On the missing axiom of Quantum Mechanics

Giacomo Mauro D'Ariano

QUIT *Group, Dipartimento di Fisica "A. Volta", via Bassi 6, I-27100 Pavia, Italy,* http://www.qubit.it
Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL 60208

Abstract. The debate on the nature of quantum probabilities in relation to Quantum Non Locality has elevated Quantum Mechanics to the level of an *Operational Epistemic Theory*. In such context the quantum superposition principle has an extraneous non epistemic nature. This leads us to seek purely operational foundations for Quantum Mechanics, from which to derive the current mathematical axiomatization based on Hilbert spaces.

In the present work I present a set of axioms of purely operational nature, based on a general definition of "the experiment", the operational/epistemic archetype of information retrieval from reality. As we will see, this starting point logically entails a series of notions [state, conditional state, local state, pure state, faithful state, instrument, propensity (i.e. "effect"), dynamical and informational equivalence, dynamical and informational compatibility, predictability, discriminability, programmability, locality, a-causality, rank of the state, maximally chaotic state, maximally entangled state, informationally complete propensity, etc.], along with a set of rules (addition, convex combination, partial orderings, ...), which, far from being of quantum origin as often considered, instead constitute the universal *syntactic manual* of the operational/epistemic approach. The missing ingredient is, of course, the quantum superposition axiom for probability amplitudes: for this I propose some substitute candidates of purely operational/epistemic nature.

Keywords: Foundations, Axiomatics, Measurement Theory
PACS: 03.65.-w

1. INTRODUCTION

Quantum Mechanics is not as any other physical theory. It applies to the entire physical domain, from micro to macrophysics, independently of the size and the energy scale, from particle physics, to nuclear, atomic, molecular, solid state physics, from the tiniest particle, to cosmology. Despite such generality, Quantum Mechanics still lacks a physical axiomatization—a quite embarrassing situation when we teach the theory to students. Why so abstract mathematical objects such as "Hilbert spaces" stay at the core axiomatic level of our most general physical theory? We are used to answer: "This is the *quantum superposition principle*, which entails complementarity and wave-particle dualism". That way we save our face.

In its very essence Quantum Mechanics addresses, for the first time, the core problem of Physics: that of *Measurement*. More generally, I would say, Quantum Mechanics deals with the description of the *Physical Experiment*. The probabilistic framework, which, in such context, is generally dictated by the obvious need of working in the presence of incomplete knowledge, contrarily to our original intentions turns out to be not of *epistemic* nature, but is truly *ontic*. This is the lesson of nonlocal EPR correlations. Incredibly, "God actually plays dice!" Now, this makes the situation even more embarrassing: on the basis of the quantum superposition principle of probability amplitudes we "physicists" preach the ontic nature of probability, and elevate Quantum Mechanics to a "Theory of Knowledge"!

Clearly, in this new view, the quantum superposition principle is not an acceptable starting point anymore: for a Theory of Knowledge we should seek operational axioms of epistemic nature, and be able to derive the usual mathematical axiomatization from such operational axioms. Shortly: for a Theory of Knowledge we need Axioms of Knowledge.

In the present work my starting point for this axiomatization is the definition of *what an experiment is*. Indeed, "the experiment" is the archetype of the *cognitive act*, being the prototype interaction with reality able to get information on it. As we will see, adopting a general definition of experiment that includes all possible interactions and information exchanges with reality, is a very seminal starting point, which logically entails a series of notions—such as that of state, conditional state, local state, pure state, faithful state, instrument, propensity (i.e. "effect"), dynamical and informational equivalence, dynamical and informational compatibility, predictability, discriminability, programmability, locality, a-causality, rank of a state, etc.]—along with a set of rules (addition, convex combination, partial orderings, ...), which, far from being of quantum origin as often considered, instead constitute the universal

CP810, *Quantum Theory: Reconsideration of Foundations) 3*,
edited by G. Adenier, A. Yu. Khrennikov, and Th. M. Nieuwenhuizen

syntactic manual of the cognitive/operational approach. The missing ingredient is still, of course, the quantum superposition axiom, and for this I will propose at the end some substitute candidates of purely cognitive/operational nature.

In the present attempt some expert readers will recognize similarities with the program of other authors during the seventies, following the Ludwig school [1], which were seeking operational principles to select the structure of quantum states from all possible convex structures [see, for example, the papers of U. Krause [2], H. Neumann [3], and E. Størmer [4] collected in the book [5]]. Why these work didn't have a followup? I think that, besides the fact that the convex structure by itself is not sufficiently rich mathematically for deriving an underlying Hilbert space structure, concepts as *entanglement* and *informationally complete measurements* (i. e. quantum tomography [6]) were still not familiar in those days. Recently it has been shown that it is possible to make a complete quantum calibration of a measuring apparatus [7] or of a quantum operation [8] by using a single pure bipartite state. I think that this gives us a new unique opportunity for deriving the Hilbert space structure from the convex structure in terms of *calibrability axioms*, which relies on the special link between the convex set of transformations and that of states which occurs in Quantum Mechanics, and which make the transformations of a single system resemble closely states of a bipartite system [9, 10]

2. AXIOMS FOR THE EXPERIMENT

It is the theory which decides what we can observe!

— Einstein to Heisenberg

General axiom 1 (On inductive-deductive science) *In any experimental inductive-deductive science we make* experiments *to get* information *on the* state *of a* objectified physical system. *Knowledge of such a state will allow us to predict the results of forthcoming experiments on the same object system. Since we necessarily work with only partial* a priori *knowledge of both system and experimental apparatus, the rules for the experiment must be given in a probabilistic setting.*

Notice that the information is of the *state* of the system, not of the system itself. In fact, in order to set the experiment we need some prior information on the physical system, e. g. if it is an electric current, a field, or a particle, what is its charge, etc. The goal of the experiment is to determine something unknown (or imprecisely known) about the system: logically this should enter in the notion of state, as will be given in Def. 2. The boundary between what is the object and what is its state will depend on the context of the particular experiment, e. g. the charge of a particle can be a property defining the object system—and used to design the measuring apparatus; if unknown, a property could be the object of the experiment itself, and, as such, it would enter the definition of state. Again we emphasize that our purpose is to give only the syntactic manual of the empirical approach, not the semantics, i. e. the specific physical context.

General axiom 2 (On what is an experiment) *An experiment on an object system consists in having it interact with an apparatus. The interaction between object and apparatus produces one of a set of possible transformations of the object, each one occurring with some probability. Information on the "state" of the object system at the beginning of the experiment is gained from the knowledge of which transformation occurred, which is the "outcome" of the experiment signaled by the apparatus.*

It is clear that here "object" and "apparatus" are both physical systems, and the asymmetry between object and apparatus is just an asymmetry in prior knowledge, namely the apparatus is the system of which the experimenter has more prior information. Clearly the knowledge gained about the state of the object depends also on the knowledge of details of the transformation undergone by the object system, and, generally, also on preexisting knowledge of the system "state" itself. In other words, the experiment can be always regarded as a *refinement* of knowledge on the object system.

One should convince himself that the above definition of experiment is very general, and includes all possible situations. For example, at first sight it may seem that it doesn't include the case in which the object is not under the experimenter's control (e. g. astronomical observations), in the sense that in such case one cannot establish an interaction with the object system. However, here also there is an interaction between the object of interest (e. g. the astronomical object) and another object (e. g. the light) which should be regarded as a part of the apparatus (i. e. telescope+light). Such cases can also be regarded as "indirect experiments", namely the experiment is performed on an auxiliary "object" (e. g. the light) which is supposed to have experienced a previous interaction with the ultimate object of interest, and whose state depends on properties/quantities of it. Also, the customary case in which a "quantity" or a "quality" is measured without in any way affecting the system corresponds to the case in which all states are left invariant by the transformations corresponding to each outcome.

Performing a different experiment on the same object obviously corresponds to the use of a different experimental apparatus or, at least, to a change of some settings of the apparatus. Abstractly, this corresponds to change the set $\{\mathscr{A}_j\}$ of possible transformations, $\mathscr{A}_j$, that the system can undergo. Such change could actually mean really changing the "dynamics" of the transformations, but it may simply mean changing only their probabilities, or, just their labeling outcomes. Any such change actually corresponds to a change of the experimental setup. Therefore, the set of all possible transformations $\{\mathscr{A}_j\}$ will be identified with the choice of experimental setting, i. e. with the *experiment* itself—or, equivalently, with the *action* of the experimenter: this will be formalized by the following definition

Definition 1 (Actions/experiments and outcomes) *An* **action** *or* **experiment** *on the object system is given by the set $\mathbb{A} \equiv \{\mathscr{A}_j\}$ of possible transformations $\mathscr{A}_j$ having overall unit probability, with the apparatus signaling the* **outcome** *j labeling which transformation actually occurred.*

Thus the action/experiment is just a *complete* set of possible transformations that can occur in an experiment. As we can see now, in a general probabilistic framework the *action* $\mathbb{A}$ is the "cause", whereas the *outcome* j labeling the transformation $\mathscr{A}_j$ that actually occurred is the "effect". The *action* has to be regarded as the “cause”, since it is the option of the experimenter, and, as such, it should be viewed as deterministic (at least one transformation $\mathscr{A}_j \in \mathbb{A}$ will occur with certainty), whereas the outcome j—i. e. which transformation $\mathscr{A}_j$ occurs—is probabilistic. The special case of a deterministic transformation $\mathscr{A}$ corresponds to a *singleton action/experiment* $\mathbb{A} \equiv \{\mathscr{A}\}$.

In the following, wherever we consider a nondeterministic transformation $\mathscr{A}$ by itself, we always regard it in the context of an experiment, namely for any nondeterministic transformation there always exists a at least complementary one $\mathscr{B}$ such that $\omega(\mathscr{A}) + \omega(\mathscr{B}) = 1$ for all states ω.

3. STATES

According to General Axiom 1 by definition the knowledge of the state of a physical system allows us to predict the results of forthcoming possible experiments on the system, or, more generally, on another system in the same physical situation. Then, according to the General Axiom 2 a precise knowledge of the state of a system would allow us to evaluate the probabilities of any possible transformation for any possible experiment. It follows that the only possible definition of state is the following

Definition 2 (States) *A state ω for a physical system is a rule that provides the probability for any possible transformation, namely*

$$\omega : \textbf{\textit{state}}, \quad \omega(\mathscr{A}) : \textit{probability that the transformation } \mathscr{A} \textit{ occurs}. \tag{1}$$

We assume that the identical transformation $\mathscr{I}$ occurs with probability one, namely

$$\omega(\mathscr{I}) = 1. \tag{2}$$

This corresponds to a kind of *interaction picture*, in which we don't consider the free evolution of the system (the scheme could be easily generalized to include a free evolution). Mathematically, a state will be a map ω from the set of physical transformations to the interval $[0,1]$, with the normalization condition (2). Moreover, for every action $\mathbb{A} = \{\mathscr{A}_j\}$ one has the normalization of probabilities

$$\sum_{\mathscr{A}_j \in \mathbb{A}} \omega(\mathscr{A}_j) = 1 \tag{3}$$

for all states ω of the system. As already noticed, in order to include also non-disturbing experiments, one must conceive situations in which all states are left invariant by each transformation (see also Remark 4 in the following).

The fact that we necessarily work in the presence of partial knowledge about both object and apparatus requires that the specification of the state and of the transformation could be given incompletely/probabilistically, entailing a convex structure on states and an addition rule for coexistent transformations. The convex structure of states is given more precisely by the rule

Rule 1 (Convex structure of states) *The possible states of a physical system comprise a convex set: for any two states ω_1 and ω_2 we can consider the state ω which is the* mixture *of ω_1 and ω_2, corresponding to have ω_1 with probability λ and ω_2 with probability $1-\lambda$. We will write*

$$\omega = \lambda\omega_1 + (1-\lambda)\omega_2, \quad 0 \leq \lambda \leq 1, \tag{4}$$

and the state ω will correspond to the following probability rule for transformations $\mathscr{A}$

$$\omega(\mathscr{A}) = \lambda\omega_1(\mathscr{A}) + (1-\lambda)\omega_2(\mathscr{A}). \tag{5}$$

Generalization to more than two states is obtained by induction. In the following the convex set of states will be denoted by $\mathfrak{S}$. We will call *pure* the states which are the extremal elements of the convex set, namely which cannot be obtained as mixture of any two states, and we will call *mixed* the non-extremal ones. As regards transformations, the addition of coexistent transformations and the convex structure will be considered in Rules 5 and 7.

Recall that for the convex set of states, as for any convex set, one can define partial orderings as follows.

Definition 3 (Partial ordering of states) *For $\omega, \zeta \in \mathfrak{S}$, $\alpha \in [0,1]$, denote by*

1. *$\omega \prec_\alpha \zeta$ if there exists a $\theta \in \mathfrak{S}$ such that $\zeta = \alpha\omega + (1-\alpha)\theta$;*
2. *$\omega \sim_\alpha \zeta$ if $\omega \prec_\alpha \zeta$ and $\zeta \prec_\alpha \omega$;*
3. *$\omega \prec \zeta$ if there exists $\alpha > 0$ such that $\omega \prec_\alpha \zeta$;*
4. *$\omega \sim \zeta$ if $\omega \prec \zeta$ and $\zeta \prec \omega$.*

For example, we can "read" the definition of $\prec$ in the following way: $\omega \prec \zeta$ means that ω belongs to an ensemble for ζ.

Definition 4 (Minimal decomposition of a state) *A minimal convex decomposition of a state is a convex expansion of the state in a minimal set of extremal states.*

Definition 5 (Caratheodory rank of a state) *The Caratheodory rank* $\mathrm{rank}(\omega)$ *of the state $\omega \in \mathfrak{S}$ (or simply rank) is the minimum number of extremal states in terms of which we can write the state as convex combination. This is also given by* $\dim[\mathrm{Fc}(\omega)] + 1$*, where* $\mathrm{Fc}(\omega) \subseteq \partial\mathfrak{S}$ *is the "face" to which the state ω belongs.*

Definition 6 (Caratheodory dimension) *We call the maximal rank of a state in $\mathfrak{S}$ the* Caratheodory dimension *of $\mathfrak{S}$, denoted by* $\mathrm{cdim}(\mathfrak{S})$.

Remark 1 According to the Caratheodory's theorem, for a convex set of real affine dimension n (i. e. embedded in $\mathbb{R}^n$) one needs at most $n+1$ extremal points to specify any point of the convex set as convex combination. However, for the convex sets of Quantum Mechanics one needs much fewer extremal points, precisely only $\sqrt{\dim(\mathfrak{S})+1}$ (the convex sets of states in Quantum Mechanics have real affine dimension $\dim(\mathfrak{S}) = k^2 - 1$, k being the dimension of the Hilbert space). Therefore, only $\sqrt{\dim(\mathfrak{S})+1}$ pure states are necessary to express each state as a convex combination. Such states are also a maximal set of perfectly discriminable states (see the following).

Remark 2 It is worth noticing that the dimension of the faces of the full convex set of quantum states $\mathfrak{S}$ for given finite dimension of the underlying Hilbert space decreases discontinuously in quadratic ladders. For example, the 8 dimensional convex set of states (corresponding to Hilbert space dimension $d = 3$) has faces that are 3-d Bloch spheres. Therefore, the faces of a complete set of quantum states are themselves complete set of quantum states (for lower dimension of the underlying Hilbert space). Each face of the complete convex set of states is itself a complete convex set of states at lower Hilbert space dimension. This lead us to consider also the following rule

Rule 2 *The faces of a "complete" set of states are themselves "complete" sets of states.*

The above rule needs a definition of what we mean by "completeness", and a possible route could be via the action of all possible invertible dynamical maps, i. e. the isometric indecomposable transformations of the set of states, namely the equivalent of unitary transformations (see the following). Notice, however, that the notion of *completeness* is not strictly operational, and for this reason we will not pursue this axiomatic route.

Using the partial ordering on the convex set of states we can easily define the maximally chaotic state as follows

Definition 7 (Maximally chaotic state) *The maximally chaotic state $\chi(\mathfrak{S})$ of $\mathfrak{S}$ is the most mixed state of $\mathfrak{S}$, in the sense that*

$$\forall\theta \in \mathfrak{S} \qquad \max\{\alpha \in [0,1] : \theta \succ_\alpha \chi(\mathfrak{S})\} \geq \max\{\beta \in [0,1] : \chi(\mathfrak{S}) \succ_\beta \theta\}. \tag{6}$$

An alternative definition is that of barycenter-state

Definition 8 (Alternative definition of maximally chaotic state) *The maximally chaotic state $\chi(\mathfrak{S})$ of the convex set $\mathfrak{S}$ is the barycenter of the set, i. e. it can be obtained by averaging over all pure states with the uniform measure, namely*

$$\chi(\mathfrak{S}) \doteq \int_{\mathrm{Extr}\,\mathfrak{S}} \mathrm{d}\psi\, \psi \tag{7}$$

where $\mathrm{Extr}\,\mathfrak{S}$ *denotes the set of extremal points of $\mathfrak{S}$, and $\mathrm{d}\psi$ is the measure that is invariant under isomorphisms of $\mathfrak{S}$.*

From Definition 7 it follows that the maximally chaotic state is full-rank, i. e. $\mathrm{rank}[\chi(\mathfrak{S})] = \sqrt{\dim(\mathfrak{S})+1}$. On the other hand, from Definition 8 it follows that the group of isomorphisms of $\mathfrak{S}$ leaves the state $\chi(\mathfrak{S})$ invariant (but generally $\chi(\mathfrak{S})$ is not the only invariant state).

4. TRANSFORMATIONS AND CONDITIONED STATES

Rule 3 (Transformations form a monoid) *The composition $\mathscr{A} \circ \mathscr{B}$ of two transformations $\mathscr{A}$ and $\mathscr{B}$ is itself a transformation. Consistency of composition of transformations requires* associativity*, namely*

$$\mathscr{C} \circ (\mathscr{B} \circ \mathscr{A}) = (\mathscr{C} \circ \mathscr{B}) \circ \mathscr{A}. \tag{8}$$

There exists the identical transformation $\mathscr{I}$ which leaves the physical system invariant, and which for every transformation $\mathscr{A}$ satisfies the composition rule

$$\mathscr{I} \circ \mathscr{A} = \mathscr{A} \circ \mathscr{I} = \mathscr{A}. \tag{9}$$

Therefore, transformations make a semigroup with identity, i. e. a monoid.

Definition 9 (Independent systems and local experiments) *We say that two physical systems are* independent *if on each system we can perform* local experiments *that don't affect the other system for any joint state of the two systems. This can be expressed synthetically with the commutativity of transformations of the local experiments, namely*

$$\mathscr{A}^{(1)} \circ \mathscr{B}^{(2)} = \mathscr{B}^{(2)} \circ \mathscr{A}^{(1)}, \tag{10}$$

where the label $n = 1,2$ of the transformations denotes the system undergoing the transformation.

In the following, when we have more than one independent system, we will denote local transformations as ordered strings of transformations as follows

$$(\mathscr{A}, \mathscr{B}, \mathscr{C}, \ldots) \doteq \mathscr{A}^{(1)} \circ \mathscr{B}^{(2)} \circ \mathscr{C}^{(3)} \circ \ldots \tag{11}$$

i. e. the transformation in parentheses corresponds to the local transformation $\mathscr{A}$ on system 1, $\mathscr{B}$ on system 2, etc.

Rule 4 (Bayes) *When composing two transformations $\mathscr{A}$ and $\mathscr{B}$, the probability $p(\mathscr{B}|\mathscr{A})$ that $\mathscr{B}$ occurs conditional on the previous occurrence of $\mathscr{A}$ is given by the Bayes rule*

$$p(\mathscr{B}|\mathscr{A}) = \frac{\omega(\mathscr{B} \circ \mathscr{A})}{\omega(\mathscr{A})}. \tag{12}$$

The Bayes rule leads to the concept of *conditional state*:

Definition 10 (Conditional state) *The* conditional state $\omega_{\mathscr{A}}$ *gives the probability that a transformation $\mathscr{B}$ occurs on the physical system in the state ω after the transformation $\mathscr{A}$ has occurred, namely*

$$\omega_{\mathscr{A}}(\mathscr{B}) \doteq \frac{\omega(\mathscr{B}\circ\mathscr{A})}{\omega(\mathscr{A})}. \tag{13}$$

Remark 3 (Linearity of evolution) At this point it is worth noticing that the present definition of "state", which logically follows from the definition of experiment, leads to a *notion of evolution as state-conditioning*. In this way, each transformation acts linearly on the state space. In addition, since states are probability functionals on transformations, by dualism (equivalence classes of) transformations are linear functionals over the state space. This clarifies the common misconception according to which it is impossible to mimic Quantum Mechanics as a mere classical probabilistic mechanics on a phase space viewed as a probability space since Quantum Mechanics admits linear evolutions only, whereas classical mechanics also admits nonlinear evolutions.

In the following we will make extensive use of the functional notation

$$\omega_{\mathscr{A}} \doteq \frac{\omega(\cdot\circ\mathscr{A})}{\omega(\mathscr{A})}, \tag{14}$$

where the centered dot stands for the argument of the map. Therefore, the notion of conditional state describes the most general *evolution*.

For the following it is convenient to extend the notion of state to that of *weight*, namely nonnegative bounded functionals $\tilde{\omega}$ over the set of transformations with $0 \leq \tilde{\omega}(\mathscr{A}) \leq \tilde{\omega}(\mathscr{I}) < +\infty$ for all transformations $\mathscr{A}$. To each weight $\tilde{\omega}$ it corresponds the properly normalized state

$$\omega = \frac{\tilde{\omega}}{\tilde{\omega}(\mathscr{I})}. \tag{15}$$

Weights make the convex cone $\tilde{\mathfrak{S}}$ which is generated by the convex set of states $\mathfrak{S}$. We are now in position to introduce the concept of operation.

Definition 11 (Operation) *To each transformation $\mathscr{A}$ we can associate a linear map* $\mathrm{Op}_{\mathscr{A}}: \mathfrak{S} \longrightarrow \tilde{\mathfrak{S}}$*, which sends a state ω into the unnormalized state* $\tilde{\omega}_{\mathscr{A}} \doteq \mathrm{Op}_{\mathscr{A}}\,\omega \in \tilde{\mathfrak{S}}$*, defined by the relation*

$$\tilde{\omega}_{\mathscr{A}}(\mathscr{B}) = \omega(\mathscr{B}\circ\mathscr{A}). \tag{16}$$

Similarly to a state, the linear form $\tilde{\omega}_{\mathscr{A}} \in \tilde{\mathfrak{S}}$ for fixed $\mathscr{A}$ maps from the set of transformations to the interval $[0,1]$. It is not strictly a state only due to lack of normalization, since $0 < \tilde{\omega}_{\mathscr{A}}(\mathscr{I}) \leq 1$. The operation Op gives the conditioned state through the state-reduction rule

$$\omega_{\mathscr{A}} = \frac{\tilde{\omega}_{\mathscr{A}}}{\omega(\mathscr{A})} \equiv \frac{\mathrm{Op}_{\mathscr{A}}\,\omega}{\mathrm{Op}_{\mathscr{A}}\,\omega(\mathscr{I})}. \tag{17}$$

The concept of conditional state naturally leads to the following category of transformations

Definition 12 (Purity of transformations) *A transformation is called* pure *if it preserves purity of states, namely if $\omega_{\mathscr{A}}$ is pure for ω pure.*

In contrast, we will call *mixing* a transformation which is not pure. We will also call *pure* an action made only of pure transformations and *mixing* an action containing at least one mixing transformation.

5. DYNAMICAL AND INFORMATIONAL EQUIVALENCE

From the Bayes rule, or, equivalently, from the definition of conditional state, we see that we can have the following complementary situations:

1. there are different transformations which produce the same state change, but generally occur with different probabilities;
2. there are different transformations which always occur with the same probability, but generally affect a different state change.

The above observation leads us to the following definitions of dynamical and informational equivalences of transformations.

Definition 13 (Dynamical equivalence of transformations) *Two transformations $\mathscr{A}$ and $\mathscr{B}$ are dynamically equivalent if $\omega_{\mathscr{A}} = \omega_{\mathscr{B}}$ for all possible states ω of the system.*

Definition 14 (Informational equivalence of transformations) *Two transformations $\mathscr{A}$ and $\mathscr{B}$ are informationally equivalent if $\omega(\mathscr{A}) = \omega(\mathscr{B})$ for all possible states ω of the system.*

Definition 15 (Complete equivalence of transformations/experiments) *Two transformations/experiments are completely equivalent iff they are both dynamically and informationally equivalent.*

Notice that even though two transformations are completely equivalent, in principle they can still be different experimentally, in the sense that they are achieved with different experimental apparatus. However, we emphasize that outcomes in different experiments corresponding to equivalent transformations always provide the same information on the state of the object, and, moreover, the corresponding transformations of the state are the same.

6. INFORMATIONAL COMPATIBILITY

The concept of dynamical equivalence of transformations leads one to introduce a convex structure also for transformations. We first need the notion of informational compatibility.

Definition 16 (Informational compatibility or coexistence) *We say that two transformations $\mathscr{A}$ and $\mathscr{B}$ are* coexistent *or* informationally compatible *if one has*

$$\omega(\mathscr{A}) + \omega(\mathscr{B}) \leq 1, \quad \forall\omega \in \mathfrak{S}, \tag{18}$$

The fact that two transformations are coexistent means that, in principle, they can occur in the same experiment, namely there exists at least an action containing both of them. We have named the present kind of compatibility "informational" since it is actually defined on the informational equivalence classes of transformations. Notice that the relation of coexistence is symmetric, but is not reflexive, since a transformation can be coexistent with itself only if $\omega(\mathscr{A}) \leq 1/2$. The present notion of coexistence is the analogous of that introduced by Ludwig [1] for the "effects". This notion is also related to that of "exclusive" transformations, since they correspond to exclusive outcomes [see also Ref. [11] in regards "exclusive" implies "coexistent", but generally not the reverse].

We are now in position to define the "addition" of coexistent transformations.

Rule 5 (Addition of coexistent transformations) *For any two coexistent transformations $\mathscr{A}$ and $\mathscr{B}$ we define the transformation $\mathscr{S} = \mathscr{A}_1 + \mathscr{A}_2$ as the transformation corresponding to the event $e = \{1,2\}$, namely the apparatus signals that either $\mathscr{A}_1$ or $\mathscr{A}_2$ occurred, but doesn't specify which one. By definition, one has the distributivity rule*

$$\forall\omega \in \mathfrak{S} \qquad \omega(\mathscr{A}_1 + \mathscr{A}_2) = \omega(\mathscr{A}_1) + \omega(\mathscr{A}_2), \tag{19}$$

whereas the state conditioning is given by

$$\forall\omega \in \mathfrak{S} \qquad \omega_{\mathscr{A}_1+\mathscr{A}_2} = \frac{\omega(\mathscr{A}_1)}{\omega(\mathscr{A}_1+\mathscr{A}_2)}\omega_{\mathscr{A}_1} + \frac{\omega(\mathscr{A}_2)}{\omega(\mathscr{A}_1+\mathscr{A}_2)}\omega_{\mathscr{A}_2}. \tag{20}$$

Notice that the two rules in Eqs. (19) and (20) completely specify the transformation $\mathscr{A}_1+\mathscr{A}_2$, both informationally and dynamically. Eq. (20) can be more easily restated in terms of operations as follows:

$$\forall\omega\in\mathfrak{S}\qquad \mathrm{Op}_{\mathscr{A}_1+\mathscr{A}_2}\,\omega=\mathrm{Op}_{\mathscr{A}_1}\,\omega+\mathrm{Op}_{\mathscr{A}_2}\,\omega. \tag{21}$$

Addition of compatible transformations is the core of the description of partial knowledge on the experimental apparatus. Notice also that the same notion of coexistence can extended to "propensities" as well (see Definition 18).

Definition 17 (Indecomposable transformation) *We call a transformation $\mathscr{T}$* indecomposable*, if there are no coexistent transformations summing to it.*

From the above definition we can see that the equivalent of quantum unitary transformations could be defined in terms of indecomposable isometric transformations.

Rule 6 (Multiplication of a transformation by a scalar) *For each transformation $\mathscr{A}$ the transformation $\lambda\mathscr{A}$ for $0\le\lambda\le 1$ is defined as the transformation which is dynamically equivalent to $\mathscr{A}$, but which occurs with probability $\omega(\lambda\mathscr{A})=\lambda\omega(\mathscr{A})$.*

Remark 4 (No-information from identity transformations) At this point a warning is in order, as regards the transformations that are dynamically equivalent to the identity, namely the *probabilistic identity transformations*. According to the Rule 6 for multiplication of transformations by a scalar, a probabilistic identity transformation will be of the form $p\mathscr{I}$, where p is the probability that the transformation occurs, namely $p=\omega(p\mathscr{I})$. One could now imagine an hypothetical situation of a "classical" experiment which leaves the object identically undisturbed, independently of its state, but still with many different outcomes j that are signaled by the apparatus. If such an experiment had an action of the form $\mathbb{A}=\{p_j\mathscr{I}\}$, it would provide no information on the state ω of the object, since by definition the probabilities of the outcomes will be independent on ω, because $\omega(p_j\mathscr{I})=p_j$. Therefore, a "classical" experiment makes sense only for an action $\mathbb{A}=\{\mathscr{A}_j\}$ made of non-identical transformations, but with the set of states restricted to be all invariant under $\mathbb{A}$.

It is now natural to introduce a norm over transformations as follows.

Theorem 1 (Norm for transformations) *The following quantity*

$$\|\mathscr{A}\|=\sup_{\omega\in\mathfrak{S}}\omega(\mathscr{A}), \tag{22}$$

is a norm on the set of transformations. In terms of such norm all transformations are contractions.

Proof. The quantity in Eq. (22) satisfy the sub-additivity relation $\|\mathscr{A}+\mathscr{B}\|\le\|\mathscr{A}\|+\|\mathscr{B}\|$, since

$$\|\mathscr{A}+\mathscr{B}\|=\sup_{\omega\in\mathfrak{S}}[\omega(\mathscr{A})+\omega(\mathscr{B})]\le\sup_{\omega\in\mathfrak{S}}\omega(\mathscr{A})+\sup_{\omega'\in\mathfrak{S}}\omega'(\mathscr{B})=\|\mathscr{A}\|+\|\mathscr{B}\|. \tag{23}$$

Moreover, it obviously satisfies the identity

$$\|\lambda\mathscr{A}\|=\lambda\|\mathscr{A}\|. \tag{24}$$

It is also clear that, by definition, for each transformation $\mathscr{A}$ one has $\|\mathscr{A}\|\le 1$, namely transformations are contractions. ■

Obviously the multiplication of a transformation $\mathscr{A}$ by a scalar is more generally defined for a scalar $\lambda\le\|\mathscr{A}\|^{-1}$, which can be larger than unity. In terms of the norm (22) one can equivalently define coexistence (informational compatibility) using the following corollary

Corollary 1 *Two transformations $\mathscr{A}$ and $\mathscr{B}$ are coexistent iff $\mathscr{A}+\mathscr{B}$ is a contraction.*

Proof. If the two transformations are coexistent, then from Eqs. (18) and (22) one has that $\|\mathscr{A}+\mathscr{B}\|\le 1$. On the other hand, if $\|\mathscr{A}+\mathscr{B}\|\le 1$, this means that Eq. (22) is satisfied for all states, namely the transformations are coexistent. ■

Corollary 2 *The transformations $\lambda\mathscr{A}$ and $(1-\lambda)\mathscr{B}$ are compatible for any couple of transformations $\mathscr{A}$ and $\mathscr{B}$.*

Proof. Clearly $\|\lambda\mathscr{A}+(1-\lambda)\mathscr{B}\| \leq \lambda\|\mathscr{A}\|+(1-\lambda)\|\mathscr{B}\| \leq 1.$ ■

The last corollary implies the rule

Rule 7 (Convex structure of transformations) *Transformations form a convex set, namely for any two transformations $\mathscr{A}_1$ and $\mathscr{A}_2$ we can consider the transformation $\mathscr{A}$ which is the* mixture *of $\mathscr{A}_1$ and $\mathscr{A}_2$ with probabilities λ and $1-\lambda$. Formally, we write*

$$\mathscr{A}=\lambda\mathscr{A}_1+(1-\lambda)\mathscr{A}_2, \quad 0 \leq \lambda \leq 1, \tag{25}$$

with the following meaning: the transformation $\mathscr{A}$ is itself a probabilistic transformation, occurring with overall probability

$$\omega(\mathscr{A})=\lambda\omega(\mathscr{A}_1)+(1-\lambda)\omega(\mathscr{A}_2), \tag{26}$$

meaning that when the transformation $\mathscr{A}$ occurred we know that the transformation dynamically was either $\mathscr{A}_1$ with (conditioned) probability λ or $\mathscr{A}_2$ with probability $(1-\lambda)$.

We have seen that the transformations form a convex set, more specifically, a spherically truncated convex cone, namely we can always add transformations or multiply a transformation by a positive scalar if the result is a contraction. In the following we will denote the spherically truncated convex cone of transformations as $\mathfrak{T}$.

We should be aware that extremality of transformations in relation to their convex structure is not equivalent to the concept of purity in Definition 12, since a pure transformation is not necessarily extremal (just consider the convex combination of two different transformations that map to the same pure state), and vice-versa the fact that a transformation is mixing doesn't logically imply that it can be always regarded as a convex combination of extremal transformations.

Remark 5 (Banach algebra of transformations) The convex cone of transformations can be extended (on the embedding affine space) to a real Banach algebra equipped with the norm given in Theorem 1, the closure corresponding to an approximation criterion for transformations.

An obvious consequence of the rule 7 is that actions too form a convex set, namely

Rule 8 (Convex structure of actions) *Actions make a convex set, namely for any two actions $\mathbb{A}=\{\mathscr{A}_j\}$ and $\mathbb{B}=\{\mathscr{B}_j\}$ we can consider the action $\mathbb{C}$ which is the* mixture *of $\mathbb{A}$ and $\mathbb{B}$ with probabilities λ and $1-\lambda$*

$$\mathbb{C}=\lambda\mathbb{A}+(1-\lambda)\mathbb{B}=\{\lambda\mathscr{A}_j,(1-\lambda)\mathscr{B}_i\}, \quad 0 \leq \lambda \leq 1,, \tag{27}$$

with the following meaning: the action $\mathbb{C}$ has the union of outcomes of actions $\mathbb{A}$ and $\mathbb{B}$, and contains the transformations $\lambda\mathscr{A}_j$ and $(1-\lambda)\mathscr{B}_j$ which are dynamically equivalent to those of actions $\mathbb{A}$ and $\mathbb{B}$.

7. PROPENSITIES

Informational equivalence allows one to define equivalence classes of transformations, which we may want to call *propensities*, since they give the occurrence probability of a transformation for each state, i. e. its "disposition" to occur.

Definition 18 (Propensities) *We call* **propensity** *an informational equivalence class of transformations.*

It is easy to see that the present notion of propensity corresponds closely to the notion of "effect" introduced by Ludwig [1]. However, we prefer to keep a separate word, since the "effect" has been identified with a quantum mechanical notion and a precise mathematical object (i. e. a positive contraction). In the following we will denote propensities with underlined symbols as $\underline{\mathscr{A}}$, $\underline{\mathscr{B}}$, etc., and we will use the notation $[\mathscr{A}]$ for the propensity containing the transformation $\mathscr{A}$, and also write $\mathscr{A}'\in[\mathscr{A}]$ to say that $\mathscr{A}'$ is informationally equivalent to $[\mathscr{A}]$. It is clear that $\lambda\mathscr{A}$ and $\lambda\mathscr{B}$ belong to the same equivalence class iff $\mathscr{A}$ and $\mathscr{B}$ are informationally equivalent. This means that also for propensities multiplication by a scalar can be defined as $\lambda[\mathscr{A}]=[\lambda\mathscr{A}]$. Moreover, since for $\mathscr{A}'\in[\mathscr{A}]$ and $\mathscr{B}'\in[\mathscr{B}]$ one has $\mathscr{A}'+\mathscr{B}'\in[\mathscr{A}+\mathscr{B}]$, we can define addition of propensities as $[\mathscr{A}]+[\mathscr{B}]=[\mathscr{A}+\mathscr{B}]$ for any choice of representatives $\mathscr{A}$ and $\mathscr{B}$ of the two added propensities. Also, since all transformations of the same equivalence class have the same norm, we can extend the definition (22) to propensities as $\|[\mathscr{A}]\| = \|\mathscr{A}\|$ for any representative $\mathscr{A}$ of the class. It is easy to check sub-additivity on classes, which implies that it is indeed a norm. In fact, one has

$$\|[\mathscr{A}]+[\mathscr{B}]\| = \|\mathscr{A}+\mathscr{B}\| \leq \|\mathscr{A}\|+\|\mathscr{B}\| = \|[\mathscr{A}]\|+\|[\mathscr{B}]\|. \tag{28}$$

Therefore, it follows that also propensities form a spherically truncated convex cone (which is a convex set), and in the following we will denote it by $\mathfrak{P}$.

With the present norm for propensities, Ludwig [1] introduces the notion of "ensembles with maximal absorption", corresponding to the state achieving the norm of the propensity $l(\omega) = \|l\|$ and of "ensembles totally absorbed" when $l(\omega) = 1$.

Remark 6 (Duality between the convex sets of states and of propensities) From the Definition 2 of state it follows that the convex set of states $\mathfrak{S}$ and the convex sets of propensities $\mathfrak{P}$ are dual each other, and the latter can be regarded as the set of positive linear contractions over the set of states, namely the set of positive functionals l on $\mathfrak{S}$ with unit upper bound, and with the functional $l_{[\mathscr{A}]}$ corresponding to the propensity $[\mathscr{A}]$ being defined as

$$l_{[\mathscr{A}]}(\omega) \doteq \omega(\mathscr{A}). \tag{29}$$

In the following we will often identify propensities with their corresponding functionals, and denote them by lowercase letters $a, b, c, \ldots$, or $l_1, l_2, \ldots$. Finally, notice that the notion of coexistence (informational compatibility) extends naturally to propensities.

Remark 7 (Dual cone notation) We can write the propensity linear functionals on $\mathfrak{S}$ with the equivalent pairing notations

$$l_{\underline{\mathscr{A}}}(\omega) \doteq \omega(\underline{\mathscr{A}}) \equiv (\underline{\mathscr{A}}, \omega). \tag{30}$$

Definition 19 (Observable) *We call observable a set of propensities $\mathbb{L} = \{l_i\}$ which is informationally equivalent to an action $\mathbb{L} \in \underline{\mathbb{A}}$, namely such that there exists an action $\mathbb{A} = \{\mathscr{A}_j\}$ for of which one has $l_i \in \underline{\mathscr{A}_j}$.*

Clearly, the generalized observable is normalized to the constant unit functional, i. e. $\sum_i l_i = 1$.

Definition 20 (Informationally complete observable) *An observable $\mathbb{L} = \{l_i\}$ is informationally complete if each propensity can be written as a linear combination of the of elements of $\mathbb{L}$, namely for each propensity l there exist coefficients $c_i(l)$ such that*

$$l = \sum_i c_i(l) l_i. \tag{31}$$

Clearly, using an informationally complete observable one can reconstruct any state ω from just the probabilities $l_i(\omega)$, since one has

$$\omega(\mathscr{A}) = \sum_i c_i(l_{\underline{\mathscr{A}}}) l_i(\omega). \tag{32}$$

Rule 9 (Partial ordering between propensities) *For two propensities $l_1, l_2 \in \mathfrak{P}$ we write $l_1 \leq l_2$ when $l_1(\omega) \leq l_2(\omega)$ $\forall \omega \in \mathfrak{S}$.*

In Ref. [1] the present partial ordering is interpreted saying that l_2 is *more sensitive* than l_1.

8. DYNAMIC COMPATIBILITY

Regarding the dynamical face of the concept of "transformation", we can introduce another notion of compatibility, which is closer to the one usually considered in quantum mechanics.

Definition 21 (Dynamical compatibility) *We say that two transformations $\mathscr{A}$ and $\mathscr{B}$ are dynamically compatible if they commute, namely $\mathscr{A} \circ \mathscr{B} = \mathscr{B} \circ \mathscr{A}$.*

An example of dynamically compatible transformations is provided by a couple of local transformations on independent object systems.

9. COMPATIBILITY OF EXPERIMENTS

The concept of dynamical compatibility naturally extends to actions as follows.

Definition 22 (Compatible experiments) *We call two experiments made with two different apparatuses compatible—i. e. they can be performed contextually on the same object system—when their relative order is irrelevant, namely their action are made of transformations that are dynamically compatible.*

The above definitions means that the actions $\mathbb{A} = \{\mathscr{A}_j\}$ and $\mathbb{B} = \{\mathscr{B}_i\}$ of two compatible experiments are such that $\mathscr{A}_j \circ \mathscr{B}_i = \mathscr{B}_i \circ \mathscr{A}_j$ for all transformations of $\mathbb{A}$ and $\mathbb{B}$. This allows one to define the contextually joint experiment, with action $\mathbb{C} = \mathbb{A}\&\mathbb{B}$ and $\mathbb{C} = \{\mathscr{C}_{ij}\}$, where now the possible outcomes are the product events ij corresponding to transformations $\mathscr{C}_{ij} = \mathscr{A}_j \circ \mathscr{B}_i \equiv \mathscr{B}_i \circ \mathscr{A}_j$. Notice that when joining contextually two experiments, generally their outcomes are correlated, namely $\omega(\mathscr{B}_i \circ \mathscr{A}_j) \neq \omega(\mathscr{B}_i)\omega(\mathscr{A}_j)$, and compatibility only implies the identity

$$\frac{\omega_{\mathscr{A}_j}(\mathscr{B}_i)}{\omega_{\mathscr{B}_i}(\mathscr{A}_j)} = \frac{\omega(\mathscr{B}_i)}{\omega(\mathscr{A}_j)}. \tag{33}$$

The present definition of contextuality may look artificial, but it is in line with the "a-temporal" scenario of our definition of experiment, where "time" refers only to the before-after ordering between the action—the "cause"—and the transformation of the object system—the "effect". In this fashion, the only logical way of defining contextually joint experiments is to consider them as equivalent for any choice of their ordering. Clearly, in any practical definition of contextual joint experiments, at least we need to have the apparatuses as independent systems themselves. On the other hand, for incompatible experiments with actions $\mathbb{A}$ and $\mathbb{B}$ one can always define the experiment corresponding to the cascade of the previous two on the same object system, with action $\mathbb{B} \circ \mathbb{A} = \{\mathscr{B}_i \circ \mathscr{A}_j\}$.

Notice how the present definition of compatible experiments is deeply related to that of independent systems. Indeed, if there exists a nonempty commutant for a complete set of transformations, this will allow one to define two subsystems, at least in the sense of “virtual subsystems” [12].

The informational counterpart of compatible experiments will be the following

Definition 23 (Informational compatibility of experiments) *We say that two experiments with actions $\mathbb{A} = \{\mathscr{A}_j\}$ and $\mathbb{B} = \{\mathscr{B}_i\}$ are informationally compatible when there exists a third experiment whose action $\mathbb{C}$ has marginals informationally equivalent to $\mathbb{A}$ and $\mathbb{B}$, namely we can partition the outcomes in such a way that we can write $\mathbb{C} = \{\mathscr{C}_{ij}\}$ with $\sum_i \mathscr{C}_{ij} \in [\mathscr{A}_j]$ and $\sum_j \mathscr{C}_{ij} \in [\mathscr{B}_i]$.*

Notice that dynamically compatible experiments are always informationally compatible, since one has

$$\begin{aligned} \sum_i \omega(\mathscr{B}_i \circ \mathscr{A}_j) &= \sum_i \omega(\mathscr{A}_j)\omega_{\mathscr{A}_j}(\mathscr{B}_i) \equiv \omega(\mathscr{A}_j), \\ \sum_j \omega(\mathscr{B}_i \circ \mathscr{A}_j) &= \sum_j \omega(\mathscr{A}_j \circ \mathscr{B}_i) = \sum_j \omega(\mathscr{B}_i)\omega_{\mathscr{B}_i}(\mathscr{A}_j) \equiv \omega(\mathscr{B}_i), \end{aligned} \tag{34}$$

whereas, generally, for the cascade of experiments $\mathbb{B} \circ \mathbb{A} = \{\mathscr{B}_i \circ \mathscr{A}_j\}$, one has only $\sum_i \mathscr{B}_i \circ \mathscr{A}_j \in [\mathscr{A}_j]$, but generally $\sum_j \mathscr{B}_i \circ \mathscr{A}_j \notin [\mathscr{B}_i]$.

10. PREDICTABILITY AND DISTANCES BETWEEN STATES

Definition 24 (Predictability and resolution) *We will call a transformation $\mathscr{A}$—and likewise its propensity—* predictable *if there exists a state for which $\mathscr{A}$ occurs with certainty and some other state for which it never occurs. The transformation (propensity) will be also called* resolved *if the state for which it occurs with certainty is unique—whence pure. An action will be called* predictable *when it is made only of predictable transformations, and* resolved *when all transformations are resolved.*

The present notion of predictability for propensity corresponds to that of "decision effects" of Ludwig [1]. For a predictable transformation $\mathscr{A}$ one has $\|\mathscr{A}\| = 1$. Notice that a predictable transformation is not deterministic, and it can generally occur with nonunit probability on some state ω. Predictable propensities $\mathscr{A}$ correspond to affine functions $f_{\mathscr{A}}$ on the state space $\mathfrak{S}$ with $0 \leq f_{\mathscr{A}} \leq 1$ achieving both bounds. Their set will be denoted by $\mathfrak{P}_p$.

Via propensities, we can also introduce notions of *distance* and of *orthogonality* on the state space $\mathfrak{S}$.

Definition 25 (Distance between states) *Let $\mathfrak{P}$ denote the set of propensities on the convex set of states $\mathfrak{S}$. Define the "distance" between states $\omega, \zeta \in \mathfrak{S}$ as follows*

$$d(\omega,\zeta) = \sup_{l\in\mathfrak{P}} l(\omega) - l(\zeta). \tag{35}$$

Theorem 2 *The function (35) is a metric on $\mathfrak{S}$.*

Proof. For every propensity l, $1-l$ is also a propensity, whence

$$d(\omega,\zeta) = \sup_{l\in\mathfrak{P}}(l(\omega)-l(\zeta)) = \sup_{l'\in\mathfrak{P}}((1-l')(\omega)-(1-l')(\zeta)) = \sup_{l'\in\mathfrak{P}}(l'(\zeta)-l'(\omega)) = d(\zeta,\omega), \tag{36}$$

namely d is symmetric. On the other hand, $d(\omega,\zeta)=0$ implies that $\zeta=\omega$, since the two states must give the same probabilities for all transformations. Finally, one has

$$d(\omega,\zeta) = \sup_{l\in\mathfrak{P}}(l(\omega)-l(\theta)+l(\theta)-l(\zeta)) \le \sup_{l\in\mathfrak{P}}(l(\omega)-l(\theta)) + \sup_{l\in\mathfrak{P}}(l(\theta)-l(\zeta)) = d(\omega,\theta)+d(\theta,\zeta), \tag{37}$$

namely it satisfy the triangular inequality, whence d is a metric. ■

One can see that, by construction, the distance is bounded as $d(\omega,\zeta)\le 1$, since the maximum value of $d(\omega,\zeta)$ is achieved for $l(\omega)=1$ and $l(\zeta)=0$. Moreover, since for a linear function on a convex domain both maximum and minimum are achieved on facets (i. e. convex hulls of some extremal points), this means that the bound $d(\omega,\zeta)=1$ can be achieved only when ω and ζ lie on different facets of the convex set. Finally, for convex combinations we have the following

Lemma 1 *Mixing reduces distances linearly.*

Proof. For any convex combination $\theta = \alpha\omega + (1-\alpha)\zeta$ one has $d(\theta,\zeta) = \alpha d(\omega,\zeta)$, since

$$d(\theta,\zeta) = \sup_{l\in\mathfrak{P}}(\alpha l(\omega) + (1-\alpha)l(\zeta) - l(\zeta)) = \sup_{l\in\mathfrak{P}}(\alpha l(\omega) - \alpha l(\zeta)) = \alpha d(\omega,\zeta). \tag{38}$$

Definition 26 (Orthogonality of states) *Two states $\omega,\zeta\in\mathfrak{S}$ are called* orthogonal *(denoted as $\omega\perp\zeta$) if $d(\omega,\zeta)=1$.*

Definition 27 (Metrical dimensionality) *The metric dimensionality is the maximum number of pairwise orthogonal states according to Definition 26.*

For example, the metric dimensionality of any N-hypersphere is 2, since the set of predictable propensity is made of the linear functions $f_{\vec{m}}(\vec{n}) = \frac{1}{2}(1+\vec{n}\cdot\vec{m})$ where $\vec{m}$ is a unit vector, and the metric is $d(\vec{n},\vec{n}') = \max_{\vec{m}} \frac{1}{2}\vec{m}\cdot(\vec{n}-\vec{n}') \equiv \frac{1}{2}|\vec{n}-\vec{n}'|$, whence one sees that only antipodal points have distance 1.

Example 1 Consider the trace-norm distance on the convex set of density operators over the Hilbert space H $d(x,y) = \frac{1}{2}\|x-y\|_1$. For pure states one has $d(x,y) = \sqrt{1-|\langle\psi_x|\psi_y\rangle|^2}$. Therefore, the metric structure of H is rediscovered via the inner metric of the state-space, and orthogonality in H means maximal inner distance $d(x,y)=1$ in the state space.

Definition 28 (Isometric transformations) *A transformation $\mathscr{U}$ is called* isometric *if it preserves the distance between states, namely*

$$d(\omega_{\mathscr{U}},\zeta_{\mathscr{U}}) \equiv d(\omega,\zeta), \qquad \forall\omega,\zeta\in\mathfrak{S}. \tag{39}$$

Isometric transformations are isomorphisms of the convex of states $\mathfrak{S}$. On the other hand, isomorphisms of the convex set of propensities $\mathfrak{P}$ are also isometric transformations of states, since

$$\sup_{l_{\mathscr{A}}\in\mathfrak{P}} \omega(\mathscr{A}\circ\mathscr{U}) - \zeta(\mathscr{A}\circ\mathscr{U}) = \sup_{l_{\mathscr{A}\circ\mathscr{U}}\in\mathfrak{P}} \omega(\mathscr{A}) - \zeta(\mathscr{A}) = d(\omega,\zeta). \tag{40}$$

Definition 29 (Perfectly discriminable set of states) *We call a set of states $\{\omega_n\}_{n=1,N}$* perfectly discriminable *if there exists an action $\mathbb{A} = \{\mathscr{A}_j\}_{j=1,N}$ with transformations $\mathscr{A}_j \in l_j$ corresponding to a set of predictable propensities $\{l_n\}_{n=1,N}$ satisfying the relation*

$$l_n(\omega_m) = \delta_{nm}. \tag{41}$$

Definition 30 (Informational dimensionality) *We call the* informational dimension *of the convex set of states* $\mathfrak{S}$*, denoted by* $\mathrm{idim}(\mathfrak{S})$*, the maximal cardinality of perfectly discriminable set of states in* $\mathfrak{S}$*.*

Theorem 3 *Two orthogonal states are perfectly discriminable.*

Proof. If the two states, say ω_1 and ω_2, are orthogonal, then this means that $1 = d(\omega_1, \omega_2) = \sup_{l\in\mathfrak{P}}(l(\omega_1) - l(\omega_2))$, namely there exists a propensity l_1 such that $l_1(\omega_1) = 1$ and $l_1(\omega_2) = 0$. Now, consider the propensity $l_2 = 1 - l_1$, and this will satisfy by definition $l_2(\omega_1) = 0$ and $l_2(\omega_2) = 1$. Now, construct an apparatus with action $\mathbb{A} = \{\mathscr{A}_1, \mathscr{A}_2\}$, with $\mathscr{A}_n \in l_n$, for $n = 1, 2$, and you are done.

Remark 8 Note that it seems that the above theorem doesn't generalize to more than two mutually orthogonal states. In fact, if there are $N > 2$ states that are orthogonal to each other, then we only know that for each of the $\frac{1}{2}N(N-1)$ couples of states, say ζ_1 and ζ_2, there exists a predictable propensity l for which $l(\zeta_1) = 1$ and $l(\zeta_2) = 0$. This does not even guarantee that if a state ω is orthogonal to both ζ_1 and ζ_2, then it should be orthogonal also to any their convex linear combination. In fact, orthogonality implies the existence of two propensities l_1 and l_2 such that $l_1(\omega) = l_2(\omega) = 1$ and $l_1(\zeta_1) = l_2(\zeta_2) = 0$. Now, the distance of ω from the convex combination $\alpha\zeta_2 + (1-\alpha)\zeta_1$ is given by

$$d(\omega, \alpha\zeta_2 + (1-\alpha)\zeta_1) = \sup_{l\in\mathfrak{P}_p}[l(\omega) - \alpha l(\zeta_2) - (1-\alpha)l(\zeta_1)] = \sup_{l\in\mathfrak{P}_p} \alpha[l(\omega) - l(\zeta_2)] + (1-\alpha)[l(\omega) - l(\zeta_1)], \quad (42)$$

which is equal to one if and only if one has both $l(\zeta_2) = l(\zeta_1) = 0$. Therefore, in order to preserve orthogonality for convex combination, we need a functional achieving $l(\omega) = 1$, and for which $l(\zeta) = 0$ for all states $\zeta \perp \omega$: it seems that the existence of such functional is not implied by the existence of many functionals l_ζ, with $l_\zeta(\zeta) = 1$ and $l_\zeta(\omega) = 0$ for all states $\omega \perp \zeta$. Also convex combination of the propensities doesn't help. In fact, consider a linear combination of the propensities $h = \beta l_{\zeta_1} + (1-\beta)l_{\zeta_2}$ on the mixture $\alpha\zeta_1 + (1-\alpha)\zeta_2$. One has $h[\alpha\zeta_1 + (1-\alpha)\zeta_2] = \beta(1-\alpha)l_{\zeta_1}(\zeta_2) + (1-\beta)\alpha l_{\zeta_2}(\zeta_1)$ which we want to vanish for all α, giving the following value for β

$$\beta = \frac{\alpha l_{\zeta_2}(\zeta_1)}{\alpha l_{\zeta_2}(\zeta_1) - (1-\alpha)l_{\zeta_1}(\zeta_2)}, \quad (43)$$

which not necessarily satisfies $0 \leq \beta \leq 1$.

The above considerations lead us to restrict the notion of joint orthogonality as follows

Definition 31 (Joint orthogonality) *We say that a set of states* S *is* jointly orthogonal *to a given state* ω *if each state of their convex hull* $\mathsf{Co}(\mathsf{S})$ *is orthogonal to* ω*.*

Clearly, the definition of joint orthogonality to a state extends to joint orthogonality to a (convex) set of states. We will denote the convex set of states in $\mathfrak{S}$ jointly orthogonal to ω by $\mathfrak{S}^{\perp}_{\omega}$, and the convex set of states in $\mathfrak{S}$ jointly orthogonal to the set S by $\mathfrak{S}^{\perp}_{\mathsf{S}}$.

Definition 31 is also equivalent to

Theorem 4 *A state* ω *is jointly orthogonal to a set of states* S *if and only if there exists a predictable propensity* l *achieving* $l(\omega) = 1$ *and which vanishes identically over the whole set* S*.*

The above theorem also implies the following corollary

Corollary 3 *Any set* $\mathfrak{S}^{\perp}_{\mathsf{S}}$ *is a planar section of* $\mathfrak{S}$*.*

Definition 32 (Discriminating observable) *An observable* $\mathbb{L} = \{l_j\}$ *is* discriminating for $\mathfrak{S}$ *when it discriminates a set of states with cardinality equal to the informational dimension* $\mathrm{idim}(\mathfrak{S})$ *of* $\mathfrak{S}$*.*

Remark 9 It is natural to conjecture that a resolved predictable action (see Definition 24) is the same as a discriminating observable. In fact, by definition, each transformation of a resolved predictable action must be predictable. On the other hand, if it is not resolved, then there will be at least an unresolved transformation, which will occur with probability one for at least two different states. These states could in principle be resolved by another transformation, but there is no guarantee that such transformation exists. Therefore, it is not obvious whether the cardinality of all resolved predictable actions are the same, whence it would coincide with $\mathrm{idim}(\mathfrak{S})$.

Remark 10 (Different dimensionalities for $\mathfrak{S}$) We have introduced three different dimensionalities for the convex set of states $\mathfrak{S}$: 1) the Caratheodory's dimension $\mathrm{cdim}(\mathfrak{S})$; 2) the metrical dimension $\mathrm{mdim}(\mathfrak{S})$; and 3) the informational dimension $\mathrm{idim}(\mathfrak{S})$. In Quantum Mechanics they all coincide. However, in general it seems that there are no definite reasons why they should have the same value. Let's analyze the possible relation between different definitions.

In order to establish a relation between Caratheodory's and metrical dimensionalities, one should first establish if: (a) for any state there always exists a minimal convex decompositions into pure states that are pairwise orthogonal; (b) any convex combination of pairwise orthogonal states is minimal for the resulting mixed state. Clearly, assertion (a) would imply that the maximal rank of a state is smaller than the maximal number of pairwise orthogonal states, namely: $\mathrm{cdim}(\mathfrak{S}) \leq \mathrm{mdim}(\mathfrak{S})$. On the other hand, assertion (b) would imply that $\mathrm{mdim}(\mathfrak{S})$ is the maximal rank of a state, whence the two dimensions coincide, i. e. $\mathrm{mdim}(\mathfrak{S}) = \mathrm{cdim}(\mathfrak{S})$.

As regards a relation between informational and metrical dimensionalities, we have noticed in Remark 8 that pairwise orthogonal states are not necessarily discriminable, whereas, obviously the reverse is true, namely perfectly discriminable states are pairwise orthogonal. Therefore, the maximal number of perfectly discriminable states is bounded by the maximal number of pairwise orthogonal states, whence $\mathrm{idim}(\mathfrak{S}) \leq \mathrm{mdim}(\mathfrak{S})$.

11. LOCAL STATE

Definition 33 (Local state) *In the presence of many independent systems in a joint state Ω, we define the* **local state** *$\omega^{(n)}$ of the n-th system the state that gives the probability for any local transformation $\mathscr{A}$ on the n-th system, with all other systems untouched, namely*

$$\omega^{(n)}(\mathscr{A}) \doteq \Omega(\mathscr{I}, \ldots, \mathscr{I}, \underbrace{\mathscr{A}}_{nth}, \mathscr{I}, \ldots). \tag{44}$$

For example, for two systems only, (which is equivalent to group $n-1$ systems into a single one), we just write $\omega^{(1)}(\mathscr{A}) = \Omega(\mathscr{A}, \mathscr{I})$. Notice that generally the commutativity Rule 9 doesn't imply that the occurrence of a transformation $\mathscr{B}$ on system 2 doesn't change the probability of occurrence of any other transformation $\mathscr{A}$ on system 1, namely, generally

$$\mathscr{A}^{(1)} \circ \mathscr{B}^{(2)} = \mathscr{B}^{(2)} \circ \mathscr{A}^{(1)} \not\Longrightarrow \frac{\Omega(\cdot, \mathscr{B})}{\Omega(\mathscr{I}, \mathscr{B})} = \Omega(\cdot, \mathscr{I}). \tag{45}$$

In other words, the occurrence of the transformation $\mathscr{B}$ on system 2 generally affects the conditioned local state on system 1, namely one has

$$\Omega_{\mathscr{B}^{(2)}}(\cdot, \mathscr{I}) \doteq \frac{\Omega(\cdot, \mathscr{B})}{\Omega(\mathscr{I}, \mathscr{B})} \neq \Omega(\cdot, \mathscr{I}) \equiv \omega^{(1)}. \tag{46}$$

Therefore, in order not to violate the relativity principle, for independent systems (e. g. space-like separated) we need to require explicitly the acausality principle:

Rule 10 (Acausality of local transformations) *Any local action on a system is equivalent to the identity transformation when viewed from an independent system, namely, in terms of states one has*

$$\forall \mathbb{A} \qquad \sum_{\mathscr{A}_j \in \mathbb{A}} \Omega(\cdot, \mathscr{A}_j) = \Omega(\cdot, \mathscr{I}) \equiv \omega^{(1)} \tag{47}$$

The acausality of local transformations Rule 10 along with the existence of inequivalent actions imply the existence of indistinguishable incompatible mixtures.

Corollary 4 (Existence of equivalent incompatible mixtures) *For any two incompatible actions $\mathbb{A} = \{\mathscr{A}_j\}$ and $\mathbb{B} = \{\mathscr{B}_i\}$, the following mixtures are the same state*

$$\sum_j p_j \omega_j = \sum_i p'_i \omega'_i \equiv \omega, \tag{48}$$

where

$$\begin{aligned}
\omega_j &= \frac{\omega(\cdot,\mathscr{A}_j)}{\omega(\mathscr{I},\mathscr{A}_j)}, \quad p_j = \omega(\mathscr{I},\mathscr{A}_j),\\
\omega'_i &= \frac{\omega(\cdot,\mathscr{B}_i)}{\omega(\mathscr{I},\mathscr{B}_i)}, \quad p'_i = \omega(\mathscr{I},\mathscr{B}_i),\\
\omega &\doteq \omega(\cdot,\mathscr{I}).
\end{aligned} \tag{49}$$

Consider now a couple of independent physical systems, say 1 and 2. As we have seen in Eq. (46), a probabilistic transformation $\mathscr{A}$ that occurred on 2 generally affects the local state of 1, which then depends on $\mathscr{A}$ as follows

$$\Omega_{\mathscr{A}^{(2)}}(\cdot,\mathscr{I}) \doteq \frac{\Omega(\cdot,\mathscr{A})}{\Omega(\mathscr{I},\mathscr{A})} = \omega^{(1)}_{\mathscr{A}^{(2)}}. \tag{50}$$

Finally it is worth mentioning that it is possible to define a "maximally entangled state" for a two-partite system on purely operational grounds as follows

Definition 34 (Maximally entangled state) *A maximally entangled state for two identical independent systems is a pure state Ω for which the local state on each system is maximally chaotic, namely*

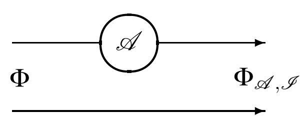
$$\Omega(\cdot,\mathscr{I}) = \Omega(\mathscr{I},\cdot) = \chi(\mathfrak{S}). \tag{51}$$

12. FAITHFUL STATE

Definition 35 (Dynamically faithful state) *We say that a state Φ of a composite system is* dynamically faithful *for the nth component system when acting on it with a transformation $\mathscr{A}$ results in an (unnormalized) conditional state that is in one-to-one correspondence with the dynamical equivalence class $[\mathscr{A}]$ of $\mathscr{A}$, namely the following map is 1-to-1:*

$$\tilde{\Phi}_{\mathscr{I},\ldots,\mathscr{I},\mathscr{A},\mathscr{I},\ldots} \leftrightarrow [\mathscr{A}]_{dyn}, \tag{52}$$

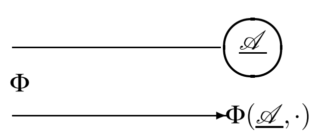
where in the above equation the transformation $\mathscr{A}$ acts locally only on the nth component system.



Definition 36 (Informationally faithful state) *We say that a state Φ of a composite system is* informationally faithful *for the nth component system when acting on it with a transformation $\mathscr{A}$ results in an (unnormalized) conditional local state on the remaining systems that is in one-to-one correspondence with the informational equivalence class $\underline{\mathscr{A}}$ of $\mathscr{A}$ (i. e. its propensity), namely the following map is 1-to-1:*

$$\Phi(\cdots,\mathscr{A},\cdots) \leftrightarrow \underline{\mathscr{A}}, \tag{53}$$

where in the above equation the transformation $\mathscr{A}$ acts locally only on the nth component system.



In the following for simplicity we restrict attention to two component systems, and take the first one for the nth. Using the definition 10 of conditional state, we see that the state Φ is dynamically faithful when the map $\Phi(\cdot\circ[\mathscr{A}]_{dyn},\mathscr{I})$ is invertible over the set of dynamical equivalence classes of transformations, namely when

$$\forall\mathscr{A},\ \Phi(\mathscr{B}_1\circ\mathscr{A},\mathscr{I}) = \Phi(\mathscr{B}_2\circ\mathscr{A},\mathscr{I}) \quad\Longleftrightarrow\quad \mathscr{B}_1 \in [\mathscr{B}_2]_{dyn}. \tag{54}$$

On the other hand, one can see that the state Φ is informationally faithful when the map $\Phi(\underline{\mathscr{A}}, \cdot)$ is invertible over the set of informationally equivalence classes of transformations, namely when

$$\forall \mathscr{A}, \; \Phi(\mathscr{B}_1, \mathscr{A}) = \Phi(\mathscr{B}_2, \mathscr{A}) \quad \Longleftrightarrow \quad \mathscr{B}_1 \in \underline{\mathscr{B}_2}. \tag{55}$$

Definition 37 (Preparationally faithful state) *We will call a state Φ of a bipartite system* preparationally faithful *if all states of one component can be achieved by a suitable local transformation of the other, namely for every state ω of the first party there exists a local transformation $\mathscr{T}_\omega$ of the other party for which the conditioned local state coincides with ω, namely*

$$\forall \omega \in \mathfrak{S} \qquad \exists \mathscr{T}_\omega : \qquad \frac{\Phi(\mathscr{T}_\omega, \cdot)}{\Phi(\mathscr{T}_\omega, \mathscr{I})} \equiv \omega. \tag{56}$$

13. IN SEARCH FOR AN OPERATIONAL AXIOM

In the following I list some possible candidates of operational axioms from which to derive the quantum superposition principle, namely from which we should be able to determine if a convex set of states is quantum. We will call a convex set of states $\mathfrak{S}$ *complete quantum convex of states* (CQCS or complete QCS) when it coincides with a complete convex set of quantum states on a given Hilbert space. For example, the Bloch sphere is a CQCS, whereas the unit disk is a QCS, but not a CQCS. For $n > 3$ the n-dimensional hypersphere is not a QCS. Similarly, a tetrahedron is a QCS, but not a CQCS. Notice that the metric is relevant, i. e. an ellipsoid is not equivalent to the Bloch sphere, since the antipodal states do not have fixed unit distance.

Clearly deriving completeness in terms of an "operational consistency" is the difficult part of the problem, and indeed assuming a kind of completeness for transformations could be just a restatement of the superposition principle. Following Hardy[13] we could at most assume that (a) for any state $\omega \in \mathfrak{S}$ of a CQCS $\mathfrak{S}$, the convex set $\mathfrak{S}_\omega^\perp$ is a CQCS too, and (b) all pure states in $\mathfrak{S}$ are connected by an isometric indecomposable transformation, and these form a continuous group. This, however, leaves out the main problem of deriving the tensor product structure for independent systems. One would be tempted to use the following easy axiom

Conjecture 1 (Existence of maximally entangled states) *A convex set of bipartite states $\mathfrak{S}^{\times 2}$ is a QCS if there exist maximally entangled states according to Definition 34.*

However, this is not of a truly operational nature. An operational axiom could be a calibrability axiom of the kind

Conjecture 2 (Dynamic calibrability) *For any bipartite system there exists a pure joint state that is dynamically faithful for one of the two systems.*

We also conjecture that as a consequence such state is also informationally faithful and preparationally faithful, or else

Conjecture 3 (Informational calibrability) *For any bipartite system there exists a pure joint state that is informationally faithful for one of the two systems.*

On the other hand, a preparability axiom could be

Conjecture 4 (Preparability) *For any bipartite system there exists a pure joint state that is preparationally faithful for one of the two systems.*

From one the above calibrability/preparability conjectures the aim would be to prove something as follows

Conjecture 5 (Dimensionality of composite systems) *The informational dimensionality of a composite system is the product of their informational dimensionalities.*

This should follow via the equivalence of the dimensionality of the convex cone of transformations/propensities and that of unnormalized states.

Another assertion that is certainly true in the quantum case is

Conjecture 6 (Informationally complete discriminating observables) *On any bipartite system there exists a discriminating observable that is informationally complete for one of the components for almost all preparations of the other component.*

The above discriminating observable are just the so-called *Bell measurements*. Another candidate for an operational axiom could be the possibility of achieving teleportation of states

Conjecture 7 (Teleportation) *There exist a joint bipartite state* Φ*, a joint bipartite (discriminating) observable* $\mathbb{L} = \{l_j\}$ *and a set of deterministic indecomposable transformations* $\{\mathscr{U}_j\}$ *by which one can teleport all states as follows*

$$\frac{\omega^{(1)}\Phi^{(2,3)}(l_j^{(1,2)}, \cdot\mathscr{U}_j^{(3)})}{\omega^{(1)}\Phi^{(2,3)}(l_j^{(1,2)}, \mathscr{I})} = \omega^{(3)}. \qquad (57)$$

Conjecture 8 (Preparability of transformations) *It is possible to achieve (probabilistically) any dynamical equivalence class of transformations using only a fixed action* $\mathbb{A} = \{\mathscr{A}^{(1,2)}, \ldots\}$ *for a fixed outcome and a fixed partite state* Φ*, as follows*

$$\exists \mathbb{A} = \{\mathscr{A}^{(1,2)}, \ldots\}: \qquad \frac{\omega^{(1)}\Phi_{\mathscr{B}}^{(2,3)}(\mathscr{A}^{(1,2)}, \cdot)}{\omega^{(1)}\Phi_{\mathscr{B}}^{(2,3)}(\mathscr{A}^{(1,2)}, \mathscr{I})} = \omega_{\mathscr{B}}^{(3)}. \qquad (58)$$

As working hypothesis I would like to consider the following combined axioms

Conjecture 9 (The minimal "lab") *On any bipartite system there exists:*

a) a discriminating observable that is informationally complete for one of the components for almost all preparations of the other component.

b) a pure joint state which, for the same component system in (a) is: dynamically, informationally, and preparationally faithful.

Another working hypothesis could be that obtained by combining Conjectures 7 and 8, but I think that Conjecture 9 represents the axiom of the most genuine operational/epistemic nature.

ACKNOWLEDGMENTS

Support from the Italian Minister of University and Research (MIUR) is acknowledged under program Prin 2003. This work has been possible during my last summer visits at Northwestern University, thanks to the kind ospitality of prof. H. P. Yuen. I thank P. Perinotti, G. Chiribella, C. Fuchs, K. Svozil, and G. Jaeger for interesting discussions on the first version of the present manuscript, in particular P. Perinotti, G. Chiribella for a critical analysis, and G. Jaeger for a careful reading. Definition 7 is due to P. Perinotti.

REFERENCES

1. G. Ludwig, *An Axiomatic Basis for Quantum Mechanics I: Derivation of Hilbert Space Structure*, Springer, SPR:adr, 1985.
2. U. Krause, *The inner orthonormality of convex sets in axiomatic quantum mechanics*, pp. 269–280, vol. 29 of [5] (1974).
3. H. Neumann, *The structure of ordered Banach spaces in Axiomatic Quantum Mechanics*, pp. 116–121, vol. 29 of [5] (1974).
4. E. Størmer, *Positive Linear Maps of C^*-algebras*, pp. 85–106, vol. 29 of [5] (1974).
5. A. Hartkämper, and H. Neumann, editors, *Foundations of Quantum Mechanics and Ordered Linear Spaces*, vol. 29, Springer, SPR:adr, 1974.
6. G. M. D'Ariano, *Tomographic methods for universal estimation in quantum optics*, IOS Press, Amsterdam, 2002, pp. 385–406, scuola "E. Fermi" on *Experimental Quantum Computation and Information.*
7. G. M. D'Ariano, P. L. Presti, and L. Maccone, **93**, 250407 (2004).
8. G. M. D'Ariano, and P. L. Presti, **86**, 4195 (2001).
9. M.-D. Choi, *Linear Algebra Appl.*, **10**, 285–290 (1975).
10. A. Jamiolkowski, *Rep. Math. Phys.*, **3**, 275 (1972).
11. K. Kraus, *Operations and Effects in the Hilbert Space Formulation of Quantum Theory*, pp. 206–229, vol. 29 of [5] (1974).
12. P. Zanardi, **87**, 077901 (2001).
13. L. Hardy, *LANL arXive eprint*, pp. 1–34 (2001), `quant-ph/0101012`.