

Informational axioms for quantum theory

G. Chiribella*, G. M. D'Ariano[†] and Paolo Perinotti[†]

*Perimeter Institute, 31 Caroline Street North, Waterloo, Ontario N2L2Y5, Canada

[†]QUIT Group, Dipartimento di Fisica "A. Volta", via Bassi 6, 27100 Pavia, Italy

Abstract. It was recently proved that quantum theory can be derived from six axioms about information processing. Here we review these axioms, discussing various facets of their information-theoretical nature, and illustrating the general picture of quantum physics that emerges from them.

Keywords: Quantum Theory, Axiomatization, Reconstruction, Information Theory, Information Processing, Causality, Purification

PACS: 03.65.-w, 03.65.Aa, 03.65.Ta, 03.67.-a

INTRODUCTION

The questions about foundations of Quantum Theory raised laborious research programs and passionate debates for almost a century. All the solutions that were proposed to the main questions opened by the advent of Quantum Physics were only partially satisfactory, mixing in different proportions interpretational changes or technical modifications of the basic elegant formalism exposed by von Neumann [1, 2], which still remains the most convincing presentation of the theory, and the one that is taught to students. In a simplified picture of the state of the art, the attempts at giving a natural and convincing look to Quantum Theory can be classified in the following two families: The first one includes all approaches retaining the reality of microscopic physical systems, starting from the idea that matter cannot be made but with smaller pieces of matter. The main concern for such attempts is to solve the measurement problem [3, 4, 5]. The second class includes those approaches that are more oriented to accepting the allegedly paradoxical features of quantum systems as a matter of fact, and to looking at the macroscopic reality as an emergent phenomenology based on radically different laws governing the microscopic scale. The proposals in this class—maybe less ambitiously than the former ones—generally seek a non-technical, physically motivated way of introducing the abstract formalism of Hilbert spaces or C*-algebras, without modifying it or changing its physical interpretation [6, 7, 8, 9]. This line of research was recently revived by the influence of Quantum Information Theory [10, 11], which provided an astonishing picture of the information processing capabilities of quantum systems. Several attempts at an information-theoretical axiomatization have been proposed in recent years [12, 13, 14], sharing the crucial feature of promoting information processing capabilities to the role of axioms for a *reconstruction* of quantum theory [15, 16].

In the present paper we will provide a solution to the main problem in the second class of approaches, namely a formulation of Quantum Theory that does not rely on mathematical axioms, but on simple general principles that can be formulated using elementary notions. Our principles perfectly fit the recent trend pioneered by J. A.

Wheeler [17] of looking at Quantum Theory—and more generally to physical theories—as rooted in information-theoretical notions. As a consequence of this formulation, the mathematical structure of finite dimensional Hilbert spaces can be thoroughly derived, as discussed in Ref. [18]. The key axiom in the present formulation is the *purification postulate*, whose consequences were discussed in detail in a very general context in Ref. [19], where the principle was introduced for the first time. This result brings to conclusion G. M. D’Ariano’s axiomatization program, whose progresses were reported in the proceedings of the Växjö conference for many years [20, 21].

THE OPERATIONAL AND PROBABILISTIC FRAMEWORK

The basic notions that we need in order to formulate quantum theory are borrowed from the operational language. Here is the starting glossary [19] on which we build the language in which the axioms are stated.

- *Test and event*: a test is a procedure that the observer can choose to follow at a given step in an experiment. A test has many outcomes, corresponding to many different events, but generally their occurrence is not under the control of the observer.
- *Outcome*: an outcome is directly registered by an observer as the result of following a procedure. The outcome is interpreted as information heralding the occurrence of the corresponding event.
- *System*: tests can be operated in a sequence according to precise rules. The rules are stated through the specification of an *input* and *output* label for every test, and by the simple prescription that two tests can be composed in a sequence whenever the output label of the first test coincides with the input label of the second one. Systems are identified with such labels.

From the description of our elementary concepts, one can figure out the main traits of the picture of physics resulting from our formulation of Quantum Theory: physical concepts are related to experimental procedures where an observer performs tests on the system. The observer registers outcomes, that are bits of information. Systems are information carriers, and it is only because of our habit that we think of them as *particles*—little pieces of matter that often behave like waves. The only inference we are allowed to do conditionally on the occurrence of an outcome, is the event—a transformation of the observed system.

Quantum Theory is then derived as a theory about information processing. More precisely, we will specify in the following a more detailed language, and use it to formulate six principles, in the form of rules about the ways in which we can process information through tests. These principles are sufficient to single out the mathematical structure of Quantum Theory.

The operational language

We will denote an outcome in a space X of possible outcomes by a lower case index, as $i \in X$, and the corresponding event as \mathcal{E}_i . A test is identified with the collection of its events and denoted as $\{\mathcal{E}_i\}_{i \in X}$. Tests, systems and events have a graphical representation as follows

$$\begin{array}{c} \text{A} \quad \text{B} \\ \text{---} \boxed{\{\mathcal{E}_i\}_{i \in X}} \text{---} \end{array}$$

where A is the input system and B is the output. The test $\{\mathcal{E}_i\}_{i \in X}$ (the event \mathcal{E}_i) is of type $A \rightarrow B$. System types A, B, \dots include the *trivial* system denoted by I , which is the input system of every *preparation-test* and the output system of any *observation-test*, corresponding to tests where the observer neglects previous or subsequent processing. These particular tests are denoted as follows

$$\boxed{\{\rho_i\}_{i \in X}} \text{---}^{\text{A}} := \text{---}^{\text{I}} \boxed{\{\rho_i\}_{i \in X}} \text{---}^{\text{A}}, \quad \text{---}^{\text{A}} \boxed{\{a_i\}_{i \in X}} \text{---} := \text{---}^{\text{A}} \boxed{\{a_i\}_{i \in X}} \text{---}^{\text{I}}$$

Tests can be composed in cascades through the *sequential composition*, an operation that allows one to define a new test $\{\mathcal{F}_j\}_{j \in Y} \circ \{\mathcal{E}_i\}_{i \in X}$ of type $A \rightarrow C$, with outcome space $X \times Y$, denoted as $\{\mathcal{F}_j \circ \mathcal{E}_i\}_{(i,j) \in X \times Y}$, from any couple of tests $\{\mathcal{E}_i\}_{i \in X}$ of type $A \rightarrow B$ and $\{\mathcal{F}_j\}_{j \in Y}$ of type $B \rightarrow C$, namely whenever the output system of the first test coincides with the input system of the second one.

For every system type X there is a unique singleton test $\{\mathcal{I}_X\}$ called *identity test* such that the *identity event* \mathcal{I}_X satisfies $\mathcal{I}_B \circ \mathcal{E} = \mathcal{E} \circ \mathcal{I}_A$ for every event \mathcal{E} of type $A \rightarrow B$.

Two arbitrary systems A and B can form a composite system, which is represented by the *parallel composition*, and denoted by $A \otimes B$, with the following properties

$$A \otimes B = B \otimes A, \quad A \otimes I = I \otimes A = A. \quad (1)$$

Arbitrary tests $\{\mathcal{E}_i\}_{i \in X}$ of type $A \rightarrow B$ and $\{\mathcal{F}_j\}_{j \in Y}$ of type $C \rightarrow D$ can be run in parallel obtaining the *parallel composition* $\{\mathcal{E}_i\}_{i \in X} \otimes \{\mathcal{F}_j\}_{j \in Y} = \{\mathcal{E}_i \otimes \mathcal{F}_j\}_{(i,j) \in X \times Y}$ of type $(A \otimes B) \rightarrow (C \otimes D)$.

The parallel composition enjoys the following property

$$(\mathcal{C} \otimes \mathcal{D}) \circ (\mathcal{A} \otimes \mathcal{B}) = (\mathcal{C} \circ \mathcal{A}) \otimes (\mathcal{D} \circ \mathcal{B}), \quad (2)$$

and, as a consequence, $(\mathcal{A} \otimes \mathcal{I}) \circ (\mathcal{I} \otimes \mathcal{B}) = \mathcal{A} \otimes \mathcal{B} = (\mathcal{I} \otimes \mathcal{B}) \circ (\mathcal{A} \otimes \mathcal{I})$.

These rules—that allow one to build arbitrary circuits—are borrowed from quantum circuits, but are also the faithful transcription of the axioms for a strict symmetric monoidal category. This structure is the core of the categorical approach to quantum foundations, recently developed by B. Coecke and collaborators (see e.g. [22]), that proved itself very fruitful in exploring many conceptual features of quantum theory and making it more intuitive through pictures that are very similar to the circuitual diagrams used in our approach. This fact anticipates one of the key messages that come from our approach: Many theorems of quantum information theory actually only depend on very general facts about the structure of quantum theory, and do not require the details of the C^* -algebraic (or Hilbert space) formalism.

The outcomes of a test can be merged into a single outcome, representing the operation of neglecting some information. On formal grounds, this corresponds to the following rule. For every test $\{\mathcal{E}_i\}_{i \in X}$ of type $A \rightarrow B$, and every disjoint partition $\{X_j\}_{j \in X'}$ of X there exists a test $\{\mathcal{E}'_j\}_{j \in X'}$ of type $A \rightarrow B$, called *coarse-graining* of the test $\{\mathcal{E}_i\}_{i \in X}$, such that for two tests $\{\mathcal{E}_i\}_{i \in X}$ and $\{\mathcal{F}_j\}_{j \in Y}$ the coarse grainings of tests $\{\mathcal{E}_i \circ \mathcal{F}_j\}_{(i,j) \in X \times Y}$ and $\{\mathcal{E}_i \otimes \mathcal{F}_j\}_{(i,j) \in X \times Y}$ corresponding to partitions $\{X_h \times Y_k\}_{(h,k) \in X' \times Y'}$ of $X \times Y$ coincide with the coarse grainings with the tests $\{\mathcal{E}'_h \circ \mathcal{F}'_k\}_{(h,k) \in X' \times Y'}$ and $\{\mathcal{E}'_h \otimes \mathcal{F}'_k\}_{(h,k) \in X' \times Y'}$. In words, the operation of coarse-graining commutes with both parallel and sequential composition. If the test $\{\mathcal{E}'_j\}_{j \in X'}$ is a coarse graining for the test $\{\mathcal{E}_i\}_{i \in X}$, equivalently we say that the test $\{\mathcal{E}_i\}_{i \in X}$ is a *refinement* for the tests $\{\mathcal{E}'_j\}_{j \in X'}$. We say that a test $\{\mathcal{E}_i\}_{i \in X}$ is deterministic if $|X| = 1$. A test is deterministic if and only if it is the coarse graining for a test $\{\mathcal{E}_i\}_{i \in X}$ corresponding to the trivial partition $\{X\}$.

The probabilistic structure

Up to now we did not introduce any probabilistic structure in our language. The whole circuitual diagrammatic apparatus just provides a formal translation of the operational grammar ruling the construction of experiments. However, the very notion of test that we provide—a collection of events—naturally suggests a picture where the events are mutually exclusive possibilities, whose occurrence provides information about the system. We then need a rule to determine the probabilities of events in order to fully specify the theory. The necessary structure to this purpose is provided through the requirement that a test for the trivial system I is a probability distribution: $\{\mathcal{E}_i\}_{i \in X} = \{p_i\}_{i \in X}$ with $p_i \geq 0$, $\mathcal{I}_1 = 1$. Parallel and sequential composition coincide, and their events are provided by the product rule $p_i \circ p_j = p_i \otimes p_j = p_i p_j$. Finally, the coarse graining for a test of type $I \rightarrow I$ is provided by the sum: $p'_j = \sum_{i \in Y_j} p_i$. These assumptions provide the interpretation of tests as collections of mutually exclusive probabilistic events, whose probability is determined only within a circuit of type $I \rightarrow I$.

The circuits become thus the tool to propagate the initial information, which is contained in the specification of the preparation-test. What is evolved from the input to the output of a circuit is thus information, rather than matter. In particular, systems carry information and allow one to evaluate probabilities of events provided an initial amount of information. The physical world emerges as a consequence of correlations between events. The basic theory of physics is thus a theory of information processing.

Before stating the axioms, it is worth noticing that from the assumptions about probabilities we can conclude that preparation-events and observation-events are real-valued functionals on each other. It is then natural to identify all those events that correspond to the same functionals, because they are operationally indistinguishable. We define the corresponding equivalence classes for system A as *states* of A ($\text{St}(A)$) and *effects* of A ($\text{Eff}(A)$), respectively. This simple fact immediately provides the sets of states and effects with a vector space structure, denoted by $\text{St}_{\mathbb{R}}(A)$ and $\text{Eff}_{\mathbb{R}}(A)$, respectively: Real-valued functionals can be multiplied by a real scalar and can be summed to define new real-valued functionals. To the system type A then corresponds an integer number D_A , which is the dimension of the vector spaces $\text{St}_{\mathbb{R}}(A)$ and $\text{Eff}_{\mathbb{R}}(A)$.

The states $\text{St}(A)$ and the effects $\text{Eff}(A)$ define the cones $\text{St}_+(A)$ and $\text{Eff}_+(A)$, that are obtained multiplying the functionals corresponding to physical events by non negative scalars. Clearly, $\text{Eff}_+(A) \subseteq \text{St}_+(A)^*$ and $\text{St}_+(A) \subseteq \text{Eff}_+(A)^*$. Even though the axioms that we will state in Sec. are satisfied by infinite-dimensional quantum systems as well, the derivation of Ref. [18] exploits the assumption that the spaces $\text{St}_{\mathbb{R}}(A) \simeq \text{Eff}_{\mathbb{R}}(A)$ are finite-dimensional. The generalization to the infinite-dimensional case is in progress.

Similarly to the case of states and effects, it is easy to check that every event \mathcal{E} of type $A \rightarrow B$ induces a well-defined linear map from $\text{St}(A)$ to $\text{St}(B)$. However, in order to characterize the event \mathcal{E} one needs the entire collection of linear maps corresponding to $\mathcal{E} \otimes \mathcal{I}_B$ for every system B . Events can thus be collected in equivalence classes defined by the following relation: $\mathcal{E} \sim \mathcal{F}$ if and only if the linear map corresponding to $\mathcal{E} \otimes \mathcal{I}_C$ coincides with the one corresponding to $\mathcal{F} \otimes \mathcal{I}_C$ for all systems C . We will call the equivalence classes *transformations*. Notice that in general two events can belong to different transformations even if $\mathcal{E} \circ \rho = \mathcal{F} \circ \rho$ for every $\rho \in \text{St}(A)$: this is because one can still have $(\mathcal{E} \otimes \mathcal{I}_C) \circ \sigma \neq (\mathcal{F} \otimes \mathcal{I}_C) \circ \sigma$ for some bipartite state $\sigma \in \text{St}(AC)$

As a consequence of the linear structure of transformations, the coarse graining $\{\mathcal{E}'_j\}_{j \in Y}$ of a test $\{\mathcal{E}_i\}_{i \in X}$ corresponding to the partition $\{X_j\}_{j \in Y}$ is provided by $\mathcal{E}'_j = \sum_{i \in X_j} \mathcal{E}_i$. A refinement of a transformation \mathcal{E} is a subset $\{\mathcal{E}_i\}_{i \in X_j}$ of a test $\{\mathcal{E}_i\}_{i \in X}$ such that $\mathcal{E} = \sum_{i \in X_j} \mathcal{E}_i$. If the transformation \mathcal{F} belongs to one refinement of the transformation \mathcal{E} , we say that \mathcal{F} *refines* \mathcal{E} , and write $\mathcal{F} \prec \mathcal{E}$. The set $R_{\mathcal{E}} := \{\mathcal{D} \mid \mathcal{D} \prec \mathcal{E}\}$ is called *refinement set* of \mathcal{E} . A transformation \mathcal{E} is *atomic* if $\mathcal{F} \prec \mathcal{E}$ implies $\mathcal{F} = \lambda \mathcal{E}$ for some $0 \leq \lambda \leq 1$. Atomic states are called *pure*. All states that are not pure are called *mixed*. A state ρ of system A is *completely mixed* if R_{ρ} is complete in $\text{St}_{\mathbb{R}}(A)$.

A couple of deterministic states ρ_0 and ρ_1 is *perfectly discriminable* if there exists a binary observation-test $\{a_0, a_1\}$ such that $a_i \circ \rho_j = \delta_{i,j}$.

A transformation \mathcal{E} corresponding to a deterministic test is called *channel*. A transformation \mathcal{E} of type $A \rightarrow B$ along with a channel \mathcal{D} of type $B \rightarrow A$ with $D_B \leq D_A$ provide a *lossless compression scheme* for the preparation test $\{\rho_i\}_{i \in X} \subseteq \text{St}(A)$ if $\mathcal{D} \circ \mathcal{E} \circ \rho_i = \rho_i$ for every $i \in X$. A lossless compression scheme for ρ is a compression scheme $(B, \mathcal{E}, \mathcal{D})$ that is lossless for every preparation test $\{\rho_i\}_{i \in X}$ with $\sum_i \rho_i = \rho$. A compression scheme for ρ is *efficient* if for every state $\sigma \in \text{St}(C)$ there exists a state $\rho' \in R_{\rho}$ such that $\sigma = \mathcal{E} \circ \rho'$. An *ideal* compression scheme for ρ is a compression scheme that is both lossless and efficient.

Finally, in order to state the axioms we need the notion of *reversible transformation*, that is any transformation \mathcal{U} of type $A \rightarrow B$ for which there exists \mathcal{U}^{-1} of type $B \rightarrow A$ such that $\mathcal{U}^{-1} \circ \mathcal{U} = \mathcal{I}_A$, and $\mathcal{U} \circ \mathcal{U}^{-1} = \mathcal{I}_B$.

THE AXIOMS

We are now ready to state the principles. Every principle is a rule about possibility or impossibility of performing a particular information processing task. Here is the full list.

- A1 Causality.** It is not possible to change the probability of a preparation event by choosing an observation-test.

- A2 **Local discriminability.** If two deterministic states Ψ and Ψ' in $\text{St}(A \otimes B)$ are different, there exists a local observation-test on which they give different probabilities.
- A3 **Atomic composition.** The sequential composition of atomic transformations is an atomic transformation.
- A4 **Perfect distinguishability.** Every state that is not maximally mixed can be perfectly distinguished from some other state.
- A5 **Compression.** For every state ρ there exists an ideal compression scheme.
- P6 **Purification.** For every state $\rho \in \text{St}(A)$ there exists a purifying system B and pure state $\Psi \in \text{St}(A \otimes B)$ such that it is possible to simulate the preparation ρ by composing the preparation Ψ with a deterministic effect on B . For a fixed purifying system B , all purifications of the same state ρ differ by a reversible transformation on B .

The first five principles are called axioms. We can think of a standard set of theories that satisfy them, including classical information theory. The sixth principle has a different status, and we call it postulate: Among the standard information processing theories, purification is the principle that singles out Quantum Theory.

Causality

Causality in science is usually related to the cause-and-effect principle, and is often naively misinterpreted as a restatement of determinism. From this point of view, it can be surprising that we take causality as the first axiom of Quantum Theory, which is popularly known as the theory of the uncertainty principle. However, in our precise formulation, causality only implies that communication in an operational probabilistic theory cannot occur from the output to the input. Indeed, the statement of the principle implies that in an experiment like the following

$$p(i, j) := \left(\rho_i \right) \xrightarrow{A} \left(a_j \right)$$

the choice of the observation-test $\{a_j\}_{i \in Y}$ cannot influence the probability distribution of outcomes $p_a(i) := \sum_{j \in Y} p(i, j)$ in the preparation-test. If this was the case, the observer performing the preparation-test could infer a non-null information from occurrence of one outcome, and this would provide a communication scheme from the observer performing the observation-test to the one performing the preparation-test. The impossibility of performing this task implies that the input-output relation reflects the direction of information flow within an operational circuit. This fact in turn makes sense of the interpretation of input and output as time-ordered systems. In order to better illustrate this point, it is useful to introduce the following equivalent condition for causality.

Theorem 1 *A theory is causal if and only if the deterministic effect e_A for a system type A is unique.*

Using this result, we can exhibit a theory that is not causal. For this purpose, we need to introduce the Choi correspondence between bipartite quantum states and linear maps from system A to system B . Reminding that states of A are positive $d_A \times d_A$ matrices

with $\text{Tr}[\sigma] \leq 1$, and $\text{St}_{\mathbb{R}}(\mathbf{A})$ are Hermitian $d_{\mathbf{A}} \times d_{\mathbf{A}}$ matrices, one can prove that linear maps \mathcal{M} of type $\mathbf{A} \rightarrow \mathbf{B}$ are in correspondence with Hermitian $d_{\mathbf{A}}d_{\mathbf{B}} \times d_{\mathbf{A}}d_{\mathbf{B}}$ matrices M as follows

$$M := \mathcal{M} \otimes \mathcal{I}_{\mathbf{A}}(|\Omega\rangle\langle\Omega|), \quad \mathcal{M}(\rho) := \text{Tr}_{\mathbf{A}}[(I_{\mathbf{B}} \otimes \rho^T)M], \quad (3)$$

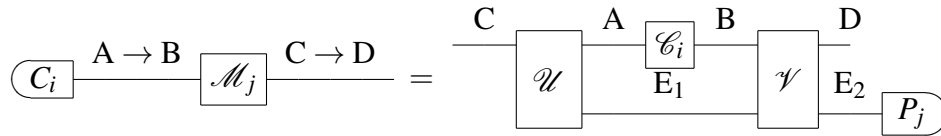
where $|\Omega\rangle := \sum_{i=1}^{d_{\mathbf{A}}} |i\rangle|i\rangle$ and X^T denotes the transpose of X in the basis $\{|i\rangle\}_{1 \leq i \leq d_{\mathbf{A}}}$. Transformations in Quantum Theory are represented by positive definite Choi matrices E satisfying $\text{Tr}_{\mathbf{B}}[E] \leq I_{\mathbf{A}}$, while channels are characterized by the identity $\text{Tr}_{\mathbf{B}}[C] = I_{\mathbf{A}}$.

Consider now the theory whose system types are $\mathbf{A} \rightarrow \mathbf{B}$, states of type $\mathbf{A} \rightarrow \mathbf{B}$ are Choi representatives of quantum operations from \mathbf{A} to \mathbf{B} , and transformations of type $(\mathbf{A} \rightarrow \mathbf{B}) \rightarrow (\mathbf{C} \rightarrow \mathbf{D})$ are completely positive maps on the Choi matrices such that whenever C represents a $\mathbf{A} \rightarrow \mathbf{B}$ channel, then $\mathcal{M}(C)$ represents a $\mathbf{C} \rightarrow \mathbf{D}$ channel. In Ref. [23] it was proved that the most general map of this kind is represented by a positive $d_{\mathbf{A}}d_{\mathbf{B}}d_{\mathbf{C}}d_{\mathbf{D}} \times d_{\mathbf{A}}d_{\mathbf{B}}d_{\mathbf{C}}d_{\mathbf{D}}$ matrix $E \leq T$ for some matrix T satisfying

$$\text{Tr}_{\mathbf{D}}[T] = I_{\mathbf{B}} \otimes T', \quad \text{Tr}_{\mathbf{A}}[T'] = I_{\mathbf{C}}. \quad (4)$$

In particular, effects in such a theory are all operators $0 \leq E \leq I_{\mathbf{B}} \otimes \rho$ for some normalized state ρ . The deterministic effect is not unique, since for every state $\rho \in \text{St}(\mathbf{A})$ the effect $I_{\mathbf{B}} \otimes \rho$ is deterministic.

The violation of causality reflects the impossibility of identifying the input-output direction with a definite time direction, because information flows both from input to the output and backwards. Indeed, we have the following picture of the information flow thanks to the underlying structure of Quantum Theory coming from Eq. (4) (see [23])



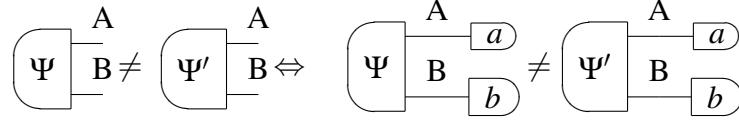
A remarkable consequence of the causality axiom is *no-signalling without interaction*, that can be viewed as the operational probabilistic analogue of Einstein's causality. The precise formulation of this principle is the following. It is impossible to influence the probability distribution of outcomes of an observation-test on system \mathbf{A} by choosing any preparation-test on system \mathbf{B} , unless systems \mathbf{A} and \mathbf{B} interact via a bipartite channel.

Indeed, the input-to-output structure of events induces a causal structure in every circuit, in which one can find the notions of light-cone, of space-like separated events, of time-like separated events. However it is only thanks to the causality axiom that communication between every two space-like separated events is forbidden. Moreover, the causal relation between any two events is invariant under a change of foliation on the circuit, corresponding to a change of reference frame.

Finally, we would like to remark that while in the general framework probabilities of events are defined only within a circuit of type $\mathbf{I} \rightarrow \mathbf{I}$, in causal theories events have probabilities also within circuits of type $\mathbf{I} \rightarrow \mathbf{A}$ for arbitrary \mathbf{A} . The transformations process probability distributions, and in the presence of an outcome we have the transition from a prior to a conditional probability. The input-output processing can thus be identified with the information flow, which is not the case in non-causal theories. Moreover, thanks to causality it is possible to identify the input-output orientation with the time arrow.

Local discriminability

The second principle that we impose is local discriminability. The axiom can be pictorially represented as follows



for some choice of local effects $a \in \text{Eff}(A)$ and $b \in \text{Eff}(B)$.

Local discriminability implies that even if a state is entangled—namely it does not correspond to a mixture of local preparations—the information it contains can be extracted by local measurements. It is due to local discriminability that $\{a \otimes b\}$ is a spanning set for bipartite effects $\text{Eff}_{\mathbb{R}}(A \otimes B)$, and thus $D_{A \otimes B} = D_A D_B$. Finally, by simple dimensionality check, this implies that also local states $\{\rho \otimes \sigma\}$ are a spanning set for $\text{St}_{\mathbb{R}}(A \otimes B)$. Then, as a consequence of local discriminability we can write any effect $A \in \text{Eff}(A \otimes B)$ and any state $\Sigma \in \text{St}(A \otimes B)$ as follows

$$A = \sum_{i,j} A_{ij}(a_i \otimes b_j), \quad \Sigma = \sum_{ij} \Sigma_{ij}(\rho_i \otimes \nu_j), \quad (5)$$

for some real-valued matrices (A_{ij}) and (Σ_{ij}) . The role of local discriminability is crucial also for the following consequence.

Theorem 2 *In a theory with local discriminability, transformations of type $A \rightarrow B$ are fully characterized by the linear map $\mathcal{E} : \text{St}_{\mathbb{R}}(A) \rightarrow \text{St}_{\mathbb{R}}(B)$.*

This result further highlights the importance of local discriminability, because it is the principle that allows one to characterize a transformation by a local test, without the need of checking it on every type of bipartite system. In a sense, local discrimination is the ingredient that makes the characterization of transformations practically meaningful.

Atomic composition

The axiom of atomic composition implies the principle of *pure conditioning* [24], namely the local composition of a pure bipartite state $\Psi \in \text{St}(A \otimes B)$ with an atomic effect $b \in \text{Eff}(B)$ produces a pure state $b \circ \Psi \in \text{St}(A)$. Atomic composition implies that if we have maximal knowledge of some events $\mathcal{A}_1, \dots, \mathcal{A}_k$ (i.e. the events $\mathcal{A}_1, \dots, \mathcal{A}_k$ are unrefinable), then we have maximal knowledge also of the history consisting of their sequence, namely the event $\mathcal{A}_k \circ \dots \circ \mathcal{A}_1$.

Perfect distinguishability

This axiom, like the causality axiom, makes a quantitative statement about probabilities. Combining perfect distinguishability with the other principles one can prove that

all the probabilities in the theory will be derived from the probabilities of the discriminating tests for non-internal states. More precisely, all effects can be written as linear combinations of perfectly discriminating effects for non-internal states.

Notice that without perfect distinguishability there would be no principle granting the possibility to encode classical bits on physical systems, and then the theory would be all in all inconsistent with the elementary notion of classical information. Moreover, in a fully probabilistic setting, the perfect distinguishability axiom recovers the logic of propositions. Non-internal states encode truth values for the propositions, that can be identified with the corresponding perfectly discriminating effects. However, as was pointed out already in Birkhoff's and von Neumann's paper [6], the structure of propositions emerging from our construction will not be classical logic, but the lattice of projections of quantum logic. Actually, the lattice of projections plays an important role in the derivation of Quantum Theory.

Compression

In classical information theory, compression is a crucial notion, that through Shannon's first theorem provides an operational interpretation of the Shannon entropy. The compression in Shannon's Theorem has a quantum counterpart in Schumacher's compression theorem, which regards the asymptotic regime where a vanishingly small decoding error is allowed and provides an operational interpretation of the von Neumann entropy. The notion of entropy in operational probabilistic theories raises tough problems, starting from the fact that entropy of a state is not uniquely defined. There are two intuitive and equivalent definitions in the quantum case, but these are inequivalent in a general scenario [25, 26, 27]. More severely, the interpretation of any of these quantities as the asymptotic information compression rates fails in the general case [25].

To avoid these technical problems, in our compression axiom we focus on the case of *ideal* compression, namely a compression that is perfectly lossless (no vanishingly small errors) and maximally efficient for a single use of the scheme. In particular, the axiom makes a statement about the systems existing in our theory, requiring the existence of systems that provide a lossless and maximally efficient carrier for any information source. The compression postulate, on more technical grounds, provides a very restrictive constraint on the convex sets of states, imposing that the all faces of a given set of states be in one-to-one correspondence with sets of states of smaller systems.

Purification

As we already mentioned, the first five axioms in our system provide the standard for information-processing theories. It is the sixth principle, the purification postulate, that captures all the features singling out Quantum Theory. Purification with causality and local discriminability alone already produces an astonishingly wide range of crucial theorems like no-cloning, no-programming, information-disturbance tradeoff, the existence

of a probabilistic teleportation scheme, the reversible dilation theorem for channels, the Choi correspondence, the structure theorem for memory channels, the necessary and sufficient conditions for error correction. For the proofs see Ref. [19].

The main consequence of purification for the purpose of deriving the Hilbert space formalism is the analogue of Choi correspondence, that is derived from the existence of a probabilistic entanglement swapping scheme. The correspondence reads

Theorem 3 (States-transformations correspondence). *Let Θ be a causal theory with local discriminability and purification. For every system type A of Θ there exists a system type B and a pure state $\Psi \in \text{St}(A \otimes B)$ such that the correspondence between transformations \mathcal{T} of type $A \rightarrow C$ and states $R \in \text{St}(C \otimes B)$ defined as*

$$R_{\mathcal{T}} := (\mathcal{T} \otimes \mathcal{I}_B) \circ \Psi, \quad (6)$$

is bijective.

As a consequence of this correspondence the sets of transformations of any type are completely specified provided all state sets are specified. As a consequence we have the following crucial theorem

Theorem 4 (States specify the theory). *Let Θ and Θ' be two causal theories with local discriminability and purification, and with the same system types. Then if the state sets of Θ and Θ' are the same, the theories coincide.*

The importance of this theorem for the derivation is easily understood: If we prove that our axioms imply that the states of our theory are positive definite matrices with trace smaller than one, then we can conclude that our theory is Quantum Theory. More precisely, states are density matrices, effects are positive matrices bounded by the identity, and transformations are completely positive trace-non-increasing maps.

Finally, another very important consequence of purification is the dilation theorem for transformations.

Theorem 5 (Reversible dilation of tests). *For every test $\{\mathcal{C}_i\}_{i \in X}$ of type $A \rightarrow B$, there exist a system C , a pure state $\Psi \in \text{St}(C \otimes B)$, a reversible channel \mathcal{U} of type $(A \otimes B \otimes C) \rightarrow (A \otimes B \otimes C)$, and an observation-test $\{c_i\}_{i \in X}$ on C such that*

$$\begin{array}{c} \text{A} \quad \boxed{\mathcal{C}_i} \quad \text{B} \\ \hline \end{array} = \begin{array}{c} \text{C} \quad \text{C} \\ \text{B} \quad \text{A} \\ \text{A} \quad \text{B} \\ \hline \end{array} \quad (7)$$

It is possible to prove that the statement of Theorem 5 is in fact equivalent to the purification postulate.

One might be tempted to say that the purification postulate, and in particular Theorem 5, shows a bias in favor of the many worlds interpretation of quantum theory [7]. However, two important remarks are in order. The first one is that standard Quantum Theory *does satisfy* the purification postulate, and thus the many world interpretation is not favored by theories with purification more than it is favored by standard Quantum

Theory. The second remark is that the purification postulate should be interpreted as a postulate about information processing and not about the ontology of processes: every test *can be perfectly simulated* by a scheme as in Eq. (7). This does not imply that any test *is* necessarily of this form. As far as no further axiom is assumed about the mechanics (like reversibility of the dynamics of closed systems), then the many world interpretation is as acceptable as any formulation of Quantum Mechanics in which a spontaneous collapse is assumed. From this point of view, our formulation of Quantum Theory is interpretation-neutral.

CONCLUSION

In conclusion, it is worth stressing that our work consisted in a derivation of the Hilbert space formalism for Quantum Theory, which provides the syntax to describe experiments in the quantum world. However, we did not address semantics questions, like "Which physical quantity is measured in this test?" or "How should we choose the Hamiltonian for a given system?". For example, it is easy to show from our axioms that every one-parameter group of reversible transformations will be represented by unitary operators of the form $U_t = \exp[-iHt]$ for some self-adjoint operator H . However, the interpretation of H as the energy of the system does not follow from the axioms. In other words, our derivation allows one to reconstruct the kinematics of Schrödinger's equation, but not the mechanics associated to it: The mechanics should instead emerge from an equivalent informational derivation of quantum field theory [28].

REFERENCES

1. D. Hilbert, J. von Neumann, L. Nordheim, *Mathematische Annalen* **98**, 1 (1928).
2. J. von Neumann, *Mathematical Foundations of Quantum Mechanics*, (Princeton Univ. Press, Princeton, 1932).
3. D. Bohm, *Phys. Rev.* **85**, 166 (1952).
4. J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics*, (Cambridge University Press, Cambridge, 1987).
5. G. C. Ghirardi, A. Rimini, and T. Weber, in *Quantum Probability and Applications II*, L. Accardi and W. von Waldenfels eds. (Lect. Notes in Mathematics, Vol. 1136) (Springer, Berlin, 1985).
6. G. Birkhoff and J. von Neumann, *Ann. Math.* **37**, 823 (1936).
7. H. Everett III, *Rev. Mod. Phys.* **29**, 454 (1957).
8. G. W. Mackey, *Amer. Math. Monthly* **64**, 45-57, (1957).
9. G. Ludwig, *An Axiomatic Basis for Quantum Mechanics*, (Springer-Verlag, Berlin, 1985).
10. C. A. Fuchs, in *Quantum Theory: Reconstruction of Foundations*, ed. A. Khrenikov (Växjö: Växjö University Press, 2002).
11. G. Brassard, *Nature Physics* **1**, 2 (2005)
12. L. Hardy, arXiv:quant-ph/0101012.
13. B. Dakic and Č. Bruckner, in *Deep Beauty*, ed. H. Halvorson (Cambridge University Press, New York, 2011).
14. L. Masanes and M. P. Müller, *New J. Phys.* **13** 063001 (2011).
15. A. Grinbaum, *Brit. J. Phil. Sci.* **58**, 387-408 (2007).
16. L. Hardy, arXiv:1104.2066.
17. J. A. Wheeler, in *Complexity, Entropy, and the Physics of Information*, ed. W. H. Zurek (Addison-Wesley, Redwood City, CA 1990).
18. G. Chiribella, G. M. D'Ariano, and P. Perinotti, *Phys. Rev. A* **84**, 012311, (2011).

19. G. Chiribella, G. M. D'Ariano, and P. Perinotti, *Phys. Rev. A* **81**, 062348, (2011).
20. G. M. D'Ariano, *AIP Conf. Proc.* **810**, 114 (2006).
21. G. M. D'Ariano, in *Philosophy of Quantum Information and Entanglement*, Eds. A. Bokulich and G. Jaeger, Cambridge University Press, Cambridge UK (2010).
22. B. Coecke, *AIP Conf. Proc.* **810**, 81 (2006).
23. G. Chiribella, G. M. D'Ariano, and P. Perinotti, *EPL* **83**, 30004 (2008).
24. A. Wilce, *Found. Phys.* **40**, 434-462 (2010).
25. H. Barnum, J. Barrett, L. O. Clark, M. Leifer, R. Spekkens, N. Stepanik, A. Wilce, and R. Wilke, *New J. Phys.* **12**, 033024 (2010).
26. A. J. Short and S. Wehner, *N. J. Phys.* **12**, 033023 (2010).
27. G. Kimura, K. Nuida, and H. Imai, *Rep. Math. Phys.* **66**, 175 (2010).
28. G. M. D'Ariano, on this same proceedings volume.