Generalized Probabilistic Theories in a New Light

Raed Shaiia (raedshaiia@gmail.com)
The Center for Advanced Sciences (CAS)

Article

Keywords: generalized-probabilistic theories, operational-probabilistic theories, quantum foundations, quantum information, quantum computing, measurement problem

Posted Date: June 21st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-636142/v1

Generalized Probabilistic Theories in a New Light

Raed Shaiia

June 17, 2021

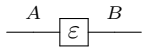
Abstract

In this paper we will present a modified formulation of generalized probabilistic theories that will always give rise to the structure of Hilbert space of quantum mechanics, in any finite outcome space, and give the guidelines to how to extend this work to infinite dimensional Hilbert spaces. Moreover, this new formulation which we will call extended operational-probabilistic theories, applies not only to quantum systems, but also equally well to classical systems, without violating Bell's theorem, and at the same time solves the measurement problem. This is why we will see that the question of why our universe is quantum mechanical rather than classical is misplaced. The only difference that exists between a classical universe and a quantum mechanical one lies merely in which observables are compatible and which are not. Besides, this extended probability theory which we present in this paper shows that it is non-determinacy, or to be more precise, the non-deterministic description of the universe, that makes the laws of physics the way they are. In addition, this paper shows us that what used to be considered as purely classical systems and to be treated that way are in fact able to be manipulated according to the rules of quantum mechanics –with this new understanding of these rules- and that there is still a possibility that there might be a deterministic level from which our universe emerges, which if understood correctly, may open the door wide to applications in areas such as quantum computing. In addition to all that, this paper shows that without the use of complex vector spaces, we cannot have any kind of continuous evolution of the states of any system.

Key words: generalized-probabilistic theories, operational-probabilistic theories, quantum foundations, quantum information, quantum computing, measurement problem.

1 Definitions

-A collection of sets and their subsets determine what we call a system. And the particular type of theory is what specifies which are the systems defined in it.
For example, an elementary particle is described by a set of the possible values of its energy, a set of the possible values of its momentum, ... etc.
A coin is represented by a set of the possible outcomes of the experiment of throwing it, a singleton set that contains the value of its mass, ... etc.
-Each collection of all the sets of some system that have the same cardinality defines what we call a state of the system.
-An event is an arrow connecting two sets of the system.
An event of type $A \to B$ represents a possible transformation in the description of the system from a description in terms of A to a description in terms of B.

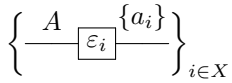


(1)

The set to the left of the diagram is called *the input set of the event*, while the set to the right is called *the output set of the event*.
For example, before throwing a coin, the sample space of the experiment done on the coin is $A = \{H, T\}$. But after we throw it and let us say it stabilizes on H, the sample space of the experiment of reading the result will be $B = \{H\}$. In this case, the event of type $A \to B$ represents a possible transformation of the description of the coin which becomes a fact if we really throw the coin and it stabilizes on H.
-An experiment on the system which measures one of its sets, let us say A, is a collection of events that have A as their input set, and the union of their distinct output sets is A, such that each output set is a singleton of A.
For example, the collection of two events of types $\{H, T\} \to \{H\}$ and $\{H, T\} \to \{T\}$ represents an experiment to measure $\{H, T\}$ if we were to throw the coin.
We notice that in this sense, we are merely talking about a potential experiment up until now.
We can write the set A as:

$$A := \{a_i | i \in X\} \tag{2}$$

Where X is an index set (we will call it from now on the outcome space) that indexes the elements of A, meaning $|X| = |A|$. thus, we can represent the experiment as:



(3)

We say that each outcome space represents a property of the system, and each value of the index is a value of that property.
-When two events belong to the same experiment we say that they are coexisting.
- The sets that are the Cartesian product of the sets of two systems, together with their events composed in parallel (what we mean by that will be defined below), assign what we call a composite system of the two systems.
For example, in the case of two coins, the set $\{H, T\} \times \{H, T\} = \{(H, H), (H, T), (T, H), (T, T)\}$ represents one of the sets that describe the composite system of the two coins.
- When the input and the output of an event represent a composite system, we draw boxes with multiple wires. For example, the box:

(4)

represents an event of type $(A \times C \to B \times D)$

1.1 Composition of events

Events can be connected into networks through the following operations.

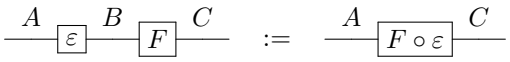
1.1.1 Sequential composition

An event of type $A \to B$ can be connected with an event of type $B \to C$, yielding an event of type $A \to C$.

1.1.2 Parallel composition

An event of type $A \to A'$ can be composed with an event of type $B \to B'$, yielding an event of type $(A \times B \to A' \times B')$

The sequential composition of two events ε and F of matching types is denoted by $F \circ \varepsilon$ and is represented graphically as:

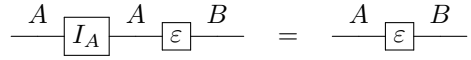


(5)

This graphical notation is justified by the requirement that sequential composition be associative, namely

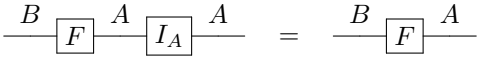
$$G \circ (F \circ \varepsilon) = (G \circ F) \circ \varepsilon \tag{6}$$

for arbitrary events ε, F and G. In addition to associativity, sequential composition is required to have an identity element for every set. The identity on set A, denoted by I_A, is the special event of type $A \to A$ identified by the conditions



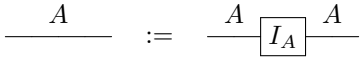
(7)

and



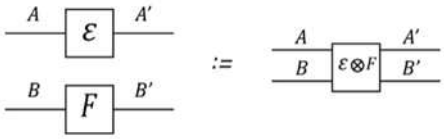
(8)

required to be valid for arbitrary sets A, B and arbitrary events ε and F of types $A \to B$ and $B \to A$, respectively. Consistently, we use the graphical notation



(9)

Mathematically, conditions (6),(7), and (8) impose that the events form a category [4], in which the sets are the objects, and the events are the arrows.
Let us consider parallel composition. The parallel composition of two events ε and F is denoted as $\varepsilon \otimes F$ and is represented graphically as



(10)

The graphical notation is justified by the requirement of the following condition

$$(\varepsilon \otimes F) \circ (G \otimes H) = (\varepsilon \circ G) \otimes (F \circ H) \tag{11}$$

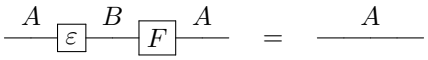
where ε, F, G, and H are arbitrary events.
In addition to equations (7) and (8), parallel composition is required to satisfy the condition

$$I_{A \times B} = I_A \otimes I_B \tag{12}$$

We denote such category by Transf.

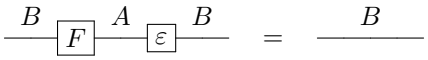
1.2 Reversible events

An event ε of type $A \to B$ is reversible iff there exists another event F, of type $B \to A$, such that



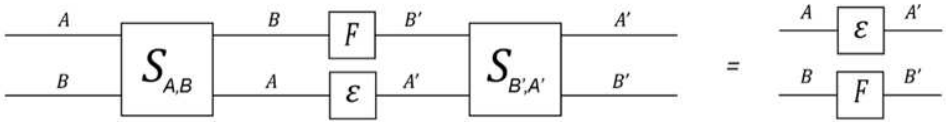
(13)

and



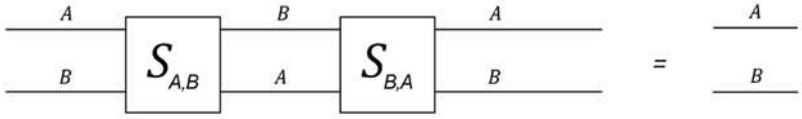
(14)

When this is the case, we write $F = \varepsilon^{-1}$ and we say that sets A and B are operationally equivalent (or simply equivalent).
We denote by RevTransf $(A \to B)$ the set of reversible events of type $A \to B$. Such set (which may be empty) depends on the specific theory. In general, we require the existence of a reversible event that swaps pairs of sets. Given two sets A and B, the swap of A with B — denoted by $S_{A,B}$ —is a reversible event of type $(A \times B) \to (B \times A)$ satisfying the condition



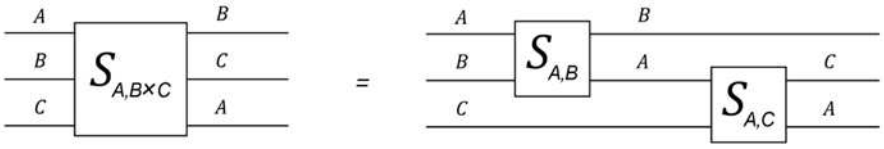
(15)

for arbitrary sets A, B, A', B' and arbitrary events ε, F, as well as the conditions



(16)

and



(17)

1.3 Summary about the operational structure

Summarizing the ideas introduced so far, an operational structure consists of a triple

$$\text{Op} = (\text{Transf, Outcomes, Experiments})$$

where Transf is a category, Outcomes is a collection of sets (the outcome spaces) closed under Cartesian product, and Experiments are sets of events as defined above.

2 Probabilistic structure

In order to make predictions on the outcomes of the experiment, we need a rule assigning a probability to the events of such experiment. The rule is provided by the probabilistic

structure of the theory, which we will give in a different manner than introduced in [1].
Definition 1 (Probabilistic structure). Let Op be an operational structure. We say that Op is provided with a probabilistic structure iff we can define on it for every experiment ε that has an outcome space X , a map

$$P : \rho(X) \to [0, 1], \tag{18}$$

With $\rho(X)$ being the power set of X, and this map satisfies the following two requirements:

1.
$$P(X) = 1, \tag{19}$$

2. For any number of disjoint sets $A_1, A_2, A_3, \ldots$ that are subsets of X, we have:

$$P(U_i A_i) = \sum_i P(A_i) \tag{20}$$

whether this number is finite or infinite (in an infinite outcome space). The map P needs not be surjective: for example, in a deterministic theory the range of P contains only the values 0 and 1.
From now on, we will use the following notation:

$$P(i) := P(\{i\}) \tag{21}$$

For any $i \in X$.
Definition 2. An extended operational-probabilistic theory Θ is a pair (Op, P) consisting of an operational structure Op and of a probabilistic structure for Op as given above.

2.1 Finite outcome spaces

2.1.1 Experiments that have input sets with the same cardinality

In this section, we will take the cardinality of all outcome spaces of the experiments to be equal to the same positive integer N.

2.1.1.1 Representing experiments by vectors From now on we will use Dirac notation for general vector spaces although these vector spaces need not be Hilbert spaces, unless we prove that.
Let us assume we have some experiment $\boldsymbol{\varepsilon}$ with an outcome space X:

$$\boldsymbol{\varepsilon} = \left\{ A \xrightarrow{\boxed{\varepsilon_i}} \{a_i\} \right\}_{i \in X} \tag{22}$$

Where $|X| = N$.
Since P is a probability function on X, then for any $B \subseteq X$ we have[5][6].

$$P(B) = \sum_{i \in B} P(i), \qquad 0 \leq P(i) \leq 1 \tag{23}$$

Particularly:

$$P(X) = \sum_{i \in X} P(i) = 1 \tag{24}$$

On the other hand, in any inner-product vector space with a dimensionality N, where the inner product is positive-definite, and let us call this space V, we can always build an orthonormal basis for it using Gram-Schmidt theorem [7]. Let this basis be $\{|u_i\rangle\}_{i=1}^{N}$ then, for any vector $|C\rangle$ with a less than one square magnitude:

$$|C\rangle = \sum_{i=1}^{N} c_i \, |u_i\rangle \tag{25}$$

And:

$$\langle C|C\rangle = \sum_{i=1}^{N} |c_i|^2, \qquad 0 \leq |c_i|^2 \leq 1 \tag{26}$$

In particular, for a normalized vector $|D\rangle$ we have:

$$\langle D|D\rangle = \sum_{i=1}^{N} |c_i|^2 = 1 \tag{27}$$

Comparing equations (23), (24), (26), and (27) we see that probability has the same behavior as the square magnitude of a vector in V. This suggests another way for representing experiments, meaning, we can associate each output set of the experiment with a vector in such a vector space V, such that the squared-norm of the vector is equal to the probability of the output set. And to distinguish different experiments from eachother, we will represent different experiments with different sets of vectors. As we said, we want each output set to be represented by a vector, this is why we need a set of N different vectors to represent each experiment. We also want to represent different experiments by different sets of vectors. We can do that if we choose the dimensionality of the vector space to be equal to the cardinality of the input sets, meaning N, and represent each experiment by an ordered basis in V, such as $\{|u_i\rangle\}_{i=1}^{N}$ in a way that each output set is represented by a vector along one of the basis vectors with a squared-norm equal to its probability, with no two different experiments represented by the same basis.
Since we need the inner product to be positive-definite, that gives us two natural choices for the vector space V: either to choose it as a real vector space or as a complex vector space. For the moment, and for the sake of generality, we will choose it to be a complex vector space, then we will see later whether we can relieve this condition.
We will choose all the bases to be orthonormal, meaning, for any basis $\{|u_i\rangle\}_{i=1}^{N}$ we have:

$$\langle u_i|u_j\rangle = \delta_{ij} \tag{28}$$

Now we can define the function:

$$f : X \to V : i \mapsto c_i \, |u_i\rangle : |c_i|^2 = P(i), i \in X \tag{29}$$

Actually, it is evident from the last definition that f is not unique. In fact, there is an infinite number of functions satisfying the previous definition in a complex vector space [6]. Building on that, we can now define another function as follows:

$$g : \rho(X) \to V : Y \mapsto |Y\rangle = \sum_{i \in Y} f(i) = \sum_{i \in Y} c_i \, |u_i\rangle \tag{30}$$

Where $\rho(X)$ is the power set of X.
We immediately see that $g(\emptyset)$ is an empty sum, which means that the empty set is represented by the zero vector:

$$|\emptyset\rangle = 0 \tag{31}$$

While X itself will be represented by a normalized vector, and that is due to the fact that:

$$|X\rangle = \sum_{i \in X} f(i) = \sum_{i \in X} c_i \, |u_i\rangle \tag{32}$$

Thus:

$$\langle X|X\rangle = \sum_{i \in X} |c_i|^2 = \sum_{i \in X} P(i) = 1 \tag{33}$$

Where we can write the last equality due to the requirements of the probabilistic structure above.
In fact, as we see in [6], we can deduce from these definitions some very useful mathematical results. We will list some of them here without proofs, because their proofs can be found in [6].
We can see in [6] that for any two elements of $\rho(X)$, for example A and B, we have:

$$P(A) = \langle A|A\rangle = \langle A|X\rangle = \langle X|A\rangle \tag{34}$$

$$P(A \cap B) = \langle A \cap B|A\rangle = \langle A|A \cap B\rangle = \langle A \cap B|B\rangle = \langle B|A \cap B\rangle = \langle A|B\rangle = \langle B|A\rangle \tag{35}$$

$$A \cap B = \phi \Rightarrow |A\rangle \perp |B\rangle \tag{36}$$

$$|A\rangle = |A \backslash B\rangle + |A \cap B\rangle \tag{37}$$

$$|A \cup B\rangle = |A\rangle + |B\rangle - |A \cap B\rangle \tag{38}$$

And if $|A\rangle = \sum_{r=1}^{m} a_r \, |u_r\rangle$ and $|B\rangle = \sum_{s=1}^{l} b_s \, |u_s\rangle$, then

$$|A \cap B\rangle = \sum_{r=1}^{m} \langle B|u_r\rangle \, |u_r\rangle = \sum_{s=1}^{l} \langle A|u_s\rangle \, |u_s\rangle \tag{39}$$

And if we define $c(A) = X \backslash A$, then

$$|c(A)\rangle = |X\rangle - |A\rangle \tag{40}$$

Since any finite dimensional inner-product vector space is a Hilbert space [8], we see immediately that this formulation of probability theory inevitably give rise to Hilbert space structure.

2.1.1.2 Observables We will call any function from the events (labelled by outcomes) of an experiment to the real numbers an observable associated with that experiment, and we call the range of this function the *spectrum of this observable*. More specifically, if we have some experiment $\boldsymbol{\varepsilon} = \{\varepsilon_i\}_{i \in X}$, then an *observable associated with this experiment* is any function of the form:

$$A : \boldsymbol{\varepsilon} = \{\varepsilon_i\}_{i \in X} \to \mathbb{R} : \varepsilon_i \mapsto a_i \tag{41}$$

It is obvious that there is an infinite number of such functions [6].
And we have chosen the domain to be the experiment itself rather than the outcome space

because what distinguishes an experiment is its events, not the outcome space.
We call any two observables that are associated with the same experiment compatible, otherwise we call them incompatible.
If A was injective, we say that its spectrum is non-degenerate. Otherwise we say that it is degenerate.
We see that we can always build the following matrix out of the values of the spectrum of A:

$$diag(a_1, ..., a_N) \tag{42}$$

Which is a Hermitian matrix. Moreover, this matrix is the matrix representing the operator $\hat{A}$ in the vector space V which satisfies:

$$\hat{A} |u_i\rangle = a_i |u_i\rangle \tag{43}$$

Since the matrix representing $\hat{A}$ is Hermitian, then $\hat{A}$ is a Hermitian operator, which as evident from the previous equation has the eigen-vectors $\{|u_i\rangle\}_{i=1}^N$ that correspond to the eigen-values $\{a_i\}_{i=1}^N$. So we can always associate any observable associated to an experiment with a Hermitian operator which spectrum is the same as that of the observable, and its eigen-vectors are the basis vectors that represent the experiment. This is why from now on when we say observable, we mean the Hermitian operator, unless otherwise stated. We notice that for this experiment we are talking about, all compatible observables share this set of basis vectors$\{|u_i\rangle\}_{i=1}^N$ that means they all commute, meaning, for any two compatible observables $\hat{A}$ and $\hat{B}$ we have[7]:

$$[\hat{A}, \hat{B}] = 0 \tag{44}$$

If we take any one of these observables, let it be $\hat{A}$, then we can think of the experiment as giving us one eigen-value of this observable. And we will say that this observable is *measured using this experiment*.
Since this is true for every one of the observables compatible with $\hat{A}$ as we saw, then we will say that *compatible observables can be measured together with a single experiment*.
Since observables associated with different experiments will not have the same sets of eigen-vectors, this means that the Hermitian operators representing them do not commute, and we say that *these observables cannot be measured simultaneously with the same experiment*.
Since we have chosen the experiment arbitrarily, the previous results hold for any experiment.
Now let us take a different experiment with an outcome space that has the same cardinality. We have said that we will represent it with a different basis. What is the relation between the bases representing different experiments?
Since all outcome spaces are represented by normalized vectors, which are linear combinations of the unit vectors representing their respective experiments, we will choose to represent all the outcome spaces by the same normalized vector, which has different components on different bases in V representing different experiments, such that it gives right probabilities according to (29), and the reason we can do that is due to the fact that we can give a vector with a fixed length any components we like by choosing the right basis in a complex vector space. The justification of the previous statement is the following lemma.
<u>*Lemma*</u>
In a finite dimensional inner-product complex vector space with dimensionality greater than or equals to two, we can always find an orthonormal basis such that we can give a vector with a fixed magnitude, any set of components.

Proof
First of all, from Gram-Schmidt theorem we know that in any inner-product vector space, we can find orthonormal bases [7].
Now, let V be a finite inner-product space with a dimensionality N.
Suppose that the vector $|X\rangle$ has the set of components $\{c_1, ..., c_N\}$ on some orthonormal basis $\{|u_i\rangle\}_{i=1}^N$, And we want to find another orthonormal basis$\{|t_j\rangle\}_{j=1}^N$ in the space such that the vector $|X\rangle$ has the components $\{b_1, ..., b_N\}$ on this new basis.
let us assume the new basis vectors are given in terms of the old ones by

$$|t_j\rangle = \sum_{i=1}^{N} f_{ji} |u_i\rangle \tag{45}$$

Since the set $\{f_{ji}\}$has N^2 elements for all the possible values of i and j, and by noticing that each f_{ji} is a complex number in general, thus it has two real numbers, this makes the number of unknowns that we need to find is $2N^2$.
Our task is then to find these unknowns. Let us count the number of equations we have.
First we have the following $2N$ equations (in the real unknowns):

$$|X\rangle = \sum_{i=1}^{N} c_i |u_i\rangle = \sum_{j=1}^{N} b_j |t_j\rangle \tag{46}$$

And the new basis should satisfy the following condition:

$$\langle t_i|t_j\rangle = \delta_{ij} \tag{47}$$

But since

$$\langle t_i|t_j\rangle = \langle t_j|t_i\rangle^* \tag{48}$$

Then the number of independent equations dictated by the inner product is

$$\frac{N(N+1)}{2} \tag{49}$$

So the total number of equations is

$$\frac{N(N+1)}{2} + 2N = \frac{N(N+5)}{2} \tag{50}$$

So to be able to find such a basis as the one we are looking for, the number of unknowns must be greater than or equals to the number of equations, meaning

$$2N^2 \geq \frac{N(N+5)}{2} \Rightarrow N(N - \frac{5}{3}) \geq 0 \tag{51}$$

Which is satisfied for any dimensionality of a vector space that is greater than or equals to two.
Notice that had we used a finite dimensional real vector space, the number of unknowns would have been N^2 while the number of equations would have been $\frac{N(N+3)}{2}$. Hence, the previous lemma would have been only valid if $N \geq 3$. Thus, we see the need to work in a complex vector space to make our formulation a general one.

Corollary (1)
Since we deal with probabilistic experiments with a sample space which has two or more elements, the number of variables will always exceed the number of equations. Thus, we will have an infinite number of distinct orthonormal bases in which the vector representing sample space will have the right sets of components in all of them. This makes us able to represent even different experiments that have identical probability distributions, with different bases.
Corollary (2)
If we apply the previous lemma to the vectors in ordinary space, we see that in order for the number of the unknowns to equal the number of equations in a finite dimensional space, the number of spatial dimensions must be 3 (or zero but this case is trivial). If $N = 3$, then there will be a unique coordinate system in which a certain point can have a specific set of coordinates.
This might be one of the reasons why three special dimensions are special!
From now on, we will call the normalized vector that represents all outcome spaces with the same cardinality *a state vector*, and we must stress that the same state vector represents all experiments with equal cardinality, this is why we will be saying *the state vector representing these experiments*.
And from now on, we say that all experiments that have outcome spaces with equal cardinality N, form a class C_N. And what we will call the state vector representing the class C_N is the state vector representing experiments that their outcome spaces have the same cardinality N.

2.1.1.3 Is the state vector unique? We will replicate a derivation that is done in [6] just to make this paper self-contained. Is the state vector unique? In other words, can we use for a given class of experiments with the same cardinality, more than one vector as a state vector?
If it is not unique, then we must find the same probability distributions for all experiments of this class, whether we used $|X\rangle$ or –if exist- the other vector/vectors that can be used as state vectors.
Again, we will assume that the cardinality of outcome spaces of the experiments we are talking about is N.
Suppose that a state vector representing these experiments is $|X\rangle$.
We will take an experiment $\boldsymbol{\varepsilon_1}$ and assume that the state vector is written using the basis representing the previous experiment according to (32) as:

$$|X\rangle = \sum_{l=1}^{N} c_l \, |u_l\rangle \tag{52}$$

Now Let us take the vector $|X'\rangle$ which is:

$$|X'\rangle = \sum_{l=1}^{N} c'_l \, |u_l\rangle : c'_l = c_l z_l \tag{53}$$

Where z_l are complex numbers which we will write in the form:

$$z_l = A_l e^{i\theta_l} : A_l \in [0, \infty), \theta_l \in \mathbb{R} \tag{54}$$

And $i = \sqrt{-1}$
For $|X'\rangle$ to be a state vector, all the probabilities of each single outcome in the outcome space (thus all the probabilities of sets in the power set of the outcome space since the probability of such a set is equal to the sum of the probabilities of the outcomes that constitute it) for any experiment from this class must be the same as given by$|X\rangle$. So, the probabilities of the outcomes of $\boldsymbol{\varepsilon_1}$ do not change.
So the following equation must hold:

$$|c_l'|^2 = |c_l|^2 \tag{55}$$

Thus:

$$\begin{aligned} &|A_l e^{i\theta_l} c_l|^2 = |c_l|^2 \Rightarrow \\ &|A_l|^2 |c_l|^2 = |c_l|^2 \end{aligned} \tag{56}$$

And because the former condition is true even if we choose the experiment to satisfy: $c_l \neq 0$ for all values of c_l, because our choice of $\boldsymbol{\varepsilon_1}$ is arbitrary, we must have:

$$|A_l|^2 = 1 \tag{57}$$

So we have the condition:

$$A_l = 1 \tag{58}$$

Which means that:

$$z_l = e^{i\theta_l} \tag{59}$$

And that:

$$c_l' = c_l z_l = c_l e^{i\theta_l} \tag{60}$$

But that is not enough, because the condition that probabilities must not change must be true for any other experiment of the same class and not just $\boldsymbol{\varepsilon_1}$, because we are talking about state vectors here.
Let us take a different experiment $\boldsymbol{\varepsilon_2}$ of the same class. We know that it must be represented by another basis, let us say $\{|t_j\rangle\}_{j=1}^N$. We must have:

$$|X\rangle = \sum_{j=1}^{N} b_j |t_j\rangle \tag{61}$$

We now have:

$$|X\rangle = \sum_{j=1}^{N} b_j |t_j\rangle = \sum_{l=1}^{N} c_l |u_l\rangle \tag{62}$$

Where:

$$b_j = \langle t_j|X\rangle = \sum_{l=1}^{N} \langle t_j|u_l\rangle \langle u_l|X\rangle = \sum_{l=1}^{N} \langle t_j|u_l\rangle c_l \tag{63}$$

For $|X'\rangle$we must have:

$$|X'\rangle = \sum_{j=1}^{N} b_j' |t_j\rangle \tag{64}$$

So:

$$b'_j = \langle t_j|X'\rangle = \sum_{l=1}^{N} \langle t_j|u_l\rangle \langle u_l|X'\rangle$$
$$= \sum_{l=1}^{N} \langle t_j|u_l\rangle c'_l = \sum_{l=1}^{N} \langle t_j|u_l\rangle c_l e^{i\theta_l} \tag{65}$$

We saw that the probabilities associated with the experiment $\boldsymbol{\varepsilon_1}$ do not change. But to reach our goal, which is that we want $|X'\rangle$to be a state vector too, then the probabilities associated with $\boldsymbol{\varepsilon_2}$ must not change. So, we must have:

$$|b'_j|^2 = |b_j|^2$$
$$b'_j b'^*_j = b_j b^*_j$$
$$\left(\sum_{l=1}^{N} \langle t_j|u_l\rangle c_l e^{i\theta_l}\right)\left(\sum_{k=1}^{N} \langle u_k|t_j\rangle c^*_k e^{-i\theta_k}\right) = \left(\sum_{l=1}^{N} \langle t_j|u_l\rangle c_l\right)\left(\sum_{k=1}^{N} \langle u_k|t_j\rangle c^*_k\right)$$

So we must have:

$$\sum_{l,k} \langle t_j|u_l\rangle \langle u_k|t_j\rangle c_l c^*_k e^{i(\theta_l-\theta_k)} = \sum_{l,k} \langle t_j|u_l\rangle \langle u_k|t_j\rangle c_l c^*_k \tag{66}$$

The former equation must be true for any c_k and c_l because we are speaking of arbitrary experiments with arbitrary probability distributions. It is true when we fix the bases whatever c_l were (we can fix the two bases and define an infinite number of experiments on them, meaning, whatever c_l and c^*_k were). So their coefficients must be the same, which means:

$$\langle t_j|u_l\rangle \langle u_k|t_j\rangle e^{i(\theta_l-\theta_k)} = \langle t_j|u_l\rangle \langle u_k|t_j\rangle \Rightarrow$$
$$\langle t_j|u_l\rangle \langle u_k|t_j\rangle [e^{i(\theta_l-\theta_k)} - 1] = 0 \tag{67}$$

It must be true for all experiments so for all bases even when:$\langle t_j|u_l\rangle \neq 0$ for any l and j. Thus:

$$e^{i(\theta_l-\theta_k)} = 1$$

Which means that:

$$e^{i\theta_l} = e^{i\theta_k} \Rightarrow \theta_l = \theta_k + 2\pi n, n \in \mathbb{Z} \tag{68}$$

And that is for any l and k. So we have:

$$z_l = z_k \tag{69}$$

So we see that:

$$z_1 = z_2 = ... = z_N \tag{70}$$

And since all of them are pure phases, then we can write:

$$z_1 = z_2 = ... = z_N \equiv e^{i\theta}$$

So:

$$|X'\rangle = \sum_{l=1}^{N} c_l e^{i\theta} |u_l\rangle = e^{i\theta} \sum_{l=1}^{N} c_l |u_l\rangle \Rightarrow$$
$$|X'\rangle = e^{i\theta} |X\rangle \tag{71}$$

So, for $|X'\rangle$ to be a state vector too, it must be of the former form. From the above we see that we can multiply $|X\rangle$ by any pure phase and still get another state vector. Immediately we can see that were we to work with a real vector space, the state vector would have been either $(|X\rangle)$ for $\theta = 0$ or $(-|X\rangle)$ for $\theta = \pi$.

2.1.2 Coarse-grained measurements and mixed states

Since we have seen that the same mathematical structure of quantum mechanics is valid for any extended probability theory whether it describes classical or quantum systems, it will be helpful if we import some useful definitions from the mathematical structure of quantum mechanics to see what their interpretations are in the light of these new extended probabilistic theories.
Given vectors and dual vectors we can define operators of the form (in this section, we will follow to a large extent the steps taken in [9]):

$$|\varphi\rangle \langle\psi| \tag{72}$$

And if $|\psi\rangle$ is a state vector, the projection operator for this state is written as:

$$\hat{P}_\psi = |\psi\rangle \langle\psi| \tag{73}$$

We will call it a density operator for a pure state.
We can even define a more general type of states, still described by density operators, by introducing 'mixtures' of pure states:

$$\hat{P} = \sum_{k=1}^{M} \lambda_k |\phi_k\rangle \langle\phi_k| \tag{74}$$

where $\{|\phi_k\rangle\}$ is some set of pure states, not necessarily orthogonal. The number M could be anything, and is not limited by the dimension of the Hilbert space. The M numbers (or 'weights') λ_k are nonzero and satisfy the relations:

$$0 < \lambda_k \leq 1; \sum_{k=1}^{M} \lambda_k = 1 \tag{75}$$

The normalization of the weights p_k expresses the condition $Tr(\hat{P}) = 1$.
Since $\hat{P}$ is Hermitian, we can diagonalize it, such that

$$\hat{P} = \sum_{k=1}^{N} p_k |\psi_k\rangle \langle\psi_k| \tag{76}$$

Where the states $\{|\psi_k\rangle\}$ are orthogonal. The numbers p_k satisfy

$$0 \leq p_k \leq 1; \sum_{k=1}^{N} p_k = 1 \tag{77}$$

The numbers p_k are, in fact, nothing but the eigenvalues of $\hat{P}$. They sum to one because of normalization. There are exactly $N = d$ of these numbers, where d is the dimension of the

Hilbert space.
Since this is the same familiar structure of quantum mechanics, we can use these two simple tests to determine whether $\hat{P}$ describes a pure state or not:
pure state: $\hat{P}^2 = \hat{P}$; mixed state: $\hat{P}^2 \neq \hat{P}$.
Or:
pure state: $Tr[\hat{P}^2] = 1$; mixed state: $Tr[\hat{P}^2] < 1$.
In fact, $P \equiv Tr[\hat{P}^2]$ is called the purity of a state. A state is pure when its purity equals 1, and mixed otherwise.
we can certainly prepare a mixed state in a probabilistic way. If we prepare with probability p_k a pure state $|\psi_k\rangle$, and then forget which pure state we prepared, the resulting mixed state is $\hat{P} = \sum_k p_k |\psi_k\rangle \langle\psi_k|$. In this case, p_k certainly has the meaning of probability.
For example, if we have an experiment (for instance throwing an ordinary die) with a state vector:

$$|\psi\rangle = \sum_{k=1}^{6} \sqrt{p_k} |k\rangle , 0 \leq p_k \leq 1, \sum_{k=1}^{6} p_k = 1 \tag{78}$$

we can immediately see that it is a pure state:

$$\hat{P} = |\psi\rangle \langle\psi| \tag{79}$$

because, since $|\psi\rangle$ is normalized, we have

$$\hat{P}^2 = |\psi\rangle \langle\psi|\psi\rangle \langle\psi| = |\psi\rangle \langle\psi| = \hat{P} \tag{80}$$

But if we take any other vector that represents a set which is a proper subset of the sample space, it will not be normalized, hence we will get $\hat{P}^2 \neq \hat{P}$, thus it represents a mixed state. For example, in the previous die example, if we take the vector that represents having an odd number, it will be represented by the vector

$$|\varphi\rangle = \sum_{n=1}^{3} \sqrt{p_n} |2n-1\rangle \tag{81}$$

which will correspond to a mixed state. This vector gives us the probability to measure an odd number and this probability is $\langle\varphi|\varphi\rangle$ as we have seen in [6]. Have we done this measurement and know that we really got an odd number, then we would have done a coarse-grained measurement if this is all the information we got. So mixed states can be used to represent coarse-graining measurements.
The density matrix operator in this case is

$$\hat{P} = \frac{1}{3} |1\rangle \langle 1| + \frac{1}{3} |3\rangle \langle 3| + \frac{1}{3} |5\rangle \langle 5| \tag{82}$$

So the density matrix is

$$\rho = \frac{1}{3} diag(1, 0, 1, 0, 1, 0) \tag{83}$$

Thus

$$\rho^2 = \frac{1}{9} diag(1, 0, 1, 0, 1, 0) \tag{84}$$

Hence

$$P = Tr[\hat{P}^2] = \frac{1}{3} < 1 \tag{85}$$

2.1.3 Entropy

Building on the previous coin example, we see that if we have a vector that represents a subset of the sample space, and we write it in the form

$$|\psi\rangle = \sum_{n=1}^{M} c_n |u_n\rangle \tag{86}$$

With $M \leq N$ where N is the cardinality of the sample space, the density operator will be

$$\hat{P} = \sum_{n=1}^{M} p_n |u_n\rangle \langle u_n| = \sum_{n=1}^{M} |c_n|^2 |u_n\rangle \langle u_n| \tag{87}$$

Which means that the ij-element of the density matrix that represents this operator is

$$\rho_{ij} = p_i \delta_{ij} \tag{88}$$

Where the index i runs over the basis vectors that appear in the expansion of $|\psi\rangle$ while the index j runs over all the basis vectors.
Hence, the ij-element of the matrix that represents $\log[\hat{P}]$ is

$$\delta_{ij} \log p_i \tag{89}$$

Thus, the ij-element of the matrix representing $\hat{P}\ \log[\hat{P}]$ will be

$$\delta_{ij} p_i \log p_i \tag{90}$$

Which means that

$$-Tr(\hat{P} \log[\hat{P}]) = -\sum_{i=1}^{M} p_i \log p_i = S \tag{91}$$

Where S is the entropy!
Now this means that the entropy of a pure state is zero, which corresponds to maximum knowledge, but it is a probabilistic knowledge about a class of probabilistic experiments that have the same number of outcomes as we saw. This means that if there is a deeper deterministic description of the situation, then there maby another knowledge hidden from us due to our probabilistic discribtion, that we can only get by knowing the deterministic description accurately.

2.1.4 Composition of experiments

Not all collections of events are "experiments". Whether or not a specific collection is an experiment is determined by the theory, compatibly with the basic requirement that the set of experiments must be closed under sequential and parallel composition.

2.1.4.1 Parallel composition The parallel composition of two experiments

$$\boldsymbol{\psi} = \left\{ \xrightarrow{A} \boxed{\varepsilon_i} \xrightarrow{\{a_i\}} \right\}_{i \in X} \quad , \quad \boldsymbol{\chi} = \left\{ \xrightarrow{B} \boxed{F_j} \xrightarrow{\{b_j\}} \right\}_{j \in Y} \tag{92}$$

is defined to be

$$\boldsymbol{\psi} \otimes \boldsymbol{\chi} := \left\{ \begin{array}{c} A \\ \hline \varepsilon_i \\ \hline B \\ \hline F_j \end{array} \begin{array}{c} \{a_i\} \\ \\ \{b_j\} \end{array} \right\}_{(i,j)\in X\times Y} = \left\{ \begin{array}{c} A\times B \\ \hline \varepsilon_i \otimes F_j \end{array} \begin{array}{c} \{(a_i,b_j)\} \\ \\ \end{array} \right\}_{(i,j)\in X\times Y} \tag{93}$$

and represents two possible non-deterministic processes occurring in parallel. The composition of experiments induces a composition of their outcome spaces via the Cartesian product. As a consequence, the set of all outcome spaces must be closed under this operation. We will denote such a set by Outcomes.
We see that the parallel composition can be interpreted as *an experiment* according to the definition we gave above for an experiment, and we will call this experiment *a composite experiment*.
For example, in the experiment of throwing two coins on parallel, if we assume that the experiment of throwing one coin is

$$\boldsymbol{F} = \left\{ \begin{array}{c} \{H,T\} \\ \hline \varepsilon_1 \end{array} \begin{array}{c} \{H\} \\ \\ \end{array} \quad , \quad \begin{array}{c} \{H,T\} \\ \hline \varepsilon_2 \end{array} \begin{array}{c} \{T\} \\ \\ \end{array} \right\} \tag{94}$$

Then the composite experiment is

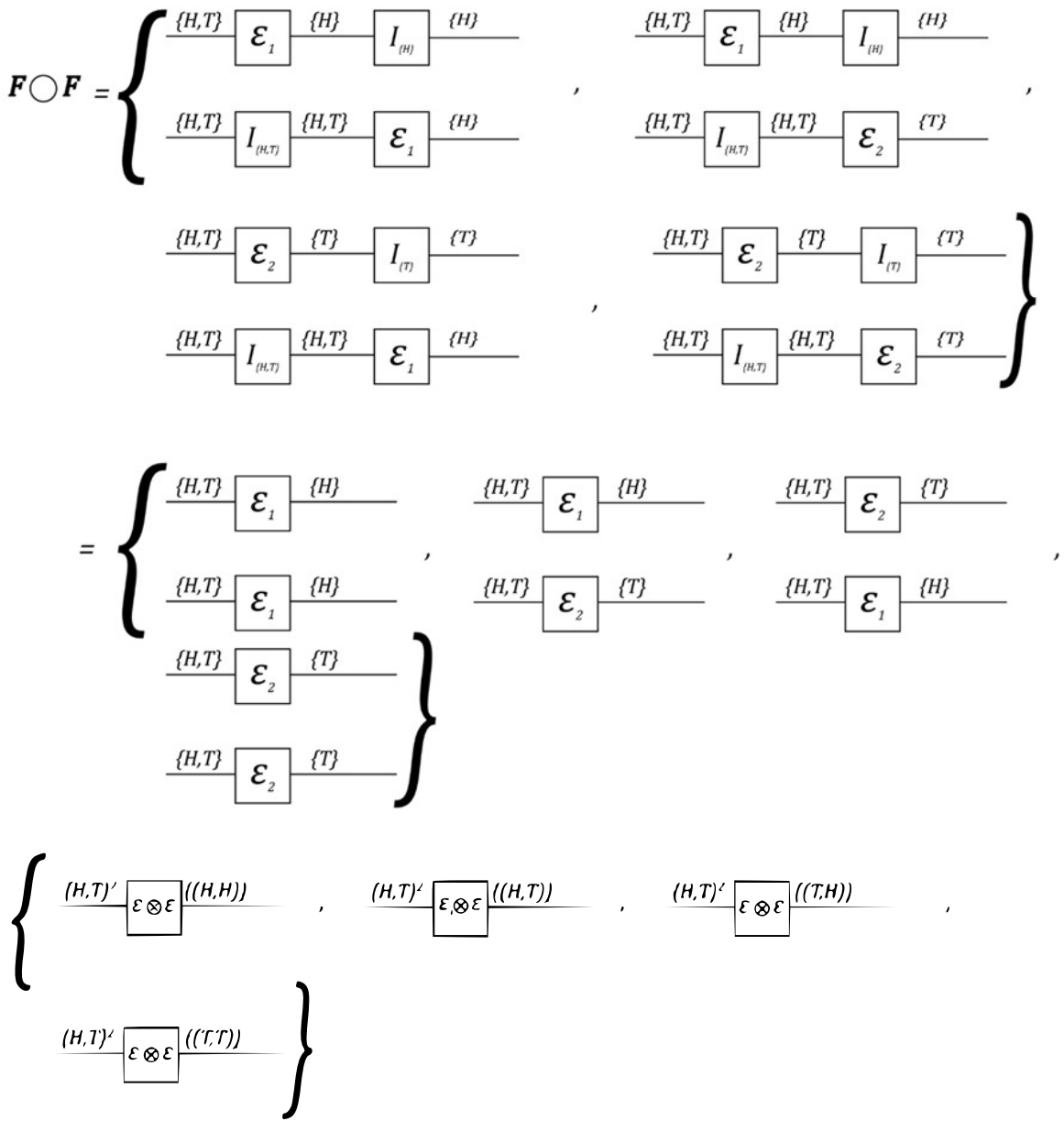
$$\boldsymbol{F} \otimes \boldsymbol{F} = \left\{ \begin{array}{c} \{H,T\} \\ \hline \varepsilon_1 \end{array} \begin{array}{c} \{H\} \\ \\ \end{array} \quad , \quad \begin{array}{c} \{H,T\} \\ \hline \varepsilon_2 \end{array} \begin{array}{c} \{T\} \\ \\ \end{array} \right\} \otimes \left\{ \begin{array}{c} \{H,T\} \\ \hline \varepsilon_1 \end{array} \begin{array}{c} \{H\} \\ \\ \end{array} \quad , \quad \begin{array}{c} \{H,T\} \\ \hline \varepsilon_2 \end{array} \begin{array}{c} \{T\} \\ \\ \end{array} \right\}$$

$$= \left\{ \begin{array}{c} \{H,T\}^2 \\ \hline \varepsilon_1 \otimes \varepsilon_1 \end{array} \begin{array}{c} \{(H,H)\} \\ \\ \end{array} \quad , \quad \begin{array}{c} \{H,T\}^2 \\ \hline \varepsilon_1 \otimes \varepsilon_2 \end{array} \begin{array}{c} \{(H,T)\} \\ \\ \end{array} \quad , \quad \begin{array}{c} \{H,T\}^2 \\ \hline \varepsilon_2 \otimes \varepsilon_1 \end{array} \begin{array}{c} \{(T,H)\} \\ \\ \end{array} \quad , \quad \begin{array}{c} \{H,T\}^2 \\ \hline \varepsilon_2 \otimes \varepsilon_2 \end{array} \begin{array}{c} \{(T,T)\} \\ \\ \end{array} \right\} \tag{95}$$

2.1.4.2 Sequential composition First, we will explain the concept using the simple example of throwing a coin two consecutive times. Again, let us assume that the experiment of throwing a coin is

$$\boldsymbol{F} = \left\{ \begin{array}{c} \{H,T\} \\ \hline \varepsilon_1 \end{array} \begin{array}{c} \{H\} \\ \\ \end{array} \quad , \quad \begin{array}{c} \{H,T\} \\ \hline \varepsilon_2 \end{array} \begin{array}{c} \{T\} \\ \\ \end{array} \right\} \tag{96}$$

We can think of the composite experiment in this case as two experiments running in parallel: the experiment of throwing the coin in the first one while parallel to it, nothing happened in the second one, then the experiment of throwing the coin in the second one, while nothing changes regarding the first one:



(97)

In general, if we have the two experiments:

$$\psi = \left\{ \; \begin{array}{c} A \\ \hline \end{array} \boxed{\mathcal{E}_i} \begin{array}{c} \{a_i\} \\ \hline \end{array} \right\}_{i \in X} \quad , \quad \chi = \left\{ \; \begin{array}{c} B \\ \hline \end{array} \boxed{F_j} \begin{array}{c} \{b_j\} \\ \hline \end{array} \right\}_{j \in Y} \tag{98}$$

Then we define their sequential composition as

$$\boldsymbol{\chi} \circ \boldsymbol{\psi} := \left\{ \begin{array}{l} A \quad \boxed{\varepsilon_i} \quad \{a_i\} \quad \boxed{I_{\{a_i\}}} \quad \{a_i\} \\ B \quad \boxed{I_B} \quad B \quad \boxed{F_j} \quad \{b_j\} \end{array} \right\}_{(i,j)\in X\times Y} = \left\{ \begin{array}{l} A \quad \boxed{\varepsilon_i} \quad \{a_i\} \\ B \quad \boxed{F_j} \quad \{b_j\} \end{array} \right\}_{(i,j)\in X\times Y}$$

$$= \left\{ A\times B \quad \boxed{\varepsilon_i \otimes F_j} \quad \{(a_i,b_j)\} \right\}_{(i,j)\in X\times Y}$$

$$= \boldsymbol{\psi} \otimes \boldsymbol{\chi} \tag{99}$$

Thus, we can see that also in the sequential composition we can get *an experiment* according to the definition we gave above for an experiment, and we will also call this experiment *a composite experiment*.
We can define a direction of time flow from top to bottom in the diagram, and we can define time steps using the number of identity events that appear in the previous equation, in this case two, which in turn is the number of experiments we compose sequentially. Thus we see that by default in this definition, the present cannot affect the past, thus causality is inevitable.
We also see that we got the remarkable result:

$$\boldsymbol{\chi} \circ \boldsymbol{\psi} = \boldsymbol{\psi} \otimes \boldsymbol{\chi} \tag{100}$$

This is a remarkable result, because it uncovers a strange and indeed deep correlation between space and time: it says that if for example, we throw a die in position A, and on parallel we throw a coin in position B, then this is equivalent from the point of view of category theory let us say to throwing the die, then throwing the coin afterwards. And if we exchange the initial positions of the coin and the die, then what is equivalent to that is throwing them in the reverse order in time! So, from category theory perspective, flipping the positions in space is equivalent to flipping the order in time!
Now we can see how this can remove some redundancy in probability theory: if we have two experiments whose sample spaces are Ω_1 and Ω_2 respectively, then we will have two options to represent the sample space of the composite experiment made out of the previous ones, and those options are either $\Omega = \Omega_1 \times \Omega_2$ or $\Omega' = \Omega_2 \times \Omega_1$. Thus, if we use one to represent parallel composition, then we can use the other to represent sequential composition!

2.1.5 The representation of composite experiments in vector spaces

Let us take two experiments, where the first experiment $\boldsymbol{\varepsilon_1}$ is from the class C_N and the second experiment $\boldsymbol{\varepsilon_2}$ is from the class C_M. Of course in general, these two experiments need not be on the same system.
We represent each experiment in its own vector space: $\boldsymbol{\varepsilon_1}$ in V_1 and $\boldsymbol{\varepsilon_2}$ in V_2.
Let us suppose we represent $\boldsymbol{\varepsilon_1}$ in V_1 by:

$$|X_1\rangle = \sum_{i=1}^{N} c_i \, |u_i\rangle \tag{101}$$

And that we represent $\boldsymbol{\varepsilon_2}$ in V_2 by:

$$|X_2\rangle = \sum_{j=1}^{M} b_j |t_j\rangle \tag{102}$$

We know from the previous section that whether we composed the previous experiments sequentially or in parallel, the outcome space of the resulting experiment will be the Cartesian product of their outcome spaces. Thus, it will have the cardinality $N.M$, this is why we need to represent the composite experiment in a vector space with dimensionality $N.M$. Since the vector space:

$$V = V_1 \otimes V_2 \tag{103}$$

has this dimensionality, and so does the vector space $V' = V_2 \otimes V_1$, this means we can represent the composite experiment by a vector in one of them (we can for example represent the parallel composition in V and the sequential composition in V'), let us call this vector $|X\rangle$. Let us say we represent the composite experiment in V. Since we can choose any orthonormal basis in V to represent this experiment, and since $\{|u_i t_j\rangle\}$ are orthonormal basis in this space, we can write:

$$|X\rangle = \sum_{i,j} f_{ij} |u_i t_j\rangle \in V_1 \otimes V_2 = V \tag{104}$$

Where $\langle X|X\rangle = 1$ and $f_{ij} |u_i t_j\rangle$ representing $\{(u_i, t_j)\}$, thus:

$$p(u_i, t_j) = |f_{ij}|^2 \tag{105}$$

If

$$|X\rangle = |X_1\rangle \otimes |X_2\rangle , \tag{106}$$

then we call it a product state. Otherwise, we call it an entangled state.
A product state represents two non-interacting systems, while the opposite is true for the entangled state, as is extensively explained in [6].
We can easily generalize this to any finite number of experiments.

2.1.6 Composite systems

From our definition above of the composite system, and from what we said about composing experiments, we see that we can treat an experiment on a composite system as a composite experiment on the subsystems that constitute the system, and treat it according to the previous mathematical structure.

2.1.7 Measurement

We will denote the collection of all experiments by Experim. Then, we will define the following function

$$\boldsymbol{M}_k : Experim \to Experim : \boldsymbol{\psi} = \left\{ \xrightarrow{A} \boxed{\varepsilon_i} \xrightarrow{\{a_i\}} \right\}_{i \in X} \longmapsto \left\{ \xrightarrow{\{a_k\}} \boxed{I_{\{a_k\}}} \xrightarrow{\{a_k\}} \right\} \tag{107}$$

for a specific value $k \in X$. We say that *this function represents making a measurement of the property X for the system under consideration, and getting specifically as a result k.*

We notice that we could have gotten for the experiment $\boldsymbol{\psi}$, a number of results which is $N = |X|$ if we try to measure X.
Thus, the measurement for some property (which is the outcome space of some experiment) gives us one event of the experiment, corresponding to one value of the property.
Now, if we try to measure the *same property* again, meaning we want to measure X, thus to take the image of

$$\boldsymbol{M}_k(\boldsymbol{\psi}) = \left\{ \xrightarrow{\{a_k\}} \boxed{I_{\{a_k\}}} \xrightarrow{\{a_k\}} \right\} \tag{108}$$

Then we can calculate the image only using the function $\boldsymbol{M}_k$ out of all the functions $\boldsymbol{M}_i$ where $i \in X$ because we have one event and one outcome in this experiment, and this outcome is k, and in this case we will get

$$\boldsymbol{M}_k(\boldsymbol{M}_k(\boldsymbol{\psi})) = \boldsymbol{M}_k(\boldsymbol{\psi}) \tag{109}$$

Thus, if we duplicate the same measurement, we will get the same result.
Now, let

$$\boldsymbol{\chi} = \left\{ \xrightarrow{B} \boxed{F_j} \xrightarrow{\{b_j\}} \right\}_{j \in Y} \tag{110}$$

Be another experiment that we want to do after the measurement that gave us k. In this case we will have

$$\boldsymbol{\chi} \circ \boldsymbol{M}_k(\boldsymbol{\psi}) = \boldsymbol{M}_k(\boldsymbol{\psi}) \otimes \boldsymbol{\chi} = \left\{ \begin{array}{l} \xrightarrow{\{a_k\}} \boxed{I_{\{a_k\}}} \xrightarrow{\{a_k\}} \\ \xrightarrow{B} \boxed{F_j} \xrightarrow{\{b_j\}} \end{array} \right\}_{j \in Y} = \left\{ \xrightarrow{\{a_k\} \times B} \boxed{I_{\{a_k\}} \otimes F_j} \xrightarrow{\{(a_k, b_j)\}} \right\}_{j \in Y} \tag{111}$$

And we can generalize this for any number of times.

2.1.8 The measurement problem (observer-system composite system)

We will start by an example, then generalize.
Suppose we take the composite system of (coin-coin tosser). Let the experiment of tossing the coin be

$$\boldsymbol{\varepsilon} = \left\{ \xrightarrow{\{H,T\}} \boxed{\varepsilon_1} \xrightarrow{\{H\}} \quad , \quad \xrightarrow{\{H,T\}} \boxed{\varepsilon_2} \xrightarrow{\{T\}} \right\} \tag{112}$$

While the experiment that represents what the tosser may see after tossing the coin is

$$\boldsymbol{O} = \left\{ \xrightarrow{\{O_H, O_T\}} \boxed{F_1} \xrightarrow{\{O_H\}} \quad , \quad \xrightarrow{\{O_H, O_T\}} \boxed{F_2} \xrightarrow{\{O_T\}} \right\} \tag{113}$$

Meaning that this observer may find the coin heads, which is represented by O_H, or the observer may find the coin tails, which is represented by O_T.
Now, after the toss, we cannot have simultaneously H and O_T, or T and O_H, because what the observer will see is defined by what the result of the toss is. Hence, the experiment that represents the potential change on the (coin/coin tosser) system will be

$$\boldsymbol{T} = \left\{ \xrightarrow{\{(H,O_H),(T,O_T)\}} \boxed{T_1} \xrightarrow{\{(H,O_H)\}} \quad , \quad \xrightarrow{\{(H,O_H),(T,O_T)\}} \boxed{T_2} \xrightarrow{\{(T,O_T)\}} \right\} \tag{114}$$

We notice that we cannot get $\boldsymbol{T}$ neither by parallel nor by sequential composition of $\boldsymbol{\varepsilon}$ and $\boldsymbol{O}$, which can be verified easily by simply taking their parallel and sequential compositions!

There is yet one more thing that we need to mention. Every time the coin is tossed and we get H, we simultaneously get O_H for the observer, and (H, O_T) for the composite system. The same is true if we replaced H by T.
Thus we must have

$$p(H, O_H) = p(O_H) = p(H) \tag{115}$$

And

$$p(T, O_T) = p(O_T) = p(T) \tag{116}$$

Now we will talk about the general case.
Supposing we have an experiment

$$\boldsymbol{\psi} = \left\{ \xrightarrow{A} \boxed{\varepsilon_i} \xrightarrow{\{a_i\}} \right\}_{i \in X} \tag{117}$$

We say that the set O is one of the sets that describe an observer of the previous experiment if and only if the following conditions are met:

1. we can build the following two experiments that have the same outcome space as $\boldsymbol{\psi}$:

$$\boldsymbol{O} = \left\{ \xrightarrow{O} \boxed{F_i} \xrightarrow{\{O_i\}} \right\}_{i \in X} \tag{118}$$

and

$$\boldsymbol{\chi} = \left\{ \xrightarrow{\{(a_j, O_j) | j \in X\}} \boxed{\chi_i} \xrightarrow{\{(a_i, O_i)\}} \right\}_{i \in X} \tag{119}$$

2. using the probabilistic structures for the previous experiments, we have

$$p(a_i, O_i) = p(a_i) = p(O_i) \tag{120}$$

We say that $\boldsymbol{\chi}$ represents the experiment done on the composite system of the observer and the original system, that corresponds to $\boldsymbol{\psi}$.
The state vectors of the three experiments respectively will be

$$|\psi\rangle = \sum_{i=1}^{N} c_i |a_i\rangle \in V_1 \tag{121}$$

$$|O\rangle = \sum_{i=1}^{N} b_i |O_i\rangle \in V_2 \tag{122}$$

$$|\chi\rangle = \sum_{i=1}^{N} d_i |a_i O_i\rangle \in V_1 \otimes V_2 \tag{123}$$

Where the following condition must be satisfied:

$$|c_i|^2 = |b_i|^2 = |d_i|^2 \tag{124}$$

So that the probabilities are the same as we have mentioned above.
We immediately see that

$$|\chi\rangle \neq |\psi\rangle \otimes |O\rangle \tag{125}$$

So, the measurement is an entanglement between the system and the observer.
Furthermore, we notice that if we make a measurement and get the value i of the property X, we will have the following three equations:

$$\boldsymbol{M_i(\psi)} = \left\{ \overset{\{a_i\}}{\text{———}} \boxed{I_{\{a_i\}}} \overset{\{a_i\}}{\text{———}} \right\} \tag{126}$$

$$\boldsymbol{M_i(O)} = \left\{ \overset{\{O_i\}}{\text{———}} \boxed{I_{\{O_i\}}} \overset{\{O_i\}}{\text{———}} \right\} \tag{127}$$

and

$$\boldsymbol{M_i(\chi)} = \left\{ \overset{\{(a_i,O_i)\}}{\text{———————}} \boxed{I_{\{(a_i,O_i)\}}} \overset{\{(a_i,O_i)\}}{\text{———————}} \right\}$$

$$= \left\{ \begin{array}{l} \overset{\{a_i\}}{\text{———}} \boxed{I_{\{a_i\}}} \overset{\{a_i\}}{\text{———}} \\ \underset{\{O_i\}}{\text{———}} \boxed{I_{\{O_i\}}} \overset{\{O_i\}}{\text{———}} \end{array} \right\}$$

$$= \boldsymbol{M_i(\psi)} \otimes \boldsymbol{M_i(O)} \tag{128}$$
$$= \boldsymbol{M_i(O)} \circ \boldsymbol{M_i(\psi)} \tag{129}$$

Now the question that arises is, how to represent $\boldsymbol{M}_k(\boldsymbol{\psi})$ by a vector, since it only has one outcome, and the lemma in section 2.1.1.2 was clear that our formulation is only valid for vector spaces with a dimensionality that is greater than or equal to two? In fact, since the only outcome that is in the experiment $\boldsymbol{M}_k$ is k, and we want this experiment to convey the meaning of measurement, and as we have seen according to (128), we can think of the experiments $\boldsymbol{M}_i(\boldsymbol{\psi})$ and $\boldsymbol{M}_i(\boldsymbol{O})$ as being done together, and once we have gotten an outcome for one of them, we get the corresponding outcome for the other, thus they have the same probability distribution after doing the measurement which is one for the outcome we got and zero for the others, and each experiment of them is a singleton, this is why we will represent them (up to a pure phase) by the following vectors respectively:

$$|\boldsymbol{M}_i(\boldsymbol{\psi})\rangle = \sum_{k=1}^{N} \delta_{ki} |a_k\rangle = |a_i\rangle \tag{130}$$

$$|\boldsymbol{M}_i(\boldsymbol{O})\rangle = \sum_{k=1}^{N} \delta_{ki} |O_k\rangle = |O_i\rangle \tag{131}$$

And since we also get one outcome in the composite experiment after the measurement, we

can here too write

$$|\boldsymbol{M}_i(\boldsymbol{\chi})\rangle = \sum_{k=1}^{N} \delta_{ki} |a_k O_k\rangle = |a_i O_i\rangle \tag{132}$$

We can call this a collapse in the state vector. But we also see that there is nothing mysterious here, for we just have a change in probability distribution after the measurement.
Another way to express the above is, that if $\hat{A}$ is an observable that the experiment measures (as we have mentioned, that means a Hermitian operator that has $\{|a_i\rangle\}_{i=1}^{N}$ as its eigenvectors), then the system after the measurement will be in an eigen state of $\hat{A}$ corresponding to the eigenvalue of it that we will measure[6].

2.1.9 Evolution of the states of a system

Let us assume that we have a system described by a state vector:

$$|\psi\rangle = \sum_m c_m |u_m\rangle \tag{133}$$

In some orthonormal basis $\{|u_m\rangle\}_{m=1}^{N}$. As we have seen in section 2.1.1.3, any other vector of the form:

$$|\varphi\rangle = \sum_m c_m e^{i\theta_m} |u_m\rangle \tag{134}$$

is another state vector for the system, which describes the same state only if all $\theta_m s$ are equal, otherwise it will describe a state in which the probability distributions of at least some experiments are different. But in general, they are not.
This enables us to deal with two kinds of evolution for states: continuous evolution, and discrete evolution.

2.1.9.1 Continuous evolution If we allow θ_m to vary continuously, then we can always write it as:

$$\theta_m = -g_m q \tag{135}$$

Where g_m is some real constant that depends only on m while q is a continuous variable. As we have seen in section 2.1.1.2, we can always build a Hermitian operator $\hat{G}$ as follows:

$$\hat{G} |u_m\rangle = g_m |u_m\rangle \tag{136}$$

Thus we have:

$$\begin{aligned}
\frac{\partial |\varphi\rangle}{\partial q} &= \frac{\partial}{\partial q} \sum_m c_m e^{-iqg_m} |u_m\rangle \\
&= \sum_m -ig_m c_m e^{-iqg_m} |u_m\rangle \\
&= \sum_m -ic_m e^{-iqg_m} \hat{G} |u_m\rangle \\
&= -i\hat{G} \sum_m c_m e^{-iqg_m} |u_m\rangle \\
&= -i\hat{G} |\varphi\rangle
\end{aligned}$$

Therefore:

$$\hat{G}\left|\varphi\right\rangle = i\frac{\partial\left|\varphi\right\rangle}{\partial q} \tag{137}$$

we can call $\hat{G}$ a generator for q translation. In fact, the previous equation is no other than Shrodinger's equation in its most general form.
Of course we could have written θ_m using any other continuous variable, for example:

$$\theta_m = -f_m r \tag{138}$$

with a continuous variable r and a constant f_m that depends only on m.
We can now build a new Hermitian operator $\hat{F}$ as follows:

$$F\left|u_m\right\rangle = f_m\left|u_m\right\rangle \tag{139}$$

and since $\hat{F}$ and $\hat{G}$ have the same eigenvectors, they commute.
Following the same procedure as before we find:

$$\hat{F}\left|\varphi\right\rangle = i\frac{\partial\left|\varphi\right\rangle}{\partial r} \tag{140}$$

and $\hat{F}$ is a generator for r translation.
Hence, we can start from any given state, and follow its continuous evolution using these generators.
We must assert here that had we used real Hilbert spaces, then the only principal values that θ_m can take are 0 or π. Hence, we would have not been able to speak of any kind of continuous evolution of states at all! Thus for example, we would not have neither a continuous space nor a continuous time, because each one of them will enter as the continuous variable in the Shrodinger's equation above, that has the right corresponding generator (the Hamiltonian in the case of time and the momentum in the case of space).

2.1.9.2 discrete evolution We can think of the transformation of a general state vector immediately into a state vector after the measurement as discussed above, as a discrete evolution of the state vector. Hence, the measurement, as discussed above, represents a discrete evolution in the state vector.

2.1.10 The uncertainty principle, entanglement and Bell's theorem

If quantum mechanics is just probability theory, then we must find classical counterparts for superposition, the uncertainty principle, and entanglement, without coming to conflict with Bell's theorem. In addition, we have to solve the measurement problem.
And in fact, it has been extensively explained in [6], how we can find counterparts to superposition, the uncertainty principle, and how this interpretation is in agreement with Bell's theorem, and it solves the measurement problem. But to make this paper as self-contained as possible, we will give a brief summary about some of te most important aspects to us here, with an emphasis in particular on what this interpretation tells us about entanglement specifically.
We saw in section 2.1.7, that in this formulation, the value that some property takes is defined through the act of measurement according to eq (107). That means, a property is defined to take a certain value only after measurement itself, or in other words, through the

act of measurement.
To understand this, let us take the example of a coin toss. What do we mean when we say that the coin is either heads or tails? We can define a unit vector $\hat{C}$ that is perpendicular to the surface of the coin and is pointing from the face that represents tails, to the face that represents heads. The coin is perpendicular to a previously known unit vector (usually it is taken as a normal vector to the surface of the earth, but not always), let us denote it by $\hat{n}$, and we say that the coin is heads when $\hat{C}$ and $\hat{n}$ are parallel, while we say that the coin is tails when they are antiparallel, as in the following figure

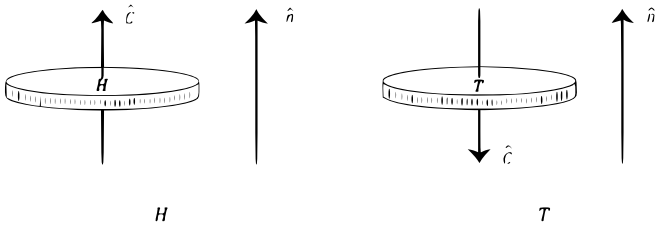


Figure 1: Defining heads and tails

Thus, the labels heads and tails have no meaning unless we determine relative to what $\hat{n}$ we are assigning them. Here we notice two things: first, we cannot assign the coin to be heads on two different directions, thus it cannot be *measured* to be heads on more than one direction, and this represents the uncertainty principle in this case. Second, since the result of the throw of the coin is not defined until after it becomes parallel/antiprallel to some direction after it is thrown, this means that this is in agreement with Bell's theorem, as is explained extensively in [6]. Furthermore, as we have seen above in the example of (coin/the person observing the coin), the state of the composite system is an entangled state. Before tossing the coin by someone, we cannot say that the coin was heads nor tails, neither that the observer sees it heads or tails (by sees it we mean that it is assigned some label in their universe). Assuming now that someone carried the coin in his/her hand and traveled to some faraway place. Nothing has changed so far, meaning, no one can assign a specific label to the coin whether heads or tails, since it is not thrown yet. When the person holding the coin throws it, and assigns it a specific label relative to some $\hat{n}$, and let us say the label was heads, then immediately in the universe of the observer, we can say that it sees (in the mentioned meaning of the word above) the coin heads. So, nothing mysterious here: what is happening in two entangled systems is that the results of one experiment done on one system, is defined using the results of the experiment done on the other!

2.2 Infinite dimensional outcome spaces

Since we have established the Hilbert space structure needed in quantum mechanics for finite dimensional cases, we can make it our starting point, and extend it in the same manner done in [6] to the infinite dimensional case. This is why we will not delve more into this matter in the current paper.

3 Final thoughts

The result of this work is that quantum mechanics is just probability theory cast in another language.
This leads to the following conclusions:

1. We do not have to assume the existence of infinite number of universes in order to answer the question of why the laws of physics are the way they are as Max Tegmark has done [10]. Because the answer to this question will simply be "because the universe is non-deterministic, thus it will be described using probability theory, hence quantum mechanics". On a personal level, I find this answer extremely philosophically pleasing, because if the deepest layer of reality is non-deterministic, then there is no meaning in questions like why the universe began if it has a beginning for the same reason that makes the question of why a radioactive atom decayed in a particular moment of time void of meaning. Nonetheless, I still prefer to look if the laws of quantum mechanics are the statistical description of a deeper deterministic layer of reality in the same sense that probability theory is a valid description for a Newtonian universe if we do not have enough information about the system that is under study. Finally, even if the theory that describes the universe is deterministic, it can be described using this structure as we have shown above in section 2.

2. Since we can use probability theory to describe "classical systems", thus, classical systems can be described using this structure of quantum mechanics but in this new understanding, hence we can apply quantum algorithms to them, which gives us more freedom in choosing the components out of which we can build quantum computers, which may give us new paths that make it more practical.

3. Since we can now simulate a quantum computer using a classical system (in principle at least), what this work reveals is that building algorithms using this probabilistic structure is more efficient than the conventional algorithms.

4. What this paper shows regarding the discussion of determinacy vs nondeterminacy of measured physical properties like position for example, is that even if there is some deterministic level from which our universe emerges, that does not necessarily mean there are some hidden variables which if we know, will make the position deterministic. Because the position -as a result of a probabilistic experiment- is in itself a part of a statistical description, and what might be deterministic in such a case are different sets of degrees of freedom that are different from the phyisical properties we usually measure!

References

[1] G. Chiribella, G.M. D'Ariano, P. Perinotti, Quantum from Principles, in Quantum Theory: Informational Foundations and Foils, Springer, 2016, pp. 3-9. doi: 10.1007/978-94-017-7303-4

[2] G. Chiribella, Dilation of states and processes in operational-probabilistic theories, Electronic Proceedings in Theoretical Computer Science 172, 2014 pp. 2–4.

[3] G. Chiribella, G.M. D'Ariano, P. Perinotti, Quantum Theory, Namely the Pure and Reversible Theory of Information, Entropy 2012, 14(10), 1877-1893, pp. 3-4. doi:10.3390/e14101877

[4] B. Coecke, What is category theory? Advanced Studies in Mathematics and Logic, volume 30, Polimetrica Publishing, 2006, p. 2.

[5] F.M. Dekking, C. Kraaikamp, H.P. Lopuhaä, L.E. Meester, A Modern Introduction to Probability and Statistics, springer, 2005, p. 16.

[6] R.M. Shaiia, A New Enhanced Version of The Ensemble Interpretation, 3 Mar 2020, pp. 1-16, arXiv:2003.05283 [physics.gen-ph]

[7] R. Shankar, principles of quantum mechanics, second edition, Plenum Press, 1994, p. 10.

[8] M. Saha, Inner product spaces and Hilbert spaces, pp. 4-5.

[9] S. van Enk, Mixed states and pure states, April 9, 2009, pp. 1-7.

[10] M. Tegmark, The Multiverse Hierarchy, 8 May 2009, pp. 10-13. arXiv:0905.1283 [physics.pop-ph]