, each projection matrix $\mathbf{P}_i$, having unit trace, is rank-one, and is therefore factorizable as the outer product of an $N \times 1$ column matrix, or *vector*, $\mathbf{e}_i$ and its adjoint $1 \times N$ row matrix, or *dual vector*, $\mathbf{e}_i^\dagger$:

$$\mathbf{P}_i = \mathbf{e}_i \mathbf{e}_i^\dagger. \tag{67}$$

The vector $\mathbf{e}_i$ is defined only up to an overall complex phase factor:

$$\mathbf{e}_i \sim e^{i\theta} \mathbf{e}_i. \tag{68}$$

The PVM conditions (51)–(54) then translate into the statement that the vectors $\mathbf{e}_1, \dots, \mathbf{e}_N$ constitute an orthonormal basis, meaning that they satisfy

$$\mathbf{e}_i^\dagger \mathbf{e}_i = 1 \quad \text{[unit norm]}, \tag{69}$$

$$\mathbf{e}_i^\dagger \mathbf{e}_j = \delta_{ij} \quad \text{[orthogonality]}, \tag{70}$$

$$\sum_{i=1}^{N} \mathbf{e}_i \mathbf{e}_i^\dagger = \mathbf{1} \quad \text{[completeness/resolution of the identity].} \tag{71}$$

Each such basis vector $\mathbf{e}_i$ is an eigenvector of the self-adjoint matrices $\mathbf{A} = \mathbf{A}^\dagger$ that represent all the observables in the context of the given PVM,

$$\mathbf{A}\mathbf{e}_i = a_i \mathbf{e}_i, \tag{72}$$

and the spectral decompositions (49) for observables and (55) for density matrices respectively become

$$\mathbf{A} = \sum_{i=1}^{N} a_i \mathbf{e}_i \mathbf{e}_i^\dagger, \tag{73}$$

$$\boldsymbol{\rho} = \sum_{i=1}^{N} p_i \mathbf{e}_i \mathbf{e}_i^\dagger. \tag{74}$$

We can now think of matrices like $\mathbf{A}$ and $\boldsymbol{\rho}$ as *linear operators* acting on vectors, which make up an N-dimensional vector space called the system's *Hilbert space.*

If it happens that the present-moment possible states of affairs belong to a specific context, and the probability variable ρ for that context is pure, (15), then the density matrix $\boldsymbol{\rho}$ representing it is equal to just one projector, $\mathbf{P}_i = \mathbf{e}_i \mathbf{e}_i^\dagger$. In this special, contingent case, one can define a distinguished *state vector*,

$$\boldsymbol{\Psi} = \begin{pmatrix} \Psi_1 \\ \vdots \\ \Psi_N \end{pmatrix} \equiv \mathbf{e}_i, \tag{75}$$

which is likewise only defined up to an overall complex phase factor,

$$\boldsymbol{\Psi} \sim e^{i\theta} \boldsymbol{\Psi}. \tag{76}$$

Its individual entries $\Psi_1, \ldots, \Psi_N$ can be understood as the range of a mapping $i \mapsto \Psi(i) \equiv \Psi_i$ from the index set $\{1, \ldots, N\}$ to the complex numbers that is called a *wave function.* The Born rule in its general form $p'_i = \text{tr}(\mathbf{P}'_i \boldsymbol{\rho})$ from (63) then reduces to

$$p'_i = |\mathbf{e}'^\dagger_i \boldsymbol{\Psi}|^2. \tag{77}$$

Even more specifically, if the 'primed' matrix PVM is given by a diagonal representation, then

$$p'_i = |\Psi(i)|^2, \tag{78}$$

which is the Born rule in close to the form originally proposed by Max Born (1926).

Far from introducing wave functions as the *protagonists* of the story of quantum theory, we therefore see that wave functions appear as *derived* pieces of tertiary mathematics many layers deep into various forms of representation: *first* we represented possible states of affairs in terms of propositional logic and probability, *then* we represented propositional logic and probability in terms of probability models and contexts in contextual probability models, *then* we represented contextual probability models in terms of an overall matrix representation consisting of self-adjoint matrices that were noncommutative across contexts, *then* we decomposed the rank-one projection matrix representing a contingently pure probability variable as a self-outer-product of a vector that was uniquely defined only up to an undetermined complex phase factor, and *then*, finally, we identified a wave function with the collective set of entries making up that vector. It is difficult, from this perspective, to see why wave functions, which show up only contingently and multiple levels deep into the formalism, should be regarded as matters of substance, rather than as mere matters of representation. (For further reasons to doubt the physical reality of wave functions, see Barandes 2026.)

It is a simple mathematical fact that any two matrix PVMs $\mathbf{P}_1, \ldots, \mathbf{P}_N$ and $\mathbf{P}'_1, \ldots, \mathbf{P}'_N$ from different contexts are related by an $N \times N$ *unitary matrix* $\mathbf{V}$ defined by

$$\mathbf{V} \equiv \begin{pmatrix} v_{11} & v_{12} & \\ v_{21} & \ddots & \\ & & v_{NN} \end{pmatrix} \equiv \sum_{i=1}^{N} \mathbf{e}_i \mathbf{e}_i'^{\dagger} = \mathbf{V}^{-1\dagger}. \tag{79}$$

Specifically,

$$\mathbf{P}'_i = \mathbf{V}^{\dagger} \mathbf{P}_i \mathbf{V}. \tag{80}$$

It then follows from a simple calculation that for any two contexts, the corresponding transition matrix $\mathbf{\Gamma}$ defined in (60) is *unistochastic* (Horn 1954; Thompson 1989; Nylen, Tam, Uhlig 1993; Bengtsson 2004), meaning that each of its entries Γ_{ij} is the modulus-square of the corresponding entry of a unitary matrix:

$$\Gamma_{ij} = \mathrm{tr}\big(\mathbf{P}'_i \mathbf{P}_j\big) = \mathrm{tr}(\mathbf{V}^{\dagger} \mathbf{P}_i \mathbf{V} \mathbf{P}_j) = |v_{ij}|^2. \tag{81}$$

Unistochastic transition matrices describe stochastic processes that generically feature *indivisibility*, a form of *non-Markovianity* (Milz, Modi 2021; Barandes 2023, 2025). Specifically, if Γ_{ij} describes conditional transition probabilities $p(i \leftarrow j) = p(i, t'|j, t)$ across a finite time interval from t to t', then there will generically fail to be first-order conditional transition probabilities that divide up that time interval into subintervals in the sense of a Markov chain, and there will also generically fail to be *uniquely defined* higher-order conditional probabilities dividing up that time interval into subintervals, either. This indivisibility leads to the appearance of various exotic-seeming phenomena found in quantum systems, including *interference* and *entanglement* (Barandes 2025).

10 The Dirac-von Neumann Axioms

Now we are ready to derive the *Dirac-von (DvN) Neumann axioms* that underwrite *orthodox* or '*textbook*' quantum theory.

We have already established the *first* DvN axiom: the contingent probabilistic information of a given quantum system is represented at each time by a positive-semidefinite, unit-trace density matrix $\boldsymbol{\rho}$, that, if rank-one, is expressible in terms of a unit-norm state vector $\boldsymbol{\Psi}$ that is uniquely defined only up to overall complex phase. All the possible vectors make up an inner-product space over the complex numbers, where this inner-product space is called the system's *Hilbert space*. (Some versions of this first DvN axiom also specify mereological conditions on how subsystems join together to make up composite systems, but I will leave that discussion to future work, due to limitations of space.)

If a quantum system is allowed to evolve from an earlier time t to a later time t' in a manner that keeps the probability distribution unchanged in form, then combining the generalized Born rule $p'_i = \mathrm{tr}(\mathbf{P}'_i\boldsymbol{\rho})$ from (63) with the unitary relationship (80) between the matrix PVMs at the two times t and t', we have

$$p'_i = \mathrm{tr}(\mathbf{V}^\dagger\mathbf{P}_i\mathbf{V}\boldsymbol{\rho}) = \mathrm{tr}(\mathbf{P}_i\mathbf{V}\boldsymbol{\rho}\mathbf{V}^\dagger).$$

With the notation $p_i(t) \equiv p'_i$, $\mathbf{U}(t' \leftarrow t) \equiv \mathbf{V}$, and $\boldsymbol{\rho}(t) \equiv \boldsymbol{\rho}$, together with the additional definition

$$\boldsymbol{\rho}(t') \equiv \mathbf{U}(t' \leftarrow t)\boldsymbol{\rho}(t)\mathbf{U}^\dagger(t' \leftarrow t), \tag{82}$$

the generalized Born rule (63) takes the form

$$p_i(t') = \mathrm{tr}\big(\mathbf{P}_i\boldsymbol{\rho}(t')\big). \tag{83}$$

The rule (82) for the change in the density matrix with time is called *unitary time evolution*, or just *unitarity*, and constitutes the *second* DvN axiom. If the density matrix $\boldsymbol{\rho}(t)$ is pure, so that we can assign the system a state vector $\boldsymbol{\Psi}(t)$, and if the time-evolution operator $\mathbf{U}(t' \leftarrow t)$ is a sufficiently smooth ('strongly continuous') function of its final-time argument t', then we can consider incremental increments $t' = t + dt$ in time, and we end up with a differential equation called the *Schrödinger equation* for the system's state vector $\boldsymbol{\Psi}(t)$.

We have already established the *third* DvN axiom, which is just that each observable A of a quantum system is represented by a self-adjoint matrix $\mathbf{A}$ whose possible outcomes, as found in measurements by an external observer, correspond to the eigenvalue spectrum of $\mathbf{A}$.

We have also established the *fourth* DvN axiom, which just states that if an external observer carries out a measurement that shifts a quantum system's context and reveals the outcome of an observable defined in that context, then the probability for that measurement outcome is given by the Born rule (63).

Now suppose that an external observer carries out a measurement on an observable A' that yields a potentially degenerate eigenvalue λ, corresponding to some projection matrix $\mathbf{P}'$ that could be non-elementary, meaning that it could have rank potentially greater than 1:

$$\mathrm{tr}\big(\mathbf{P}'\big) \geq 1. \tag{84}$$

Then, like the identity matrix $\mathbf{1}$, this projection matrix $\mathbf{P}'$ may not single out a unique context. In principle, a highly disturbing measurement could shift the quantum system into any of a potentially large number of possible new contexts. However, for an *ideal measurement* that is as non-disturbing as possible, the shift in context should be as minimal as possible, in the sense that in the case that $\boldsymbol{\rho}$ already belongs to a context compatible with $\mathbf{P}'$ (such as in the hypothetical limit $\mathbf{P}' \to \mathbf{1}$), there should be no shift in context at all. Letting $\mathbf{P}'_1, \ldots, \mathbf{P}'_N$ be the matrix PVM for whatever context the quantum system ends up in, the quantum system's new density matrix $\boldsymbol{\rho}'$ has the spectral decomposition

$$\boldsymbol{\rho}' = \sum_{i=1}^{N} p'_i \mathbf{P}'_i = \frac{1}{\mathrm{tr}(\mathbf{P}'\boldsymbol{\rho})} \sum_{k=1}^{M} \mathbf{P}'_{i_k} \boldsymbol{\rho} \mathbf{P}'_{i_k}, \tag{85}$$

where for some $i_1, \ldots, i_M$, the non-elementary projection matrix $\mathbf{P}'$ has spectral decomposition

$$\mathbf{P}' = \mathbf{P}'_{i_1} + \cdots + \mathbf{P}'_{i_M}. \tag{86}$$

A matrix PVM $\mathbf{P}'_1, \ldots, \mathbf{P}'_N$ satisfying the necessary ideal-measurement criteria described above is then characterized by the key property that the quantum system's new density matrix (85) is *equivalently* given by the *Lüders projection* (Luders 1950), which depends solely on $\mathbf{P}'$ and $\boldsymbol{\rho}$:

$$\boldsymbol{\rho}' = \frac{\mathbf{P}'\boldsymbol{\rho}\mathbf{P}'}{\mathrm{tr}(\mathbf{P}'\boldsymbol{\rho})}. \tag{87}$$

The stochastic 'collapse' to this new density matrix, which occurs with overall measurement-outcome probability $\mathrm{tr}(\mathbf{P}'\boldsymbol{\rho}) = \sum_k \mathrm{tr}\big(\mathbf{P}'_{i_k}\boldsymbol{\rho}\big)$, constitutes the *fifth* and *final* DvN axiom.

11 Conclusions and Future Work

What we learn from the construction in this paper is that a quantum system can be regarded as a particular kind of mathematical structure, called a contextual probability model, that is based on ordinary notions of probability, albeit after giving up some 'classical' assumptions, like divisibility and Markovianity of transition probabilities.

The contextual probability models presented in this paper, and the quantum systems constructed from them, were limited to finite numbers $N < \infty$ of possible states of affairs. Extending the construction here to the countably infinite case would be a reasonably straightforward exercise, leading to the separable Hilbert spaces commonly used in formulating most kinds of quantum systems, as well as to abstractions like noncommutative C*-algebras. Working with non-separable Hilbert spaces, as well as Type II and Type III von Neumann algebras, would present a more challenging future goal.

From the standpoint of the construction laid out in this paper, the *indivisible formulation* of quantum theory (Barandes 2025) can be understood as picking out contextual probability models in which all the contexts share a *fixed, common* sample space Ω that contains not only the quantum systems being studied, but also all measuring devices, external observers, and environments, treated as physical parts of the overall system. That approach claims to resolve the measurement problem by unifying the time evolution of quantum theory into a single, unistochastic dynamical law.

Beyond quantum theory, contextual probability models may have other interesting uses in statistical modeling. It would be well worth studying and developing such applications in future work.

References

[1] J. A. Barandes. "The Stochastic-Quantum Theorem", 2023. URL: https://arxiv.org/abs/2309.03085;https://doi.org/10.48550/arXiv.2309.03085, arXiv:2309.03085, doi:10.48550/arXiv.2309.03085.

[2] J. A. Barandes. "The Stochastic-Quantum Correspondence". *Philosophy of Physics*, 3(1):8, June 2025. arXiv:2302.10778, doi:10.31389/pop.186.

[3] J. A. Barandes. "Pilot-Wave Theories as Hidden Markov Models". 2026. arXiv:2602.10569, doi:10.48550/arXiv.2602.10569.

[4] I. Bengtsson. "The Importance of Being Unistochastic", March 2004. URL: https://arxiv.org/abs/quant-ph/0403088, arXiv:quant-ph/0403088v1.

[5] M. Born. "Zur Quantenmechanik der Stoßvorgänge ('On the Quantum Mechanics of Collision Processes')". *Zeitschrift für Physik*, 37(12):863–867, 1926. doi:10.1007/BF01397477.

[6] A. Horn. "Doubly Stochastic Matrices and the Diagonal of a Rotation Matrix". *American Journal of Mathematics*, 76(3):620–630, 1954. doi:10.2307/2372705.

[7] G. Lüders. "Über die Zustandsänderung durch den Meßprozeß". *Annalen der Physik*, 443(5-8):322–328, 1950. doi:10.1002/andp.19504430510.

[8] G. W. Mackey. "Induced Representations of Locally Compact Groups I". *Annals of Mathematics*, 55(1):101–139, 1952. URL: https://www.jstor.org/stable/1969423, doi:10.2307/1969423.

[9] G. W. Mackey. "Quantum Mechanics and Hilbert Space". *American Mathematical Monthly*, 64(8):45–57, 1957. URL: https://www.jstor.org/stable/2308516, doi:10.2307/2308516.

[10] C. Menzel. "Possible Worlds". In E. N. Zalta and U. Nodelman, editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2025 edition, 2025. URL: https://plato.stanford.edu/archives/fall2025/entries/possible-worlds.

[11] S. Milz and K. Modi. "Quantum Stochastic Processes and Quantum Non-Markovian Phenomena". *PRX Quantum*, 2:030201, May 2021. arXiv:2012.01894v2, doi:10.1103/PRXQuantum.2.030201.

[12] P. Nylen, T.-Y. Tam, and F. Uhlig. "On the Eigenvalues of Principal Submatrices of Normal, Hermitian and Symmetric Matrices". *Linear and Multilinear Algebra*, 36(1):69–78, 1993. doi:10.1080/03081089308818276.

[13] M. Textor. "States of Affairs". In E. N. Zalta and U. Nodelman, editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2025 edition, 2025. URL: https://plato.stanford.edu/archives/win2025/entries/states-of-affairs.

[14] R. C. Thompson. "Lecture Notes from a Johns Hopkins University Lecture Series". Unpublished lecture notes, 1989.