arXiv:1506.00398v1 [quant-ph] 1 Jun 2015

Quantum from principles

Giulio Chiribella

Center for Quantum Information, Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, 100084

Giacomo Mauro D'Ariano

QUIT Group, *Dipartimento di Fisica, Università di Pavia, via Bassi 6, 27100 Pavia, Italy*

INFN sezione di Pavia, via Bassi 6, 27100 Pavia, Italy

Paolo Perinotti

QUIT Group, *Dipartimento di Fisica, Università di Pavia, via Bassi 6, 27100 Pavia, Italy*

INFN sezione di Pavia, via Bassi 6, 27100 Pavia, Italy

Abstract

Quantum theory was discovered in an adventurous way, under the urge to solve puzzles—like the spectrum of the blackbody radiation—that haunted the physics community at the beginning of the 20th century. It soon became clear, though, that quantum theory was not just a theory of specific physical systems, but rather a new language of universal applicability. Can this language be reconstructed from first principles? Can we arrive at it from logical reasoning, instead of *ad hoc* guesswork? A positive answer was provided in Refs. [1, 2], where we put forward six principles that identify quantum theory uniquely in a broad class of theories. We first defined a class of "theories of information", constructed as extensions of probability theory in which events can be connected into networks. In this framework, we formulated the six principles as rules governing the control and the accessibility of information. Directly from these rules, we reconstructed a number of quantum information features, and eventually, the whole Hilbert space framework. In short, our principles characterize quantum theory as the theory of information that allows for maximal control of randomness.

1. Introduction

Quantum foundations is an old field—as old as quantum mechanics itself. Among the early works stand out the seminal papers by Einstein, Podolski, and Rosen [3] and Schrödinger [4], who addressed quantum entanglement for the first time, exploring quantum mechanics within the Hilbert space formulation. Almost at the same time, Birkhoff and von Neumann [5] looked at the theory

in a wider framework allowing for alternative theories. From that angle, it was natural to ask what is special about quantum mechanics and why Nature obeys its peculiar laws. The take of Birkhoff and von Neumann was that quantum theory should be regarded as a new form of logic, whose laws could be derived from physically motivated axioms. This programme gave rise to the tradition of quantum logic [6, 7, 8, 9, 10], whose ramifications are still object of active research [11].

Researchers in quantum logic managed to derive a significant part of the quantum framework from logical axioms. However, there is a general consensus that the axioms put forward in this context are not as insightful as one would have hoped. For both experts and non-experts, it is hard to figure out what is the moral of the quantum-logic axiomatizations. What is special about quantum theory after all? Why should quantum theory be preferred to alternative theories? Not many answers can be found in the popular accounts of quantum logic (see e.g. the Wikipedia entry [12]) and even understanding what the axioms are requires delving into a highly specialized literature.

The ambition to find a more insightful axiomatization reemerged with the rise of quantum information. The new field showed that the mathematical axioms of quantum theory imply striking operational consequences, such as quantum key distribution [13, 14], quantum algorithms [15, 16], no-cloning [17, 18], quantum teleportation [19] and dense coding [20]. A natural question is: Can we reverse the implication and *derive* the mathematics of quantum theory from some of its operational consequences? This question is at the core of a research programme launched by Fuchs [21] and Brassard [22], which can be synthesized by the motto *"quantum foundations in the light of quantum information"* [23] [1]. The ultimate goal of the programme is to reconstruct the whole structure of quantum theory from few simple principles of information-theoretic nature.

One may wonder why quantum information theorists should be more successful than their predecessors in the axiomatic endeavour. A good reason is the following: In the pre-quantum information era, quantum theory was viewed like an impoverished version of classical theory, lacking the ability to make deterministic predictions about the outcomes of experiments. Clearly, this perspective offered no vantage point for explaining why the world should be quantum. Contrarily, quantum information provided plenty of positive reasons for preferring quantum theory to its classical counterpart. Turning some of these reasons into axioms then appeared as a promising route towards a compelling axiomatization. Pioneering works along this route are those by Hardy [25] and D'Ariano [26, 27]. More recently, the programme flourished, leading to an explosion of new axiomatizations [2, 28, 29, 30, 31, 32, 33, 34].

Here we review the axiomatization of Refs. [2]. In this work, quantum theory is derived from six principles, formulated in a general framework of theo-

[1] This was also the title of one influential conference, held in May 2000 at the Université de Montréal [24], which kickstarted the new wave of quantum axiomatizations.

ries of information. The first five principles—Causality, Purity of Composition, Local Discriminability, Perfect State Discrimination, and Ideal Compression—express ordinary properties that are shared by quantum and classical information theory: such principles define what we call a *standard* theory of information. Among all standard theories of information, the sixth principle—Purification—identifies quantum theory uniquely. Purification states that every random preparation can be simulated via non-random preparation procedure, in which the system is manipulated together with an environment. An agent that has access to both the system and the environment would then have maximal control of the preparation—*maximal* in the sense that no other agent could conceivably have higher control. The moral of our work is that *Quantum Theory is the theory that allows maximal control of randomness*, giving us—at least in principle—the power to control all possible preparations and all possible dynamics.

The chapter is structured as follows: in section 2 we provide an introduction to the framework of *operational-probabilistic theories*—general theories of information arising from the combination of the circuit framework with probability theory. Then, section 3 presents the background to the reconstruction, discussing the main standing assumptions—finite-dimensionality, non-determinism, and closure under limits—and introducing a few basic operational tasks: signalling, collecting side information, doing state tomography, distinguishing states, compressing information, and simulating preparations. The principles are then analyzed in section 4. Section 5 provides a guided tour through the main results in our reconstruction, showing how the main features of quantum theory can be derived directly from the principles. Finally, the conclusions are drawn in section 6.

2. Operational-probabilistic theories

In order to reconstruct quantum theory and the features of quantum information, one needs a framework capable to describe a variety of alternative theories. Different frameworks have been proposed for this scope, under the broad name of *general probabilistic theories* [25, 35, 26, 27, 36, 37, 1, 2, 28, 38, 39]. Our reconstruction is based on a specific variant of general probabilistic theories, which we call *operational-probabilistic theories (OPTs)* [1, 2]. OPTs are an extension of probability theory, in which events can be connected into circuits. Technically, OPTs arise from the combination of the categorical framework of Abramsky and Coecke [40, 41, 42] with the toolbox of elementary probability theory. We regard such a combination as the natural mathematical object describing a "general theory of information". In the following we present a concise summary of the OPT framework.

2.1. Operational structure

2.1.1. Systems

Systems are labels, which determine how different events can be connected to one another. We denote systems by capital letters, such as $\mathrm{A}, \mathrm{B}, \mathrm{C}$, and so

on. The letter I will be reserved for the *trivial system*, representing "nothing" [2]. The set of all systems under consideration will be denoted by Sys.

Every two systems A and B can be considered together as a composite system, denoted by $\mathrm{A} \otimes \mathrm{B}$. The composition of systems is associative, namely

$$\mathrm{A} \otimes (\mathrm{B} \otimes \mathrm{C}) = (\mathrm{A} \otimes \mathrm{B}) \otimes \mathrm{C} \qquad \forall \mathrm{A}, \mathrm{B}, \mathrm{C} \tag{1}$$

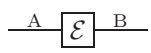
and has the trivial system as identity element, namely

$$\mathrm{A} \otimes \mathrm{I} = \mathrm{I} \otimes \mathrm{A} = \mathrm{A} \qquad \forall \mathrm{A}\,. \tag{2}$$

The second condition means that considering system A together with "nothing" is the same as considering system A alone.

2.1.2. Events

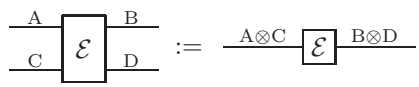
An event of type $\mathrm{A} \to \mathrm{B}$ represents the occurrence of a transformation that converts the input system A into the output system B. An event $\mathcal{E}$ of type $\mathrm{A} \to \mathrm{B}$ will be represented graphically as



.

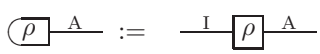
The set of all events of type $\mathrm{A} \to \mathrm{B}$ will be denoted by $\mathsf{Transf}(\mathrm{A} \to \mathrm{B})$, identifying events with the corresponding transformations.

When the input and output systems are composite systems, we draw boxes with multiple wires. For example, the box

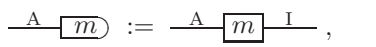


represents an event of type $(\mathrm{A} \otimes \mathrm{C}) \to (\mathrm{B} \otimes \mathrm{D})$.

Some types of events are particularly important and deserve a name of their own. An event of type $\mathrm{I} \to \mathrm{A}$ is a *preparation-event* (or simply, a *preparation*), that is, an event that makes system A available to further processing. An event of type $\mathrm{A} \to \mathrm{I}$ is an *observation-event* (or simply, an *observation*), after which system A is no longer available. Preparation- and observation-events will be represented as



and



,

respectively. We will often use the Dirac-like notation $(a|$ and $|\rho)$ the observation a and the preparation ρ, respectively.

Events of type $\mathrm{I} \to \mathrm{I}$ will be called *scalars* [40]. Scalars will be represented "out of the box", as

$s :=$

.

Later, scalars will be associated to probabilities. For the moment, however, they are just a special type of events.

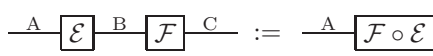
[2] More precisely, "nothing that the theory cares to describe".

2.1.3. Composition of events

Events can be connected into networks through the following operations

1. *Sequential composition:* an event of type $\mathrm{A} \to \mathrm{B}$ can be connected to an event of type $\mathrm{B} \to \mathrm{C}$, yielding an event of type $\mathrm{A} \to \mathrm{C}$.
2. *Parallel composition:* an event of type $\mathrm{A} \to \mathrm{A}'$ can be composed with an event of type $\mathrm{B} \to \mathrm{B}'$, yielding an event of type $(\mathrm{A} \otimes \mathrm{B}) \to (\mathrm{A}' \otimes \mathrm{B}')$.

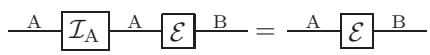
Intuitively, the sequential composition represents two events happening at "subsequent time steps" [3]. The sequential composition of two events $\mathcal{E}$ and $\mathcal{F}$ of matching types is denoted by $\mathcal{F} \circ \mathcal{E}$ and is represented graphically as



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

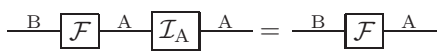
$$\mathcal{G} \circ (\mathcal{F} \circ \mathcal{E}) = (\mathcal{G} \circ \mathcal{F}) \circ \mathcal{E} \,, \tag{3}$$

for arbitrary events $\mathcal{E}, \mathcal{F}$ and $\mathcal{G}$ of matching types. In addition to associativity, sequential composition is required to have an identity element for every system. The *identity on system* A, denoted by $\mathcal{I}_{\mathrm{A}}$, is the special event of type $\mathrm{A} \to \mathrm{A}$ identified by the conditions



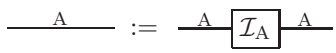
(4)

and

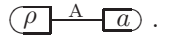


(5)

required to be valid for arbitrary systems A, B and arbitrary events $\mathcal{E}$ and $\mathcal{F}$ of types $\mathrm{A} \to \mathrm{B}$ and $\mathrm{B} \to \mathrm{A}$, respectively. The intuitive content of the above equations is that $\mathcal{I}_{\mathrm{A}}$ represents the process that "does nothing on the system". Consistently, we use the graphical notation



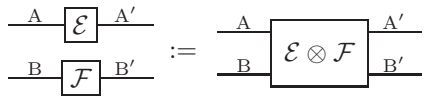
Mathematically, conditions (3), (4), and (5) impose that the events form a *category* [43, 44], in which the systems are the objects and the events are the arrows. For the sequential composition of a preparation and an observation we will often use the Dirac-like notation,



(6)

[3] *Per se*, the mathematical formalism does not force us to interpret the order of sequential composition as an order in time. Nevertheless, composition in time is the reference situation that we will have in mind when phrasing our axioms.

Let us consider parallel composition. The parallel composition of two events $\mathcal{E}$ and $\mathcal{F}$ is denoted as $\mathcal{E} \otimes \mathcal{F}$ and is represented graphically as



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

$$(\mathcal{E} \otimes \mathcal{F}) \circ (\mathcal{G} \otimes \mathcal{H}) = (\mathcal{E} \circ \mathcal{G}) \otimes (\mathcal{F} \circ \mathcal{H}) \,, \tag{7}$$

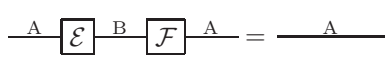
where $\mathcal{E}, \mathcal{F}, \mathcal{G}$, and $\mathcal{H}$ are arbitrary events of matching types. Such condition is necessary for the graphical notation to make sense, since in graphical notation the two sides of Eq. (7) look exactly the same. In addition to Eq. (7), parallel composition is required to satisfy the condition

$$\mathcal{I}_{\mathrm{A}\otimes\mathrm{B}} = \mathcal{I}_{\mathrm{A}} \otimes \mathcal{I}_{\mathrm{B}} \,. \tag{8}$$

Mathematically, the presence of parallel composition turns the category of events into a strict monoidal category, whose key properties are summarized by Eqs. (1), (2), (7), and (8). We denote such category by Transf.

2.1.4. Reversible events

An event $\mathcal{E}$ of type $\mathrm{A} \to \mathrm{B}$ is *reversible* iff there exists another event $\mathcal{F}$, of type $\mathrm{B} \to \mathrm{A}$, such that



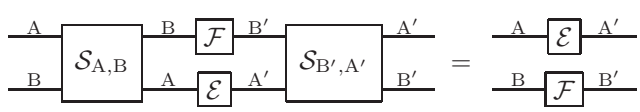
(9)

and

(10)

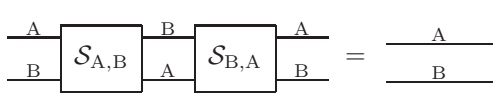
When this is the case, we write $\mathcal{F} = \mathcal{E}^{-1}$ and we say that systems A and B are *operationally equivalent* (or simply *equivalent*).

We denote by $\mathsf{RevTransf}(\mathrm{A} \to \mathrm{B})$ the set of reversible events of type $\mathrm{A} \to \mathrm{B}$. Such set (which may be empty) depends on the specific theory. In general, we require the existence of a reversible event that swaps pairs of systems. Given two systems A and B, the *swap of* A *with* B—denoted by $\mathcal{S}_{\mathrm{A},\mathrm{B}}$—is a reversible event of type $(\mathrm{A} \otimes \mathrm{B}) \to (\mathrm{B} \otimes \mathrm{A})$ satisfying the condition



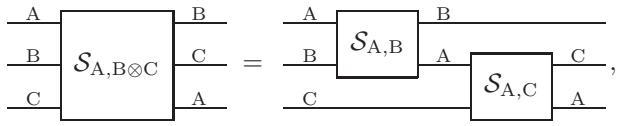
(11)

for arbitrary systems $\mathrm{A}, \mathrm{B}, \mathrm{A}', \mathrm{B}'$ and arbitrary events $\mathcal{E}, \mathcal{F}$, as well as the conditions



(12)

and



(13)

The presence of the swap, with the related equations (11), (12), and (13), turns the strict monoidal category into a strict *symmetric* monoidal category [45, 46] (strict SMC, for short).

2.1.5. Tests

A test represents a process, which can generally be non-deterministic—i. e. it can result in multiple alternative events. Specifically, a test of type $\mathrm{A} \to \mathrm{B}$ is collection of events of type $\mathrm{A} \to \mathrm{B}$, labelled by outcomes in a suitable outcome set X. The test $\boldsymbol{\mathcal{E}} := \{\mathcal{E}_x\}_{x\in\mathrm{X}}$ is represented graphically as

.

When two events/transformations belong to the same test, we say that they are *coexisting*.

The set of tests of type $\mathrm{A} \to \mathrm{B}$ with outcomes in X will be denoted by $\mathsf{Tests}(\mathrm{A} \to \mathrm{B}, \mathrm{X})$. We will restrict our attention to tests with a *finite* outcome set.

Tests with $|\mathrm{X}| = 1$ are called *deterministic*, because only one event can take place. We will often identify a deterministic test $\{\mathcal{E}_{x_0}\}$ with the corresponding event $\mathcal{E}_{x_0}$, saying that $\mathcal{E}_{x_0}$ is a *deterministic event* (or a *deterministic transformation*). The deterministic transformations form a strict symmetric monoidal subcategory of Transf, denoted by $\mathsf{DetTransf}$.

Some types of tests are particularly important and deserve a name of their own. A test of type $\mathrm{I} \to \mathrm{A}$ is a *preparation-test* (or an *ensemble*), which prepares system A in a non-deterministic way, with the possible preparations labelled by different outcomes. A test of type $\mathrm{A} \to \mathrm{I}$ is an *observation-test*, corresponding to a demolition measurement that absorbs system A while producing an outcome.

2.1.6. Composition of tests

Not all collections of events of are "tests". Whether or not a specific collection is a test is determined by the theory, compatibly with two basic requirements:

1. the set of tests must be closed under sequential and parallel composition
2. the set of tests must contain deterministic tests corresponding to reversible events.

Let us discuss these requirements in more detail:

1. The sequential composition of two tests $\boldsymbol{\mathcal{E}} = \{\mathcal{E}_x\}_{x\in\mathrm{X}}$ and $\boldsymbol{\mathcal{F}} = \{\mathcal{F}_y\}_{y\in\mathrm{Y}}$ of matching types is defined as

$$\boldsymbol{\mathcal{F}} \circ \boldsymbol{\mathcal{E}} := \{\mathcal{F}_y \circ \mathcal{E}_x\}_{(x,y)\in\mathrm{X}\times\mathrm{Y}} \ .$$

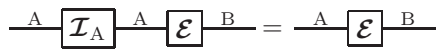
The test $\boldsymbol{\mathcal{F}} \circ \boldsymbol{\mathcal{E}}$ represents a cascade of two (generally non-deterministic) processes, wherein each process can result in a number of alternative events. Similarly, the parallel composition of two tests is defined as

$$\boldsymbol{\mathcal{E}} \otimes \boldsymbol{\mathcal{F}} := \{\mathcal{E}_x \otimes \mathcal{F}_y\}_{(x,y)\in \mathrm{X}\times\mathrm{Y}}$$

and represents two non-deterministic processes occurring in parallel. The composition of tests induces a composition of their outcome spaces via the Cartesian product. As a consequence, the set of all outcome spaces must be closed under this operation. We will denote such a set by $\mathsf{Outcomes}$.

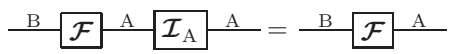
2. If $\mathcal{U}$ is a reversible event of type $\mathrm{A} \to \mathrm{B}$, we require that there exists a deterministic test $\boldsymbol{\mathcal{U}} := \{\mathcal{U}\}$. In particular, there must be a deterministic test $\boldsymbol{\mathcal{I}}_{\mathrm{A}} := \{\mathcal{I}_{\mathrm{A}}\}$ corresponding to the identity on system A and a deterministic test $\boldsymbol{\mathcal{S}}_{\mathrm{A,B}} := \{\mathcal{S}_{\mathrm{A,B}}\}$ corresponding to the swap of systems A and B.

Note that all the basic equations valid for events can be lifted to tests: for example, the identity test acts as identity element with respect to the composition, that is, one has



(14)

and



(15)

for arbitrary systems A, B and for arbitrary tests $\boldsymbol{\mathcal{E}}$ and $\boldsymbol{\mathcal{F}}$ of types $\mathrm{A} \to \mathrm{B}$ and $\mathrm{B} \to \mathrm{A}$, respectively. Since events form a strict SMC, also the tests form a strict SMC, which we denote by Tests.

2.1.7. Summary about the operational structure

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

$$\mathsf{Op} = (\mathsf{Transf}, \mathsf{Outcomes}, \mathsf{Tests})\,,$$

where Transf is a strict symmetric monoidal category, $\mathsf{Outcomes}$ is a collection of sets closed under Cartesian product, and Tests is a strict symmetric monoidal category, related to Transf and $\mathsf{Outcomes}$ as described in the previous paragraph. Intuitively, the operational structure describes

1. what can be done (connecting tests)
2. what can be observed (outcomes), and
3. what can happen (events).

2.2. Probabilistic structure

The goal of a physical theory is not only to *describe* a class of experiments, but also to *make predictions* about the outcomes of such experiments. In the following we show how this can be accomplished by adding a probabilistic structure on top of the operational structure.

2.2.1. Assigning probabilities

An *experiment* consists in sequence of tests that starts from a preparation-test and ends with an observation-test, leaving no open wires, as in the following example

. (16)

If we compose all the tests involved in an experiment, we obtain a single test, which transforms the trivial system into itself. In order to make predictions on the outcomes of the experiment, we need a rule assigning a probability to the events of such test. The rule is provided by the probabilistic structure of the theory:

Definition 1 (Probabilistic structure)**.** *Let* Op *be an operational structure. A* probabilistic structure *for* Op *is a map* $\mathsf{Prob} : \mathsf{Transf}(\mathrm{I} \to \mathrm{I}) \to [0,1]$*, which associates a given scalar* s *to a probability* $\mathsf{Prob}(s)$*, in accordance to the following two requirements:*

1. Consistency: $\sum_{x\in \mathrm{X}} \mathsf{Prob}(s_x) = 1$ *for every outcome set* $\mathrm{X} \in \mathsf{Outcomes}$ *and for every test* $\boldsymbol{s} \in \mathsf{Tests}(\mathrm{I} \to \mathrm{I}, \mathrm{X})$
2. Independence: $\mathsf{Prob}(s \otimes t) = \mathsf{Prob}(s)\,\mathsf{Prob}(t)$ *for every pair of scalars* s *and* t*.*

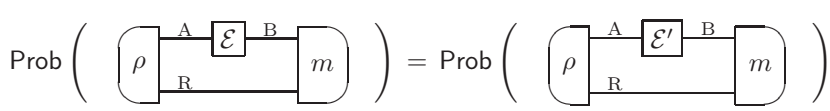
The consistency requirement guarantees that we can interpret $\mathsf{Prob}(s_x)$ as the probability of the outcome $x \in \mathrm{X}$. The independence requirement guarantees that experiments that involve only independent tests on two systems give rise to uncorrelated outcomes. As observed by Hardy [28, 38], independence is equivalent to the requirement that probabilities can assigned to the outcomes of an experiment in a way that is independent of the context in which the experiment is performed. Note that the map Prob needs not be surjective: for example, in a *deterministic* theory the range of Prob are only the values 0 and 1.

We are now ready to give the formal definition of OPT:

Definition 2. *An* operational-probabilistic theory Θ *is a pair* $(\mathsf{Op}, \mathsf{Prob})$ *consisting of an operational structure* Op *and of a probabilistic structure for* Op*.*

2.2.2. Statistically equivalent events

Once probabilities are introduced, it is natural to identify events that give rise to the same probabilities in all possible circuits. Precisely, we say that two events of type $\mathrm{A} \to \mathrm{B}$, say $\mathcal{E}$ and $\mathcal{E}'$, are *statistically equivalent* iff



for every system R, every preparation-event $\rho \in \mathsf{Transf}(\mathrm{I} \to \mathrm{A} \otimes \mathrm{R})$ and every observation-event $m \in \mathsf{Transf}(\mathrm{B} \otimes \mathrm{R} \to \mathrm{I})$. We denote by $[\mathcal{E}]$ the equivalence class of the event $\mathcal{E}$.

Equivalence classes can be composed in sequence and parallel in the obvious way

$$[\mathcal{F}] \circ [\mathcal{E}] := [\mathcal{F} \circ \mathcal{E}]\,, \qquad\qquad [\mathcal{E}] \otimes [\mathcal{F}] := [\mathcal{E} \otimes \mathcal{F}]$$

and it is easily verified that both definitions are well-posed. Furthermore, $[\mathcal{I}_{\mathrm{A}}]$ and $[\mathcal{S}_{\mathrm{A,B}}]$ behave like the identity on A and the swap between A and B, respectively. As a result, the equivalence classes of events form a strict SMC, which we denote by $[\mathsf{Transf}]$.

Similar considerations apply to tests: the equivalence class of a test $\boldsymbol{\mathcal{E}} = \{\mathcal{E}_x\}_{x\in\mathrm{X}}$ is defined as $[\boldsymbol{\mathcal{E}}] := \{[\mathcal{E}_x]\}_{x\in\mathrm{X}}$ and the sequential/parallel composition of equivalence classes of tests are induced by the sequential/parallel composition of events:

$$[\boldsymbol{\mathcal{F}}] \circ [\boldsymbol{\mathcal{E}}] := [\boldsymbol{\mathcal{F}} \circ \boldsymbol{\mathcal{E}}]\,, \qquad\qquad [\boldsymbol{\mathcal{E}}] \otimes [\boldsymbol{\mathcal{F}}] := [\boldsymbol{\mathcal{E}} \otimes \boldsymbol{\mathcal{F}}]\,.$$

Again, the equivalence class of $[\boldsymbol{\mathcal{I}}_{\mathrm{A}}]$ and $[\boldsymbol{\mathcal{S}}_{\mathrm{A,B}}]$ behave like the identity and the swap. As a result, the equivalence classes of tests form a strict SMC, which we denote by $[\mathsf{Tests}]$.

2.2.3. The quotient OPT

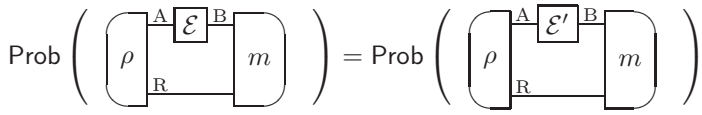
The notion of statistical equivalence allowed us to transform the original operational structure $\mathsf{Op} = (\mathsf{Transf}, \mathsf{Outcomes}, \mathsf{Tests})$ into a new operational structure $[\mathsf{Op}] := ([\mathsf{Transf}]\,, \mathsf{Outcomes}, [\mathsf{Tests}])$, which we call the *quotient operational structure*. The operational structure $[\mathsf{Op}]$ comes with an obvious probabilistic structure $[\mathsf{Prob}]$, defined as

$$[\mathsf{Prob}]\,(\,[s]\,) := \mathsf{Prob}(s) \qquad\qquad \forall s \in \mathsf{Transf}(\mathrm{I} \to \mathrm{I})\,.$$

It is indeed immediate to verify that the consistency and independence conditions in definition 1 are satisfied. As a result, the original OPT $\Theta = (\mathsf{Op}, \mathsf{Prob})$ has been turned into a new OPT $[\Theta] := ([\mathsf{Op}], [\mathsf{Prob}])$, which we call the *quotient OPT*. Intuitively, the quotient OPT contains all the information that is statistically relevant, disregarding those distinctions that have no consequences for the purpose of making probabilistic predictions.

In the following we will focus on quotient OPTs: by default, an OPT will be a *quotient* OPT. Accordingly, we will omit the symbol of equivalence class everywhere and write $\Theta = (\mathsf{Op}, \mathsf{Prob})$, assuming that equivalence classes have been already taken from the start. This is equivalent to requiring the following *separation property* [47]:

Definition 3. *An OPT satisfies the* separation property *iff for every pair of systems* A *and* B *and every pair of events* $\mathcal{E}$ *and* $\mathcal{E}'$ *of type* $\mathrm{A} \to \mathrm{B}$ *the condition*



$$\forall \mathrm{R} \in \mathsf{Sys}$$
$$\forall \rho \in \mathsf{Transf}(\mathrm{I} \to \mathrm{A} \otimes \mathrm{R})$$
$$\forall m \in \mathsf{Transf}(\mathrm{B} \otimes \mathrm{R} \to \mathrm{I})$$

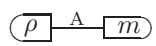
implies $\mathcal{E} = \mathcal{E}'$.

In a quotient OPT preparation-events (respectively, observation-events) will be called *states* (respectively, *effects*) and we will use the notation $\mathsf{St}(\mathrm{A}) := \mathsf{Transf}(\mathrm{I} \to \mathrm{A})$ (respectively, $\mathsf{Eff}(\mathrm{A}) := \mathsf{Transf}(\mathrm{A} \to \mathrm{I})$).

2.2.4. Vector space representation of an OPT

OPTs satisfying the separation property have a convenient representation in terms of ordered vector spaces and positive maps. The construction proceeds in four steps:

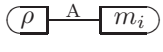
1. The separation property guarantees that a scalar s can be identified with its probability $\mathsf{Prob}(s)$. Hence, from now on we will omit Prob and will identify the set of scalars $\mathsf{Transf}(\mathrm{I} \to \mathrm{I})$ with a subset of the real interval $[0, 1]$.
2. By the separation property, a state $\rho \in \mathsf{St}(\mathrm{A})$ can be identified with the real-valued function $\hat{\rho} : \mathsf{Eff}(\mathrm{A}) \to \mathbb{R}$ defined by

   $\widehat{\rho}(m) :=$ 

   (indeed, one has $\rho = \sigma$ if and only if $\widehat{\rho} = \widehat{\sigma}$). Since real-valued functions form a vector space, we can define the vector (sub)space spanned by the states of system A as

   $$\mathsf{St}_{\mathbb{R}}(\mathrm{A}) := \mathsf{Span}_{\mathbb{R}} \{\rho \mid \rho \in \mathsf{St}(\mathrm{A})\} \ .$$

   Limiting ourselves to linear combination with positive coefficients we obtain the proper cone $\mathsf{St}_{+}(\mathrm{A})$, which turns $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$ into an ordered vector space.
3. Every effect $m \in \mathsf{Eff}(\mathrm{A})$ defines a *linear* function $\widehat{m} : \mathsf{St}_{\mathbb{R}}(\mathrm{A}) \to \mathbb{R}$, via the relation

   $\widehat{m}\left(\sum_i c_i\, \rho_i\right) := \sum_i c_i$ 

   $, \qquad \forall \{c_i\} \subset \mathbb{R}\,, \quad \forall \{\rho_i\} \subset \mathsf{St}(\mathrm{A})\,.$

   It is immediate to see that the definition is well-posed, namely $\widehat{m}\left(\sum_i c_i\, \rho_i\right) = \widehat{m}\left(\sum_j c'_j\, \rho'_j\right)$ whenever $\sum_i c_i\, \rho_i = \sum_j c'_j\, \rho'_j$. Again, the effect m can be identified with the linear function $\widehat{m}$ thanks to the separation property. Taking linear combinations of effects we obtain the vector space

   $$\mathsf{Eff}_{\mathbb{R}}(\mathrm{A}) := \mathsf{Span}_{\mathbb{R}} \{m \mid m \in \mathsf{Eff}(\mathrm{A})\} \ ,$$

   while restricting to positive linear combinations we obtain the proper cone $\mathsf{Eff}_{+}(\mathrm{A})$. As a result, also $\mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$ is an ordered vector space.
4. Every event $\mathcal{E}$ of type $\mathrm{A} \to \mathrm{B}$ induces a linear map $\widehat{\mathcal{E}} : \mathsf{St}_{\mathbb{R}}(\mathrm{A}) \to \mathsf{St}_{\mathbb{R}}(\mathrm{B})$, via the definition

   $$\widehat{\mathcal{E}}\left(\sum_i c_i\, \rho_i\right) := \sum_i c_i\, \left(\mathcal{E} \circ \rho_i\right) \ , \qquad \forall \{c_i\} \subset \mathbb{R}\,, \quad \forall \{\rho_i\} \subset \mathsf{St}(\mathrm{A})\,.$$

Again, it is not hard to see that the definition is well-posed, namely that $\widehat{\mathcal{E}}\left(\sum_i c_i\,\rho_i\right) = \widehat{\mathcal{E}}\left(\sum_j c'_j\,\rho'_j\right)$ whenever $\sum_i c_i\,\rho_i = \sum_j c'_j\,\rho'_j$. Note that the map $\widehat{\mathcal{E}}$ is not only linear, but also *positive*: indeed, it sends elements of the cone $\mathsf{St}_+(\mathrm{A})$ to elements of the cone $\mathsf{St}_+(\mathrm{B})$. We call $\widehat{\mathcal{E}}$ the *state change* associated to $\mathcal{E}$.

At this point, a few remarks are in order:

1. *Linearity vs convexity.* Traditionally, the linearity of state changes has been argued from the assumption that the state space $\mathsf{St}(\mathrm{A})$ is convex. However, our argument shows that such assumption is *not* needed: the probabilistic structure alone suffices to define the linear map $\widehat{\mathcal{E}}$.
2. *Finite vs infinite dimensional systems.* For a given system A, we define D_{A} to be the dimension of the vector space $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$ and we say that system A is *finite dimensional* if D_{A} is finite. For finite systems, one has the equality $\mathsf{Eff}_{\mathbb{R}}(\mathrm{A}) = \mathsf{St}_{\mathbb{R}}(\mathrm{A})^*$, where $\mathsf{St}_{\mathbb{R}}(\mathrm{A})^*$ is the vector space of all linear functionals on $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$. For infinite dimensional systems, such an equality may not hold.
3. *The no-restriction hypothesis.* Since effects are identified with positive linear functions, one has the inclusion $\mathsf{Eff}_+(\mathrm{A}) \subseteq \mathsf{St}_+(\mathrm{A})^*$, where $\mathsf{St}_+(\mathrm{A})^*$ denotes the dual cone of $\mathsf{St}_+(\mathrm{A})$

$$\mathsf{St}_+(\mathrm{A})^* := \{\, m \in \mathsf{St}_{\mathbb{R}}(\mathrm{A})^* \mid m(\rho) \geq 0 \quad \forall \rho \in \mathsf{St}_+(\mathrm{A})\,\}\,. \tag{17}$$

   Even for finite dimensional systems, the inclusion $\mathsf{Eff}_+(\mathrm{A}) \subseteq \mathsf{St}_+(\mathrm{A})^*$ may not be an equality. The assumption $\mathsf{Eff}_+(\mathrm{A}) = \mathsf{St}_+(\mathrm{A})^*$ is known as *No-Restriction Hypothesis* [1]. We stress that such an assumption is *not* made in our derivation.
4. *Transformations vs linear maps.* Unlike in the case of states and effects, the correspondence between the transformation $\mathcal{E}$ and the linear map $\widehat{\mathcal{E}}$ may not be one-to-one. The reason for this is that the difference between two transformations $\mathcal{E}$ and $\mathcal{E}'$ may show up when one applies them locally on a part of a composite system: one can have $\widehat{\mathcal{E}\otimes\mathcal{I}_{\mathrm{R}}} \neq \widehat{\mathcal{E}'\otimes\mathcal{I}_{\mathrm{R}}}$ for some $\mathrm{R} \in \mathsf{Sys}$ even if $\widehat{\mathcal{E}} = \widehat{\mathcal{E}'}$. This problem disappears if one assumes the axiom of Local Tomography, as we will see later in this chapter. In the lack of Local Tomography, however, the transformation $\mathcal{E}$ can still be identified with a linear map: for this purpose, one can choose the linear map $\widehat{\mathcal{E}}_{\oplus}$ defined by [47]

$$\widehat{\mathcal{E}}_{\oplus} := \bigoplus_{\mathrm{R}\in\mathsf{Sys}} \widehat{\mathcal{E}\otimes\mathcal{I}_{\mathrm{R}}}\,. \tag{18}$$

   The map $\widehat{\mathcal{E}}_{\oplus}$ transforms elements of the (infinite-dimensional) vector space $\mathsf{St}_{\mathbb{R},\oplus}(\mathrm{A}) := \bigoplus_{\mathrm{R}\in\mathsf{Sys}} \mathsf{St}_{\mathbb{R}}(\mathrm{A}\otimes\mathrm{R})$ into elements of the (infinite-dimensional) vector space $\mathsf{St}_{\mathbb{R},\oplus}(\mathrm{B}) := \bigoplus_{\mathrm{R}\in\mathsf{Sys}} \mathsf{St}_{\mathbb{R}}(\mathrm{B}\otimes\mathrm{R})$. Then, the separation property guarantees that the correspondence between $\mathcal{E}$ and $\widehat{\mathcal{E}}_{\oplus}$ is one-to-one.

5. *The vector space of transformations.* So far we have defined the vector spaces of states and effects. A vector space of transformations can be defined using the one-to-one correspondence with the linear maps in Eq. (18) and setting

$$\mathsf{Transf}_{\mathbb{R}}(\mathrm{A} \to \mathrm{B}) := \mathsf{Span}_{\mathbb{R}}\{\mathsf{Transf}(\mathrm{A} \to \mathrm{B})\}\,. \tag{19}$$

Again, a proper cone $\mathsf{Transf}_{+}(\mathrm{A} \to \mathrm{B})$ can be defined by restricting the attention to linear combinations with positive coefficients. Note that, in general, the vector space $\mathsf{Transf}_{\mathbb{R}}(\mathrm{A} \to \mathrm{B})$ and the cone $\mathsf{Transf}_{+}(\mathrm{A} \to \mathrm{B})$ can be infinite-dimensional *even if both systems* A *and* B *are finite dimensional*. However, this is not the case when the theory satisfies the Local Tomography.

2.2.5. Closure under coarse-graining

A key notion that comes with the probabilistic structure is the notion of *coarse-graining*: given a test $\boldsymbol{\mathcal{T}} = \{\mathcal{T}_x\}_{x\in\mathrm{X}}$, one can decide to identify some outcomes, thus obtaining another, coarse-grained test. Mathematically, a coarse-graining is defined by partitioning the outcome set X into mutually disjoint subsets $\{\mathrm{X}_y\}_{y\in\mathrm{Y}}$. Relative to such partition, the coarse-graining of the test $\boldsymbol{\mathcal{T}}$ is the test $\boldsymbol{\mathcal{T}}' = \{\mathcal{T}'_y\}_{y\in\mathrm{Y}}$ defined by[4]

$$\mathcal{T}'_y := \sum_{x\in\mathrm{X}_y} \mathcal{T}_x\,, \tag{20}$$

setting $\mathcal{T}'_y = 0$ for $\mathrm{X}_y = \emptyset$, where 0 is the zero element of the vector space $\mathsf{Transf}_{\mathbb{R}}(\mathrm{A} \to \mathrm{B})$. Note that, by calling $\boldsymbol{\mathcal{T}}'$ a *test* we have implicitly made two assumptions, namely that

1. the set Y belongs to $\mathsf{Outcomes}$
2. the collection $\{\mathcal{T}'_y\}_{y\in\mathrm{Y}} \subset \mathsf{Transf}_{\mathbb{R}}(\mathrm{A} \to \mathrm{B})$ belongs to $\mathsf{Tests}(\mathrm{A} \to \mathrm{B}, \mathrm{Y})$.

From now on, we will require that our OPT is *closed under coarse-graning* meaning that the above conditions are satisfied.

By coarse-graining over all outcomes of a test $\boldsymbol{\mathcal{T}} \in \mathsf{Tests}(\mathrm{A} \to \mathrm{B}, \mathrm{X})$ one obtains a deterministic test, identified with the deterministic transformation $\mathcal{T} := \sum_{x\in\mathrm{X}} \mathcal{T}_x \in \mathsf{DetTransf}(\mathrm{A} \to \mathrm{B})$. In particular, when a preparation test $\boldsymbol{\rho} \in \mathsf{Tests}(\mathrm{I} \to \mathrm{A}, \mathrm{X})$ satisfies $\sum_{x\in\mathrm{X}} \rho_x = \rho$ we say that the test $\boldsymbol{\rho}$ is an *ensemble decomposition of* ρ.

2.2.6. Summary of the OPT framework

Let us sum up the main points discussed so far. We defined an OPT as a pair $\Theta = (\mathsf{Op}, \mathsf{Prob})$, consisting of an operational structure $\mathsf{Op} = (\mathsf{Transf}, \mathsf{Outcomes}, \mathsf{Tests})$ and of a probabilistic structure Prob that assigns probabilities to scalars. We

[4] Note that the summation is well-defined thanks to the vector space structure of $\mathsf{Transf}_{\mathbb{R}}(\mathrm{A} \to \mathrm{B})$.

restricted our attention to OPTs that satisfy the Separation Property (definition 3), which implies that one can identify scalars with probabilities, states with elements of suitable vector spaces, and effects with linear functionals over them. Transformations with nontrivial input and output induce linear maps on the corresponding state spaces. Finally, in agreement with the probabilistic interpretation, we demanded that the theory Θ is closed under coarse-graining.

3. Background of the quantum reconstruction

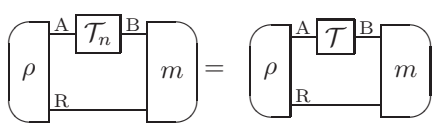
In this section we provide some background that will be useful for our reconstruction of quantum theory. We start by reviewing three standing assumptions: finite-dimensionality, non-determinism, and closure under operational limits. We will then review the operational tasks that motivate our axioms.

3.1. Standing assumptions

Here we introduce three standing assumptions that will be made in the rest of the chapter. These assumptions are common to all recent axiomatizations of quantum theory, and could also be even incorporated in the OPT framework. We keep them separate from the rest, both for clarity of presentation and for the sake of maintaining the OPT framework as flexible as possible. The assumptions are the following:

1. *Finite dimensionality.* We restrict our attention to finite systems, i. e. systems with finite dimensional state spaces. Operationally, this means that the state of every system can be identified from the statistics of a *finite number of finite-outcome measurements*. Of course, the implicit assumption here is that finite systems exist and form a sub-theory of our theory, meaning that the operational structure Op contains a non-trivial substructure $\mathsf{FiniteOp}$, consisting of transformations, outcome sets, and tests involving only finite systems.
2. *Non-determinism.* While the OPT framework accommodates a variety of theories, here we focus on OPTs that are non-deterministic, meaning that there exists at least one experiment for which the outcome is not determined a priori. Mathematically, this means that the range of the probability function Prob is not just $\{0, 1\}$. Note that non-determinism is a weaker assumption than convexity of the state spaces: there exist indeed examples of theories—such as Spekkens' toy theory [48]—that are non-deterministic and yet do not have convex state spaces.
3. *Closure under operational limits.* Suppose that $(\mathcal{T}_n)_{n\in\mathbb{N}}$ is a sequence of transformations of type $\mathrm{A} \to \mathrm{B}$ and that $\mathcal{T}$ is an element of the vector space $\mathsf{Transf}_{\mathbb{R}}(\mathrm{A} \to \mathrm{B})$ such that

$$\lim_{n\to\infty}$$



$$\forall \mathrm{R} \in \mathsf{Sys}$$
$$\forall \rho \in \mathsf{Transf}(\mathrm{I} \to \mathrm{A} \otimes \mathrm{R})$$
$$\forall m \in \mathsf{Transf}(\mathrm{B} \otimes \mathrm{R} \to \mathrm{I})\,,$$

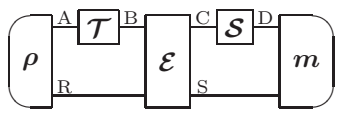
meaning that the probability of every experiment involving $\mathcal{T}_n$ converges to the probability of an hypothetical experiment involving $\mathcal{T}$. When this is the case, we assume that $\mathcal{T}$ belongs to $\mathsf{Transf}(\mathrm{A} \to \mathrm{B})$. Operationally, one can think of the sequence $(\mathcal{T}_n)_{n\in\mathbb{N}}$ as a *limit procedure* to implement the transformation $\mathcal{T}$.

3.2. Basic operational tasks

We now give a brief list of the operational notions on which our axioms are based.

3.2.1. Signalling

When a number of devices are connected into a network, it is natural to ask whether one node of the network can signal to another. For example, given the experiment



(21)

one can ask whether the choice of the test $\boldsymbol{\mathcal{T}}$ can influence the outcome of the test $\boldsymbol{\mathcal{S}}$. Precisely, the question is whether or not the marginal probability distribution for the outcomes of $\boldsymbol{\mathcal{S}}$ (obtained by summing over the outcomes of all the other tests in the network) depends on $\boldsymbol{\mathcal{T}}$. Denoting the marginal probability distribution by $p(x|\boldsymbol{\mathcal{T}})$, $x \in \mathrm{X}$, we say that the node occupied by the test $\boldsymbol{\mathcal{T}}$ *does not signal* to the node occupied by the test $\boldsymbol{\mathcal{S}}$ iff

$$p(x|\boldsymbol{\mathcal{T}}_0) = p(x|\boldsymbol{\mathcal{T}}_1) \qquad \forall x \in \mathrm{X}\,,$$

for every possible choice of tests $\mathcal{T}_0$ and $\mathcal{T}_1$. Similarly, one can ask whether the node occupied by the test $\boldsymbol{\mathcal{S}}$ can signal to the node occupied by the test $\boldsymbol{\mathcal{T}}$. Now, note that the test $\boldsymbol{\mathcal{S}}$ is performed *after* the test $\boldsymbol{\mathcal{T}}$: if the node occupied by $\boldsymbol{\mathcal{S}}$ can signal to the node occupied by $\boldsymbol{\mathcal{T}}$ we say that the circuit of Eq. (21) allows for *signalling from the future to the past*.

3.2.2. Collecting side information

Suppose that the test $\boldsymbol{\mathcal{T}} = \{\mathcal{T}_x\}_{x\in\mathrm{X}}$ is obtained from the test $\boldsymbol{\mathcal{T}}' = \{\mathcal{T}_z\}_{z\in\mathrm{Z}}$ via coarse-graining, namely

$$\mathcal{T}_x = \sum_{z\in\mathrm{Z}_x} \mathcal{T}'_z \qquad \forall x \in \mathrm{X}\,,$$

where $\{\mathrm{Z}_x\}_{x\in\mathrm{X}}$ is a partition of Z into disjoint subsets. In this case we say that $\boldsymbol{\mathcal{T}}'$ *refines* $\boldsymbol{\mathcal{T}}$. Now, it is convenient to relabel the outcomes of $\boldsymbol{\mathcal{T}}'$ as $z = (x, y)$, with $x \in \mathrm{X}$ and $y \in \mathrm{Z}_x$, and to write $\mathcal{T}'_{x,y}$ in place of $\mathcal{T}'_z$. In this way, we can think of the random variable y as a *side information*, which is not accessible to the agent Alice performing the test $\boldsymbol{\mathcal{T}}$, but may be accessible to some other agent Eve. This picture is particularly relevant to cryptographic scenarios, wherein Eve could be an eavesdropper attempting to collect as much information as possible. In all such scenarios, a special role is played by those transformations that do not leak any useful side information. We call such transformations *pure*:

Definition 4. *We say that a transformation $\mathcal{E}$ is* pure[5] *iff for every test $\boldsymbol{\mathcal{T}}$ containing $\mathcal{E}$ and for every test $\boldsymbol{\mathcal{T}}'$ refining $\boldsymbol{\mathcal{T}}$ one has*

$$\mathcal{T}'_{x_0,y} = p_y \, \mathcal{T}_{x_0} \,, \tag{22}$$

where x_0 is the outcome such that $\mathcal{T}_{x_0} = \mathcal{E}$ and $\{p_y\}$ is a probability distribution.

Informally, the purity condition (22) states that the side information possessed by Eve is uncorrelated with the transformation $\mathcal{E}$ taking place in Alice's laboratory. We denote the set of pure transformations of type $\mathrm{A} \to \mathrm{B}$ by $\mathsf{PurTransf}(\mathrm{A} \to \mathrm{B})$. In the special case of transformations with trivial input we will use the notation $\mathsf{PurSt}(\mathrm{A})$ (respectively, $\mathsf{PurEff}(\mathrm{A})$), referring to *pure states* (respectively, *pure effects*). An *pure test* is a test consisting of pure transformations.

Transformations that are not necessarily pure will be called *mixed*. Among the mixed transformations, the ones that are in the interior of the cone $\mathsf{Transf}_+(\mathrm{A} \to \mathrm{B})$ play an important role. They are defined as follows:

Definition 5. *A transformation $\mathcal{E} \in \mathsf{Transf}(\mathrm{A} \to \mathrm{B})$ is called* internal *iff for every transformation $\mathcal{F} \in \mathsf{Transf}(\mathrm{A} \to \mathrm{B})$ there exists a transformation $\mathcal{G}$ and a scaling constant $\lambda > 0$ such that*

1. $\mathcal{E} = \lambda \, \mathcal{F} + \mathcal{G}$
2. *$\lambda \, \mathcal{F}$ and $\mathcal{G}$ coexist in a test* [6]*.*

Roughly speaking, an internal transformation is compatible with the occurrence of any other transformation of the same type. Internal transformations with trivial input (output) will be called *internal states* (*internal effects*).

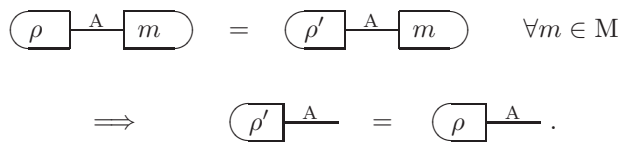
3.2.3. State tomography

The task of state tomography consists in identifying the state of a system from the statistics of a restricted set of observations. Suppose that an experimenter is able to perform a set of observation-tests and let M be the set of all effects appearing in such tests.

[5] In previous works, we used different names for transformations that do not allow for side information: in Refs. [1, 2] they were called *atomic*, while in the popularized version of Ref. [49] they were called *fine-grained*. We apologize with our readers for the changes of terminology, due to an ongoing search for the word that best captures this operational concept. In this chapter, we adopted the word *pure*, because *i)* this term is the standard one in the case of states and *ii)* using the same term for transformations should hopefully ease the reading. Still, a warning is in order: when the set of transformations $\mathsf{Transf}(\mathrm{A} \to \mathrm{B})$ is convex, the pure transformations $\mathsf{PurTransf}(\mathrm{A} \to \mathrm{B})$ *may not coincide* with the extreme points of $\mathsf{Transf}(\mathrm{A} \to \mathrm{B})$. For example, in quantum theory the identity effect I_{A} is an extreme point of the set of effects, but is not pure in the sense of our definition because it can be decomposed e. g. as $I_{\mathrm{A}} = \sum_{n=1}^{d_{\mathrm{A}}} P_n$, where the effects $\{P_n = |n\rangle\langle n| \mid n = 1, \ldots, d_{\mathrm{A}}\}$ represent a projective measurement on some orthonormal basis $\{|n\rangle \mid n = 1, \ldots, d_{\mathrm{A}}\}$.

[6] Note that, in principle, our definition of "internal transformations" may not include all the transformations in the interior of the cone, because the $\lambda \, \mathcal{F}$ and $\mathcal{G}$ may fail to coexist in a test. However, this annoying discrepancy disappears under the mild assumption that the set of transformations is convex. Later, we will justify this assumption on the basis of the Causality axiom.

Definition 6. *We say that the effects in* M *are* tomographically complete *for system* A *iff, for every pair of states* ρ *and* ρ' *of system* A*, one has the implication*



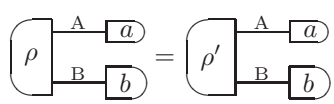
In the contrapositive: if two states are different, then the difference can be detected from the statistics of some effect in M.

Let us consider state tomography for composite systems. Suppose that two experimenters Alice and Bob perform measurements on two systems A and B, respectively, and that Alice (Bob) is able to perform the set of measurements with effects M (N). Then, by coordinating their choices of measurements and by communicating the outcomes to each other, Alice and Bob can observe the statistics of all product measurements. Hence, their set of measurement effects will be

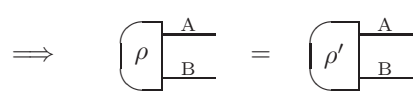
$$\mathrm{M} \otimes \mathrm{N} := \{\mathbf{m} \otimes \mathbf{n} \mid \mathbf{m} \in \mathrm{M}\,, \quad \mathbf{n} \in \mathrm{N}\}\,.$$

Now the question is: is there a choice of measurement effects M and N such that the set $\mathrm{M} \otimes \mathrm{N}$ tomographically complete? In the affirmative case, we say that system $\mathrm{A} \otimes \mathrm{B}$ *allows for local tomography*:

Definition 7. *System* $\mathrm{A} \otimes \mathrm{B}$ allows for local tomography *iff, for every pair of states* $\rho, \rho' \in \mathsf{St}(\mathrm{A} \otimes \mathrm{B})$*, one has the implication*



$$\forall a \in \mathsf{Eff}(\mathrm{A})\,, \qquad \forall b \in \mathsf{Eff}(\mathrm{B}) \tag{23}$$



(24)

More generally, we have the following

Definition 8. *An* K*-partite system* $\mathrm{A} = \bigotimes_{k=1}^{K} \mathrm{A}_k$ allows for local tomography *iff for every* $k \in \{1, \ldots, K\}$ *there exists a set of measurement effects* M_k *on system* A_k *such that the set* $\bigotimes_{k=1}^{K} \mathrm{M}_k$ *is tomographically complete.*

For a given OPT, it is easy to see that the following conditions are equivalent:

1. every multipartite system allows for local tomography
2. every bipartite system allows for local tomography.

In other words, the possibility of local tomography for arbitrary composite systems can be established by just checking bipartite systems.

3.2.4. State discrimination

The task of state discrimination can be presented as a game featuring a player and a referee. The referee prepares a physical system A in a state ρ_x, belonging to some set $\{\rho_x \mid x \in \mathrm{X}\}$ known to the player. The player is asked to guess the label x. In order to do that, she performs a measurement $\mathbf{m}$ with outcomes in X: upon finding the outcome x', she will guess that the state was $\rho_{x'}$. If the player guesses right all the times, we say that the states are perfectly distinguishable:

Definition 9. *The states* $\{\rho_x \mid x \in \mathrm{X}\}$ *are* perfectly distinguishable *iff there exists a measurement* $\mathbf{m}$ *such that*

$$(m_x|\rho_{x'}) = \delta_{x,x'} \qquad \forall x, x' \in \mathrm{X}\,.$$

When this is the case, we say that $\boldsymbol{m}$ *is a* discriminating measurement*.*

Note that, in order to be perfectly distinguishable, the states must be

1. *normalized*, namely $\|\rho_x\| = 1\,\forall x \in \mathrm{X}$, where $\|\cdot\|$ is the operational norm [1] given by $\|\rho\| = \sup_{a\in\mathsf{Eff}(\mathrm{A})}(a|\rho)$
2. *non-internal*: indeed, if a state $\rho_{x'}$ is internal, then $(m_x|\rho_{x'}) = 0$ implies $m_x = 0$, in contradiction with the condition $(m_x|\rho_x) = 1$.

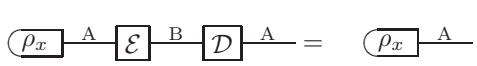
Note that *a priori* an OPT may not have any distinguishable states at all. However, the existence of distinguishable states is essential if we want our theory to include classical computation and classical information theory.

3.2.5. Ideal compression

A preparation-test $\boldsymbol{\rho} \in \mathsf{Tests}(\mathrm{I} \to \mathrm{A}, \mathrm{X})$ can be thought as describing a *source of information*. An interesting question is how well such information can be transferred from the original system to another physical support, say system B. An *encoding* of the preparation-test $\boldsymbol{\rho}$ is a deterministic transformation $\mathcal{E} \in \mathsf{DetTransf}(\mathrm{A} \to \mathrm{B})$, which transforms $\boldsymbol{\rho}$ into a new preparation-test $\boldsymbol{\rho}' := \{\mathcal{E} \circ \rho_x\}_{x\in\mathrm{X}}$. The states $\{\mathcal{E} \circ \rho_x \mid x \in \mathrm{X}\}$ are called *codewords*.

The ideal property of an encoding is to be lossless, in the following sense:

Definition 10. *An encoding* $\mathcal{E} \in \mathsf{DetTransf}(\mathrm{A} \to \mathrm{B})$ *is* lossless for the preparation-test $\boldsymbol{\rho} \in \mathsf{Tests}(\mathrm{I} \to \mathrm{A}, \mathrm{X})$ *iff there exists a deterministic transformation* $\mathcal{D} \in \mathsf{DetTransf}(\mathrm{B} \to \mathrm{A})$*, called the* decoding*, such that*



$$\forall x \in \mathrm{X}\,. \tag{25}$$

We say that

- $\mathcal{E}$ *is a* lossless encoding for the state $\rho \in \mathsf{DetSt}(\mathrm{A})$ *iff* $\mathcal{E}$ *is a lossless encoding for every ensemble decomposition of* ρ*.*
- $\mathcal{E}$ *a* lossless encoding of system A into system B *iff* $\mathcal{E}$ *is a lossless encoding for all states* $\rho \in \mathsf{DetSt}(\mathrm{A})$*.*

The notion of encoding offers an operational way to compare the size of different systems: naturally, we can say that system A is *no larger* than system B iff there exists a lossless encoding of A into B.

Among all possible encodings, we now consider the compressions:

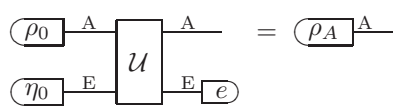
Definition 11. *A* compression *of system* A *into system* B *is an encoding* $\mathcal{E} \in \mathsf{DetTransf}(\mathrm{A} \to \mathrm{B})$ *where* B *is no larger than* A.

How much can we compress a given state? The ultimate limit to compression is when *every* state of system B is proportional to a codeword, i. e. when every state $\sigma \in \mathsf{St}(\mathrm{B})$ can be written as $\sigma = \lambda\, \mathcal{E} \rho_{x_0}$, for some scaling constant $\lambda \geq 0$ and some state ρ_{x_0} belonging to some ensemble decomposition of ρ. When this is the case, we say that the compression $\mathcal{E}$ is *maximally efficient*. Summing up, we have the following

Definition 12. *A transformation* $\mathcal{E} \in \mathsf{DetTransf}(\mathrm{A} \to \mathrm{B})$ *is* ideal compression of the state $\rho \in \mathsf{DetSt}(\mathrm{A})$ *iff it is lossless and maximally efficient.*

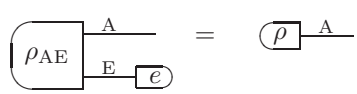
3.2.6. Simulating preparations

A state can be prepared in many different ways. For example, a state ρ_{A} could be prepared by a circuit that involves many auxiliary systems, which interact with A and are finally discarded. We refer to these systems as the *environment* and describe them collectively as a single system E. Assuming that the system and the environment are initially uncorrelated, the fact that the circuit prepares the state ρ_{A} is expressed by the diagram



(26)

where ρ_0 and η_0 are the initial states of system and environment, respectively, $\mathcal{U}$ is a transformation representing all the system-environment interaction, and e is a some effect. By defining the state $\rho_{\mathrm{AE}} := \mathcal{U}(\rho_0 \otimes \eta_0)$ the circuit of Eq. (26) can be simplified to



(27)

To capture the idea that the environment is discarded, we require the effect e to be *deterministic*:

Definition 13. *A* simulation *of the preparation* ρ_{A} *is a triple* $(\mathrm{E}, \rho_{\mathrm{AE}}, e)$ *where* E *is a system,* ρ_{AE} *is a state of* $\mathrm{A} \otimes \mathrm{E}$*, and* e *is a deterministic effect satisfying Eq. (27). If the state* ρ_{AE} *is pure, we say that* $(\mathrm{E}, \rho_{\mathrm{AE}}, e)$ *is a* pure simulation*—or, more concisely, a* purification*—of* ρ_{A}.

Purifications arise, for example, when we start from a *pure* product state $\alpha_0 \otimes \eta_0 \in \mathsf{PurSt}(\mathrm{A} \otimes \mathrm{E})$ and evolve it through a *reversible* transformation $\mathcal{U}$.

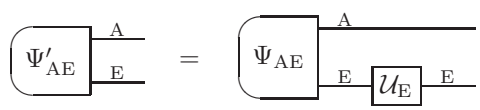
A purification gives the agent maximal control over the process of preparation: indeed, an agent possessing systems A and E can be sure that no side information can hide outside her laboratory.

Given the importance of purifications, it is important to ask how many of them can be found for a given state. From a purification there are two trivial ways to generate new ones:

1. by transforming the environment with a reversible transformation $\mathcal{U}_{\mathrm{E}}$ such that $(e|\mathcal{U} = (e|$, and
2. by appending a dummy system D to the environment, prepared in a pure deterministic state δ_{D} such that $\rho_{\mathrm{AE}} \otimes \delta_{\mathrm{D}}$ is pure.

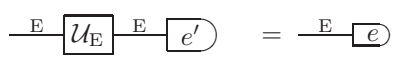
We say that a pure simulation is *essentially unique* if it is unique up to trivial transformations:

Definition 14. *A state* ρ_{A} *has an* essentially unique purification *iff for every two purifications* (E, Ψ, e) *and* (E', Ψ', e') *with* $\mathrm{E} = \mathrm{E}'$ *one has*



(28)

and [7]



(29)

for some reversible transformation $\mathcal{U}_{\mathrm{E}}$.

4. The principles

We are now ready to state our principles for quantum theory. We divide them into five *Axioms* and one *Postulate* [8]. The five axioms are

A1 **Causality.** No signal can be sent from the future to the past.

A2 **Purity of Composition.** No side information can hide in the composition of two pure transformations.

A3 **Local Tomography.** State tomography can be performed with only local measurements.

A4 **Perfect State Discrimination.** Every normalized non-internal state can be perfectly distinguished from some other state.

A5 **Ideal Compression.** Every state can be compressed in an ideal way.

[7] It turns out that the second condition is automatically satisfied if the theory satisfies the Causality axiom—see the next section.

[8] We differentiate the names in order to highlight the different roles of these principles in our reconstruction. Mathematically, there is no difference between axioms, postulates, background assumptions, and requirements in the OPT framework (all of them are "axioms"). The point of using different names is just to provide a more intuitive picture.

The five Axioms express generic and rather unsurprising features, which are common to classical and quantum theory. We regard the theories satisfying these axioms as *standard*. The Postulate is

P6 **Purification.** Every preparation can be simulated via a pure preparation in an essentially unique way.

Purification brings in a radically non-classical feature: the idea that randomness can be simulated through the preparation of pure states. We will see that this feature singles out quantum theory uniquely among all standard OPTs.

4.1. Causality

Causality states that signals cannot be sent from the future to the past. To check this condition, it is sufficient to look at a special class of circuits, consisting of a single preparation-test, followed by a single observation-test. Precisely, we have the following

Proposition 1. *An OPT satisfies Causality if and only if for every system* $\mathrm{A} \in \mathsf{Sys}$*, every preparation-test* $\boldsymbol{\rho} \in \mathsf{Tests}(\mathrm{I} \to \mathrm{A}, \mathrm{X})$ *and every pair of observation-tests* $\mathbf{m_0} \in \mathsf{Tests}(\mathrm{A} \to \mathrm{I}, \mathrm{Y}_0)$ *and* $\mathbf{m_1} \in \mathsf{Tests}(\mathrm{A} \to \mathrm{I}, \mathrm{Y}_1)$ *one has*

$$p(x|\boldsymbol{m}_0) = p(x|\boldsymbol{m}_1) \qquad \forall x \in \mathrm{X}\,,$$

with $p(x|\boldsymbol{m}_i) := \sum_{y_i \in \mathrm{Y}_i} (m_{y_i}|\rho_x)$.

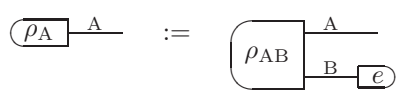
An even simpler condition for causality is given by

Proposition 2. *A theory satisfies Causality if and only if every system* A *has a unique deterministic effect* $e_{\mathrm{A}} \in \mathsf{DetEff}(\mathrm{A})$.

In categorical terms, the uniqueness of the deterministic effect can be phrased as "terminality of the tensor unit" in the category of deterministic transformations $\mathsf{DetTransf}$. Categories where the tensor unit is terminal have been introduced by Coecke and Lal [50, 51], who named them *causal categories*.

Recall that deterministic effects can be used to describe "discarding operations", whereby a physical system is eliminated from the description. Now, Causality is equivalent to the statement that every physical system can be discarded in a unique way. Thanks to Causality, we can define the marginals of a bipartite state in a canonical way

Definition 15. *Let* ρ_{AB} *be a state of system* $\mathrm{A} \otimes \mathrm{B}$*. The* marginal *of* ρ_{AB} *on system* A *is the state* ρ_{A} *defined as*



4.1.1. Causality and No-Signalling

An important consequence of Causality is the impossibility to signal without interaction: in the lack of any interaction between system A an system B, it is impossible to influence the probability distribution of a test on system A by performing tests on system B. The precise statement is the following

Proposition 3. *For every state* ρ_{AB} *and every triple of tests* $\boldsymbol{\mathcal{A}} \in \mathsf{Tests}(\mathrm{A} \to \mathrm{A}', \mathrm{X})$, $\boldsymbol{\mathcal{B}}_0 \in \mathsf{Tests}(\mathrm{B} \to \mathrm{B}'_0, \mathrm{Y}_0)$ *and* $\boldsymbol{\mathcal{B}}_1 \in \mathsf{Tests}(\mathrm{B} \to \mathrm{B}'_1, \mathrm{Y}_1)$ *one has*

$$p\left(x|\boldsymbol{\mathcal{B}}_0\right) = p\left(x|\boldsymbol{\mathcal{B}}_1\right) \qquad \forall x \in \mathrm{X}\,,$$

with $p(x|\boldsymbol{\mathcal{B}}_i) := \sum_{y_i \in \mathrm{Y}_i} \left(e_{\mathrm{B}_i} |\, \mathcal{A}_x \otimes \mathcal{B}_{i,y_i} \,|\rho_x\right)$, $i \in \{0,1\}$.

4.1.2. Causality and conditional tests

We introduced Causality as a negative statement:

C: the choice of tests performed in the future *cannot* affect the outcome probabilities of tests performed in the past.

The axiom can be reformulated in a positive, and slightly stronger way:

C′: the outcomes of tests performed in the past *can* affect the choice of tests performed in the future.

Technically, Condition **C′** establishes the possibility of performing *conditional tests*, defined as follows:

Definition 16. *Given a test* $\boldsymbol{\mathcal{T}} \in \mathsf{Tests}(\mathrm{A} \to \mathrm{B}, \mathrm{X})$ *and a collection of tests* $\{\boldsymbol{\mathcal{S}}_x \in \mathsf{Tests}(\mathrm{B} \to \mathrm{C}, \mathrm{Y}_x) \mid x \in \mathrm{X}\}$, *the* conditional test *associated to them is the collection of transformations*

$$\{\boldsymbol{\mathcal{S}}_x\} \odot \boldsymbol{\mathcal{T}} := \left\{ \xrightarrow{A} \boxed{\mathcal{T}_x} \xrightarrow{B} \boxed{\mathcal{S}^x_{y_x}} \xrightarrow{C} \;\middle|\; x \in \mathrm{X}\,, y_x \in \mathrm{Y}_x \right\}.$$

Condition **C′** states that such collection is actually a *test*, meaning that

1. the set $\mathrm{Z} = \bigcup_{x \in \mathrm{X}} \{x\} \times \mathrm{Y}_x$ belongs to $\mathsf{Outcomes}$, and
2. the collection $\{\boldsymbol{\mathcal{S}}_x\} \odot \boldsymbol{\mathcal{T}}$ belongs to $\mathsf{Tests}(\mathrm{A} \to \mathrm{C}, \mathrm{Z})$.

The relation between **C** and **C′** is the following:

1. **C′** *implies* **C**,
2. **C** implies that the theory can be enlarged to another theory satisfying **C′**: thanks to **C**, all conditional tests can be included without losing the consistency of the probabilistic structure [1].

Since conditional tests can be included, we will always assume that they *are* included, i. e. we will take the validity of **C′** as part of the Causality package.

4.1.3. Convexity

The ability to perform conditional tests brings naturally to convexity of the sets of physical transformations. This result can be obtained in two steps:

1. Under the standing assumptions that the theory is not deterministic and that the set $\mathsf{Transf}(\mathrm{I} \to \mathrm{I})$ is closed, we obtain that $\mathsf{Transf}(\mathrm{I} \to \mathrm{I})$ is the whole interval $[0, 1]$. In other words, every number in the interval $[0, 1]$ can be interpreted as the probability of some outcome in some test allowed by the theory.
2. Given two transformations $\mathcal{T}_0, \mathcal{T}_1 \in \mathsf{Transf}(\mathrm{A} \to \mathrm{B})$, the convex combination $p\,\mathcal{T} + (1-p)\,\mathcal{T}'$ can be generated by
   (a) performing a binary test with the outcomes 0 and 1 generated with probabilities $p_0 = p$ and $p_1 = 1 - p$
   (b) conditionally on the occurrence of the outcome i, performing a test $\boldsymbol{\mathcal{T}}_i$ containing the transformation $\mathcal{T}_i$
   (c) coarse-graining over the appropriate outcomes of the conditional test.

The above observations show that convexity needs not be assumed from the start, but can be derived from non-determinism and Causality (in the positive formulation $\mathbf{C}'$), under the standard assumption that the set of probabilities generated by tests in the theory is closed.

4.1.4. Rescaling

In addition to convexity, conditional tests guarantee that every state is proportional to a *normalized* state. Specifically, given a state ρ of a generic system A, one can define the normalized state $\widetilde{\rho} := \rho/(e_{\mathrm{A}}|\rho)$. An approximate way to prepare the state $\widetilde{\rho}$ is to

1. pick a binary test $\{\rho_0, \rho_1\}$ such that $\rho_1 = \rho$
2. perform it N times, generating a string of outcomes $(x_1, x_2, \ldots, x_N)$
3. perform a conditional test that discards $N - 1$ systems, keeping only a system i such that $x_i = 1$, if such a system exists, or otherwise keeping only the first system
4. coarse-grain over all outcomes, thus obtaining the deterministic state

$$\rho_N := (1 - p_N)\,\widetilde{\rho} + p_N\,\widetilde{\rho}_0 \qquad\qquad p_N = (e_{\mathrm{A}}|\rho_0)^N\,.$$

Clearly, the state ρ_N converges to $\widetilde{\rho}$ when N goes to infinity. Hence, the standard assumption that the set of states is closed guarantees that $\widetilde{\rho}$ is a state allowed by the theory.

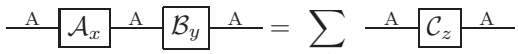
4.2. Purity of Composition

Purity of Composition is a very primitive rule about how information propagates in time. Mathematically, the axiom consists of the implication

$$\begin{aligned} &\mathcal{A} \in \mathsf{PurTransf}(\mathrm{A} \to \mathrm{B})\,,\ \mathcal{B} \in \mathsf{PurTransf}(\mathrm{B} \to \mathrm{C}) \\ &\qquad\qquad \Longrightarrow \qquad \mathcal{B} \circ \mathcal{A} \in \mathsf{PurTransf}(\mathrm{A} \to \mathrm{C})\,, \end{aligned}$$

required to be valid for all systems $\mathrm{A}, \mathrm{B}, \mathrm{C} \in \mathsf{Sys}$ and for all pure transformations $\mathcal{A}$ and $\mathcal{B}$.

Think of a world where this were not the case. In that world, an agent Alice could perform a test $\boldsymbol{\mathcal{A}} \in \mathsf{Tests}(\mathrm{A} \to \mathrm{B}\,, \mathrm{X})$ with such degree of control that, upon knowing the outcome, she could not possibly know better what happened to her system. Immediately after, another agent Bob could perform another test $\boldsymbol{\mathcal{B}} \in \mathsf{Tests}(\mathrm{B} \to \mathrm{C}\,, \mathrm{Y})$ also having maximal knowledge of the system's conditional evolution. Still, some of the resulting transformations $\mathcal{B}_y \mathcal{A}_x$ may not be pure. This means that $\mathcal{B}_y \mathcal{A}_x$ can be simulated by a third party—Charlie—by performing one test $\{\mathcal{C}_z\}_{z \in \mathrm{Z}}$ and joining together the outcomes in a suitable subset $\mathrm{S}_{xy} \subset \mathrm{Z}$



(30)

Although this scenario is logically conceivable, it rises a puzzling question: What is the extra information about? Which physical parameters correspond to the outcome z? Surely the information is not about what happened in the first step, because Alice already had maximal knowledge about this. Nor it is about what happened in the second step, because Bob has maximal information about that. The outcome z has to specify a feature of how the two time steps interacted together—in a sense, a kind of information that is *non-local in time*. Quantum theory is non-local, but not in such an extreme way! Indeed, pure transformations in quantum theory are described by completely positive maps with a single Kraus operator, i. e. of the form $\mathcal{A}_x(\cdot) = A_x \cdot A_x^\dagger$ and $\mathcal{B}_y(\cdot) = B_y \cdot B_y^\dagger$, and clearly the composition of two pure transformations is still pure: $\mathcal{B}_y \mathcal{A}_x(\cdot) = (B_y A_x) \cdot (B_y A_x)^\dagger$. Purity of Composition guarantees this property at the level of first principles.

4.3. Local Tomography

Local Tomography implies that even if a state is entangled, the information it contains can be extracted by local measurements. This fact reconciles the holism of entanglement and the reductionist idea that the full information about a composite system can be obtained by studying its parts [27].

Mathematically, Local Tomography states that product effects form a separating set for the vector space $\mathsf{St}_{\mathbb{R}}(\mathrm{A} \otimes \mathrm{B})$. Equivalently[9], they form a spanning set for the dual space $\mathsf{St}_{\mathbb{R}}(\mathrm{A} \otimes \mathrm{B})^* \equiv \mathsf{Eff}_{\mathbb{R}}(\mathrm{A} \otimes \mathrm{B})$. Hence, we must have the conditions

$$\mathsf{Eff}_{\mathbb{R}}(\mathrm{A} \otimes \mathrm{B}) = \mathsf{Eff}_{\mathbb{R}}(\mathrm{A}) \otimes \mathsf{Eff}_{\mathbb{R}}(\mathrm{B}) \qquad \text{and} \qquad \mathsf{St}_{\mathbb{R}}(\mathrm{A} \otimes \mathrm{B}) = \mathsf{St}_{\mathbb{R}}(\mathrm{A}) \otimes \mathsf{St}_{\mathbb{R}}(\mathrm{B})\,, \tag{31}$$

where $\otimes$ in the r.h.s. denoted the tensor product of finite dimensional vector spaces. Eq. (31) implies that the dimensions of the vector spaces in question

[9] Recall that we are assuming that the state spaces are finite-dimensional.

satisfy the product relation [25]

$$D_{\mathrm{A}\otimes\mathrm{B}} = D_{\mathrm{A}}\, D_{\mathrm{B}}\,. \tag{32}$$

Moreover, a generic state $\rho \in \mathsf{St}(\mathrm{A}\otimes\mathrm{B})$ and a generic effect $m \in \mathsf{Eff}(\mathrm{A}\otimes\mathrm{B})$ can be expanded as

$$\rho = \sum_{i,j} \rho_{ij}\ (v_i \otimes w_j) \qquad \text{and} \qquad m = \sum_{i,j} m_{ij}\ \left(v_i^* \otimes w_j^*\right)\ , \tag{33}$$

where $[\rho_{ij}]$ and $[m_{ij}]$ are real matrices, $\{v_i\}_{i=1}^{D_{\mathrm{A}}}$ and $\{w_j\}_{j=1}^{D_{\mathrm{B}}}$ are bases for the vector spaces $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$ and $\mathsf{St}_{\mathbb{R}}(\mathrm{B})$, respectively, and $\{v_i^*\}_{i=1}^{D_{\mathrm{A}}}$ and $\{w_j^*\}_{j=1}^{D_{\mathrm{B}}}$ are the dual bases, defined by the relations $(v_i^*|v_k) = \delta_{ik}$ and $(w_j^*|w_l) = \delta_{jl}$, respectively. As a result, the probability of the effect m on the state ρ can be expressed as

$$(m|\rho) = \mathrm{Tr}[m\,\rho]\,, \tag{34}$$

having committed a little abuse of notation in using the letter m (respectively, ρ) both for the effect (respectively, state) and for the corresponding matrix $[m_{ij}]$ (respectively, $[\rho_{ij}]$).

Finally, the decomposition in Eq. (33) implies the following

Theorem 1. *In a theory satisfying Local Tomography, the correspondence between a transformation $\mathcal{E} \in \mathsf{Transf}(\mathrm{A}\to\mathrm{B})$ and the linear map $\widehat{\mathcal{E}} : \mathsf{St}_{\mathbb{R}}(\mathrm{A}) \to \mathsf{St}_{\mathbb{R}}(\mathrm{B})$ is invertible.*

In other words, Local Tomography guarantees that physical transformations can be characterized in the simplest possible way: by preparing a set of input states and performing a set of measurements on the output.

A remarkable example of a theory that does not satisfy Local Tomography is quantum theory on real Hilbert spaces [52], RQT for short. In this theory, states and effects are real symmetric matrices, and transformations are represented by completely positive maps mapping symmetric matrices into symmetric matrices. The failure of the relation $D_{\mathrm{A}\otimes\mathrm{B}} = D_{\mathrm{A}}\, D_{\mathrm{B}}$ was first noted by Araki [53]. More explicitly, Wootters [54] noted that two different bipartite states can be locally indistinguishable, as in the following extreme example:

$$\rho = \frac{1}{2}|\Phi_+\rangle\langle\Phi_+| + \frac{1}{2}|\Psi_-\rangle\langle\Psi_-| \qquad\qquad \rho' = \frac{1}{2}|\Phi_-\rangle\langle\Phi_-| + \frac{1}{2}|\Psi_+\rangle\langle\Psi_+| \tag{35}$$

with $|\Phi_\pm\rangle := (|0\rangle|0\rangle \pm |1\rangle|1\rangle)/\sqrt{2}$ and $|\Psi_\pm\rangle = (|0\rangle|1\rangle \pm |1\rangle|0\rangle)/\sqrt{2}$. Here the states ρ and ρ' have orthogonal support and therefore are perfectly distinguishable. However, it is easy to check that one has

$$\rho - \rho' = \frac{1}{2}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\ ,$$

and, therefore, $\mathrm{Tr}[(\rho-\rho')(P_{\mathrm{A}}\otimes P_{\mathrm{B}})] = 0$ for every pair of real symmetric matrices P_{A} and P_{B}. In other words, ρ and ρ' give exactly the same statistics for all possible local measurements.

RQT has another, closely related quirk: two *different* transformations of system A can act in the same way on all states of A. For example, consider the qubit channels $\mathcal{C}$ and $\mathcal{C}'$, whose action on a generic 2×2 matrix is defined by

$$\mathcal{C}(M) := \frac{1}{2}\, M + \frac{1}{2}\, YMY \qquad \text{and} \qquad \mathcal{C}'(M) := \frac{1}{2}\, ZMZ + \frac{1}{2}\, XMX\,,$$

X, Y, and Z being the Pauli matrices. When acting on symmetric matrices, the two channels give exactly the same output: one has $\mathcal{C}(\tau) = \mathcal{C}'(\tau) = I/2$ for every symmetric matrix τ. On the other hand, one has

$$(\mathcal{C} \otimes \mathcal{I})(|\Phi_+\rangle\langle\Phi_+) = \rho \qquad (\mathcal{C}' \otimes \mathcal{I})(|\Phi_+\rangle\langle\Phi_+) = \rho'\,,$$

where ρ and ρ' are the two perfectly distinguishable states defined in Eq. (35) above. This means that, in fact, the two transformations $\mathcal{C}$ and $\mathcal{C}'$ are perfectly distinguishable with the help of a reference system. For a more extensive discussion of tomography in RQT we refer the reader to subsection V.A of Ref. [1] and to the work of Hardy and Wootters [55].

4.4. Perfect State Discrimination

Perfect State Discrimination is an optimistic statement about the possibility to encode bits without error. It guarantees that every state that *could* be part of a set of perfectly distinguishable states *is* indeed perfectly distinguishable from some other state.

By virtue of Perfect State Discrimination, every normalized non-internal state ρ_0 can be perfectly distinguished from some state ρ_1. As a result, the two states ρ_0 and ρ_1 can be used to encode the value of a bit without errors. It is easy to see that Quantum theory satisfies the axiom. Indeed, Aadensity matrix is internal if and only if it has full rank. Hence, a non-internal density matrix ρ_0 must have a kernel, so that every state ρ_1 with support in the kernel of ρ_0 will be perfectly distinguishable from ρ_0.

4.5. Ideal Compression

Ideal Compression expresses the idea that information is *fungible*, i. e. independent of the physical support in which it is encoded. The axiom implies non-trivial statements about the state spaces arising in the theory. For example, suppose that the theory contains a system whose space of deterministic states is a square. Then, the theory should contain also a system whose space of deterministic states is a segment—in other words, the theory should contain a classical bit. Indeed, only in this way one could encode a side of the square in a lossless and maximally efficient way. More generally, Ideal Compression imposes that the every face of the convex set of deterministic states be in one-to-one correspondence with the set of deterministic states of some smaller physical system.

Ideal Compression is clearly satisfied by quantum theory. Indeed, every density matrix of rank r can be compressed ideally to a density matrix of an

r-dimensional quantum system. For example, the two-qubit density matrix

$$\rho = \begin{pmatrix} \rho_{00,00} & 0 & 0 & \rho_{00,11} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \rho_{11,00} & 0 & 0 & \rho_{11,11} \end{pmatrix} \tag{36}$$

can be compressed ideally to the one-qubit density matrix

$$\mathcal{E}(\rho) = \begin{pmatrix} \rho_{00,00} & \rho_{00,11} \\ \rho_{11,00} & \rho_{11,11} \end{pmatrix} \tag{37}$$

with encoding and decoding channels given by

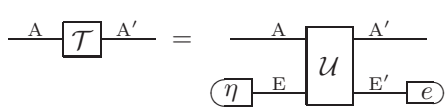
$$\mathcal{E}(\cdot) := V^\dagger\,(\cdot)\,V + \mathrm{Tr}[(I - V\,V^\dagger)\,(\cdot)\,]\,|0\rangle\langle 0| \qquad V := |0\rangle|0\rangle\langle 0| + |1\rangle|1\rangle\langle 1|$$
$$\mathcal{D}(\cdot) := V\,(\cdot)\,V^\dagger\,.$$

Note that Ideal Compression refers to a *single-shot, zero error scenario*, i. e. a scenario where the source is used only once and no decoding errors are allowed. Such a scenario is different from the asymptotic scenario considered in Shannon's [56] and Schumacher's [57] compression, wherein small decoding errors are allowed, under the condition that they vanish in the asymptotic limit of infinitely many uses of the same source.

4.6. Purification

While our first five axioms expressed standard requirements for information-processing, Purification brings in a radically new idea: at least in principle, every state can be prepared by an agent who has maximal control over all the systems involved in the preparation process. In short, Purification allows us to harness randomness by controlling the environment. The idea does not apply only to preparations, but also to arbitrary deterministic transformations: combining Purification with Causality and Local Tomography, we can prove the following

Theorem 2. *[1] For every deterministic transformation* $\mathcal{T} \in \mathsf{DetTransf}(\mathrm{A} \to \mathrm{A}')$*, there exist two systems* E *and* E'*, a pure state* $\eta \in \mathsf{PurSt}(\mathrm{E})$*, and a reversible transformation* $\mathcal{U} \in \mathsf{RevTransf}(\mathrm{A} \otimes \mathrm{E} \to \mathrm{A}' \otimes \mathrm{E}')$ *such that*



(38)

where e *is the unique deterministic effect of system* E'*.*

In other words, Purification implies that every irreversible process can be simulated through reversible interactions between the system and its environment, with the environment initialized in a pure state. This result is a necessary condition for the formulation of physical theories in which elementary processes are reversible at the fundamental level.

Purification is known to be satisfied by quantum mechanics. For example, consider a single-qubit mixed state, diagonalized as

$$\rho = p_0 |0\rangle\langle 0| + p_1 |1\rangle\langle 1| \,, \tag{39}$$

for some suitable orthonormal basis $\{|0\rangle, |1\rangle\}$. A purification of the state ρ can be obtained by adding a second qubit and by preparing the two qubits in the pure state

$$|\Psi\rangle := \sqrt{p_0}\, |0\rangle|0\rangle + \sqrt{p_1}\, |1\rangle|1\rangle \,. \tag{40}$$

Indeed, it is immediate to see that ρ is the marginal of the density matrix $|\Psi\rangle\langle\Psi|$ on the first qubit. In addition, any other purification $|\Psi'\rangle$—using a single qubit as the purifying system—must be of the form $|\Psi\rangle' = (I \otimes U)\, |\Psi\rangle$ for some unitary matrix U.

In the quantum information community, taking purifications is a standard approach to quantum communication, cryptography, and quantum error correction. The approach is familiarly known with the nickname of "going to the Church of the larger Hilbert space" [10]. Purification is known among mathematicians as the *Gelfand-Naimark-Segal construction* [59, 60].

Two important remarks are in order:

1. *Purification, entanglement, and quantum information.* Purification is intimately linked with the phenomenon of *entanglement* [4], namely the existence of pure bipartite states $\Psi_{\rm AB}$ that are not of the product form $\psi_{\rm A} \otimes \psi_{\rm B}$. In the OPT framework, the link is made precise by the following

   Proposition 4. *Let Θ be a theory satisfying Causality, Local Tomography, and Purification. Then, there are only two alternatives: either Θ is deterministic, or Θ exhibits entanglement.*

   Under our standing assumption that the theory is non-deterministic, entanglement follows from Purification as a necessary consequence.
   Entanglement is a very peculiar feature—far from what we experience in our everyday life. How can we claim that we know A *and* B if we do not know A alone? This puzzling feature had been noted already in the early days of quantum theory, when Schrödinger famously wrote: "Another way of expressing the peculiar situation is: *the best possible knowledge of a whole does not necessarily include the best possible knowledge of all its parts*" [4]. And, in the same paper: "I would not call that *one* but rather *the* characteristic trait of quantum mechanics, the one that enforces its entire departure from classical lines of thought". In a sense, our reconstruction can be considered as a mathematical proof of Schrödinger's intuition [11]: on the background of five standard axioms

[10] The expression is due to John Smolin, see e.g. the lecture notes [58].

[11] It is worth stressing that Schrödinger's paper was not just about the *existence* of entangled states, but also about how entanglement interacted with the reversible dynamics and with the process of measurement (cf. the notion of *steering*, which made its first appearance in the very same paper).

satisfied by both classical theory and quantum theory, Purification is the ingredient that allows to reconstruct the Hilbert space framework and the distinctive information-theoretic features of quantum theory. Combined with Causality and Local Tomography, Purification already reproduces an impressive list of quantum-like features, like no-cloning, no-programming, information-disturbance tradeoff, no bit commitment, conclusive teleportation and entanglement swapping, the reversible dilation of channels, the state-transformation isomorphism, the structure of error correction, and the structure of no-signalling channels [1].

2. *Purification and the Many Wolrd Interpretation.* Pondering about the meaning of Purification, one may tempted to conclude that it favours the Many Worlds Interpretation (MWI) of quantum mechanics [61]. In fact, Purification is feature of quantum theory, and, as such, it does not favour the MWI more than quantum theory itself does. Whether or not quantum theory provides any evidence for many worlds is a debatable point, but the validity of Purification is independent such interpretative issue. Furthermore, we stress that we did not phrase Purification as an ontological statement about "how processes occur in nature", but rather an operational statement about the agent's ability to simulate physical processes with maximal control. Purification is *compatible* with the idea that processes are reversible at the fundamental level, and its validity is a *necessary condition* for building up a physical description of nature in terms of pure states and reversible processes. Still, here we do not make any commitment as to how processes are realized in nature, because this would unnecessarily limit the range of application of our results.

5. The reconstruction of Quantum Theory

Here we provide a summary of the reconstruction of Refs. [1, 2], highlighting the key theorems and providing a guide to the original papers. The scope of the reconstruction is not just to derive the Hilbert space framework, but also to rebuild the key quantum features directly from first principles. Accordingly, we try to derive as much as possible of quantum theory directly from the axioms, leaving Hilbert spaces to the very end. We organize our results in six subsections:

1. Elementary facts.
2. Correlation structures.
3. Distinguishability structures.
4. Interaction between correlation and distinguishability structures.
5. Qubit features.
6. The density matrix.

5.1. Elementary facts

5.1.1. From Local Tomography

Local Tomography implies a few useful facts:

1. If $\alpha \in \mathsf{St}(\mathrm{A})$ and $\beta \in \mathsf{St}(\mathrm{B})$ are pure, then also $\alpha \otimes \beta$ is pure.
2. Let ρ_{AB} be a state of the composite system $\mathrm{A} \otimes \mathrm{B}$ and, assuming Causality, let ρ_{A} be its marginal on system A. If ρ_{A} is pure, then ρ_{AB} is a product state.
3. If $\rho_{\mathrm{A}} \in \mathsf{St}(\mathrm{A})$ and $\rho_{\mathrm{B}} \in \mathsf{St}(\mathrm{B})$ are internal states, then also $\rho_{\mathrm{A}} \otimes \rho_{\mathrm{B}}$ is an internal state.
4. Suppose that every system A has a unique *invariant state* χ_{A}, i. e. a unique state satisfying the condition $\mathcal{U}\chi_{\mathrm{A}} = \chi_{\mathrm{A}}$ for every reversible transformation $\mathcal{U}$. Then, $\chi_{\mathrm{A}\otimes\mathrm{B}} = \chi_{\mathrm{A}} \otimes \chi_{\mathrm{B}}$.

5.1.2. From Purification

Purification has a few immediate consequences. First, all pure states of a given system are connected to one another through reversible transformations:

Proposition 5. *For every system* $\mathrm{A} \in \mathsf{Sys}$ *and every pair of pure states* $\alpha, \alpha' \in \mathsf{PurSt}(\mathrm{A})$ *there exists a reversible transformation* $\mathcal{U}$ *such that* $\alpha' = \mathcal{U}\,\alpha$.

To prove this fact, it is enough to pick a system B and pure state $\beta \in \mathsf{PurSt}(\mathrm{B})$, consider the states $\Psi = \alpha \otimes \beta$ and $\Psi' = \alpha' \otimes \beta$ as purifications of β, and invoke the essential uniqueness of purification [Eq. (28)]. Mathematically, the above proposition expresses the fact that the action of the reversible transformations is *transitive* on the manifold of pure states—a requirement that played an important role in many recent reconstructions, see e. g. [25, 30, 29]. A byproduct of transitivity is

Proposition 6. *Every system* $\mathrm{A} \in \mathsf{Sys}$ *has a unique invariant state* χ_{A}.

Finally, combining Ideal Compression and Purification it is easy to see that every state has a *minimal purification*, in the following sense

Definition 17. *Let* $\Psi \in \mathsf{PurSt}(\mathrm{A} \otimes \mathrm{B})$ *be a pure state with marginals* ρ_{A} *and* ρ_{B} *on systems* A *and* B*, respectively. We say that* Ψ *is a* minimal purification *of* ρ_{A} *iff* ρ_{B} *is internal.*

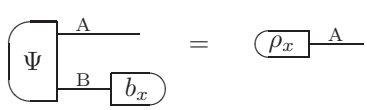
To construct a minimal purification, it is enough to take an arbitrary purification and to compress the state of the purifying system.

5.2. Correlation structures

5.2.1. Pure Steering

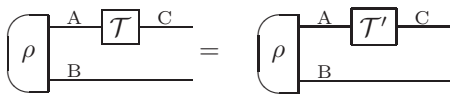
One of the most important consequences of our axioms is that pure bipartite states enable *steering*, namely the ability to remotely generate every desired ensemble decomposition of a marginal state [4, 62]:

Proposition 7 (Pure Steering)**.** *Let* Ψ *be a pure state of the composite system* $\mathrm{A} \otimes \mathrm{B}$*, let* ρ *be the marginal of* Ψ_{AB} *on system* A*, and let* $\boldsymbol{\rho} = \{\rho_x\}_{x\in\mathrm{X}}$ *be an ensemble decomposition of* ρ*. Then there exists a measurement* $\boldsymbol{b} = \{b_x\}_{x\in\mathrm{X}}$ *such that*



$$\forall x \in \mathrm{X}\,. \tag{41}$$

Pure steering is the essential ingredient for a number of major results. The first result is the existence of *pure, tomographically faithful states*. A state $\rho \in \mathsf{St}(\mathrm{A} \otimes \mathrm{B})$ is called *tomographically faithful* for system A iff the implication



$$\Longrightarrow \qquad \mathcal{T} = \mathcal{T}' \,, \tag{42}$$

holds for every system C and every pair of transformations $\mathcal{T}$ and $\mathcal{T}'$ of type $\mathrm{A} \to \mathrm{C}$. Thanks to Pure Steering and Local Tomography, we are able to construct tomographically faithful states:

Proposition 8. *Let ρ_{A} be an internal state of system* A *and let $\Psi \in \mathsf{PurSt}(\mathrm{A} \otimes \mathrm{B})$ be a purification of ρ_{A}. Then, Ψ is tomographically faithful for system* A.

The result can be improved by choosing a *minimal* purification: in this way, the pure state Ψ is faithful on both systems A and B. We call such a state *doubly faithful*.

5.2.2. Conjugate systems

A canonical choice of doubly faithful state is obtained by picking a minimal purification of the invariant state χ_{A}. We denote such purification by $\Phi \in \mathsf{PurSt}(\mathrm{A} \otimes \overline{\mathrm{A}})$ and call system $\overline{\mathrm{A}}$ the *conjugate of system* A. The name is motivated by the following facts:

1. system $\overline{\mathrm{A}}$ is uniquely defined, up to operational equivalence
2. the marginal of Φ on system $\overline{\mathrm{A}}$ is the invariant state $\chi_{\overline{\mathrm{A}}}$ (cf. Corollary 46 of [1]), meaning that we have $\overline{\overline{\mathrm{A}}} = \mathrm{A}$, up to operational equivalence.

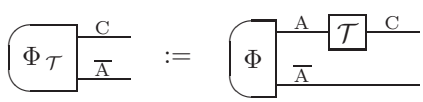
Summarizing, the state Φ satisfies the relations

$$\tag{43}$$

By analogy with quantum theory, we call Φ a *Bell state*.

5.2.3. The state-transformation isomorphism

For a given transformation $\mathcal{T}$, we define the (generally unnormalized) state



$$\tag{44}$$

and call the correspondence $\mathcal{T} \mapsto \Phi_{\mathcal{T}}$ the *state-transformation isomorphism*. Since the Bell state Φ is doubly faithful, the correspondence is one-to-one. In quantum theory, the state-transformation isomorphism concides with the Choi isomorphism [63]. By analogy, we call the state $\Phi_{\mathcal{T}}$ the *Choi state*.

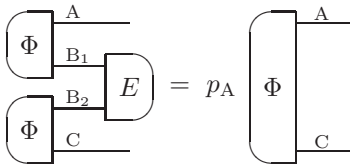
A powerful byproduct of the state-transformation isomorphism is that the normalized states completely identify the theory:

Theorem 3. *Let Θ and Θ' be two theories with the same set of systems. If the sets of normalized states of Θ and Θ' coincide for all systems, then the two theories coincide.*

Thanks to this result, deriving the density matrix representation of normalized states is sufficient to derive the whole of quantum theory.

5.2.4. Conclusive entanglement swapping

An important consequence of Pure Steering is the possibility of entanglement swapping, namely the possibility to generate entanglement remotely by performing a joint measurement. Consider, as a prototype of entangled state, the Bell state Φ. Then, it is possible to show that there exists a pure effect $E \in \mathsf{PurEff}(\overline{\mathrm{A}} \otimes \mathrm{A})$ and a non-zero probability $p_{\mathrm{A}} > 0$ such that



$$\mathrm{A} \equiv \mathrm{B}_2\,, \qquad \mathrm{B}_1 \equiv \mathrm{C} \equiv \overline{\mathrm{A}}\,. \tag{45}$$

This diagram represents an instance of *conclusive entanglement swapping*: conditionally on the occurrence of the effect E, the two systems A and C are prepared in the Bell state, consuming the initial entanglement present in the composite systems $\mathrm{A} \otimes \mathrm{B}_1$ and $\mathrm{B}_2 \otimes \mathrm{C}$.

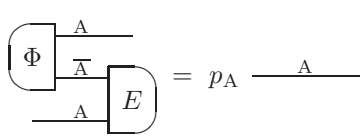
The possibility of entanglement swapping follows easily from Pure Steering: Since the states χ_{A} and $\chi_{\overline{\mathrm{A}}}$ are internal, Local Tomography implies that their product $\chi_{\mathrm{A}} \otimes \chi_{\overline{\mathrm{A}}}$ is internal. Hence, there must exist a non-zero probability $p_{\mathrm{A}} > 0$ such that

$$\chi_{\mathrm{A}} \otimes \chi_{\overline{\mathrm{A}}} = p_{\mathrm{A}}\, \Phi + (1 - p_{\mathrm{A}})\, \tau\,, \tag{46}$$

for some state τ. Applying Pure Steering (proposition 7) to the pure state $\Phi \otimes \Phi$ and to the ensemble $\{p_{\mathrm{A}}\, \Phi\,, (1 - p_{\mathrm{A}})\, \tau\}$ one can find a binary measurement $\{E, e_{\mathrm{B}_1} \otimes e_{\mathrm{B}_2} - E\}$ such that the entanglement swapping condition (45) holds. Using the fact that the state $\Phi \otimes \Phi$ is doubly faithful, it is easy to see that the effect E must be pure.

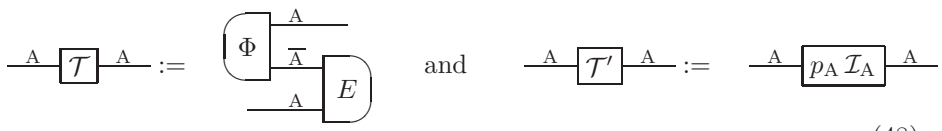
5.2.5. Conclusive teleportation

By the state-transformation isomorphism, conclusive entanglement swapping is equivalent to *conclusive teleportation* [19], expressed by the diagram



$$. \tag{47}$$

Indeed, the entanglement swapping diagram (45) is equivalent to the condition $\Phi_{\mathcal{T}} = \Phi_{\mathcal{T}'}$, with



(48)

By the state-transformation isomorphism, $\Phi_{\mathcal{T}} = \Phi_{\mathcal{T}'}$ implies $\mathcal{T} = \mathcal{T}'$, which is nothing but the teleportation diagram.

5.2.6. The teleportation upper bound

Combined with Local Tomography, the teleportation diagram allows us to upper bound the dimension of the state space. The idea is to write the teleportation diagram in matrix elements, by expanding Φ and E as

$$\Phi = \sum_{ik} \Phi_{ik}\ (v_i \otimes w_k) \qquad \text{and} \qquad E = \sum_{jl} E_{jl}\ \left(w_j^* \otimes v_l^*\right)\ , \tag{49}$$

with suitable bases $\{v_i\}_{i=1}^{D_{\mathrm{A}}}$ and $\{w_j\}_{j=1}^{D_{\overline{\mathrm{A}}}}$. In this representation, Eq. (47) becomes

$$[\,\Phi\, E\,]_{il} = p_{\mathrm{A}}\, \delta_{il}\,, \tag{50}$$

and, taking the trace,

$$\mathrm{Tr}[\Phi E] = p_{\mathrm{A}}\, D_{\mathrm{A}}\,. \tag{51}$$

On the other hand, we have

$$\mathrm{Tr}[\Phi E] = (E|\,\mathcal{S}_{\mathrm{A},\overline{\mathrm{A}}}\,|\Phi) \leq 1\,, \tag{52}$$

which combined with Eq. (51) leads to bound

$$D_{\mathrm{A}} \leq \frac{1}{p_{\mathrm{A}}}\,. \tag{53}$$

Clearly, in order to have the best bound we need to find the *maximum* probability of teleportation. To discover what the maximum is, we need to move our attention to the distinguishability structures implied by our axioms.

5.3. Distinguishability structures

5.3.1. No disturbance without information

Our first move is to derive a simple result about the structure of measurements: a measurement that extracts no information from a face of the state space can be implemented without disturbing that face. By *face of the state*

space we mean a face of the convex set of deterministic states[12]. We say that the measurement $\boldsymbol{m} \in \mathsf{Tests}(\mathrm{A} \to \mathrm{I}, \mathrm{X})$ *does not extract information* from the face F iff there exists a set of probabilities $\{p_x\}_{x \in \mathrm{X}}$ such that

$$(m_x|\tau) = p_x \qquad \forall x \in \mathrm{X}, \quad \forall \tau \in F\,.$$

Also, we say that a test $\boldsymbol{\mathcal{T}} \in \mathsf{Tests}(\mathrm{A} \to \mathrm{A}, \mathrm{X})$ *does not disturb the face* F iff $\sum_{x \in \mathrm{X}} \mathcal{T}_x|\tau) = |\tau)$ for every state $\tau \in F$.

With this terminology, our result is the following:

Proposition 9. *If a measurement* $\boldsymbol{m}$ *does not extract information from the face* F*, then there exists a test* $\boldsymbol{\mathcal{T}}$ *that realizes the measurement—namely* $(e_{\mathrm{A}}|\mathcal{T}_x = m_x$*,* $\forall x \in \mathrm{X}$*—and does not disturb* F*.*

This result has two important consequences. First, it allows us to establish whether or not a set of perfectly distinguishable set can be extended:

Proposition 10. *Let* $\mathrm{S} = \{\rho_x \mid x \in \mathrm{X}\}$ *be a set of perfectly distinguishable states and let* ω_{S} *be its barycenter, defined as*

$$\omega_{\mathrm{S}} := \frac{1}{|\mathrm{X}|} \sum_{x \in \mathrm{X}} \rho_x\,.$$

Then, the following are equivalent:

1. *the set* S *is* maximal*, i. e. no other set* $\mathrm{S}' \supset \mathrm{S}$ *can consist of perfectly distinguishable states*
2. *the barycenter of* S *is internal.*

Another important consequence is that only the *pure* maximal sets can have maximum cardinality:

Proposition 11. *Let* S *be a maximal set of perfectly distinguishable states of system* A*. If one of the states in* S *is not pure, then there exists another maximal set* $\mathrm{S}' \subset \mathsf{St}(\mathrm{A})$*, consisting only of pure states and having strictly larger cardinality* $|\mathrm{S}'| > |\mathrm{S}|$*.*

Combining the above points we have that every pure state belongs to some maximal set of perfectly distinguishable pure states. For short, we call such such sets *pure maximal sets*.

5.3.2. Duality between pure states and pure effects

For a pure maximal set S, we observe that the measurement that distinguishes the states in S must consist of *pure* effects. Hence, for every pure state $\alpha \in \mathsf{PurSt}(\mathrm{A})$ there exists an pure effect a such that $(a|\alpha) = 1$. Expanding

[12] We recall that a *face* of a convex set C is a convex subset $F \subseteq C$ satisfying the condition that, for every $x \in F$, if x is a non-trivial convex combination of x_1 and x_2 with $x_1, x_2 \in C$, then x_1 and x_2 belong to F.

on this observation, we establish a one-to-one correspondence between pure normalized states and pure normalized effects [13], denoted by $\mathsf{PurSt}_1(\mathrm{A})$ and $\mathsf{PurEff}_1(\mathrm{A})$, respectively.

Theorem 4. *For every system* $\mathrm{A} \in \mathsf{Sys}$*, there exists a one-to-one map* $\dagger : \mathsf{PurSt}_1(\mathrm{A}) \to \mathsf{PurEff}_1(\mathrm{A})$*, sending pure normalized states to pure normalized effects and satisfying the condition*

$$(\alpha^\dagger|\alpha) = 1 \qquad \forall \alpha \in \mathsf{PurSt}_1(\mathrm{A})\,.$$

The proof is rather elaborate. The two main steps are

1. proving that every pure normalized effect a *identifies* a pure state α, meaning that $(a|\rho) = 1$ if and only if $\rho = \alpha$.
2. proving that, if two pure effects identify the same state, then they must coincide.

The second step uses Pure Steering in an essential way, suggesting that the distinguishability features of quantum theory are deeply connected with its correlation features.

5.3.3. The informational dimension

An easy consequence of the state-effect duality is that every two pure normalized effects are connected by a reversible transformation, just like the pure states. In turn, this leads to a useful result

Proposition 12. *For a given system* $\mathrm{A} \in \mathsf{Sys}$*, all pure maximal sets have the same cardinality.*

The proof idea is simple: let $\boldsymbol{a} = \{a_x\}_{x\in\mathrm{X}}$ be the measurement that distinguishes among the states in a pure maximal set $\mathrm{S} = \{\alpha_x \mid x \in \mathrm{X}\}$. As we already observed, all the effects in $\boldsymbol{a}$ must be pure. Since every two pure normalized effects are connected by a reversible transformation, we must have $a_x = a \circ \mathcal{U}_x \, \forall x \in \mathrm{X}$, where a is fixed (but otherwise arbitrary) effect in $\mathsf{PurEff}_1(\mathrm{A})$ and $\mathcal{U}_x$ is a reversible transformation. Applying the effects to the invariant state χ we then obtain

$$(a_x|\chi) = (a|\chi) \qquad \forall x \in \mathrm{X}\,,$$

and summing over x we get the equality $1 = |\mathrm{X}|\,(a|\chi)$. Hence, the cardinality of the maximal set S is $|\mathrm{S}| \equiv |\mathrm{X}| = 1/(a|\chi)$. Since S is a generic pure maximal set, we proved the desired result.

In the following, the cardinality of the pure maximal sets in A be denoted by d_{A}. We call it the *informational dimension*, because it is the number of distinct classical messages that can be encoded in system A and decoded without error.

[13] We call an effect of system A *normalized* iff there exists an effect a state ρ such that $(a|\rho) = 1$.

In Quantum Theory, $d_{\rm A}$ is the dimension of the Hilbert space associated to system A.

For composite systems, the informational dimension has the product form:

Proposition 13. *For every pair of systems* A *and* B *one has* $d_{\rm A\otimes B} = d_{\rm A}\, d_{\rm B}$.

The reason is simply that the product of two pure maximal sets for systems A and B is a pure maximal set for $\rm A\otimes B$: it is pure, because the product of two pure states is pure (by Local Tomography) and it is maximal because the product of two internal states is internal (again, by Local Tomography)—hence, maximality follows by proposition 10.

5.3.4. The spectral theorem

An important consequence of the state-effect duality is the ability to decompose every state as a mixture of perfectly distinguishable pure states. The crucial step is to prove such a decomposition for the invariant state:

Lemma 1. *For every pure maximal set* $\{\alpha_x\}_{x=1}^{d_{\rm A}} \subset \mathsf{PurSt}(\rm A)$ *one has* $\chi = \frac{1}{d_{\rm A}} \sum_{x=1}^{d_{\rm A}} \alpha_x$.

This result is extremely important, because it helps us to cope with the existence of different maximal sets of pure states. To begin with, it allows us to prove the analogue of the spectral theorem:

Theorem 5 (Spectral Decomposition)**.** *For every vector* $v \in \mathsf{St}_{\mathbb{R}}(\rm A)$ *there exists a pure maximal set* $\{\alpha_x\}_{x=1}^{d_{\rm A}} \subset \mathsf{PurSt}(\rm A)$ *and a set of real coefficients* $\{c_x\}_{x=1}^{d_{\rm A}}$ *such that*

$$v = \sum_{x=1}^{d_{\rm A}} c_x\, \alpha_x\,. \tag{54}$$

Similarly, for every vector $w \in \mathsf{Eff}_{\mathbb{R}}(\rm A)$ *there exists a pure discriminating measurement* $\{a_x\}_{x=1}^{d_{\rm A}}$ *and a set of real coefficients* $\{d_x\}_{x=1}^{d_{\rm A}}$ *such that*

$$w = \sum_{x=1}^{d_{\rm A}} d_x\, a_x\,. \tag{55}$$

5.3.5. Orthogonal faces

Thanks to spectrality, it is easy to retrieve the basic structures of quantum logic. In general, the faces of a convex set C form a bounded lattice, with partial order $\preceq$ corresponding to set-theoretic inclusion and with meet and join operations defined as $F \wedge G := F \cap G$ and $F \vee G := \bigcap\{H \mid F \subseteq H\,,\, G \subseteq H\}$, respectively. The lattice is bounded, with the convex set C being the top element and the empty set $\emptyset$ being the bottom element. Hence, the set of deterministic states $C_{\rm A} := \mathsf{DetSt}(\rm A)$ in a convex theory can be seen as a lattice in the above way. However, our axioms imply much more: according to them, the faces of the state space form an *orthomodular lattice*, i. e. a lattice with an operation of orthogonal complement $\perp$ satisfying the orthomodularity condition $F \preceq G \Longrightarrow G = F \vee (G \wedge F^{\perp})$.

Let us see why this is the case. For a given face $F \subseteq C_{\mathrm{A}}$ we can pick a set of perfectly distinguishable pure states $\mathrm{S}_F = \{\alpha_x\}_{x=1}^{d_F} \subset F$ that is *maximal in* F, meaning that no other state in F can be distinguished perfectly from the states in S_F. Then, we can define the *barycenter of* F as

$$\omega_F := \frac{1}{d_F} \sum_{x=1}^{d_F} \alpha_x \,. \tag{56}$$

Since the face F can be compressed into the state space of a smaller system, lemma 1 guarantees that the definition of the state ω_F depends only on F, and not on the maximal set S_F. In other words, Eq. (56) sets up a one-to-one correspondence between faces and their barycenters.

Now, we can extend the set S_F to a pure maximal set for system A, say $\{\alpha_x\}_{x=1}^{d_{\mathrm{A}}}$. Let us define the set $\mathrm{S}_{F^\perp} := \{\alpha_x\}_{x=d_F+1}^{d_{\mathrm{A}}}$ and denote by $F^\perp$ the smallest face containing $\mathrm{S}_{F^\perp}$. By construction, it is easy to verify that the set $\mathrm{S}_{F^\perp}$ is maximal in $F^\perp$ and therefore

$$\omega_{F^\perp} = \frac{1}{d_{\mathrm{A}} - d_F} \sum_{x=d_F+1}^{d_{\mathrm{A}}} \alpha_x \,.$$

$F^\perp$ can be equivalently characterized as the face containing all the states that are perfectly distinguishable from F. Moreover, it is not hard to show that

1. $F \vee F^\perp = C_{\mathrm{A}}$
2. $F \wedge F^\perp = \emptyset$
3. $\left(F^\perp\right)^\perp \equiv F$
4. $F \preceq G \Longrightarrow G^\perp \preceq F^\perp$
5. $F \preceq G \Longrightarrow G = F \vee \left(G \wedge F^\perp\right)$,

where the last two properties are proven by picking a pure maximal set for F, extending it to a pure maximal set for G, and extending the latter to a pure maximal set for the whole convex set C_{A}. Properties 1-4 show that the operation $\perp$ is an *orthogonal complement*, while property 5 is the orthomodularity condition. Hence, we obtained that the set of faces must be an orthomodular lattice.

5.3.6. Orthogonal effects

By the state-effect duality, we can associate every face F with an effect a_F, defined as

$$a_F := \sum_{x=1}^{d_F} \alpha_x^\dagger \,, \tag{57}$$

where $\mathrm{S}_F = \{\alpha_x\}_{x=1}^{d_F}$ is a pure maximal set in F. Again, it is easy to see that the definition of a_F is independent of the choice of maximal set S_F. Indeed, by

definition one has $a_F + a_{F^\perp} = e_{\rm A}$ for every pure maximal set $S_{F^\perp}$. Varying $\rm S_F$ without varying $S_{F^\perp}$ shows that the definition of a_F depends only on F.

Thanks to the spectral theorem, a_F can be operationally characterized the only effect that happens with unit probability on F and with zero probability on $F^\perp$:

Proposition 14. *a_F is the unique effect $a \in \mathsf{Eff}(\rm A)$ satisfying the conditions*

$$(a|\rho) = 1 \qquad \forall \rho \in F$$
$$(a|\sigma) = 0 \qquad \forall \sigma \in F^\perp\,.$$

For this reason, we call a_F the *identifying effect* of the face F. The set of identifying effects inherits the structure of orthomodular lattice from the set of faces, via the following definitions

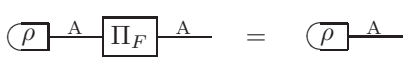
1. $a_F \preceq a_G$ iff $F \preceq G$,
2. $a_F \wedge a_G := a_{F\wedge G}$,
3. $a_F \vee a_G := a_{F\vee G}$, and
4. $a_F^\perp := a_{F^\perp}$.

In quantum theory, the lattice of identifying effects is the lattice of projectors on subspaces of the Hilbert space. It is easy to see that the partial order $\preceq$ coincides with the partial order $\leq$ induced by the probabilities, namely $a_F \preceq a_G$ if and only if $(a_F|\rho) \leq (a_G|\rho)$ for every state ρ.

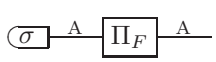
5.3.7. Orthogonal projections

Faces of the state space can also be associated with physical transformations, in the following way:

Definition 18. *A transformation $\Pi_F \in \mathsf{Transf}(\rm A \to A)$ is an* orthogonal projection *on the face $F \subseteq C_{\rm A}$ iff the following conditions are satisfied*[14]



$$\forall \rho \in F \tag{59}$$



$$= \; 0 \qquad \forall \sigma \in F^\perp\,. \tag{60}$$

[14] In the original work [2], we also required that projections be *pure*. However, in the context of our axioms, purity is implied by the two conditions in the present definition. A sketch of proof is the following: First, one can prove that for every pure state $\alpha \in F$ one must have $(\alpha^\dagger|\,\Pi_F = (\alpha^\dagger|$ (this follows from the definition and from proposition 14). As a consequence, one also has $(a_F|\,\Pi_F = (a_F|$. This implies that, for every state $\rho \in \mathsf{St}(\rm A)$, the unnormalized state $\Pi_F\,|\rho)$ is proportional to a state in F. Now, for two projections Π_F and Π'_F one must have

$$(\alpha^\dagger|\,\Pi_F|\rho) = (\alpha^\dagger|\rho) = (\alpha^\dagger|\,\Pi'_F|\rho)\,, \tag{58}$$

for every pure state $\alpha \in F$. Since the states $\Pi_F\,|\rho)$ and $\Pi'_F\,|\rho)$ are proportional to states in F and $\alpha \in F$ is a generic pure state, Ideal Compression implies $\Pi_F\,|\rho) = \Pi'_F\,|\rho)$, or equivalently, $\Pi_F = \Pi'_F$, because the state ρ is generic.

The definition is non-empty: thanks to Purification and Purity of Composition, we are able to construct a *pure* projection Π_F for every face F. Moreover, it follows from the definition that the projection Π_F is unique.

In addition to purity, projections have a number of properties, including

1. $(a_F^\perp|\,\Pi_F = 0$
2. $(a_G|\,\Pi_F = (a_G|$ whenever $G \preceq F$
3. for every input state ρ, the normalized output state $\tau := \Pi_F\,|\rho)/(e_{\rm A}|\,\Pi_F\,|\rho)$ belongs to F
4. $\Pi_G\,\Pi_F = \Pi_F \Pi_G = \Pi_G$ whenever $G \preceq F$.

5.4. Interaction between correlation and distinguishability structures

We have seen that our axioms imply peculiar features, both in the way systems correlate and in the way states can be distinguished. It is time to combine these two types of features and to explore the consequences.

5.4.1. The Schmidt bases

Combining Pure Steering and Spectral Decomposition, we are now in position to give the operational version of the Schmidt bases in quantum theory. The result can be summarized as follows:

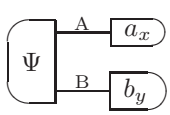
Proposition 15. *Let Ψ be a pure state of $\mathrm{A}\otimes\mathrm{B}$ and let $\rho_{\rm A}$ and $\rho_{\rm B}$ be its marginals on systems A and B, respectively. Then, for every spectral decomposition*

$$\rho_{\rm A} = \sum_{x=1}^r p_x\,\alpha_x\,,$$

there exists a set of perfectly distinguishable pure states $\{\beta_x\}_{x=1}^r \subset \mathsf{PurSt}(\mathrm{B})$ such that

$$\rho_{\rm B} = \sum_{x=1}^r p_x\,\beta_x\,. \tag{61}$$

Moreover, one has



$$= \begin{cases} p_x\,\delta_{xy} & x,y\in\{1,\dots,r\} \\ 0 & x,y\notin\{1,\dots,r\} \end{cases} \tag{62}$$

for every two measurements $\boldsymbol{a} = \{a_x\}_{x=1}^{k_{\rm A}}$ and $\boldsymbol{b} = \{b_y\}_{y=1}^{k_{\rm B}}$ satisfying $a_x = \alpha_x^\dagger$ and $b_y = \beta_y^\dagger$ for every $x \le r$ and for every $y \le r$.

In particular, applying the result to the Bell state Φ, we obtain that the invariant state $\chi_{\overline{A}}$ can be decomposed as $\chi_{\overline{A}} = \frac{1}{d_{\rm A}}\sum_{x=1}^{d_{\rm A}}\overline{\alpha}_x$, for a suitable set of perfectly distinguishable pure states $\{\overline{\alpha}_x\}_{x=1}^{d_{\rm A}}$. In particular, this implies that conjugate systems have the same informational dimension:

Corollary 1. *For every system* A*, one has* $d_{\overline{\mathrm{A}}} = d_{\mathrm{A}}$*.*

Combined with the fact that the informational dimension is multiplicative (proposition 13), the above result implies that the composite system $\mathrm{A} \otimes \overline{\mathrm{A}}$ has informational dimension

$$d_{\mathrm{A}\otimes\overline{\mathrm{A}}} = d_{\mathrm{A}}^2 \,.$$

5.4.2. The maximum probability of conclusive teleportation

In our construction of conclusive teleportation, the teleportation probability was equal to the probability of the state Φ in an ensemble decomposition of the invariant state $\chi_{\mathrm{A}} \otimes \chi_{\overline{\mathrm{A}}}$, cf. Eq. (46). Now, since $\chi_{\mathrm{A}} \otimes \chi_{\overline{\mathrm{A}}}$ is the invariant state, it can be decomposed as

$$\chi_{\mathrm{A}} \otimes \chi_{\overline{\mathrm{A}}} = \frac{1}{d_{\mathrm{A}}^2} \sum_{x=1}^{d_{\mathrm{A}}^2} \Phi_x \,,$$

for every pure maximal set $\{\Phi_x\}_{x=1}^{d_{\mathrm{A}}^2}$. The maximum probability of the Bell state in a convex decomposition of $\chi_{\mathrm{A}} \otimes \chi_{\overline{\mathrm{A}}}$ is then given by

$$p_{\mathrm{A}}^{\max} = \frac{1}{d_A^2} \,. \tag{63}$$

Inserting the above equality into the teleportation upper bound (53) we obtain the relation

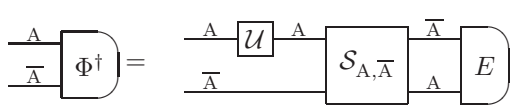
$$D_{\mathrm{A}} \leq d_{\mathrm{A}}^2 \,. \tag{64}$$

In the next paragraph we will see how to obtain the converse inequality.

5.4.3. The teleportation lower bound

Thanks to the state-effect duality, it is possible to establish a lower bound on the state space dimension. The proof is a little bit laborious and consists of two steps:

1. show that the effect $\Phi^\dagger$ that identifies the Bell state is of the form



   where E is the effect achieving maximum teleportation probability, $\mathcal{S}_{\mathrm{A},\overline{\mathrm{A}}}$ is the swap, and $\mathcal{U}$ is some reversible transformation.
2. show that, with a suitable choice of basis for the vector space $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$, every reversible transformation $\mathcal{U}$ is represented by an orthogonal matrix $M_{\mathcal{U}}$.

Once these two results are established, we can expand the Bell state Φ and the teleportation effect E as in Eq. (49), thus obtaining

$$1 = (\Phi^\dagger|\Phi) = \mathrm{Tr}[\Phi\, E\, M_{\mathcal{U}}] = p_{\mathrm{A}}^{\max}\ \mathrm{Tr}[M_{\mathcal{U}}] \leq p_{\mathrm{A}}^{\max}\, D_{\mathrm{A}}\,, \tag{65}$$

having used the teleportation equality $\Phi\, E = p_{\mathrm{A}}^{\max}\, I_{D_{\mathrm{A}}}$ and the fact that the trace of an orthogonal matrix cannot be larger than the trace of the identity. Hence, we obtained the *teleportation lower bound*

$$D_{\mathrm{A}} \geq \frac{1}{p_{\mathrm{A}}^{\max}}\,. \tag{66}$$

Combining the teleportation lower bound with Eqs. (63) and (64), we obtain the equality

$$D_{\mathrm{A}} = d_{\mathrm{A}}^2\,. \tag{67}$$

5.5. Qubit structures

So far, we avoided giving a concrete representation of our state spaces: all the quantum features that we have shown followed *directly* from the principles. We now proceed to analyze some features that are more closely related to the concrete geometrical shape of the quantum state spaces. We will first see that all two-dimensional systems in our theory have qubit state spaces. Leveraging on this fact, we will then derive two features of higher-dimensional systems: *i)* an operational version of the superposition principle, and *ii)* the fact that all systems of the same dimension are operationally equivalent.

5.5.1. Derivation of the qubit

Showing that the states of a two-dimensional system can be described by density matrices is quite easy. This can be done geometrically, by showing that the deterministic states form a 3-dimensional Euclidean ball. The 3-dimensionality is obvious from the equality $D_{\mathrm{A}} = d_{\mathrm{A}}^2$, which for $d_{\mathrm{A}} = 2$ implies that the convex set $C_{\mathrm{A}} = \mathsf{DetSt}(\mathrm{A})$ is a three-dimensional manifold [15]. Then, we can make a simple geometrical reasoning:

1. all the pure states are generated from a fixed pure state by application of reversible transformations, and, by choosing a suitable basis for the state space, such transformations act in the 3-dimensional space as orthogonal matrices.
2. all states on the border of C_{A} are pure—otherwise, Perfect State Discrimination and proposition 11 would imply $d_{\mathrm{A}} > 2$. This means that, if we move away from the invariant state χ_{A} in an arbitrary direction, at some point we will hit a pure state.

[15] In general, the dimension of the convex set C_{A} is given by $D_{\mathrm{A}} - 1$.

In the ordinary 3-dimensional space, the sphere is the only (closed) 3-dimensional convex set generated by orthogonal matrices and with only pure states on the border.

Once we established that the convex set C_{A} is a sphere, we can represent every normalized state $\rho \in C_{\mathrm{A}}$ with a density matrix S_ρ. In particular, the pure states will be of the form

$$S_\alpha = \begin{pmatrix} p & \sqrt{p(1-p)}\, e^{-i\theta} \\ \sqrt{p(1-p)}\, e^{i\theta} & 1-p \end{pmatrix} = |\alpha\rangle\langle\alpha| \tag{68}$$

$$|\alpha\rangle := \sqrt{p}\,|0\rangle + e^{i\theta}\,\sqrt{1-p}\,|1\rangle\,,$$

for some probability $p \in [0,1]$ and some phase $\theta \in [0, 2\pi)$. Once we have chosen this representation, it is obvious that every effect $a \in \mathsf{Eff}(\mathrm{A})$ must be described by a positive semidefinite matrix E_a upper bounded by the identity and that probabilities are given by the Born rule

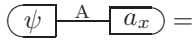
$$(a|\rho) = \mathrm{Tr}[E_a\, S_\rho]\,. \tag{69}$$

Moreover, the state-effect duality imposes that *all* such matrices represent valid effects.

5.5.2. The superposition principle

Pure states in quantum theory satisfy the so-called "superposition principle", which just means that they are in one-to-one correspondence with the rays of the underlying Hilbert space. *Per se*, this statement has hardly any operational meaning. However, one can formulate an operational version in general OPTs:

Definition 19 (Superposition Principle)**.** *We say that system* A satisfies the superposition principle *iff for every pure maximal set* $\mathrm{S} = \{\alpha_x \mid x \in \mathrm{X}\} \subset \mathsf{PurSt}(\mathrm{A})$ *and for every probability distribution* $\{p_x\}_{x\in\mathrm{X}}$ *there exists one pure state* ψ *such that*



$$= \quad p_x \qquad \forall x \in \mathrm{X}\,, \tag{70}$$

for every measurement $\boldsymbol{a} = \{a_x\}_{x\in\mathrm{X}}$ *that perfectly distinguishes among the states in the maximal set* S.

Now, in a theory satisfying our principles we know that the two-dimensional systems are quantum—and therefore satisfy the superposition principle. Thanks to Ideal Compression, it is then easy to generalize the result to systems of arbitrary dimension: given two perfectly distinguishable pure states, one can encode them into a two-dimensional system, use the Bloch sphere representation to find the superposition state, and come back with the decoding operation. Iterating this procedure, we can superpose any number of perfectly distinguishable pure states.

As a simple application of the superposition principle, we obtain the following

Proposition 16. *A state $\rho_{\rm A}$ with spectral decomposition $\rho_{\rm A} = \sum_{x=1}^r p_x\, \alpha_x$ has a purification with purifying system* B *if and only if $d_{\rm B} \geq r$.*

The "only if" part was already clear from the Schmidt decomposition. For the "if" part, it is enough to pick r perfectly distinguishable pure states of B, say $\{\beta_x\}_{x=1}^r$, and to superpose the product states $\{\alpha_x \otimes \beta_x\}_{x=1}^r$ with probabilities $\{p_x\}_{x=1}^r$. The resulting pure state $\Psi \in \mathsf{PurSt}({\rm A} \otimes {\rm B})$ is the desired purification.

5.5.3. The superposition principle for transformations

The superposition principle allows us to glue distinguishable states in any way we like. Thanks to the state-transformation isomorphism, we can extend this idea to transformations. For example, consider a set of pure transformations $\{\mathcal{A}_x \mid x \in {\rm X}\} \subset \mathsf{PurTransf}({\rm A} \to {\rm B})$ and suppose that they have *orthogonal support*, that is, that there exists a set of orthogonal faces $\{F_x \mid x \in {\rm X}\}$ such that

$$\mathcal{A}_x = \mathcal{A}_x\, \Pi_{F_x} \qquad \forall x \in {\rm X}\,. \tag{71}$$

Then, it is possible to find a pure transformation $\mathcal{A} \in \mathsf{PurTransf}({\rm A} \to {\rm B})$ such that

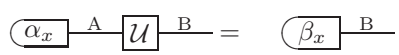
$$\mathcal{A}\, \Pi_{F_x} = \mathcal{A}_x \quad \forall x \in {\rm X}\,. \tag{72}$$

The result follows by noticing that the Choi states $\{\Phi_{\mathcal{A}_x} \mid x \in {\rm X}\}$ are proportional to pure and perfectly distinguishable states and by applying the superposition principle to corresponding the normalized states.

5.5.4. Equivalence of pure maximal sets up to reversible transformations

Using the superposition principle for transformations we can prove that all pure maximal sets of the same cardinality are equivalent:

Proposition 17. *Let $\{\alpha_x\}_{x=1}^{d_{\rm A}}$ and $\{\beta_y\}_{y=1}^{d_{\rm B}}$ be pure maximal sets for systems* A *and* B*, respectively. If $d_{\rm A} = d_{\rm B}$, then there exists a reversible transformation $\mathcal{U} \in \mathsf{Transf}({\rm A} \to {\rm B})$ such that*



$\forall x \in {\rm X}\,.$

The result follows immediately from the application of the superposition principle to the pure transformations $\mathcal{A}_x = |\beta_x)(\alpha_x^\dagger|$. As a corollary, we have that all systems of the same dimension are operationally equivalent.

5.6. The density matrix

We finally reached to the end of the reconstruction. It is now time to enter into the specific details of the Hilbert space formalism of quantum theory. Our strategy to reconstruct the Hilbert space formalism is to show that, for every system A, there exists a one-to-one linear map from the vector space $\mathsf{St}_{\mathbb{R}}({\rm A})$ to the space of $d_{\rm A} \times d_{\rm A}$ Hermitian matrices, with the property that the convex set

of deterministic states is mapped to the convex set of density matrices (non-negative matrices with unit trace).

Let us see how this can be proven. Since the dimension of the state space satisfies the relation $D_{\rm A} = d_{\rm A}^2$, every vector $v \in \mathsf{St}_{\mathbb{R}}({\rm A})$ can be represented as square $d_{\rm A} \times d_{\rm A}$ real matrix M_v. In turn, the matrix M_v can be turned into a complex Hermitian matrix S_v, applying the linear transformation

$$S_v := \left(M_v + M_v^T\right) + i\left(M_v - M_v^T\right), \tag{73}$$

where M^T denotes the transpose of M. The problem is now to find a suitable representation in which normalized states $\rho \in C_{\rm A}$ correspond to density matrices, that is $S_\rho \geq 0$ and ${\rm Tr}[\rho] = 1$. To find such a representation, we follow Hardy's method [25]: we pick a pure maximal set $\{\alpha_m\}_{m=1}^{d_{\rm A}}$ and define the diagonal elements of the matrix S_ρ as

$$[S_\rho]_{mm} := (\alpha_m^\dagger|\rho)\,,$$

In this way, we guarantee the unit-trace condition ${\rm Tr}[S_\rho] = 1$. To define the off-diagonal elements, we consider the two-dimensional faces $F_{mn} := \{\alpha_m\} \vee \{\alpha_n\}$, $n > m$. Projecting the state inside these faces, we obtain the states

$$|\rho^{mn}) = \frac{\Pi_{F_{mn}}|\rho)}{(e_{\rm A}|\,\Pi_{F_{mn}}\,|\rho)} \qquad n > m\,.$$

Since every state ρ^{mn} is belongs to a two-dimensional face, it can be encoded into a qubit system and can be associated with a density matrix τ^{mn}. The off-diagonal elements $[S_\rho]_{mn}$ and $[S_\rho]_{nm}$ are defined in term of the qubit density matrix τ^{mn}, as

$$[S_\rho]_{mn} := [\tau^{mn}]_{01} \qquad \text{and} \qquad [S_\rho]_{nm} := [\tau^{mn}]_{10}\,.$$

The matrix S_ρ defined in this way is clearly Hermitian and, with a little bit of work, one can see that the linear map $\rho \mapsto S_\rho$ is one-to-one.

At this point the problem is to guarantee that the matrix S_ρ is positive. We consider first the case of pure states $\alpha \in \mathsf{PurSt}({\rm A})$, for which one has

$$[S_\alpha]_{mn} = \sqrt{p_m\,p_n}\,e^{i\theta_{mn}}$$

where $\{p_m\}_{m=1}^{d_{\rm A}}$ is a suitable probability distribution and $\{\theta_{mn}\}$ are phases satisfying the conditions $\theta_{mm} = 0$ for every m and $\theta_{nm} = -\theta_{mn}$ for every $n > m$. This expression follows from the fact that each state $|\alpha^{mn}) = \Pi_{F_{mn}}|\alpha)/(e_{\rm A}|\Pi_{F_{mn}}|\alpha)$ is pure and, once encoded into a qubit, it has a density matrix of the form (68). In order to prove positivity, we need to show that the phases θ_{mn} are of the form $\theta_{mn} = \gamma_m - \gamma_n$, for some phases $\{\gamma_m\}$. The strategy is to prove the result first in dimension $d_{\rm A} = 3$ and then to extend it to arbitrary dimensions.

Once we have proven that pure states correspond to rank-one projectors, it remains to show that *all* such projectors correspond to pure states. This can

be done by using the superposition principle (both for states and for reversible transformations). Having proven that the set of pure states is in one-to-one correspondence with the set of rank-one projectors, it follows by convexity that the set of states is in one-to-one correspondence with the set of density matrices. In short, all state spaces are quantum.

To complete our reconstruction, we invoke theorem 3, which guarantees that the tests in our theory are in one-to-one correspondence with the test allows by quantum theory.

6. Conclusions

Quantum theory can be rebuilt from bottom to top starting from six basic principles. The principles do not refer to specific physical systems such as particles or waves: instead, they are the rules that dictate how information can be processed. The first five principles—Causality, Purity of Composition, Local Tomography, Perfect State Discrimination, and Ideal Compression—can be thought of as requirements for a standard theory of information. On the background of these five principles, the sixth—Purification—stands out as *the* quantum principle, which brings in counterintuitive features like entanglement, no cloning, and teleportation. Purification gives the agent the power to harness randomness, by simulating the preparation of every state through the preparation of a pure bipartite state. When this is done, the agent has an intrinsic guarantee that no side information can hide outside her control. The moral of our reconstruction is quantum theory is the standard theory of information that allows for maximal control of randomness.

Acknowledgements The work is supported by the Templeton Foundation under the project ID# 43796 *A Quantum-Digital Universe*, by the Foundational Questions Institute through the large grant *The fundamental principles of information dynamics* (FQXi-RFP3-1325), by the 1000 Youth Fellowship Program of China, and by the National Natural Science Foundation of China through Grants 11450110096 and 11350110207. GC acknowledges the hospitality of the Simons Center for the Theory of Computation and of Perimeter Institute for Theoretical Physics. Research at Perimeter Institute for Theoretical Physics is supported in part by the Government of Canada through NSERC and by the Province of Ontario through MRI.

References

[1] G. Chiribella, G. M. D'Ariano, P. Perinotti, Probabilistic theories with purification, Phys. Rev. A 81 (2010) 062348. doi:10.1103/PhysRevA.81.062348.
URL http://link.aps.org/doi/10.1103/PhysRevA.81.062348

[2] G. Chiribella, G. M. D'Ariano, P. Perinotti, Informational derivation of quantum theory, Phys. Rev. A 84 (2011)

012311. doi:10.1103/PhysRevA.84.012311.
URL http://link.aps.org/doi/10.1103/PhysRevA.84.012311

[3] A. Einstein, B. Podolsky, N. Rosen, Can quantum-mechanical description of physical reality be considered complete?, Phys. Rev. 47 (1935) 777–780. doi:10.1103/PhysRev.47.777.
URL http://link.aps.org/doi/10.1103/PhysRev.47.777

[4] E. Schrödinger, Discussion of probability relations between separated systems, in: Mathematical Proceedings of the Cambridge Philosophical Society, Vol. 31, Cambridge Univ Press, 1935, pp. 555–563.

[5] G. Birkhoff, J. V. Neumann, The logic of quantum mechanics, Annals of Mathematics 37 (4) (1936) pp. 823–843.
URL http://www.jstor.org/stable/1968621

[6] G. W. Mackey, Quantum mechanics and hilbert space, American Mathematical Monthly (1957) 45–57.

[7] G. Ludwig, Versuch einer axiomatischen grundlegung der quantenmechanik und allgemeinerer physikalischer theorien, Zeitschrift für Physik 181 (3) (1964) 233–260.

[8] C. Piron, Axiomatique quantique, Helvetica physica acta 37 (4-5) (1964) 439.

[9] J. Jauch, C. Piron, On the structure of quantal proposition systems, in: The Logico-Algebraic Approach to Quantum Mechanics, Springer, 1975, pp. 427–436.

[10] E. G. Beltrametti, G. Cassinelli, The logic of quantum mechanics, Vol. 15, Cambridge University Press, 2010.

[11] B. Coecke, D. Moore, A. Wilce, Current research in operational quantum logic: algebras, categories, languages, Vol. 111, Springer Science & Business Media, 2000.

[12] Quantum logic, http://en.wikipedia.org/wiki/Quantum_logic, accessed: 2015-04-30.

[13] C. H. Bennett, G. Brassard, Quantum cryptography: public key distribution and coin tossing, in: Proceedings of IEEE International Conference on Computers, Systems and Signal Processing, 1984, pp. 175–179.

[14] A. K. Ekert, Quantum cryptography based on bell's theorem, Physical review letters 67 (6) (1991) 661.

[15] P. W. Shor, Algorithms for quantum computation: Discrete logarithms and factoring, in: Foundations of Computer Science, 1994 Proceedings., 35th Annual Symposium on, IEEE, 1994, pp. 124–134.

[16] L. K. Grover, A fast quantum mechanical algorithm for database search, in: Proceedings of the Twenty-eighth Annual ACM Symposium on Theory of Computing, STOC '96, ACM, New York, NY, USA, 1996, pp. 212–219. doi:10.1145/237814.237866.
URL http://doi.acm.org/10.1145/237814.237866

[17] W. Wootters, W. Zurek, A single quantum cannot be cloned, Nature 299 (5886) (1982) 802–803.

[18] D. Dieks, Communication by epr devices, Physics Letters A 92 (6) (1982) 271–272.

[19] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, W. K. Wootters, Teleporting an unknown quantum state via dual classical and einstein-podolsky-rosen channels, Physical review letters 70 (13) (1993) 1895.

[20] C. H. Bennett, S. J. Wiesner, Communication via one-and two-particle operators on einstein-podolsky-rosen states, Physical review letters 69 (20) (1992) 2881.

[21] C. A. Fuchs, Quantum mechanics as quantum information, mostly, Journal of Modern Optics 50 (6-7) (2003) 987–1023.

[22] G. Brassard, Is information the key?, Nature Physics 1 (1) (2005) 2–4.

[23] C. A. Fuchs, et al., Quantum foundations in the light of quantum information, NATO SCIENCE SERIES SUB SERIES III COMPUTER AND SYSTEMS SCIENCES 182 (2001) 38–82.

[24] C. A. Fuchs, Coming of age with quantum information, Coming of Age With Quantum Information, by Christopher A. Fuchs, Cambridge, UK: Cambridge University Press, 2011.

[25] L. Hardy, Quantum theory from five reasonable axioms, arXiv preprint quant-ph/0101012.

[26] G. M. D'Ariano, How to derive the hilbert-space formulation of quantum mechanics from purely operational axioms, AIP Conference Proceedings 844 (2006) 101.

[27] G. M. D'Ariano, Probabilistic theories: what is special about quantum mechanics, in: A. Bokulich, G. Jaeger (Eds.), Philosophy of quantum information and entanglement, Cambridge University Press, Cambridge, 2010, pp. 85–126. doi:10.1017/CBO9780511676550.007.

[28] L. Hardy, Reformulating and reconstructing quantum theory, arXiv:1104.2066.

[29] L. Masanes, M. P. Müller, A derivation of quantum theory from physical requirements, New Journal of Physics 13 (6) (2011) 063001.

[30] B. Dakic, C. Brukner, Quantum Theory and Beyond: Is Entanglement Special?, in: H. Halvorson (Ed.), Deep Beauty: Understanding the Quantum World through Mathematical Innovation, Cambridge University Press, 2011, pp. 365–392.

[31] P. Goyal, K. H. Knuth, J. Skilling, Origin of complex quantum amplitudes and feynman's rules, Phys. Rev. A 81 (2010) 022109. doi:10.1103/PhysRevA.81.022109.
URL http://link.aps.org/doi/10.1103/PhysRevA.81.022109

[32] L. Masanes, M. P. Müller, R. Augusiak, D. Pérez-García, Existence of an information unit as a postulate of quantum theory, Proceedings of the National Academy of Sciences 110 (41) (2013) 16373–16377.

[33] A. Wilce, Conjugates, correlation and quantum mechanics, arXiv:1206.2897.

[34] H. Barnum, M. P. Mueller, C. Ududec, Higher-order interference and single-system postulates characterizing quantum theory, arXiv:1403.4147.

[35] J. Barrett, Information processing in generalized probabilistic theories, Phys. Rev. A 75 (2007) 032304. doi:10.1103/PhysRevA.75.032304.
URL http://link.aps.org/doi/10.1103/PhysRevA.75.032304

[36] H. Barnum, J. Barrett, M. Leifer, A. Wilce, Generalized no-broadcasting theorem, Physical Review Letters 99 (24) (2007) 240501. doi:10.1103/PhysRevLett.99.240501.

[37] H. Barnum, J. Barrett, M. Leifer, A. Wilce, Teleportation in General Probabilistic Theories, ArXiv e-printsarXiv:0805.3553.

[38] L. Hardy, A formalism-local framework for general probabilistic theories, including quantum theory, Mathematical Structures in Computer Science 23 (02) (2013) 399–440. doi:10.1017/S0960129512000163.

[39] H. Barnum, A. Wilce, Information processing in convex operational theories, Electronic Notes in Theoretical Computer Science 270 (1) (2011) 3–15.

[40] S. Abramsky, B. Coecke, A categorical semantics of quantum protocols, in: Logic in Computer Science, 2004. Proceedings of the 19th Annual IEEE Symposium on, IEEE, 2004, pp. 415–425. doi:10.1109/LICS.2004.1319636.

[41] S. Abramsky, B. Coecke, Categorical quantum mechanics, in: K. Engesser, D. M. Gabbay, D. Lehmann (Eds.), Handbook of quantum logic and quantum structures: quantum logic, Elsevier, 2008, pp. 261–324. doi:10.1016/B978-0-444-52869-8.50014\-1.

[42] B. Coecke, A universe of processes and some of its guises, in: H. Halvorson (Ed.), Deep Beauty: Understanding the Quantum World Through Mathematical Innovation, Cambridge University Press, 2010, pp. 129–186.

[43] B. Coecke, É. O. Paquette, Categories for the practising physicist, in: New Structures for Physics, Springer, 2011, pp. 173–286.

[44] S. Mac Lane, Categories for the working mathematician, Vol. 5, Springer Science & Business Media, 1978.

[45] B. Coecke, Quantum picturalism, Contemporary physics 51 (1) (2010) 59–83. doi:10.1080/00107510903257624.

[46] P. Selinger, A survey of graphical languages for monoidal categories, in: B. Coecke (Ed.), New Structures for Physics, Vol. 813 of Lecture Notes in Physics, 2011. doi:10.1007/978-3-642-12821-9_4.

[47] G. Chiribella, Dilation of states and processes in operational-probabilistic theories, in: B. Coecke, I. Hasuo, P. Panangaden (Eds.), Proceedings 11th workshop on Quantum Physics and Logic, Kyoto, Japan, 4-6th June 2014, Vol. 172 of Electronic Proceedings in Theoretical Computer Science, Open Publishing Association, 2014, pp. 1–14. doi:10.4204/EPTCS.172.1.

[48] R. W. Spekkens, Evidence for the epistemic view of quantum states: A toy theory, Phys. Rev. A 75 (2007) 032110. doi:10.1103/PhysRevA.75.032110. URL http://link.aps.org/doi/10.1103/PhysRevA.75.032110

[49] G. Chiribella, G. M. D?Ariano, P. Perinotti, Quantum theory, namely the pure and reversible theory of information, Entropy 14 (10) (2012) 1877–1893.

[50] B. Coecke, R. Lal, Causal categories: relativistically interacting processes, Foundations of Physics 43 (4) (2013) 458–501.

[51] B. Coecke, Terminality implies non-signalling, arXiv preprint arXiv:1405.3681.

[52] E. C. Stueckelberg, Quantum theory in real hilbert space, Helv. Phys. Acta 33 (727) (1960) 458.

[53] H. Araki, On a characterization of the state space of quantum mechanics, Communications in Mathematical Physics 75 (1) (1980) 1–24.

[54] W. K. Wootters, Local accessibility of quantum states, Complexity, entropy and the physics of information 8 (1990) 39–46.

[55] L. Hardy, W. K. Wootters, Limited holism and real-vector-space quantum theory, Foundations of Physics 42 (3) (2012) 454–473.

[56] C. Shannon, A mathematical theory of communication, Bell System Technical Journal, The 27 (3) (1948) 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.

[57] B. Schumacher, Quantum coding, Physical Review A 51 (4) (1995) 2738.

[58] C. H. Bennett, More about entanglement and cryptography, http://www.lancaster.ac.uk/users/esqn/windsor07/Lectures/Bennett2.pdf, accessed: 2014-11-14 (2007).

[59] I. M. Gelfand, M. A. Naimark, On the imbedding of normed rings into the ring of operators in hilbert space, Matematiceskij sbornik 54 (2) (1943) 197–217.

[60] I. E. Segal, Irreducible representations of operator algebras, Bulletin of the American Mathematical Society 53 (2) (1947) 73–88. doi:10.1090/S0002-9904-1947-08742-5.

[61] H. Everett III, "relative state" formulation of quantum mechanics, Reviews of modern physics 29 (3) (1957) 454.

[62] H. Barnum, C. P. Gaebler, A. Wilce, Ensemble steering, weak self-duality, and the structure of probabilistic theories, Foundations of Physics 43 (12) (2013) 1411–1427.

[63] M.-D. Choi, Completely positive linear maps on complex matrices, Linear Algebra and its Applications 10 (3) (1975) 285–290.

View publication stats