arXiv:0908.1583v5 [quant-ph] 2 Jun 2010

Probabilistic theories with purification

Giulio Chiribella*
Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Ontario, Ontario N2L 2Y5, Canada.†

Giacomo Mauro D'Ariano‡ and Paolo Perinotti§
QUIT Group, *Dipartimento di Fisica "A. Volta" and INFN Sezione di Pavia, via Bassi 6, 27100 Pavia, Italy*¶

(Dated: February 13, 2022)

We investigate general probabilistic theories in which every mixed state has a purification, unique up to reversible channels on the purifying system. We show that the purification principle is equivalent to the existence of a reversible realization of every physical process, that is, to the fact that every physical process can be regarded as arising from a reversible interaction of the system with an environment, which is eventually discarded. From the purification principle we also construct an isomorphism between transformations and bipartite states that possesses all structural properties of the Choi-Jamiołkowski isomorphism in quantum theory. Such an isomorphism allows one to prove most of the basic features of quantum theory, like e.g. existence of pure bipartite states giving perfect correlations in independent experiments, no information without disturbance, no joint discrimination of all pure states, no cloning, teleportation, no programming, no bit commitment, complementarity between correctable channels and deletion channels, characterization of entanglement-breaking channels as measure-and-prepare channels, and others, without resorting to the mathematical framework of Hilbert spaces.

PACS numbers: 03.67.-a, 03.67.Ac, 03.65.Ta

Contents

I. Introduction 2

II. Operational-probabilistic theories 3
- A. Systems and tests 3
- B. Sequential composition of tests 4
- C. Composite systems and parallel composition of tests 5
- D. Operational theories 5
- E. Relation with category theory 6
- F. Probabilistic structure: states, effects, and transformations 6
- G. Relation with the convex sets framework 8
- H. Coarse-graining and refinement 8
- I. Discrimination and distance 8
- J. Closure 10

III. Causal theories 10
- A. Definition and main properties 10
- B. Conditioning 12
- C. Distance between transformations 13
- D. Closure and convexity in causal theories 14
- E. No-restriction hypothesis in causal theories 14

IV. Local discriminability 14
- A. Definition and main properties 14
- B. Causal theories with local discriminability 16

V. Beyond local discriminability and convexity 17
- A. Relaxing local discriminability 17
- B. Relaxing convexity 18

VI. Summary of the framework 18

VII. Theories with purification 18
- A. The purification postulate 18
- B. Purification of preparation-tests 20
- C. Dynamically faithful pure states 22
- D. No information without disturbance 23
- E. No-cloning 24

VIII. Probabilistic teleportation 24
- A. Entanglement-swapping and teleportation 24
- B. Storing and probabilistic retrieving of transformations 26
- C. Systems of purifications and the link product 27

IX. Dilation of physical processes 27
- A. Reversible dilation of channels 27
- B. Reversible dilation of tests 29

X. States-transformations isomorphism 30
- A. First consequences of the isomorphism 31
- B. Entanglement breaking channels 32
- C. Completeness of theories with purification 32

XI. Error correction 34
- A. Basic definitions 34

---

[*] Electronic address: gchiribella@perimeterinstitute.ca
[†] URL: `http://www.perimeterinstitute.ca`
[‡] Electronic address: dariano@unipv.it
[§] Electronic address: paolo.perinotti@unipv.it
[¶] URL: `http://www.qubit.it`

B. Error correction and the complementarity between correctable and deletion channels 34
C. Error correction with one-way classical communication from the environment 36

XII. Causally ordered channels and channels with memory 36
A. Dilation of causally ordered channels 37
B. No bit commitment 38

XIII. Deterministic programming of reversible transformations 39

XIV. Purification with conjugate systems 40
A. Conjugate purifying systems 40
B. States-transformations isomorphism for conjugate purifying systems 40
C. Conjugated transformations 41
D. Deterministic teleportation 43

XV. Conclusions and perspectives on future work 45

Acknowledgments 45

References 45

I. INTRODUCTION

In the past two decades the field of quantum information theory has brought to light an enormous amount of protocols and tasks that originate from the structure of quantum theory and have dramatic consequences in the way information can be processed. Non-locality, no-cloning, teleportation, dense coding, quantum key distribution, quantum algorithms, and quantum error correction are only the most celebrated examples of a much longer list. An important lesson from this experience is that the abstract formalism of quantum mechanics has a huge number of operational consequences.

At the same time, the question whether quantum theory is the only conceivable theory with such operational consequences has attracted the attention of an increasing number of researchers. In a seminal paper [1], Popescu and Rohrlich showed that non-locality is not an exclusive feature of quantum theory, and that there are in fact possible theories that exhibit stronger nonlocality than quantum theory without violating relativistic no-signaling. An intense work on non-locality in general non-signaling theories has followed this observation, opening a very active line of research (see e.g. [2–5]). On the other hand, the authors of Refs. [6, 7] have analyzed tasks like cloning and broadcasting of states, showing that the impossibility of achieving them is a highly generic property, while Ref. [8] thoroughly discussed theories with a local discriminability property that share other features of quantum mechanics, like the non-unique convex decomposition of a mixed state or the non-existence of ideal non-disturbing measurements. Entanglement swapping and teleportation protocols have been considered in Refs [9, 10], where the authors noticed the remarkable fact that the no-signaling boxes of Popescu and Rohrlich do not allow for entanglement swapping, nor for teleportation. Very recently, the authors of Ref. [11] have introduced the new physical principle of information causality, showing that while the principle holds for quantum theory, it is violated by Popescu-Rohrlich boxes.

Despite the numerous advancements in the understanding of general probabilistic theories, the fundamental problem of deriving quantum mechanics from basic physical principles is still completely open. In particular, no physical principle is known that can single out quantum mechanics in the physically motivated set of *causal theories with local discriminability*. With this expression we mean probabilistic theories where *i)* the probability of outcomes of an experiment performed at a given time does not depend on the choice of experiments that will be performed at later times, and *ii)* if two bipartite states are different, then one can discriminate between them using only local devices with an error probability that is smaller than 1/2, the random guess value. In the case of classical physics, finding a description is relatively simple: among theories in the above family, classical probability is the only one where all pure states are perfectly distinguishable. On the contrary, every current description of quantum theory is a description of its mathematical apparatus: e.g. one can say that quantum theory is the theory where pure states are unit vectors in complex Hilbert spaces and probabilities are given by the Born rule, or, equivalently, that it is the theory where observables form a C*-algebra of complex matrices.

In the past there have been many attempts to find a more basic description of quantum theory, in particular by discussing it from the point of view of logic [12–15] (see also Ref. [16] and references therein). More recently, Hardy [17] has approached the problem from a different perspective, providing a characterization of quantum theory based on principles of mathematical simplicity in the interplay among dimension of the state space, structure of subsystems and subspaces, number of distinguishable states, and topology of the set of pure states. On the other hand, in recent years one of the authors has tackled the problem using physical principles related to tomography and calibration of physical devices, experimental complexity, and to the composition of elementary (atomic) transformations (see Ref. [18] for the state of the art of this project). In particular, Ref. [19] firstly introduced the concept of *dynamically and preparationally faithful state*, which will play an important role in this paper.

In this paper we introduce the purification principle "Every mixed state has a purification, unique up to reversible channels on the purifying system". The main message of our work is simple: most of the characteristic features of quantum theory can be summarized in the physical statement "quantum theory is a causal theory with purification and local discriminability". In particular, from the purification principle we derive the following features: no information without disturbance, no joint discriminability and no cloning of pure states, existence of pure entangled states with perfect correlations, probabilistic teleportation, one-to-one correspondence between transformations and bipartite states, dilation of physical processes to reversible interactions with an environment, necessary and sufficient conditions for error correction in terms of the reversible dilation, no bit commitment, no programming of reversible channels without perfectly distinguishable program states, and identification of causal channels with sequences of channels with memory, and characterization of entanglement breaking channels as measure-and-prepare channels. Moreover, we also discuss a stronger version of the purification principle: "For every system A there exists a *conjugate system* $\tilde{\mathrm{A}}$ such that every state of A has a purification in $\mathrm{A}\tilde{\mathrm{A}}$. The conjugate of $\tilde{\mathrm{A}}$ is A (symmetry), and the conjugate of a composite system AB is the composite system $\tilde{\mathrm{A}}\tilde{\mathrm{B}}$ (regularity under composition)". With this further property one can prove deterministic teleportation and show that its structure is unique: the resource state for deterministic teleportation must be a purification of the unique mixed state that is invariant under all reversible channels.

As we will show, the purification principle is equivalent to the fact that every irreversible process arises from a reversible interaction with an environment that is eventually lost. This can be viewed as a law of "conservation of information": information cannot be erased, it can only be discarded. Moreover, we will see that the purification principle has other remarkable consequences: From the structural point of view, a theory with purification is completely identified by the states of all possible systems in it. Once the states are given, all possible measurements and evolutions are fixed. Even more strongly, the purification postulate implies the completeness property "whatever transformation is mathematically admissible (in a sense that will be made precise later) must be feasible". Conversely, we can explicitly say that whatever limitation to the feasibility of a mathematically admissible map results in a limitation to the purifiability of some state. The analogue of this property in quantum information is that every trace-preserving completely positive map must be feasible.

It is important to stress that we are not claiming that we derived quantum theory. What we can say is that we "zipped" a large part of it, by reducing a long list of features to a single physical principle. In the process of doing this, we found proofs that are often simpler (or at least more intuitive) than the original quantum proofs.

In order to minimize the notational burden due to the lack of a commonly established formalism, in presenting these proofs we opted for a graphical notation, which is equivalent to formulae and replaces them in most of the paper. Since this notation is exactly the same notation used in quantum circuits, a reader with a background in quantum information can easily read the general equations without spending too much time in the introductory part of the paper. On the other hand, an extended discussion on graphical calculus can be found in the work by Penrose [20] and in the rigorous formalization by Joyal and Street within the theory of symmetric monoidal categories [21] (we also suggest the beautiful introductions in the topic by Selinger [22] and Coecke [23]). We anyway stress that in the present paper the choice of graphical notation is just the choice of a more user-friendly way of presenting formulae, and that no prerequisite on e.g. category theory is needed from the reader.

II. OPERATIONAL-PROBABILISTIC THEORIES

In this Section we introduce some basic notions that will be used in the paper. In particular, we introduce the notion of operational-probabilistic theory as a theory that *i)* describes a set of possible experiments that can be done with physical devices and *ii)* gives predictions about the probabilities of the outcomes in these experiments.

A. Systems and tests

Systems and *tests* are the primitive notions of an operational theory. Each test represents one use of a *physical device*, like a Stern-Gerlach magnet, a beamsplitter, or a photon counter. Systems play the role of labels attached to physical devices: any device has an input and an output port labeled by an *output* and an *input system*, respectively. These labels establish a rule for connecting physical devices among themselves: two devices can be connected in a sequence only if the output of the first device is a system of the same type as the input of the second.

All throughout the paper we will denote systems with capital letters, like $\mathrm{A}, \mathrm{B}, \mathrm{C}$, and so on. We reserve the letter I for the *trivial system*, which simply means "nothing". A device with input (output) system I is a device with no input (no output).

Let us now make more precise the notion of test. We already mentioned that a test represents one use of a physical device. When the physical device is used, it produces an *outcome* i in some set X, e.g. the outcome could be a sequence of digits appearing on a display, a light, or a sound emitted by the device. The outcome produced by the device heralds the fact that some *event* has occurred. These intuitive features concur in the definition of test:

Definition 1 (Test) *A test with input system* A *and output system* B *is a collection of events* $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ *labeled by outcomes in some outcome set* X. *Diagrammatically,* *the test* $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ *is represented as follows*

$$\overset{\mathrm{A}}{-\!-\!-}\boxed{\{\mathscr{C}_i\}_{i\in\mathrm{X}}}\overset{\mathrm{B}}{-\!-\!-} \tag{1}$$

while the specific event $\mathscr{C}_i$ *is represented by*

$$\overset{\mathrm{A}}{-\!-\!-}\boxed{\mathscr{C}_i}\overset{\mathrm{B}}{-\!-\!-} \tag{2}$$

We denote by $\mathfrak{T}(\mathrm{A},\mathrm{B})$ the set of all events appearing in all tests from A to B. When $\mathrm{B}\equiv\mathrm{A}$ we will write $\mathfrak{T}(\mathrm{A})$.

Tests with trivial input will be called *preparation-tests*, and the corresponding events will be called *preparation-events*. In quantum information, a preparation-test is what is called a "random source of quantum states". In analogy we will adopt for preparation-events the usual notation as for states in quantum circuits:

$$\left(\rho_i\right|\overset{\mathrm{B}}{-\!-\!-} \;:=\; \overset{\mathrm{I}}{-\!-\!-}\boxed{\mathscr{C}_i}\overset{\mathrm{B}}{-\!-\!-} \tag{3}$$

In formulae, we will often use the "Dirac-like" notation $|\rho_i)_{\mathrm{B}}$ to denote a preparation event of system B. We will denote by $\mathfrak{S}(\mathrm{A})$ the set of preparation-events for system A, namely $\mathfrak{S}(\mathrm{A}):=\mathfrak{T}(\mathrm{I},\mathrm{A})$.

Similarly, we will call tests with trivial output *observation-tests*, and the corresponding events *observation-events*. In quantum theory, an observation-test is a quantum measurement, and is represented by *positive operator valued measure (POVM)*, that is, by a collection of positive operators $\{P_i\}_{i\in\mathrm{X}}$ satisfying $\sum_{i\in\mathrm{X}}P_i=I_{\mathrm{A}}$, where I_{A} is the identity on the Hilbert space of system A. For observation-tests we will then adopt the usual notation for measurements in quantum circuits:

$$\overset{\mathrm{A}}{-\!-\!-}\left|a_j\right) \;:=\; \overset{\mathrm{A}}{-\!-\!-}\boxed{\mathscr{C}_j}\overset{\mathrm{I}}{-\!-\!-} \tag{4}$$

In formulae, we will often denote observation-events with the notation $(a_j|_{\mathrm{A}}$. We will denote by $\mathfrak{E}(\mathrm{A})$ the set of observation-events for system A, namely $\mathfrak{E}(\mathrm{A}):=\mathfrak{T}(\mathrm{A},\mathrm{I})$.

For tests from the trivial system to itself we will omit the box and the wires, as follows:

$$p_k \;:=\; \overset{\mathrm{I}}{-\!-\!-}\boxed{p_k}\overset{\mathrm{I}}{-\!-\!-} \tag{5}$$

In Subsect. II F we will interpret events from the trivial system to itself as *probabilities*.

Another important case of tests is that of *single-outcome* tests, in which the outcome space X consists of a single element: $\mathrm{X}=\{i_0\}$. Whenever a device represented by a single-outcome test is used, the experimenter is sure that only one event can take place. This motivates the following definition:

Definition 2 (Deterministic tests) *A test is* deterministic *if its outcome set has a single element, namely* $|\mathrm{X}|=1$.

B. Sequential composition of tests

Physical devices can be used in sequences, as long as the output of each device coincides with the input of the next one. When two tests are composed in a sequence we obtain a new test, as in the following

Definition 3 (Sequential composition of tests) *If* $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ *is a test from* A *to* B *and* $\{\mathscr{D}_j\}_{j\in\mathrm{Y}}$ *is a test from* B *to* C*, then their sequential composition is test from* A *to* C*, with outcomes* $(i,j)\in\mathrm{X}\times\mathrm{Y}$*, and events* $\{\mathscr{D}_j\circ\mathscr{C}_i\}_{(i,j)\in\mathrm{X}\times\mathrm{Y}}$*. Diagrammatically, the events* $\mathscr{D}_j\circ\mathscr{C}_i$ *are represented as follows*

$$\overset{\mathrm{A}}{-\!-\!-}\boxed{\mathscr{C}_i}\overset{\mathrm{B}}{-\!-\!-}\boxed{\mathscr{D}_j}\overset{\mathrm{C}}{-\!-\!-} \;:=\; \overset{\mathrm{A}}{-\!-\!-}\boxed{\mathscr{D}_j\circ\mathscr{C}_i}\overset{\mathrm{C}}{-\!-\!-} \tag{6}$$

We will say that test $\{\mathscr{D}_j\}$ "follows" test $\{\mathscr{C}_i\}$, or, equivalently, $\{\mathscr{C}_i\}$ "precedes" $\{\mathscr{D}_j\}$. For the moment, the order of composition is not necessarily temporal. The interpretation of sequential composition as a sequence of time-steps will be given in Subsect. III within the framework of causal theories.

The sequential composition of tests brings immediately the notion of *identity test*.

Definition 4 (Identity test) *The identity test for system* A *is a test with a single event* $\mathscr{I}_{\mathrm{A}}$ *such that for every system* B

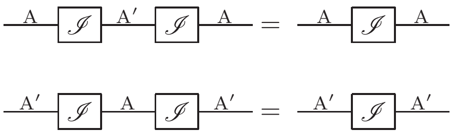
$$\begin{aligned}
&\overset{\mathrm{A}}{-\!-\!-}\boxed{\mathscr{I}}\overset{\mathrm{A}}{-\!-\!-}\boxed{\mathscr{C}_i}\overset{\mathrm{B}}{-\!-\!-} \;=\; \overset{\mathrm{A}}{-\!-\!-}\boxed{\mathscr{C}_i}\overset{\mathrm{B}}{-\!-\!-} \qquad \forall\mathscr{C}_i\in\mathfrak{T}(\mathrm{A},\mathrm{B})\\
&\overset{\mathrm{B}}{-\!-\!-}\boxed{\mathscr{D}_j}\overset{\mathrm{A}}{-\!-\!-}\boxed{\mathscr{I}}\overset{\mathrm{A}}{-\!-\!-} \;=\; \overset{\mathrm{B}}{-\!-\!-}\boxed{\mathscr{D}_i}\overset{\mathrm{A}}{-\!-\!-} \qquad \forall\mathscr{D}_j\in\mathfrak{T}(\mathrm{B},\mathrm{A})
\end{aligned} \tag{7}$$

Performing the identity test on a system just means "doing nothing" on it. We can think of the outcome of the identity test as a blank character, which provides no information.

In some protocols, such as teleportation, one wants to emphasize that one is dealing with two different systems "of the same type". For examples, in quantum theory one can have two electrons in different (spatially separated) regions. Distinguishing two systems of the same type is essentially a matter of bookkeeping. Moreover, we can have different physical systems that are "operationally equivalent", *e. g.* the polarization of a single photon and the spin of an electron in quantum theory are both represented by a *qubit*, and can be (at least in principle) converted one to another in a reversible fashion. For this reason we introduce a formal notion of operational equivalence between systems, based on their mutual convertibility:

Definition 5 (Operationally equivalent systems) *Two systems* A *and* A' *are* operationally equivalent*— denoted as* $\mathrm{A}'\simeq\mathrm{A}$*— if there exist a deterministic test*

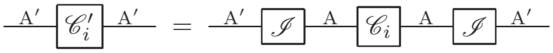
$\{\mathscr{I}_{\mathrm{A,A'}}\}$ *from* A *to* A′ *and a deterministic test* $\{\mathscr{I}_{\mathrm{A',A}}\}$ *from* A′ *to* A*, respectively, such that*



A
$\mathscr{I}$
A′
$\mathscr{I}$
A
=
A
$\mathscr{I}$
A
A′
$\mathscr{I}$
A
$\mathscr{I}$
A′
=
A′
$\mathscr{I}$
A′

(8)

Accordingly, if $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ is a test for system A, performing the "same test" on system A′ means performing the test $\{\mathscr{C}'_i\}_{i\in\mathrm{X}}$ defined by



A′
$\mathscr{C}'_i$
A′
=
A′
$\mathscr{I}$
A
$\mathscr{C}_i$
A
$\mathscr{I}$
A′

(9)

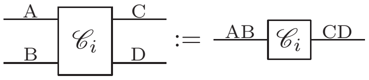
Clearly, the above notion of "same test on a different system" depends on the choice of the privileged test $\{\mathscr{I}_{\mathrm{A,A'}}\}$ used to set up the operational equivalence between A and A′. We will often drop the primes and write $\mathscr{C}_i$ instead of $\mathscr{C}'_i$.

C. Composite systems and parallel composition of tests

Given two systems A and B, one can consider them together, thus forming the corresponding *composite system*, here denoted by AB. A test with input (output) system AB (CD), represents one use of a physical device with two input (output) ports, labeled by A and B (C and D), respectively.

Definition 6 (Composite system) *If* A, B *are systems, the corresponding composite system is* AB*. Composition of systems enjoys the properties* i) A = IA = AI*,* ii) AB ≃ BA*, and* iii) A(BC) = (AB)C := ABC*.*

Diagrammatically, an event from AB *to* CD *is represented as a box with multiple wires:*

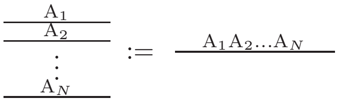


A
B
$\mathscr{C}_i$
C
D
:=
AB
$\mathscr{C}_i$
CD

(10)

The property *i)* in Def. 6 expresses the fact that system A together with "nothing" is still system A, while properties *ii)* and *iii)* express the fact that the specification of a composite system depends only on the list of component systems, and not on how the elements of the list are ordered (up to operational equivalence, implemented by a deterministic test that permutes the component systems), nor on how they are grouped.

In general, we will represent the N-partite composite system $\mathrm{A}_1\ldots\mathrm{A}_N$ with N wires, as follows:



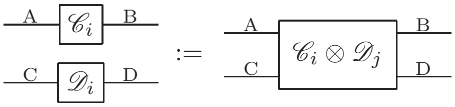
A_1
A_2
$\vdots$
A_N
:=
$\mathrm{A}_1\mathrm{A}_2\ldots\mathrm{A}_N$

(11)

In the case of trivial systems, we will typically omit the wire. In the sequential composition of two boxes with multiple wires we will always match the output wires of the first box with the input wires of the second.

Physical devices can be run in parallel on different systems, thus performing a test on the composite system, as in the following

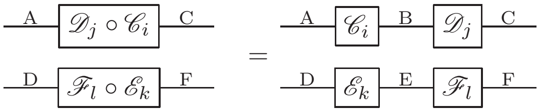
Definition 7 (Parallel composition of tests) *If* $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ *is a test from* A *to* B *and* $\{\mathscr{D}_j\}_{j\in\mathrm{Y}}$ *is a test from* C *to* D*, then their parallel composition is the test from* AC *to* BD*, with outcomes* $(i,j)\in\mathrm{X}\times\mathrm{Y}$*, and events* $\{\mathscr{C}_i\otimes\mathscr{D}_j\}_{(i,j)\in\mathrm{X}\times\mathrm{Y}}$*. Diagrammatically the events* $\mathscr{C}_i\otimes\mathscr{D}_j$ *are represented as follows*



A
$\mathscr{C}_i$
B
C
$\mathscr{D}_i$
D
:=
A
C
$\mathscr{C}_i\otimes\mathscr{D}_j$
B
D

(12)

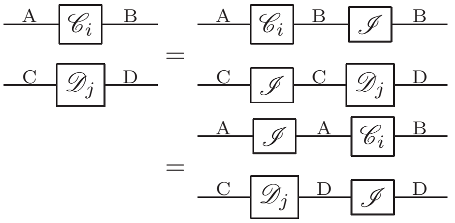
If $\mathscr{C}_i, \mathscr{D}_j, \mathscr{E}_k, \mathscr{F}_l$ *are events from* A *to* B*,* B *to* C*,* D *to* E*, and* E *to* F*, respectively, their parallel composition enjoys the property*



A
$\mathscr{D}_j\circ\mathscr{C}_i$
C
D
$\mathscr{F}_l\circ\mathscr{E}_k$
F
=
A
$\mathscr{C}_i$
B
$\mathscr{D}_j$
C
D
$\mathscr{E}_k$
E
$\mathscr{F}_l$
F

(13)

Note that property (13) implies that tests on different systems commute, that is, for every couple of events $\mathscr{C}_i, \mathscr{D}_j$



A
$\mathscr{C}_i$
B
C
$\mathscr{D}_j$
D
=
A
$\mathscr{C}_i$
B
$\mathscr{I}$
B
C
$\mathscr{I}$
C
$\mathscr{D}_j$
D
=
A
$\mathscr{I}$
A
$\mathscr{C}_i$
B
C
$\mathscr{D}_j$
D
$\mathscr{I}$
D

(14)

From now on, in diagrams like the above we will typically omit the box with identity test, leaving just a wire for the corresponding system. Also in formulae we will often omit the identity, e.g. for $\mathscr{C}\in\mathfrak{T}(\mathrm{A},\mathrm{B})$ and $\rho\in\mathfrak{S}(\mathrm{AC})$ we will often write $\mathscr{C}\,|\rho)_{\mathrm{AB}}$ in place of $(\mathscr{C}\otimes\mathscr{I}_{\mathrm{C}})\,|\rho)_{\mathrm{AC}}$.

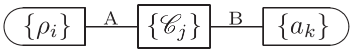
Note that the difference between parallel and sequential composition of two tests is already encoded in their input and output spaces: if the input of a test is the output of the other the composition is sequential, if all spaces are distinct the composition is parallel. For this reason, when the kind of composition is evident we will omit the symbols $\circ$ and $\otimes$. For example, if ρ is a preparation-event for A and $\mathscr{C}$ is an event from A to B we will write $\mathscr{C}\,|\rho)_{\mathrm{A}}$ in place of $\mathscr{C}\circ|\rho)_{\mathrm{A}}$, whereas if ρ and σ are preparation-events for A and B, respectively, we will write $|\rho)_{\mathrm{A}}\,|\sigma)_{\mathrm{B}}$ in place of $|\rho)_{\mathrm{A}}\otimes|\sigma)_{\mathrm{B}}$.

D. Operational theories

We are now in position to make more precise the notion "operational theory":

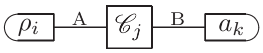
Definition 8 (Operational theory) *An operational theory is specified by a collection of systems, closed under composition, and by a collection of tests, closed under parallel and sequential composition.*

In an operational theory one can draw circuits that *i)* represent the connections of physical devices in an experiment, like e.g. the circuit



(15)

and *ii)* can also represent which specific set of events took place in the experiment, like e.g. the circuit



(16)

In particular, the latter circuit represents the preparation-event ρ_i followed by the event $\mathscr{C}_j$ from system A to system B, which is in turn followed by the observation-event a_k on system B. The whole sequence can be seen as single event $p_{kji} := (a_k|_{\mathrm{B}} \mathscr{C}_j \, |\rho_i)_{\mathrm{A}}$ from the trivial system to itself.

E. Relation with category theory

In the previous Subsections we presented in an informal way the basic notions pertaining to the use of physical devices in sequences and in parallel. More formally, these notions can be summarized with the language of category theory [24], which provides the suitable mathematical framework capturing the fundamental structure presented so far. In this language, an operational theory is a category, where systems and events are respectively *objects* and *arrows*. Every arrow has an input and an output object, and arrows can be sequentially composed. A test is then a collection of arrows labeled by outcomes.

The fact that in an operational theory we have a parallel composition of systems, and that such a composition is symmetric (i.e. AB $\simeq$ BA) is expressed in technical words by saying that we have a *strict symmetric monoidal category* [24]. In the next Subsection we will specify more requirements on this category, imposing that the scalars (arrows from the trivial system to itself) are probabilities.

F. Probabilistic structure: states, effects, and transformations

An operational theory is a language, whose words are diagrams representing circuits. With this language one can give instructions to build up experiments or, alternatively, one can graphically represent which particular outcomes took place in an experiment. However, in a physical theory one wants more: one wants to give probabilistic predictions about the occurrence of possible outcomes. To have this, there must be a rule assigning a probability to every event from the trivial system to itself [25]. More directly, we can say that in a probabilistic theory the events from the trivial system to itself are probabilities, as in the following

Definition 9 (Operational-probabilistic theory) *An operational theory is* probabilistic *if for every test $\{p_i\}_{i\in \mathrm{X}}$ from the trivial system* I *to itself one has $p_i \in [0,1]$ and $\sum_{i\in\mathrm{X}} p_i = 1$, and the composition of two events from the trivial system to itself is given by the product of probabilities: $p_i \otimes q_j = p_i \circ q_j = p_i q_j$.*

For short, we will often refer to operational-probabilistic theories simply as probabilistic theories.

In a probabilistic theory, a preparation-event ρ_i for system A defines a function $\hat{\rho}_i$ sending observation-events of A to probabilities:

$$\hat{\rho}_i : \mathfrak{E}(\mathrm{A}) \to [0,1], \quad (a_j| \mapsto (a_j|\rho_i). \tag{17}$$

Likewise, an observation-event a_j defines a function $\hat{a}_j$ from preparation-events to probabilities

$$\hat{a}_j : \mathfrak{S}(\mathrm{A}) \to [0,1], \quad |\rho_i) \mapsto (a_j|\rho_i). \tag{18}$$

From a probabilistic point of view, two observation-events (preparation-events) corresponding to the same function are indistinguishable. This leads to the notions of states and effects (see [15, 26]):

Definition 10 (States and effects) *Equivalence classes of indistinguishable preparation-events are called* states. *Equivalence classes of indistinguishable observation-events are called* effects.

From now on we will identify preparation-events with states and observation-events with effects, without keeping the distinction between an event ρ_i (a_j) and the corresponding function $\hat{\rho}_i$ ($\hat{a}_j$). Accordingly, a preparation(observation)-test will be a collection of states (effects), and the sets $\mathfrak{S}(\mathrm{A}), \mathfrak{E}(\mathrm{A})$ will be the set of states and and the set of effects of system A, respectively.

Remark (states and effects in quantum theory). In quantum theory systems are associated with Hilbert spaces. The deterministic states of a system A are represented by density matrices on the corresponding Hilbert space: a deterministic state ρ is a matrix satisfying $\rho \geq 0$ and $\mathrm{Tr}[\rho] = 1$. A non-deterministic preparation-test $\{\rho_i\}_{i\in\mathrm{X}}$, sometimes called a quantum information source, is a collection of positive operators with the property $\sum_{i\in\mathrm{X}} \mathrm{Tr}[\rho_i] = 1$. Accordingly, the set $\mathfrak{S}(\mathrm{A})$ of all states of system A is the collection of all unnormalized density matrices ρ with $\mathrm{Tr}[\rho] \leq 1$. An effect is represented by positive operator P with $P \leq I_{\mathrm{A}}$ (I_{A} being the identity operator), and the probability resulting from the pairing between a state ρ and and effect P is given by the Born rule: $(P|\rho)_{\mathrm{A}} = \mathrm{Tr}[P\rho]$.

Notice that according to the definition of states and effects as equivalence classes, states are *separating* for effects and effects are *separating* for states, that is,

$$\begin{aligned} |\rho_0)_{\mathrm{A}} = |\rho_1)_{\mathrm{A}} &\Longleftrightarrow (a|\rho_0)_{\mathrm{A}} = (a|\rho_1)_{\mathrm{A}} \quad \forall a \in \mathfrak{E}(\mathrm{A}) \\ (a_0|_{\mathrm{A}} = (a_1|_{\mathrm{A}} &\Longleftrightarrow (a_0|\rho)_{\mathrm{A}} = (a_1|\rho)_{\mathrm{A}} \quad \forall \rho \in \mathfrak{S}(\mathrm{A}). \end{aligned} \tag{19}$$

Since states (effects) are functions from effects (states) to probabilities, one can take linear combinations of them. This defines two real vector spaces $\mathfrak{S}_{\mathbb{R}}(\mathrm{A})$ and $\mathfrak{E}_{\mathbb{R}}(\mathrm{A})$, one dual of the other (we recall that the dual of a real vector space V is the real vector space V^* of all linear functions from V to $\mathbb{R}$). In this paper we will always restrict our attention to the case of set of states that span finite dimensional vector spaces. In this case, by construction one has

$$\dim(\mathfrak{S}_{\mathbb{R}}(\mathrm{A})) = \dim(\mathfrak{E}_{\mathbb{R}}(\mathrm{A})). \tag{20}$$

Notice that a spanning set for $\mathfrak{S}_{\mathbb{R}}(\mathrm{A})$ is a separating set for $\mathfrak{E}_{\mathbb{R}}(\mathrm{A})$, while a spanning set for $\mathfrak{E}_{\mathbb{R}}(\mathrm{A})$ is a separating set for $\mathfrak{S}_{\mathbb{R}}(\mathrm{A})$.

Moreover, linear combinations with positive coefficients define two convex cones $\mathfrak{S}_+(\mathrm{A})$ and $\mathfrak{E}_+(\mathrm{A})$ (we recall that a set S is a cone if for every $x \in S$ and for every $\lambda \geq 0$ one has $\lambda x \in S$, whereas the set is convex if for every $x, y \in S$ and for every $p \in [0,1]$ one has $px + (1-p)y \in S$). Since the pairing between states and effects yields positive numbers, one has the inclusions

$$\begin{aligned} \mathfrak{E}_+(\mathrm{A}) &\subseteq \mathfrak{S}_+(\mathrm{A})^* \\ \mathfrak{S}_+(\mathrm{A}) &\subseteq \mathfrak{E}_+(\mathrm{A})^*, \end{aligned} \tag{21}$$

where $\mathfrak{S}_+(\mathrm{A})^*$ and $\mathfrak{E}_+(\mathrm{A})^*$ are the dual cones of $\mathfrak{S}_+(\mathrm{A})$ and $\mathfrak{E}_+(\mathrm{A})$, respectively. We recall that the dual of a cone S in some vector space V is the cone S^* defined by $S^* := \{\lambda \in V^*, \lambda(x) \geq 0 \ \forall x \in S\}$.

We conclude this Subsection by noting that every event $\mathscr{C}_k$ from A to B induces a linear map $\hat{\mathscr{C}}_k$ from $\mathfrak{S}_{\mathbb{R}}(\mathrm{A})$ to $\mathfrak{S}_{\mathbb{R}}(\mathrm{B})$, uniquely defined by [27]

$$\hat{\mathscr{C}}_k : |\rho) \in \mathfrak{S}(\mathrm{A}) \mapsto \mathscr{C}_k\, |\rho)_{\mathrm{A}} \in \mathfrak{S}(\mathrm{B}). \tag{22}$$

Likewise, for every system C the event $\mathscr{C}_i \otimes \mathscr{I}_{\mathrm{C}}$ induces a linear map from $\mathfrak{S}_{\mathbb{R}}(\mathrm{AC})$ to $\mathfrak{S}_{\mathbb{R}}(\mathrm{BC})$. From a statistical point of view, if two events $\mathscr{C}_i$ and $\mathscr{C}'_i$ induce the same maps for every possible system C, then they are indistinguishable.

Definition 11 (Transformations) *Equivalence classes of indistinguishable events from* A *to* B *are called* transformations *from* A *to* B.

Again, we will assume that the equivalence classes have been already done since the start, and, consequently, we will identify events with transformations, without introducing new notation. Accordingly, a test will be a collection of transformations.

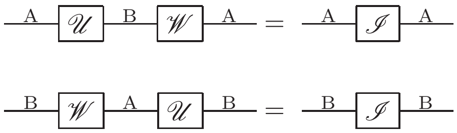
Remark (transformations and tests in quantum theory). In quantum theory, a transformation is usually called *quantum operation.* Technically speaking, a quantum operation from A to B is a linear, completely positive, trace non-increasing map sending density matrices of system A to (unnormalized) density matrices of system B. A test $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ from A to B is typically referred to as a *quantum instrument* [28], and is a collection of quantum operations with the property that $\sum_{i\in\mathrm{X}} \mathscr{C}_i$ is trace-preserving, namely $\sum_{i\in\mathrm{X}} \mathrm{Tr}[\mathscr{C}_i(\rho)] = \mathrm{Tr}[\rho]$ for every state ρ.

Remark (different transformations). Note that two transformations $\mathscr{C}, \mathscr{D} \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ can be different even if $\mathscr{C}\,|\rho)_{\mathrm{A}} = \mathscr{D}\,|\rho)_{\mathrm{A}}$ for every $\rho \in \mathfrak{S}(\mathrm{A})$: indeed to make $\mathscr{C}$ different from $\mathscr{D}$ it is enough that there exists an ancillary system C and a joint state $|\rho)_{\mathrm{AC}}$ such that $(\mathscr{C} \otimes \mathscr{I}_{\mathrm{C}})\,|\rho)_{\mathrm{AC}} \neq (\mathscr{D} \otimes \mathscr{I}_{\mathrm{C}})\,|\rho)_{\mathrm{AC}}$. We will come back on this point when discussing local discriminability in Sect. IV.

The following definitions will be used in the following

Definition 12 (Channel) *A deterministic transformation* $\mathscr{C} \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ *is called* channel.

Definition 13 (Reversible channel) *A channel* $\mathscr{U} \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ *is called* reversible *if there is another channel* $\mathscr{W} \in \mathfrak{T}(\mathrm{B},\mathrm{A})$ *such that*

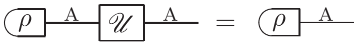


(23)

If there exists a reversible channel $\mathscr{U}$ from A to B, then the systems A and B are operationally equivalent, in the sense of Def. 5. Note that the reversible channels from A to itself form a group. We will denote this group by $\mathbf{G}_{\mathrm{A}}$.

We can now consider states that are invariant under the group of reversible transformations $\mathbf{G}_{\mathrm{A}}$:

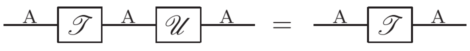
Definition 14 (Invariant states) *A state* $\rho \in \mathfrak{S}(\mathrm{A})$ *is* invariant *under the action of the group* $\mathbf{G}_{\mathrm{A}}$ *if*



$\forall \mathscr{U} \in \mathbf{G}_{\mathrm{A}}$. (24)

Similarly, we can consider channels with invariant output, that we call *twirling channels.*

Definition 15 (Twirling channels/Twirling tests) *A channel* $\mathscr{T} \in \mathfrak{T}(\mathrm{A})$ *is a* twirling-channel *if*



$\forall \mathscr{U} \in \mathbf{G}_{\mathrm{A}}$. (25)

If a test $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ *is such that* $\sum_{i\in\mathrm{X}} \mathscr{C}_i$ *is a twirling channel, we call it a* twirling test.

We will see that in a theory with purification there is a unique invariant state and a unique twirling channel for every system.

G. Relation with the convex sets framework

The standard assumption in the literature is that, since the experimenter is free to randomize the choice of devices with arbitrary probabilities, all sets of states, effects, and transformations are convex. We will call the theories satisfying this assumption "convex". The assumption of convexity will be clarified in Subsect. III D in the context of *causal theories*. Nevertheless, for many of our results the assumption of convexity is not essential, and we will discuss the validity of our results in non-convex theories, like the toy-theories considered by Spekkens in Ref. [29]. Bearing this in mind, whenever possible we will present our results in a convexity-independent language. We will add the specification "convex" to the theory for those particular results in which convexity is essential.

In addition to the convexity of all sets of states, effects, and transformations, the usual convex sets framework (see e.g. Refs. [13, 15, 26], and, more recently, Refs. [8, 17]) includes an assumption of mathematical simplicity. The assumption is that every binary probability rule describes the statistics of a possible two-outcome experiment. Precisely, with the expression "probability rule" we mean a collection of positive linear functionals $\{a_i\}_{i\in\mathrm{X}} \subset \mathfrak{S}^*_+(\mathrm{A})$ such that $\sum_{i\in\mathrm{X}} (a_i|\rho)_{\mathrm{A}} = 1$ for every deterministic state $\rho \in \mathfrak{S}(\mathrm{A})$. We will refer to this assumption as "no-restriction hypothesis", as it states that there is no restriction on the set of (binary) probability rules that can be implemented in actual experiments.

Definition 16 (No-restriction hypothesis) *A probabilistic theory satisfies the* no-restriction hypothesis *if every binary probability rule* $\{a_0, a_1\} \subset \mathfrak{S}^*_+(\mathrm{A})$ *is an observation-test.*

In this paper we will not make this assumption. However, we will discuss a few implications of it in subsections VII D and X C.

H. Coarse-graining and refinement

Here we give some definitions that will be often used in this paper.

Definition 17 (Coarse-graining) *A test* $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ *is a coarse-graining of the test* $\{\mathscr{D}_j\}_{j\in\mathrm{Y}}$ *is there is a partition of* Y *into disjoint sets* Y_i *such that* $\mathscr{C}_i = \sum_{j\in\mathrm{Y}_i} \mathscr{D}_j$ *for every* $i \in \mathrm{X}$.

Since we can always decide to join two (or more) outcomes in a single outcome, the set of all tests must be closed under coarse-graining.

The inverse of coarse-graining is refinement:

Definition 18 (Refinement of a test) *If* $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ *is a* coarse-graining *of* $\{\mathscr{D}_j\}_{j\in\mathrm{Y}}$*, we say that* $\{\mathscr{D}_j\}_{j\in\mathrm{Y}}$ *is a* refinement *of* $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$.

Definition 19 (Refinement of an event) *A* refinement *of the event* $\mathscr{C}$ *is given by a test* $\{\mathscr{D}_j\}_{j\in\mathrm{Y}}$ *and a subset* $\mathrm{Y}_0 \subseteq \mathrm{Y}$ *such that* $\mathscr{C} = \sum_{j\in\mathrm{Y}_0} \mathscr{D}_i$.

Definition 20 *We say that an event* $\mathscr{D} \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ *refines* $\mathscr{C} \in \mathfrak{T}(\mathrm{A},\mathrm{B})$*, and write* $\mathscr{D} \prec \mathscr{C}$*, if there exist a refinement of* $\mathscr{C}$ *such that* $\mathscr{D} \in \{\mathscr{D}_j\}_{j\in\mathrm{Y}_0}$.

Definition 21 (Refinement set) *The refinement set* $D_{\mathscr{C}}$ *of an event* $\mathscr{C} \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ *is the set of all events* $\mathscr{D}$ *that refine* $\mathscr{C}$*, namely* $D_{\mathscr{C}} := \{\mathscr{D} \in \mathfrak{T}(\mathrm{A},\mathrm{B}) \mid \mathscr{D} \prec \mathscr{C}\}$.

Definition 22 (Atomic vs refinable events) *An event* $\mathscr{C}$ *is called* atomic *if it admits only trivial refinements,—equivalently, if* $\mathscr{D} \prec \mathscr{C}$ *implies* $\mathscr{D} = \lambda\mathscr{C}$ *for some* $\lambda \in [0,1]$. *An event is* refinable *if it is not atomic.*

In the case of preparation-events the notion of refinement gives rise to the definitions of pure and mixed states:

Definition 23 (Pure vs mixed states) *An atomic preparation-event* $\rho \in \mathfrak{S}(\mathrm{A})$ *is called* pure state. *A refinable preparation-event is called* mixed state.

Clearly, in a convex theory a state ρ is pure if and only if it is an extreme point of the convex set $\mathfrak{S}(\mathrm{A})$. Moreover, in a convex theory the refinement set D_ρ is a convex subset of the state space. For example, in quantum theory the refinement set of a density matrix ρ is the set of all (unnormalized) density matrices σ such that $\sigma \leq \rho$, and is clearly convex. Note that the condition $\sigma \leq \rho$ implies that the support of σ is contained in the support of ρ. In fact, any density matrix σ with $\mathrm{Supp}(\sigma) \subseteq \mathrm{Supp}(\rho)$, is proportional to a matrix in D_ρ. In particular, if the support of ρ is the whole Hilbert space (that is, if ρ is a full-rank matrix), then any density matrix is proportional to a matrix in D_ρ. In this case D_ρ is a spanning set for the set of all hermitian operators. The analogue of a full rank density matrix in the general context is given by the notion of *internal state*:

Definition 24 (Internal state) *A state* $\omega \in \mathfrak{S}(\mathrm{A})$ *is* internal *if its refinements span the whole state space, i.e. if* $\mathsf{Span}(D_\omega) = \mathfrak{S}_{\mathbb{R}}(\mathrm{A})$.

In the probabilistic theories considered in this paper every preparation-test $\{\rho_i\}_{i\in\mathrm{X}}$ for system A admits an ultimate refinement $\{\varphi_j\}_{j\in\mathrm{Y}}$, such that each state φ_j is pure. Using the states-transformations isomorphism we will also prove in Sect. X that in a theory with purification this property is enough to imply that every test $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ from A to B admits an ultimate refinement $\{\mathscr{D}_j\}_{j\in\mathrm{Y}}$, such that each event $\mathscr{D}_j$ is atomic.

I. Discrimination and distance

By making tests one can try to discriminate between different devices. For example, imagine that we have a black box preparing one of the two deterministic states $\rho_0, \rho_1 \in \mathfrak{S}(\mathrm{A})$, and that we want to find out which one. To discriminate between the two states we can perform a binary observation-test $\{a_0, a_1\}$. The probabilities of outcomes are then given by

$$p(j|i) := (a_j|\rho_i)_\mathrm{A} \qquad i, j = 0, 1. \tag{26}$$

Assuming prior probabilities π_0, π_1 for the states ρ_0, ρ_1, respectively, we can try to maximize the (average) probability of correct discrimination, defined as $p_{succ} := \pi_0\ p(0|0) + \pi_1\ p(1|1)$. Substituting the expression for the probabilities given in Eq. (26) and using the fact probabilities sum up to unit, we obtain

$$\begin{aligned} p_{succ} &= \pi_0 + (a_1|\pi_1\rho_1 - \pi_0\rho_0)_\mathrm{A} \\ &= \pi_1 + (a_0|\pi_0\rho_0 - \pi_1\rho_1)_\mathrm{A}, \end{aligned} \tag{27}$$

and, optimizing over all binary tests,

$$\begin{aligned} p_{succ}^{(opt)} &= \pi_0 + \sup_{a_1\in\mathfrak{E}(\mathrm{A})} (a_1|\pi_1\rho_1 - \pi_0\rho_0)_\mathrm{A} \\ &= \pi_1 + \sup_{a_0\in\mathfrak{E}(\mathrm{A})} (a_0|\pi_0\rho_0 - \pi_1\rho_1)_\mathrm{A}. \end{aligned} \tag{28}$$

Summing the two expressions above we finally get

$$p_{succ}^{(opt)} = \frac{1 + \|\pi_1\rho_1 - \pi_0\rho_0\|_\mathrm{A}}{2} \tag{29}$$

where $\|\cdot\|_\mathrm{A}$ is the *operational norm* defined by

$$\|\delta\|_\mathrm{A} = \sup_{a_1\in\mathfrak{E}(\mathrm{A})} (a_1|\delta)_\mathrm{A} - \inf_{a_0\in\mathfrak{E}(\mathrm{A})} (a_0|\delta)_\mathrm{A} \qquad \delta \in \mathfrak{S}_\mathbb{R}(\mathrm{A}). \tag{30}$$

Note that the norm $\|\pi_1\rho_1 - \pi_0\rho_0\|_\mathrm{A}$ ranges between 0 (when the two states and the prior probabilities are equal) and 1 (when the two states are perfectly discriminable). For real numbers $x \in \mathfrak{S}_\mathbb{R}(\mathrm{I}) \equiv \mathbb{R}$ one has $\|x\|_\mathrm{I} = |x|$.

Remark (operational norm in quantum theory). In quantum theory the operational norm is the usual *trace-norm* $\|\cdot\|_1$: Indeed, if we denote by δ_+ and δ_- the positive and negative part of the hermitian operator $\delta = \pi_1\rho_1 - \pi_0\rho_0$, respectively, we obtain $\|\delta\|_\mathrm{A} = \mathrm{Tr}[\delta_+] - \mathrm{Tr}[\delta_-] = \|\delta\|_1$.

In addition to the defining properties of a norm, the operational norm has a simple monotonicity property:

Lemma 1 (Monotonicity of the operational norm) *If* $\mathscr{C} \in \mathfrak{T}(\mathrm{A}, \mathrm{B})$ *is a channel from* A *to* B, *then for every* $\delta \in \mathfrak{S}_\mathbb{R}(\mathrm{A})$ *one has*

$$\|\mathscr{C}\delta\|_\mathrm{B} \leq \|\delta\|_\mathrm{A}. \tag{31}$$

If $\mathscr{C}$ *is reversible one has the equality.*

Proof. By definition, $\|\mathscr{C}\delta\|_\mathrm{B} = \sup_{b_1\in\mathfrak{E}(\mathrm{B})} (b_1|_\mathrm{B}\mathscr{C}|\delta)_\mathrm{A} - \inf_{b_0\in\mathfrak{E}(\mathrm{B})} (b_0|_\mathrm{B}\mathscr{C}|\delta)_\mathrm{A}$. Since $(b_1|_\mathrm{B}\mathscr{C}$ and $(b_0|_\mathrm{B}\mathscr{C}$ are effects on system A, one has $\|\mathscr{C}\delta\|_\mathrm{B} \leq \sup_{a_1\in\mathfrak{E}(\mathrm{A})} (a_1|_\mathrm{B}|\delta)_\mathrm{A} - \inf_{a_0\in\mathfrak{E}(\mathrm{A})} (a_0|_\mathrm{A}|\delta)_\mathrm{A} = \|\delta\|_\mathrm{A}$. Clearly, if $\mathscr{C}$ is reversible one has the converse bound $\|\delta\|_\mathrm{A} = \|\mathscr{C}^{-1}\mathscr{C}\delta\|_\mathrm{A} \leq \|\mathscr{C}\delta\|_\mathrm{B}$, thus proving the equality $\|\delta\|_\mathrm{A} = \|\mathscr{C}\delta\|_\mathrm{B}$. ■

For a generic state $\rho \in \mathfrak{S}(\mathrm{A})$, Eq. (30) reduces to

$$\|\rho\|_\mathrm{A} = \sup_{e\in\mathfrak{E}(\mathrm{A})}{}'\ (e|\rho), \tag{32}$$

where sup′ denotes the supremum restricted to the set of deterministic effects. We can now give the notion of *normalized states*:

Definition 25 (Normalized states) *A state* $\rho \in \mathfrak{S}(\mathrm{A})$ *is* normalized *if* $\|\rho\|_\mathrm{A} = 1$. *We will denote the set of normalized states by* $\mathfrak{S}_1(\mathrm{A})$.

Clearly, if ρ is deterministic, then Eq. (32) implies that it is normalized (since ρ corresponds to a single-outcome preparation-test and e to a single-outcome observation-test, the probability of the only possible outcome, given by $(e|\rho)_\mathrm{A}$, must be unit). In Sect. III we will consider causal theories, where the deterministic effect $e \in \mathfrak{E}(\mathrm{A})$ is unique, and, therefore one has $\|\rho\|_\mathrm{A} = (e|\rho)_\mathrm{A}$. In this context one also has the converse: if a state is normalized, then it is deterministic.

Definition 26 (Distinguishable states, discriminating tests) *The states* $\{\rho_i\}_{i\in\mathrm{X}}$ *are* perfectly distinguishable *if there is a test* $\{a_i\}_{i\in\mathrm{X}}$ *such that*

$$(a_j|\rho_i) = \|\rho_i\|_\mathrm{A}\ \delta_{ij}. \tag{33}$$

The test $\{a_i\}_{i\in\mathrm{X}}$ *is called* discriminating test.

Remark (Distinguishable states and discriminating test in quantum theory). In quantum theory a set of distinguishable states $\{\rho_i\}_{i=1}^n$ is a set of density matrices with orthogonal support. An example of discriminating test for this set is the collection of orthogonal projectors $\{P_i\}_{i=1}^n$, where P_i is the projector on the support of ρ_i for all $i < n$, while $P_n = I - \sum_{i=1}^{n-1} P_i$. Clearly, the maximum number of distinguishable states available for a certain system is the dimension d of the corresponding Hilbert space. In this case, the distinguishable states are rank-one projectors on an orthonormal basis, and the corresponding discriminating test is the projective measurement on the same basis.

If we want a theory that can describe the exchange of classical messages, we need at least two states ρ_0 and ρ_1 that are deterministic and perfectly distinguishable. In this case, a sender can encode a classical bit $b = 0, 1$ in these two states and a receiver can decode perfectly the message by using the binary discriminating test $\{a_0, a_1\}$. Indeed, one has $p(j|i) = \delta_{ij}$. Clearly, using this encoding for any bit in a string allows perfect deterministic decoding of the whole string.

We conclude this Subsection with a simple Lemma that will be useful in the discussion of the general no-cloning theorem for probabilistic theories (see Theorem 12):

Lemma 2 *In any convex theory, if two deterministic states $\rho_0, \rho_1 \in \mathfrak{S}(\mathrm{A})$ are distinct (*i.e. $\rho_0 \neq \rho_1$*), then there exists a binary test* $\{a_0, a_1\}$ *such that*

$$p(1|0) = p(0|1) < \frac{1}{2}. \tag{34}$$

Proof. Since the states are distinct there exists at least an effect a such that $(a|\rho_0) > (a|\rho_1)$. Moreover, since the theory is convex we can choose without loss of generality $(a|\rho_1) \geq 1/2$ (if a does not meet this condition, we can replace it with the convex combination $a' = 1/2(a+e)$). Now define the binary test $\{a_0, a_1\}$ by the convex combination

$$\begin{cases} a_0 &= qa + (1-q)0 \\ a_1 &= e - a_0 \end{cases} \qquad q = \frac{1}{(a|\rho_0) + (a|\rho_1)} \tag{35}$$

where 0 is the null effect, defined by $(0|\rho)_\mathrm{A} = 0, \forall \rho \in \mathfrak{S}(\mathrm{A})$. For this test one has $p(1|0) = p(0|1) = (a|\rho_1)/[(a|\rho_0) + (a|\rho_1)] < 1/2$. ■

The above Lemma states that if two states are different, then the worst-case error probability, defined as $p_{wc} := \max\{p(1|0), p(0|1)\}$, can be reduced to a value that is strictly smaller than 1/2. In other words, if two states are different, then in the worst-case scenario we can always distinguish between them better than with a random guess.

J. Closure

The closure of $\mathfrak{S}(\mathrm{A})$ with respect to the operational norm contains all the elements of $\mathfrak{S}_\mathbb{R}(\mathrm{A})$ that can be approximated arbitrarily well by physical states: a vector $\rho \in \mathfrak{S}_\mathbb{R}(\mathrm{A})$ is in the closure if there is a sequence of states $\{\rho_n\}$ such that $\lim_{n\to\infty} \|\rho - \rho_n\|_\mathrm{A} = 0$. Since $\mathfrak{S}_\mathbb{R}(\mathrm{A})$ is finite dimensional, it is natural to assume that all such vectors correspond to physical states. We will make this assumption in the paper. In particular, assuming that the set $\mathfrak{S}(\mathrm{I})$ of states of the trivial system is closed with respect to the operational norm means assuming that the probabilities appearing in the theory form a closed subset of the interval $[0,1]$. In fact, we have the following:

Lemma 3 *If an operational-probabilistic theory is not deterministic, then $\mathfrak{S}(\mathrm{I})$ is dense in the interval* $[0,1]$.

Proof. If the theory is not deterministic there is a binary test giving outcomes 0, 1 with probabilities $q_0, q_1 \neq 0$, respectively. Now, this test provides a biased coin, which can be tossed many times, thus allowing for the approximation of any coin with bias $p \in [0,1]$ [30]. ■

Therefore, if we assume that the set of states $\mathfrak{S}(\mathrm{I})$ is closed, then the previous Lemma implies the following:

Corollary 1 *If $\mathfrak{S}(\mathrm{I})$ is closed, then it is the whole interval* $[0,1]$.

In Subsect. III D we will discuss the relation between closure and convexity in the context of causal theories.

III. CAUSAL THEORIES

In this Section we restrict our attention to causal theories, in which the probability of outcomes of an experiment at a given time does not depend on the choice of experiments performed at later times.

A. Definition and main properties

Although in the circuits discussed until now we had sequences of tests, such sequences were not necessarily *causal sequences*. The input-output arrow determined by the connections of physical devices was not necessarily the causal arrow defined a signalling structure. In fact, one can formulate operational-probabilistic theories even in the absence of a pre-defined causal arrow, and this is a crucial point to formulate a quantum theory of gravity (see e.g. Hardy in Ref. [31]). A concrete example of non-causal theory is the theory studied in Refs. [32, 33], where the states are quantum operations, and the transformations are "supermaps" transforming quantum operations into quantum operations. In this case, transforming a "state" means inserting the corresponding quantum operation in a larger circuit, and the sequence of two such transformations is not a causal sequence. However, the analysis of non-causal theories is not the scope of the present work. We now give the condition that allows us to interpret sequential composition as a causal cascade:

Definition 27 (Causal theories) *A theory is* causal *if for every preparation-test $\{\rho_i\}_{i\in\mathrm{X}}$ and every observation-test $\{a_j\}_{j\in\mathrm{Y}}$ on system* A *the marginal probability $p_i := \sum_{j\in\mathrm{Y}} (a_j|\rho_i)_\mathrm{A}$ is independent of the choice of the observation-test $\{a_j\}_{j\in\mathrm{Y}}$. Precisely, if $\{a_j\}_{j\in\mathrm{Y}}$ and $\{b_k\}_{k\in\mathrm{Z}}$ are two different observation-tests, then one has*

$$\sum_{j\in\mathrm{Y}} (a_j|\rho_i)_\mathrm{A} = \sum_{k\in\mathrm{Z}} (b_k|\rho_i)_\mathrm{A}. \tag{36}$$

Loosely speaking, we may say that the condition of Eq. (36) expresses the principle of "no-signaling from the future".

Causal theories have a simple characterization:

Lemma 4 (Characterization of causal theories) *A theory is causal if and only if for every system* A *there is a unique deterministic effect* $(e|_\mathrm{A}$.

Proof. Suppose that e and e' are two deterministic effects for system A. Since deterministic effects belong to single-outcome tests, Eq. (36) gives $(e|\rho_i)_\mathrm{A} = (e'|\rho_i)_\mathrm{A}$ for every state ρ_i. Therefore, $e = e'$. Conversely, suppose that the deterministic effect is unique and take an observation-test $\{a_j\}_{j\in\mathrm{Y}}$ on system A. Then by coarse-graining one obtains a single-outcome test, with deterministic effect $(e'|_\mathrm{A} = \sum_{j\in\mathrm{Y}} (a_j|_\mathrm{A}$, and, by uniqueness of the deterministic effect, $(e|_\mathrm{A} = (e'|_\mathrm{A} = \sum_{j\in\mathrm{Y}} (a_j|_\mathrm{A}$. Therefore, for every state ρ_i we have $\sum_{j\in\mathrm{Y}} (a_j|\rho_i)_\mathrm{A} = (e|\rho_i)_\mathrm{A}$, independently of the choice of the observation-test $\{a_j\}_{j\in\mathrm{Y}}$. This proves Eq. (36). ■

Remark (quantum theory as an example of causal theory) Ordinary quantum theory is an example of causal theory. Indeed, there is a unique deterministic effect, corresponding to the (trace with the) identity operator on the system's Hilbert space. In other words, the only operator P satisfying the equation $\mathrm{Tr}_\mathrm{A}[P\rho] = 1$ for every density matrix is $P = I_\mathrm{A}$, the identity on A.

An immediate consequence of causality is that the deterministic effect of a composite system AB is the product of the deterministic effects of A and B, as expressed by the following

Corollary 2 (Factorization of the deterministic effect on product systems) *Let* A *and* B *be two arbitrary systems. In a causal theory one has*

$$(e|_\mathrm{AB} = (e|_\mathrm{A}\,(e|_\mathrm{B}\,. \tag{37}$$

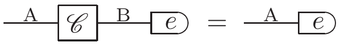
Proof. Since the parallel composition of two single-outcome tests is a single-outcome test, the effect $(e|_\mathrm{A}\,(e|_\mathrm{B}$ is deterministic, according to Def. 2. Since the deterministic effect $(e|_\mathrm{AB}$ is unique, one must have $(e|_\mathrm{A}\,(e|_\mathrm{B} = (e|_\mathrm{AB}$. ■

Note that in a causal theory there is a unique way of defining marginal states:

Definition 28 (Marginal state) *The marginal state of* $|\sigma)_\mathrm{AB}$ *on system* A *is the state* $|\rho)_\mathrm{A} := (e|_\mathrm{B}\,|\sigma)_\mathrm{AB}$.

In a causal theory the channels (deterministic transformations corresponding to single-outcome tests) are characterized as follows:

Lemma 5 (Characterization of channels) *In a causal theory a transformation* $\mathscr{C} \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ *is a channel (Def. 12) if and only if* $(e|_\mathrm{B}\mathscr{C} = (e|_\mathrm{A}$. *Diagrammatically,*



(38)

In particular, a state $\rho \in \mathfrak{S}(\mathrm{B})$ *is deterministic if and only if* $(e|\rho)_\mathrm{B} = 1$.

Proof. If $\mathscr{C}$ is a channel, then $(e|_\mathrm{B}\mathscr{C}$ is a deterministic effect. By uniqueness of the deterministic effect, Eq. (38) holds. Conversely, suppose that $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ is a test from A to B and $\mathscr{C} \equiv \mathscr{C}_{i_0}$ is a transformation such that Eq. (38) holds. By coarse-graining, we can define the channel $\mathscr{C}' := \sum_{i\in\mathrm{X}} \mathscr{C}_i$. Since $\mathscr{C}'$ is a channel, we must have $(e|_\mathrm{A} = (e|_\mathrm{B}\mathscr{C}' = (e|_\mathrm{A} + (e|_\mathrm{B}\,(\sum_{i\neq i_0} \mathscr{C}_i)$, whence $(e|_\mathrm{B}\,(\sum_{i\neq i_0} \mathscr{C}_i) = 0$. But this implies $\sum_{i\neq i_0} \mathscr{C}_i = 0$, and, therefore, $\mathscr{C} = \mathscr{C}'$. Hence, $\mathscr{C}$ is a channel. Finally, a deterministic state is nothing but a channel with trivial input system $\mathrm{A} = \mathrm{I}$. Since the deterministic effect of the trivial system I is the number 1, the normalization of Eq. (38) becomes $(e|\rho)_\mathrm{B} = 1$. ■

Lemma 5 also leads to the following

Corollary 3 (Normalization of tests) *A test* $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ *from* A *to* B *satisfies the normalization condition*

$$\sum_{i\in\mathrm{X}} (e|_\mathrm{B}\,\mathscr{C}_i = (e|_\mathrm{A}\,. \tag{39}$$

In particular, an observation-test $\{a_i\}_{i\in\mathrm{X}}$ *on system* A *must satisfy the normalization condition*

$$\sum_{i\in\mathrm{X}} (a_i|_\mathrm{A} = (e|_\mathrm{A}\,. \tag{40}$$

In quantum theory, the normalization condition of Eq. (38) means that any quantum channel must be trace-preserving (identity preserving in the Heisenberg picture). Indeed, the deterministic effect is the identity operator, and Eq. (38) implies that, for every quantum state ρ, one has $\mathrm{Tr}_\mathrm{B}[\mathscr{C}(\rho)] = \mathrm{Tr}_\mathrm{A}[\rho]$. The normalization condition for observation-tests given in Eq. 40 is instead the normalization of quantum measurements: a quantum measurement is a POVM, that is a collection of positive operators $\{A_i\}_{i\in\mathrm{X}}$ satisfying the condition $\sum_{i\in\mathrm{X}} A_i = I_\mathrm{A}$, where I_A is the identity operator on the system's Hilbert space.

Moreover, in a causal theory we have a simple characterization of the normalized states:

Corollary 4 (Characterization of normalized states) *Let* ρ *be a state of system* A. *In a causal theory the following are equivalent*

1. ρ *is normalized*
2. $(e|\rho)_\mathrm{A} = 1$
3. ρ *is deterministic.*

Proof. Since there is a unique deterministic effect, the expression of the norm given in Eq. (32) yields $\|\rho\|_\mathrm{A} = (e|\rho)_\mathrm{A}$. This proves the equivalence 1 ⇔ 2. The equivalence 2 ⇔ 3 was already proved in Lemma 5. ■

For every state $|\rho)_\mathrm{A}$ we can consider the normalized state

$$|\bar{\rho})_\mathrm{A} := \frac{|\rho)_\mathrm{A}}{(e|\rho)_\mathrm{A}}. \tag{41}$$

Operationally, this means that we can always make *rescaled preparations*: we can perform a preparation-test $\{\rho_i\}_{i\in\mathrm{X}}$, and, if the test gives outcome i_0 we can claim that we prepared the normalized state $\bar{\rho}_{i_0}$. In other words, in a causal theory any preparation-event can be promoted to a single-outcome preparation-test. Following this observation, in a causal theory there is no reason to forbid that every normalized state can be actually produced in some single-outcome test. This implies that every state is proportional to a deterministic one. In the following we will always assume this fact as a property of causal theories.

Note that also the converse is true:

Lemma 6 (Causality is necessary for rescaled preparations) *A theory where every state is proportional to a deterministic one is causal.*

Proof. Let $|\rho)_{\mathrm{A}}$ be an arbitrary state and e and e' be two deterministic effects. By hypothesis, we have $|\rho)_{\mathrm{A}} = k\,|\bar\rho)_{\mathrm{A}}$, where $\bar\rho$ is deterministic. This implies $(e|\,\rho)_{\mathrm{A}} = k = (e'|\,\rho)$, and, since ρ is arbitrary $e = e'$. By lemma 4, this implies that the theory is causal. ■

Remarkably, the causal principle of "no-signalling from the future" implies the impossibility of signalling in space without exchange of physical systems:

Theorem 1 (No-signalling without exchange of physical systems) *In a causal theory it is impossible to have signalling without exchanging systems.*

Proof. Suppose that two distant parties Alice and Bob share a bipartite state $|\Psi)_{\mathrm{AB}}$, and that Alice (Bob) performs a local test $\{\mathscr{A}_i\}_{i\in\mathrm{X}}$ ($\{\mathscr{B}_j\}_{j\in\mathrm{Y}}$) on the system at her (his) disposal. Let us define the joint probability $p_{ij} := (e|_{\mathrm{AB}}\,(\mathscr{A}_i \otimes \mathscr{B}_j)\,|\Psi)_{\mathrm{AB}}$ and its marginal $p^{(\mathrm{A})}_i := \sum_j p_{ij}$ ($p^{(\mathrm{B})}_j := \sum_i p_{ij}$) on Alice's (Bob's) side. It is immediate to verify that the marginal $p^{(\mathrm{A})}_i$ on Alice's side does not depend on the test $\{\mathscr{B}_j\}$ on Bob's side: indeed, one has

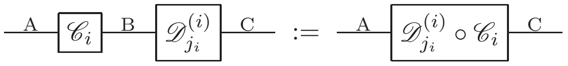
$$\begin{aligned}
p^{(\mathrm{A})}_i &= \sum_j (e|_{\mathrm{A}}\,(e|_{\mathrm{B}}\,(\mathscr{A}_i \otimes \mathscr{B}_j)\,|\Psi)_{\mathrm{AB}} \\
&= (e|_{\mathrm{A}} \left(\mathscr{A}_i \otimes \left[\sum_j (e|_{\mathrm{B}}\,\mathscr{B}_j \right] \right) |\Psi)_{\mathrm{AB}} \\
&= (e|_{\mathrm{A}}\,\mathscr{A}_i\,|\rho)_{\mathrm{A}},
\end{aligned} \tag{42}$$

having used the normalization condition $\sum_j (e|_{\mathrm{B}}\,\mathscr{B}_j = (e|_{\mathrm{B}}$ (Corollary 3), and having defined the marginal state $|\rho)_{\mathrm{A}} := (e|_{\mathrm{B}}\,|\Psi)_{\mathrm{AB}}$. The same reasoning holds for the marginal on Bob's side. ■

B. Conditioning

In a causal sequence the choice of a device can depend on the outcomes of previous devices. This gives rise to the notion of *conditioned test*, which generalizes the notion of sequential composition:

Definition 29 (Conditioned test) *If* $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ *is a test from* A *to* B *and, for every* i, $\left\{\mathscr{D}^{(i)}_{j_i}\right\}_{j_i\in\mathrm{Y}_i}$ *is a test from* B *to* C*, then the* conditioned test *is a test from* A *to* C*, with outcomes* $(i,j_i) \in \mathrm{Z} := \bigcup_{i\in\mathrm{X}} \{i\} \times \mathrm{Y}_i$*, and events* $\{\mathscr{D}^{(i)}_{j_i} \circ \mathscr{C}_i\}_{(i,j_i)\in\mathrm{Z}}$*. Diagrammatically, the events* $\mathscr{D}^{(i)}_{j_i} \circ \mathscr{C}_i$ *are represented as follows*


A
$\mathscr{C}_i$
B
$\mathscr{D}^{(i)}_{j_i}$
C
:=
A
$\mathscr{D}^{(i)}_{j_i} \circ \mathscr{C}_i$
C
(43)

The above definition of conditioning makes sense in a causal theory, where the uniqueness of the deterministic effect ensures that the test $\{\mathscr{D}^{(i)}_{j_i} \circ \mathscr{C}_i\}_{i\in\mathrm{X}, j_i\in\mathrm{Y}_i}$ satisfies the normalization condition required by Corollary 3:

$$\sum_{i\in\mathrm{X}} \sum_{j_i\in\mathrm{Y}_i} (e|_{\mathrm{C}}\,\mathscr{D}^{(i)}_{j_i} \circ \mathscr{C}_i = \sum_{i\in\mathrm{X}} (e|_{\mathrm{B}}\,\mathscr{C}_i = (e|_{\mathrm{A}}\,. \tag{44}$$

Conditioning expresses the possibility of choosing what to do at a certain step using the classical information generated in the previous steps. In a causal operational theory there is no reason to forbid an experimenter to perform conditioned tests. Accordingly, in the following we will assume that in a causal theory any conditioned test is allowed. In fact, the possibility to perform conditioned tests is essentially equivalent to causality. Indeed, one has also the converse statement:

Lemma 7 (Causality is necessary for conditioned tests) *A theory where every conditioned test is possible is causal.*

Proof. To prove that the theory is causal we show that for every system A the deterministic effect $(e|_{\mathrm{A}}$ is unique. Suppose that $(e|_{\mathrm{A}}$ and $(e'|_{\mathrm{A}}$ are two deterministic effects, and let $\rho \in \mathfrak{S}(\mathrm{A})$ be an arbitrary state. By definition, there is a preparation-test $\{\rho_i\}_{i\in\mathrm{X}}$ that contains ρ, that is, $\rho = \rho_{i_0}$ for some outcome $i_0 \in \mathrm{X}$. Moreover, using coarse-graining we obtain the two-outcome preparation-test $\{\rho_0, \rho_1\}$, where $\rho_0 = \rho$ and $\rho_1 := \sum_{i\neq i_0} \rho_i$. Now, consider the conditioned test $\{(e|\,\rho_0)_{\mathrm{A}}, (e'|\,\rho_1)_{\mathrm{A}}\}$, defined by the following procedure: first perform the preparation-test $\{\rho_0, \rho_1\}$, and then, if the outcome is 0 apply the effect $(e|_{\mathrm{A}}$, otherwise apply $(e'|_{\mathrm{A}}$. Since $\{(e|\,\rho_0)_{\mathrm{A}}, (e'|\,\rho_1)_{\mathrm{A}}\}$ is a test from the trivial system to itself one must have

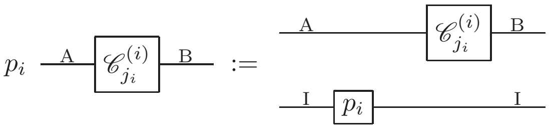
$$(e|\,\rho_0)_{\mathrm{A}} + (e'|\,\rho_1)_{\mathrm{A}} = 1 \tag{45}$$

On the other hand, since the effect e' is deterministic, one must have $(e'|\,\rho_0)_{\mathrm{A}} + (e'|\,\rho_1)_{\mathrm{A}} = 1$. By comparison, this implies $(e|\,\rho_0)_{\mathrm{A}} = (e'|\,\rho_0)_{\mathrm{A}}$, and, since ρ_0 was a generic state, $e = e'$. ■

Remark (conditioning with different outputs and "direct sum" systems). In principle, one could also consider a conditioning where the output system of each test $\left\{\mathscr{D}^{(i)}_{j_i}\right\}$ is a system C_i that depends on the outcome i. In this case the output of the conditioned test would be a "direct sum" system "$\mathrm{C} := \bigoplus_{i\in\mathrm{X}} \mathrm{C}_i$". In quantum theory, this situation can be described introducing a superselection rule, according to which the possible states of the "direct sum" system are the block-diagonal density matrices of the form $\rho = \bigoplus_{i\in \mathrm{X}} \rho_i$, where each ρ_i is a density matrix on the Hilbert space associated to system C_i. This kind of extension would also require treating the outcome spaces X as a *classical systems* that can be the input or the output of some classical information-processing device. However, we will not consider here this generalization as it is not needed for the main purpose of the paper.

A particular case of conditioning is *randomization*:

Definition 30 (Randomization) *If* $\{p_i\}_{i\in \mathrm{X}}$ *is a preparation-test for the trivial system and, for every outcome* i, $\{\mathscr{C}^{(i)}_{j_i}\}_{j_i\in \mathrm{Y}_i}$ *is a test from* A *to* B, *the* randomized test $\{p_i\mathscr{C}^{(i)}_{j_i}\}_{i\in \mathrm{X}, j_i\in \mathrm{Y}_i}$ *is the test from* A *to* B *with events defined by*



(46)

(on the left-hand side we used the fact that that the composition with trivial systems is trivial, and, therefore, one has $\mathrm{AI} = \mathrm{A}, \mathrm{BI} = \mathrm{B}$*).*

If a causal theory is not deterministic (i.e. if the possible values of probabilities are not only 0 and 1) then randomization and coarse-graining always allows one to construct an internal state (see Def. 24): it is enough to take a spanning set of states $\{\rho_i\}_{i\in \mathrm{X}}$, to randomize them with some non-zero probabilities $\{p_i\}_{i\in \mathrm{X}}$, and then to coarse-grain, thus getting the internal state $\omega = \sum_{i\in \mathrm{X}} p_i\rho_i$.

Finally, conditioning allows one to prove that a causal theory contains all possible *measure-and-prepare* channels, defined as follows

Definition 31 (Measure-and-prepare channels) *A channel* $\mathscr{C} \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ *is* measure-and-prepare *if there exists an observation-test* $\{a_i\}_{i\in \mathrm{X}}$ *on* A, *and a collection of normalized states* $\{\beta_i\}_{i\in \mathrm{X}} \subset \mathfrak{S}_1(\mathrm{B})$ *such that*

$$\mathscr{C} = \sum_{i\in \mathrm{X}} |\beta_i)_\mathrm{B}\, (a_i|_\mathrm{A}\,. \tag{47}$$

C. Distance between transformations

Here we introduce a norm for transformations that has a direct operational interpretation: it quantifies the maximum probability of success in the discrimination of two channels in a causal theory. Suppose that we are given two channels $\mathscr{C}_0, \mathscr{C}_1 \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ with prior probabilities π_0, π_1, respectively. In a causal theory, the most general way to discriminate is to prepare a bipartite input state $\rho \in \mathfrak{S}_1(\mathrm{AC})$, to apply the unknown channel, and to perform a binary test that distinguishes between the two possible output states $\mathscr{C}_0\,|\rho)_\mathrm{AC}$ and $\mathscr{C}_1\,|\rho)_\mathrm{AC}$. Optimizing over all binary tests and using Eq. (29) we obtain the success probability $p_{succ} = 1/2(1 + \|(\pi_1\mathscr{C}_1 - \pi_0\mathscr{C}_0)\rho\|_\mathrm{BC})$ Moreover, optimizing the input state and the extension we find the maximum probability of success

$$p^{opt}_{succ} = \frac{1}{2}(1 + \|\pi_1\mathscr{C}_1 - \pi_0\mathscr{C}_0\|_{\mathrm{A},\mathrm{B}}) \tag{48}$$

where the operational norm for transformations is defined by

$$\|\Delta\|_{\mathrm{A},\mathrm{B}} = \sup_{\mathrm{C}} \sup_{\rho\in\mathfrak{S}_1(\mathrm{AC})} \|\Delta\rho\|_\mathrm{BC} \qquad \Delta \in \mathfrak{T}_\mathbb{R}(\mathrm{A},\mathrm{B}). \tag{49}$$

In quantum theory our expression for the operational norm reduces to the diamond norm in Schrödinger picture [34], or equivalently, to the completely bounded (CB) norm in Heisenberg picture [35].

In the case of trivial input system A = I, Eq. (49) gives back the norm of states introduced in Eq. (30). In the case of trivial output system B = I, it provides an operational norm for effects, given by

$$\|\delta\|_{\mathrm{A},\mathrm{I}} = \sup_{\mathrm{C}} \sup_{\rho\in\mathfrak{S}_1(\mathrm{AC})} \|\delta\rho\|_\mathrm{C} \qquad \delta \in \mathfrak{E}_\mathbb{R}(\mathrm{A}). \tag{50}$$

In fact, the extension with the ancillary system C is not needed in this case:

Lemma 8 *The operational norm of an element of the vector space* $\delta \in \mathfrak{E}_\mathbb{R}(\mathrm{A})$ *spanned by the effects for system* A *is given by the expression*

$$\|\delta\|_{\mathrm{A},\mathrm{I}} = \sup_{\rho\in\mathfrak{S}_1(\mathrm{A})} |\,(\delta|\,\rho)_\mathrm{A}|. \tag{51}$$

Proof. Taking C = I in Eq. (50) yields the lower bound $\|\delta\|_{\mathrm{A},\mathrm{I}} \geq \sup_{\rho\in\mathfrak{S}_1(\mathrm{A})} \|\,(\delta|\,\rho)_\mathrm{A}\|_\mathrm{I} = \sup_{\rho\in\mathfrak{S}_1(\mathrm{A})} |\,(\delta|\,\rho)_\mathrm{A}|$, where we used the fact that the norm of a real number $x \in \mathbb{R} \equiv \mathfrak{S}_\mathbb{R}(\mathrm{I})$ is given by its modulus: $\|x\|_\mathrm{I} = |x|$. To prove the equality of Eq. (51) we now prove converse bound. By the definition of the operational norm for states in Eq. (30), for every $\sigma \in \mathfrak{S}_1(\mathrm{AC})$ we have

$$\begin{aligned}\|\delta\sigma\|_\mathrm{C} &= \sup_{c_1\in\mathfrak{E}(\mathrm{A})} (\delta|_\mathrm{A}\,(c_1|_\mathrm{C}\,|\sigma)_\mathrm{AC} - \inf_{c_0\in\mathfrak{E}(\mathrm{A})} (\delta|_\mathrm{A}\,(c_0|_\mathrm{C}\,|\sigma)_\mathrm{AC} \\ &= \sup_{\{c_0,c_1\}} (\delta|_\mathrm{A}\,(c_1 - c_0|_\mathrm{C}\,|\sigma)_\mathrm{AC},\end{aligned} \tag{52}$$

where the optimization in the last equation is over all possible binary tests $\{c_0, c_1\}$ for system C. Now, applying the observation-test $\{c_0, c_1\}$ to the bipartite state $|\sigma)_\mathrm{AC}$ we obtain a preparation-test $\{\rho_0, \rho_1\}$ for system A, defined by $|\rho_i)_\mathrm{A} = (c_i|_\mathrm{C}\,|\sigma)_\mathrm{AC}, i = 0, 1$. Defining the probabilities $p_i = (e|\,\rho_i)_\mathrm{A}$ and the normalized states $\bar\rho_i = \rho_i/\,(e|\,\rho_i)_\mathrm{A}$ we then have

$$\begin{aligned}(\delta|_\mathrm{A}\,(c_1 - c_0|_\mathrm{C}\,|\sigma)_\mathrm{AC} &= p_1\,(\delta|\,\bar\rho_1)_\mathrm{A} - p_0\,(\delta|\,\bar\rho_0)_\mathrm{A} \\ &\leq \max\{(\delta|\,\bar\rho_1)_\mathrm{A}, -(\delta|\,\bar\rho_0)_\mathrm{A}\} \\ &\leq \sup_{\rho\in\mathfrak{S}_1(\mathrm{A})} |\,(\delta|\,\rho)_\mathrm{A}|.\end{aligned} \tag{53}$$

■

In quantum theory the norm $\|D\|_{\mathrm{A,I}}$ of an hermitian operator on the Hilbert space of system A coincides with the operator norm $\|D\|_\infty = \sup_{\rho\geq 0,\mathrm{Tr}[\rho]=1}|\mathrm{Tr}[D\rho]| = \max_i\{|d_i|\}$, where $\{d_i\}$ are the eigenvalues of D.

We conclude by mentioning a monotonicity property of the operational norm of transformations:

Lemma 9 (Monotonicity of the operational norm for transformations) *If $\mathscr{C}\in\mathfrak{T}(\mathrm{A,B})$ and $\mathscr{E}\in\mathfrak{T}(\mathrm{C,D})$ are two channels, then for every $\Delta\in\mathfrak{T}_{\mathbb{R}}(\mathrm{B,C})$ one has*

$$\|\mathscr{E}\Delta\mathscr{C}\|_{\mathrm{A,D}} \leq \|\Delta\|_{\mathrm{B,C}}. \tag{54}$$

If $\mathscr{C}$ and $\mathscr{E}$ are reversible one has the equality.

Proof. Let R be an ancillary system, and $\rho\in\mathfrak{S}_1(\mathrm{AR})$ be a normalized state of AR. Then, since $|\sigma)_{\mathrm{BR}} = \mathscr{C}|\rho)_{\mathrm{AR}}$ is a normalized state of BR, we have $\|\mathscr{E}\Delta\mathscr{C}\|_{\mathrm{A,D}} = \sup_{\mathrm{R}}\sup_{\rho\in\mathfrak{S}_1(\mathrm{AR})}\|\mathscr{E}\Delta\mathscr{C}\rho\|_{\mathrm{DR}} \leq \sup_{\mathrm{R}}\sup_{\sigma\in\mathfrak{S}_1(\mathrm{BR})}\|\mathscr{E}\Delta\sigma\|_{\mathrm{DR}}$. Now, using Lemma 1 we obtain $\|\mathscr{E}\Delta\sigma\|_{\mathrm{DR}} \leq \|\Delta\sigma\|_{\mathrm{CR}}$. Hence, $\|\mathscr{E}\Delta\mathscr{C}\|_{\mathrm{A,D}} \leq \sup_{\mathrm{R}}\sup_{\sigma\in\mathfrak{S}_1(\mathrm{BR})}\|\Delta\sigma\|_{\mathrm{CR}} = \|\Delta\|_{\mathrm{B,C}}$. Clearly, if $\mathscr{C}$ and $\mathscr{E}$ are reversible, one has the converse bound $\|\Delta\|_{\mathrm{B,C}} = \|\mathscr{E}^{-1}(\mathscr{E}\Delta\mathscr{C})\mathscr{C}^{-1}\|_{\mathrm{B,C}} \leq \|\mathscr{E}\Delta\mathscr{C}\|_{\mathrm{A,D}}$, thus proving the equality. ■

D. Closure and convexity in causal theories

In Subsect. II J we saw that if a theory is not deterministic, then one can construct a circuit that simulates (with arbitrary precision) a coin with arbitrary bias $p\in[0,1]$.

In causal theories the possibility of conditioning gives directly the following:

Lemma 10 (Approximation of convex combinations) *If a causal theory is not deterministic, then any convex combination of states, effects, and transformations can be approximated with arbitrary precision.*

Proof. Let $p\in[0,1]$ be an arbitrary probability and $p_n\in\mathfrak{S}(\mathrm{I})$ be such that $|p-p_n|<1/n$ (such a probability exists because $\mathfrak{S}(\mathrm{I})$ is dense in the interval $[0,1]$, as stated by Lemma 3). Consider two arbitrary tests $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ and $\{\mathscr{D}_j\}_{j\in\mathrm{Y}}$ from A to B. By randomization, we get the test $\{p_n\mathscr{C}_i\}_{i\in\mathrm{X}}\cup\{(1-p_n)\mathscr{D}_j\}_{j\in\mathrm{Y}}$. Then, by coarse-graining we can obtain the convex combination $p_n\mathscr{C}_i+(1-p_n)\mathscr{D}_j$. The distance with the desired convex combination $p\mathscr{C}_i+(1-p)\mathscr{D}_j$ is bounded by $(\|\mathscr{C}_i\|_{\mathrm{A,B}}+\|\mathscr{D}_j\|_{\mathrm{A,B}})/n < 2/n$. ■

As a simple consequence we have the following

Corollary 5 (Closure implies convexity) *If a causal theory is not deterministic and the set of states of the trivial system is closed, then all sets of states, effects, and transformations are convex.*

In this paper for simplicity we will always work with closed sets of states. Our attention will be devoted to non-deterministic causal theories, and, therefore, by the previous Corollary 5 closure implies convexity. Note that, however, most results hold independently of the assumption of convexity, since in the context of non-deterministic causal theories any desired combination can be approximated with arbitrary precision.

E. No-restriction hypothesis in causal theories

In a causal theory the no-restriction hypothesis of Def. 16 implies that for every system A the cone generated by the effects coincides with the dual of the cone generated by the states:

Lemma 11 *In a causal theory the no-restriction hypothesis of Def. 16 implies the condition $\mathfrak{E}_+(\mathrm{A})=\mathfrak{S}^*_+(\mathrm{A})$ for every system* A.

Proof. Suppose that a is an element of $\mathfrak{S}^*_+(\mathrm{A})$ and let $\|a\|_{\mathrm{A,I}}$ be the operational norm of a, as defined in Eq. (51). If $\|a\|_{\mathrm{A,I}}=0$, then a is the null effect, which is trivially an element of $\mathfrak{E}_+(\mathrm{A})$. If $\|a\|_{\mathrm{A,I}}\neq 0$, then define the normalized effect $a_0=a/\|a\|_{\mathrm{A,I}}$. Upon defining $a_1=e-a_0$, we now have $(a_1|\rho)\geq 0$ for all $\rho\in\mathfrak{S}_+(\mathrm{A})$, i.e. $a_1\in\mathfrak{S}^*_+(\mathrm{A})$. Moreover, $(a_0|\rho)_{\mathrm{A}}+(a_1|\rho)_{\mathrm{A}}=(e|\rho)_{\mathrm{A}}=1$ for every normalized state $\rho\in\mathfrak{S}_1(\mathrm{A})$. Hence, $\{a_0,a_1\}$ is a probability rule. By the no-restriction hypothesis, we then have that $\{a_0,a_1\}$ is an observation-test, and, therefore, a_0 and a_1 are effects. This proves that every $a\in\mathfrak{S}^*_+(\mathrm{A})$ is proportional to an effect a_0, that is, $\mathfrak{S}^*_+(\mathrm{A})\subseteq\mathfrak{E}_+(\mathrm{A})$. On the other hand, all effects are positive functionals on states, and, therefore $\mathfrak{S}^*_+(\mathrm{A})\supseteq\mathfrak{E}_+(\mathrm{A})$. ■

The above condition will be useful when discussing the implications of the no-restriction hypothesis in subsections VII D and X C.

IV. LOCAL DISCRIMINABILITY

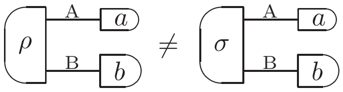
Here we discuss the property of local discriminability, which expresses the possibility of distinguishing multipartite states using only local devices.

A. Definition and main properties

A common assumption in the literature on probabilistic theories is what we will call here *local discriminability* (see e.g. Refs. [7–11, 17, 18]).

Definition 32 (Local discriminability) *A theory enjoys local discriminability if whenever two states $\rho,\sigma\in$*

$\mathfrak{S}(\mathrm{AB})$ *are distinct, there are two local effects* $a \in \mathfrak{E}(\mathrm{A})$ *and* $b \in \mathfrak{E}(\mathrm{B})$ *such that*



(55)

Note that local discriminability on bipartite states implies local discriminability on multipartite states, as can be seen by simple iteration.

The meaning of the local discriminability condition is that if two bipartite states are different, then there is a chance of distinguishing between them by using only local devices. Of course, the resulting discrimination may not be optimal, but at least it is strictly better than the random guess. Indeed, in the next Lemma we show that in a convex theory with local discriminability two parties Alice and Bob, holding systems A and B, respectively, can always find a discrimination protocol that uses only local operations and classical communication (LOCC) and outperforms the random guess.

Lemma 12 (LOCC discrimination) *In a convex theory with local discriminability, if two states* $\rho_0, \rho_1 \in \mathfrak{S}_1(\mathrm{AB})$ *are distinct, then there exists a LOCC discrimination protocol, described by a binary test* $\{A_0, A_1\}$*, such that the probability* $p_{wc} := \max\{p(0|1), p(1|0)\}$, $p(i|j) = (A_i|\rho_j)_{\mathrm{AB}}$ *is strictly smaller than* $1/2$.

Proof. If $\rho \neq \sigma$, then by local discriminability there are always two effects a, b such that $(a \otimes b|\rho)_{\mathrm{AB}} > (a \otimes b|\sigma)_{\mathrm{AB}}$. The binary test $\{A, e_{\mathrm{AB}} - A\}$ defined by $A := a \otimes b$ can be obtained by performing the local tests $\{a, e_{\mathrm{A}} - a\}$ and $\{b, e_{\mathrm{B}} - b\}$ and taking a coarse-graining. If the theory is convex, exploiting the construction of Lemma 2 (which only requires randomization and coarse-graining) we obtain a binary test $\{A_0, A_1\}$ satisfying $p(0|1) = p(1|0) < 1/2$ and, therefore $p_{wc} < 1/2$. ■

Local discriminability is an enormous advantage in experiments. For example it allows one to perform tomography of multipartite states with only local measurements. Indeed, every bipartite effect $(E|_{\mathrm{AB}}$ can be written as linear combination of product effects, and, therefore every probability $(E|\rho)_{\mathrm{AB}}$ can be computed as a linear combination of the probabilities $(a_i \otimes b_j|\rho)_{\mathrm{AB}}$ arising from a finite set of product effects:

Lemma 13 (Local tomography) *Let* $\{\rho_i\}$ *and* $\{\tilde{\rho}_j\}$ *be two bases for the vector spaces* $\mathfrak{S}_{\mathbb{R}}(\mathrm{A})$ *and* $\mathfrak{S}_{\mathbb{R}}(\mathrm{B})$*, respectively, and let* $\{a_i\}$ *and* $\{b_j\}$ *be two bases for the vector spaces* $\mathfrak{E}_{\mathbb{R}}(\mathrm{A})$ *and* $\mathfrak{E}_{\mathbb{R}}(\mathrm{B})$*, respectively. A theory enjoys local discriminability if and only if every state* $\sigma \in \mathfrak{S}(\mathrm{AB})$ *(every effect* $E \in \mathfrak{E}(\mathrm{AB})$*) can be written as*

$$|\sigma)_{\mathrm{AB}} = \sum_{i,j} A_{ij}\, |\rho_i)_{\mathrm{A}}\, |\tilde{\rho}_j)_{\mathrm{B}}$$
$$\left((E|_{\mathrm{AB}} = \sum_{i,j} B_{ij}\, (a_i|_{\mathrm{A}}\, (b_j|_{\mathrm{B}} \right) \tag{56}$$

for some suitable real matrix A_{ij} *(*B_{ij}*).*

Proof. Suppose that local discriminability holds. By definition, the product effects $a \otimes b$ are a separating set for $\mathfrak{S}_{\mathbb{R}}(\mathrm{AB})$, and, therefore, they are a spanning set for $\mathfrak{E}_{\mathbb{R}}(\mathrm{AB})$. Since states and effects span vector spaces of equal dimension, this also implies that the product states are a spanning set for $\mathfrak{S}_{\mathbb{R}}(\mathrm{AB})$. Conversely, if Eq. (56) holds, then the product effects are a spanning set for the vector space $\mathfrak{E}_{\mathbb{R}}(\mathrm{AB})$. Clearly, if $(a \otimes b|\rho)_{\mathrm{AB}} = (a \otimes b|\sigma)_{\mathrm{AB}}$ for all product effects, then one has $\rho = \sigma$, and this proves local discriminability. ■

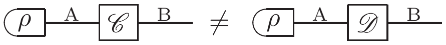
This also implies:

Theorem 2 (Product of internal states is internal) *In a causal theory with local discriminability if the states* ω_{A} *and* ω_{B} *are internal in* $\mathfrak{S}(\mathrm{A})$ *and* $\mathfrak{S}(\mathrm{B})$*, respectively, then the product* $\omega_{\mathrm{A}} \otimes \omega_{\mathrm{B}}$ *is internal in* $\mathfrak{S}(\mathrm{AB})$.

Proof. By definition, one has $\mathsf{Span}(D_{\omega_{\mathrm{A}} \otimes \omega_{\mathrm{B}}}) \supset \mathsf{Span}(D_{\omega_{\mathrm{A}}}) \otimes \mathsf{Span}(D_{\omega_{\mathrm{B}}}) = \mathfrak{S}_{\mathbb{R}}(\mathrm{A}) \otimes \mathfrak{S}_{\mathbb{R}}(\mathrm{B})$. Since local discriminability holds, this is also equal to $\mathfrak{S}_{\mathbb{R}}(\mathrm{AB})$. ■

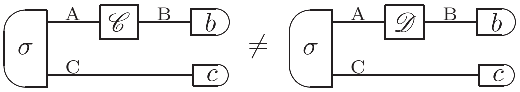
Moreover, local discriminability allows one to distinguish two different transformations $\mathscr{C}, \mathscr{D} \in \mathfrak{T}(\mathrm{A}, \mathrm{B})$ without considering their extension with an arbitrary ancilla system C:

Lemma 14 *If two transformations* $\mathscr{C}, \mathscr{D} \in \mathfrak{T}(\mathrm{A}, \mathrm{B})$ *are different and local discriminability holds, then there exist a state* $\rho \in \mathfrak{S}(\mathrm{A})$ *such that*



(57)

Proof. By definition, if $\mathscr{C}$ and $\mathscr{D}$ are different there exist a system C and a joint state $\sigma \in \mathfrak{S}(\mathrm{AC})$ such that $\mathscr{C}|\sigma)_{\mathrm{AC}} \neq \mathscr{D}|\sigma)_{\mathrm{AC}}$. Now, since local discriminability holds, there are two effects b, c on systems B, C, respectively such that



(58)

Defining $|\rho) := (c|_{\mathrm{C}}|\sigma)_{\mathrm{AC}}$ we then obtain $(b|_{\mathrm{B}}\mathscr{C}|\rho)_{\mathrm{A}} \neq (b|_{\mathrm{B}}\mathscr{D}|\rho)_{\mathrm{A}}$. This implies $\mathscr{C}|\rho)_{\mathrm{A}} \neq \mathscr{D}|\rho)_{\mathrm{A}}$. ■

B. Causal theories with local discriminability

The results of this paper can be formulated in the simplest way for causal theories that enjoy local discriminability. In this case one has the following useful properties:

Lemma 15 *Let* $|\sigma)_{\mathrm{AB}}$ *be a state of* AB *and* $|\rho)_{\mathrm{A}} := (e|_{\mathrm{B}} |\sigma)_{\mathrm{AB}}$, $|\tilde{\rho})_{\mathrm{B}} := (e|_{\mathrm{A}} |\sigma)_{\mathrm{AB}}$ *be its marginals on systems* A, B, *respectively. In a causal theory with local discriminability one has*

$$\sigma \in \mathsf{Span}(D_{\rho\otimes\tilde{\rho}}), \tag{59}$$

where $D_{\rho\otimes\tilde{\rho}}$ *is the refinement set of* $\rho\otimes\tilde{\rho}$, *as defined in Def. 21.*

Proof. Take a basis $\{\rho_i\}_{i=1}^{n}$ ($\{\tilde{\rho}_j\}_{j=1}^{\tilde{n}}$) of states for the (span of) the refinement set of ρ ($\tilde{\rho}$), and extend it to a basis $\{\rho_i\}_{i=1}^{D_{\mathrm{A}}}$ ($\{\tilde{\rho}_j\}_{j=1}^{D_{\mathrm{B}}}$) of $\mathfrak{S}_{\mathbb{R}}(\mathrm{A})$ (of $\mathfrak{S}_{\mathbb{R}}(\mathrm{B})$). By local discriminability, we can write σ as a linear combination as in Eq. (56) for some coefficients A_{ij}. Now, for every effect $(a|_{\mathrm{A}}$ the state $|\tilde{\rho}_a)_{\mathrm{B}} := (a|_{\mathrm{A}} |\sigma)_{\mathrm{AB}}$ is clearly in $D_{\tilde{\rho}}$. Therefore, we must have $A_{ij} = 0$ for all $j > \tilde{n}$. Likewise, applying an arbitrary effect $(b|_{\mathrm{B}}$ on system B we find that we must have $A_{ij} = 0$ for all $i > n$. This implies

$$|\sigma)_{\mathrm{AB}} = \sum_{i=1}^{n}\sum_{j=1}^{\tilde{n}} A_{ij}\, |\rho)_i\, |\tilde{\rho})_j, \tag{60}$$

that is, $\sigma \in \mathsf{Span}(D_{\rho\otimes\tilde{\rho}})$. ■

Since in a non-deterministic causal theory the set of states $\mathfrak{S}(\mathrm{A})$ is convex (Corollary 5 along with the assumption that $\mathfrak{S}(\mathrm{I})$ is closed), we also have the following:

Theorem 3 *Let* $|\sigma)_{\mathrm{AB}}$ *be a state of* AB *and* $|\rho)_{\mathrm{A}} := (e|_{\mathrm{B}} |\sigma)_{\mathrm{AB}}$, $|\tilde{\rho})_{\mathrm{B}} := (e|_{\mathrm{A}} |\sigma)_{\mathrm{AB}}$ *be its marginals on systems* A, B, *respectively. In a non-determinisitc causal theory with local discriminability there exists a non-zero probability* $k > 0$ *such that*

$$k\sigma \in D_{\rho\otimes\tilde{\rho}}. \tag{61}$$

The proof of the Theorem is immediate using Lemma 15 along with the following

Lemma 16 *In a non-deterministic causal theory, for every couple of states* $\sigma, \rho \in \mathfrak{S}_1(\mathrm{A})$ *one has*

$$\sigma \in \mathsf{Span}(D_\rho) \Longrightarrow k\sigma \in D_\rho, \tag{62}$$

for some non-zero probability $k > 0$.

Proof. Take a basis $\{\rho_i\}_{i=1}^{n}$ of states in D_ρ. By hypothesis, we can write $\sigma = \sum_i s_i\rho_i$ with suitable real coefficients s_i. Moreover, since we are in finite dimensions, there is surely a maximum coefficient $s_{\max} = \max_i s_i$. On the other hand, since ρ_i belongs to D_ρ, there is surely a state χ_i such that $\rho = \rho_i + \chi_i$. This implies

$$\rho = \frac{1}{n}\sum_i(\rho_i + \chi_i). \tag{63}$$

Let us define $\tau := \rho - k\sigma$, with $k = \frac{1}{2ns_{\max}}$, and normalize it as $\bar{\tau} := \tau/\,(e|\tau)_{\mathrm{A}}$. Using Eq. (63) it is easy to verify that $\bar{\tau}$ is a state, since it is a convex combination of states (recall that in a non-deterministic causal theory the set of states is convex). Moreover, we have $\rho = k\sigma + (1-k)\bar{\tau}$, which implies the thesis. ■

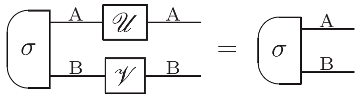
Remark. In the previous Lemma 16 we used the fact that in a non-deterministic causal theory a set of states is convex (Corollary 5 along with the assumption that $\mathfrak{S}(\mathrm{I})$ is closed). In fact, we can weaken this assumption in the proofs of Theorem 2 and Lemma 16. Indeed, in any non-deterministic causal theory we can approximate the convex combinations needed for the proof of Lemma 16 with arbitrary precision (Lemma 10), thus proving Eqs. (61) and (62) with a non-zero probability $k > 0$ that arises from a test allowed by the theory.

Theorems 2 and 3 state two very natural properties. Even when discussing the extension of our results beyond the framework of local discriminability we will assume these properties to hold.

Finally, causal theories with local discriminability enjoy a nice characterization of states that are invariant under the group of reversible transformations:

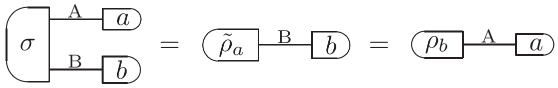
Theorem 4 *In a causal theory with local discriminability if systems* A *and* B *have unique invariant states* $|\chi)_{\mathrm{A}} \in \mathfrak{S}_1(\mathrm{A})$ *and* $|\chi)_{\mathrm{B}} \in \mathfrak{S}_1(\mathrm{B})$, *respectively, then* $|\chi)_{\mathrm{A}}\, |\chi)_{\mathrm{B}} \in \mathfrak{S}_1(\mathrm{AB})$ *is the unique locally invariant state of system* AB.

Proof. Suppose that $|\sigma)_{\mathrm{AB}}$ is a locally invariant state, namely



(64)

for all $\mathscr{U} \in \mathbf{G}_{\mathrm{A}}$ and $\mathscr{V} \in \mathbf{G}_{\mathrm{B}}$. If we apply two arbitrary effects $(a|_{\mathrm{A}}$ and $(b|_{\mathrm{B}}$ we then get

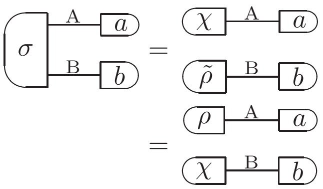


(65)

having defined $|\tilde{\rho}_a)_{\mathrm{B}} := (a|_{\mathrm{A}} |\sigma)_{\mathrm{AB}}$ and $|\rho_b)_{\mathrm{A}} := (b|_{\mathrm{B}} |\sigma)_{\mathrm{AB}}$. Now, $\tilde{\rho}_a$ and ρ_b are invariant (unnormalized) states. Since χ_{A} is the unique state of B that is invariant and normalized, one must have

$$\begin{aligned} |\chi)_{\mathrm{A}} &= \frac{|\rho_b)_{\mathrm{A}}}{(e|\rho_b)_{\mathrm{A}}} = \frac{|\rho_b)_{\mathrm{A}}}{(e\otimes b|\sigma)_{\mathrm{AB}}} := \frac{|\rho_b)_{\mathrm{A}}}{(b|\tilde{\rho})_{\mathrm{B}}} \\ |\chi)_{\mathrm{B}} &= \frac{|\tilde{\rho}_a)_{\mathrm{B}}}{(e|\tilde{\rho}_a)_{\mathrm{B}}} = \frac{|\tilde{\rho}_a)_{\mathrm{B}}}{(a\otimes e|\sigma)_{\mathrm{AB}}} := \frac{|\tilde{\rho}_a)_{\mathrm{B}}}{(a|\rho)_{\mathrm{A}}}, \end{aligned} \tag{66}$$

$|\rho)_{\mathrm{A}}$, $|\tilde{\rho})_{\mathrm{B}}$ being the marginal states on systems A, B, respectively. Inserting the above relations in Eq. (65), we then obtain



(67)

for every a, b. By local discriminability, this implies $|\sigma)_{\mathrm{AB}} = |\chi)_{\mathrm{A}}\, |\tilde{\rho})_{\mathrm{B}} = |\rho)_{\mathrm{A}}\, |\chi)_{\mathrm{B}}$, and, therefore, $|\sigma)_{\mathrm{AB}} = |\chi)_{\mathrm{A}}\, |\chi)_{\mathrm{B}}$. ■

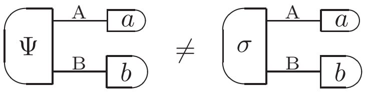
V. BEYOND LOCAL DISCRIMINABILITY AND CONVEXITY

Although the results of this paper take their simplest form for causal theories with local discriminability, most of them are valid in causal theories under weaker requirements. For example, they hold for quantum theory on real Hilbert spaces, which is a well known example of theory without local discriminability. Moreover, although convexity is very well motivated in the context of causal theories, most results of this paper hold even in non-convex theories. In this Section we briefly discuss these generalizations.

A. Relaxing local discriminability

A weaker requirement than local discriminability is local discriminability on pure states:

Definition 33 (Local discriminability on pure states) *A theory enjoys* local discriminability on pure states *if whenever two states $\Psi, \sigma \in \mathfrak{S}(\mathrm{AB})$ are different, and one of the two states (say Ψ) is pure, there are two effects $a \in \mathfrak{E}(\mathrm{A})$ and $b \in \mathfrak{E}(\mathrm{B})$ such that*



(68)

An example of theory with this property is quantum theory on real Hilbert spaces:

Lemma 17 *Quantum theory on real Hilbert spaces enjoys local discriminability on pure states.*

Proof. Let $\rho = \sum_i p_i |\Phi_i\rangle\langle\Phi_i|$ be a density matrix on the real Hilbert space $\mathscr{H}_{\mathrm{A}} \otimes \mathscr{H}_{\mathrm{B}}$ with $\mathscr{H}_{\mathrm{A}} = \mathbb{R}^m$ and $\mathscr{H}_{\mathrm{B}} = \mathbb{R}^n$ and $|\Psi\rangle \in \mathbb{R}^m \otimes \mathbb{R}^n$ be a unit vector. Suppose that $\mathrm{Tr}[(\rho - |\Psi\rangle\langle\Psi|)(a \otimes b)] = 0$ for every couple of real matrices a and b. Taking $a = |v\rangle\langle v|$ for some $v \in \mathbb{R}^m$ we then obtain $\langle v|_{\mathrm{A}}|\Phi_i\rangle_{\mathrm{AB}} = k_{i,v}\langle v|_{\mathrm{A}}|\Psi\rangle_{\mathrm{AB}}$ for some constant $k_{i,v}$. Likewise, taking $b = |w\rangle\langle w|$ for some $w \in \mathbb{R}^n$ we obtain $\langle w|_{\mathrm{B}}|\Phi_i\rangle_{\mathrm{AB}} = l_{i,w}\langle w|_{\mathrm{B}}|\Psi\rangle_{\mathrm{AB}}$ for some constant $l_{i,w}$. Putting the two things together we have

$$\begin{aligned}\langle v|_{\mathrm{A}}\langle w|_{\mathrm{B}}|\Phi_i\rangle_{\mathrm{AB}} &= k_{i,v}\langle v|_{\mathrm{A}}\langle w|_{\mathrm{B}}|\Psi\rangle_{\mathrm{AB}}\\ &= l_{i,w}\langle v|_{\mathrm{A}}\langle w|_{\mathrm{B}}|\Psi\rangle_{\mathrm{AB}}\end{aligned} \tag{69}$$

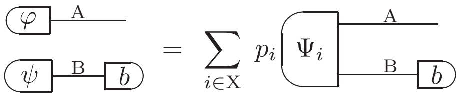
hence $k_{i,v} \equiv l_{i,w} := c_i$. Finally, $\langle v|_{\mathrm{A}}\langle w|_{\mathrm{B}}|\Phi_i\rangle_{\mathrm{AB}} = c_i\langle v|_{\mathrm{A}}\langle w|_{\mathrm{B}}|\Psi\rangle_{\mathrm{AB}}$ for every v, w implies $|\Phi_i\rangle = c_i|\Psi\rangle$, and, therefore $\sigma = |\Psi\rangle\langle\Psi|$. ■

When generalizing our results to theories without local discriminability we will always assume local discriminability on pure states along with the theses of Theorems 2, 3, and 4. Again, all these requirements are met by quantum theory on real Hilbert spaces.

An elementary property of causal theories with local discriminability on pure states is that the product of two pure states is pure, as stated in the following Lemma.

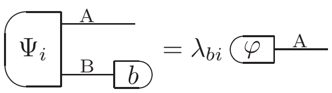
Lemma 18 (Product of pure states is pure) *In a causal theory with local discriminability on pure states, if the states $|\varphi)_{\mathrm{A}} \in \mathfrak{S}_1(\mathrm{A})$ and $|\psi)_{\mathrm{B}} \in \mathfrak{S}_1(\mathrm{B})$ are pure, then their product $|\varphi)_{\mathrm{A}}\, |\psi)_{\mathrm{B}} \in \mathfrak{S}_1(\mathrm{AB})$ is pure.*

Proof. Suppose that the product can be written as a convex combination $|\varphi)_{\mathrm{A}}\, |\psi)_{\mathrm{B}} = \sum_{i\in\mathrm{X}} p_i\, |\Psi_i)_{\mathrm{AB}}$, with $|\Psi_i)_{\mathrm{AB}} \in \mathfrak{S}_1(\mathrm{AB})$. We now show that $|\Psi_i)_{\mathrm{AB}} = |\varphi)_{\mathrm{A}}\, |\psi)_{\mathrm{B}}$ for every $i \in \mathrm{X}$. Let $(b|_{\mathrm{B}}$ be an arbitrary effect for system B. We then have



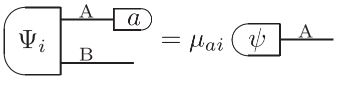
(70)

Since $|\varphi)_{\mathrm{A}}$ is pure, this implies



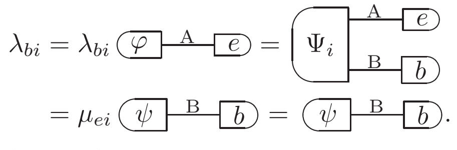
(71)

for some coefficient $\lambda_{bi} \geq 0$. Clearly, for $(b|_{\mathrm{B}} = (e|_{\mathrm{B}}$ one has $\lambda_{ei} = 1$. Similarly, if $(a|_{\mathrm{A}}$ is an arbitrary effect for system A, we obtain



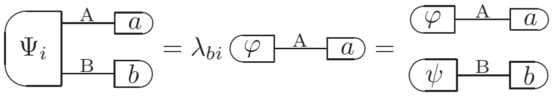
(72)

for some coefficient $\mu_{ai} \geq 0$ satisfying $\mu_{ei} = 1$. Combining the above facts, we obtain



(73)

Finally, this implies



(74)

and, by local discriminability on pure states $|\Psi_i)_{\mathrm{AB}} = |\varphi)_{\mathrm{A}}\, |\psi)_{\mathrm{B}}$. ■

Clearly, iterating the above reasoning one can also show that the product of N pure states $|\varphi_1)_{\mathrm{A}_1}\, |\varphi_2)_{\mathrm{A}_2} \dots |\varphi_N)_{\mathrm{A}_N}$ is pure.

B. Relaxing convexity

If one wants to relax convexity, it is clear from the proof of Lemma 10 and Corollary 5 that one must have at least one of the following features: *i)* the theory is deterministic, i.e. all events have either zero or unit probability, *ii)* some randomizations or some coarse-grainings are forbidden, and *iii)* the set of probabilities $\mathfrak{S}(\mathrm{I})$ of the theory is not closed. For the purposes of this paper, deterministic theories are not quite interesting, and theories with non-closed sets of transformations are just technically cumbersome, although most of the conclusions of this paper remain unchanged. Therefore, in relaxing convexity we will only consider the case in which some conditioned tests or some coarse-grained tests are forbidden. Of course, if one wants to drop a basic operational requirement like the possibility of conditioning, one has to take care that some minimal properties hold. For example, the existence of internal states, the fact that every test has an ultimate refinement, and the validity of the theses of Theorems 2 and 3 have to be explicitly postulated. One would also need to assume that is not forbidden *i)* to attach a distinguishable state $|\varphi_i)_{\mathrm{B}}$ to every state in a preparation-test $\{|\rho_i)_{\mathrm{A}}\}_{i\in\mathrm{X}}$, thus getting the new test $\{|\rho_i)_{\mathrm{A}}\,|\varphi_i)_{\mathrm{B}}\}_{i\in\mathrm{X}}$, and *ii)* to perform a discriminating test $\{a_i\}_{i\in\mathrm{X}}$ for the perfectly discriminable states $\{\rho_i\}_{i\in\mathrm{X}}$, and to re-prepare state ρ_i when the outcome is i, thus getting the "measure-and-prepare" test $\{|\rho_i)_{\mathrm{A}}\,(a_i|_{\mathrm{A}}\}_{i\in\mathrm{X}}$.

Finally, we will show that the existence of twirling tests is necessary for deterministic teleportation. If one wants to consider non-convex theories with deterministic teleportation one has also to require the existence of a twirling-test and the thesis of Theorem 4.

VI. SUMMARY OF THE FRAMEWORK

This short Section concludes the presentation of the general framework used in this paper. The standing assumptions of the paper are summarized by the following table:

In this paper, if not otherwise stated, we will consider operational-probabilistic theories satisfying the following requirements:

1. **the theory is causal (every state is proportional to a normalized one)**
2. **local discriminability holds**
3. **the set of all tests is closed under coarse-graining and conditioning**
4. **for every system, the set of states is finite-dimensional and closed in the operational norm**
5. **there exist perfectly discriminable states**
6. **the theory is not deterministic**

Note that the existence of perfectly discriminable states, needed to describe perfect classical communication, is guaranteed in the usual convex framework, which contains the no-restriction hypothesis of Def. 16. We recall that we don't make this assumption here.

In most proofs the background requirement 2. can be always weakened to:

2'. **local discriminability of pure states and the theses of Theorems 2, 4 and 3 hold**

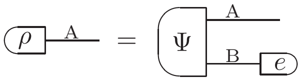
If a particular results requires local discriminability or convexity this will be mentioned explicitly in its statement.

VII. THEORIES WITH PURIFICATION

Here we introduce the purification postulate "every mixed state has a purification, unique up to reversible transformations on the purifying system", and we explore its consequences within the general framework outlined in the previous Sections.

A. The purification postulate

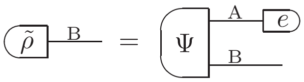
Definition 34 (Purification) *A pure state* $\Psi \in \mathfrak{S}_1(\mathrm{AB})$ *is a* purification *of* $\rho \in \mathfrak{S}_1(\mathrm{A})$ *if* $|\rho)_{\mathrm{A}} = (e|_{\mathrm{B}}\,|\Psi)_{\mathrm{AB}}$. *Diagrammatically,*



(75)

Definition 35 (Purifying system) *If system* AB *contains a purification of* $\rho \in \mathfrak{S}_1(\mathrm{A})$*, we call system* B *a* purifying system *for* ρ.

Definition 36 (Complementary state) *Let* $\Psi \in \mathfrak{S}_1(\mathrm{AB})$ *be a purification of* $\rho \in \mathfrak{S}_1(\mathrm{A})$*. The* complementary state *of ρ is the state $\tilde{\rho} \in \mathfrak{S}_1(\mathrm{B})$ defined by*

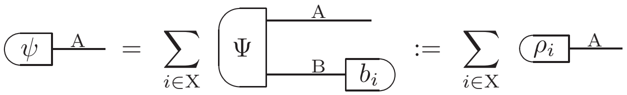


(76)

An elementary property of purification is the following

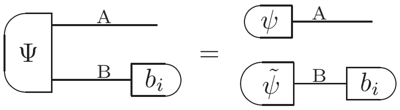
Lemma 19 *If $\psi \in \mathfrak{S}_1(\mathrm{A})$ is pure and $\Psi \in \mathfrak{S}_1(\mathrm{AB})$ is a purification of ψ, then Ψ must be of the form $\Psi = \psi \otimes \tilde{\psi}$, with $\tilde{\psi} \in \mathfrak{S}_1(\mathrm{B})$ pure.*

Proof. Take an observation-test $\{b_i\}_{i\in\mathrm{X}}$ on B. Since $\sum_i b_i = e_\mathrm{B}$ we have



(77)

namely, the states $\{\rho_i\}_{i\in\mathrm{X}}$ defined by $\rho_i := (b_i|_\mathrm{B}\, |\Psi)_\mathrm{AB}$ form a refinement of ψ. Since ψ is pure, we necessarily have $\rho_i = p_i\psi$ for some probabilities $\{p_i\}$. Precisely, we have $p_i = (e|\,\rho_i)_\mathrm{A} = (e_\mathrm{A} \otimes b_i|\,\Psi)_\mathrm{AB} = (b_i|\,\tilde{\psi}\big)_\mathrm{B}$, where $\tilde{\psi}$ is the complementary state of ψ. Therefore, we have



$\forall i \in \mathrm{X}$. (78)

The above equation implies that Ψ cannot be distinguished from $\psi \otimes \tilde{\psi}$ by any local test. Since Ψ is pure, this implies $\Psi = \psi \otimes \tilde{\psi}$. Clearly, $\tilde{\psi}$ has to be pure, otherwise we would have a non-trivial refinement of the pure state Ψ. ■

It is important to stress that purification is not a physical process: There is no physical transformation that is able to turn any arbitrary mixed state ρ into some purification Ψ of it. In quantum mechanics, this has been noted by Kleinman *et al.* in Ref. [36]. Along the same lines, it is easy to prove the following general Theorem:

Theorem 5 (No-purification of collinear states) *Let $\rho_i, i = 1,2,3$ be three distinct collinear states of system A—i.e. $\rho_1 \neq \rho_3$ and $\rho_2 = p\rho_1 + (1-p)\rho_3$ for some $0 < p < 1$. Suppose that $|\Psi_i)_\mathrm{AB}, i = 1,2,3$ is a purification of $|\rho_i)_\mathrm{A}$. Then for every finite number of copies N there is no physical transformation $\mathscr{C} \in \mathfrak{T}(\mathrm{A}^{\otimes N}, \mathrm{AB})$ such that $\mathscr{C}\,|\rho_i)_\mathrm{A}^{\otimes N} = |\Psi_i)_\mathrm{AB}$ for every $i = 1,2,3$.*

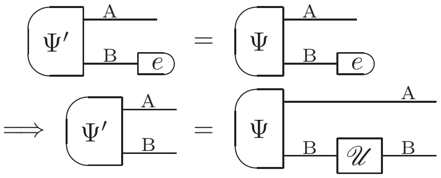
Proof. The proof is by contradiction. Suppose that such a transformation $\mathscr{C}$ exists for some finite N. Then, expanding the product $\rho_2^{\otimes N} = [p\rho_1 + (1-p)\rho_3]^{\otimes N}$, and applying the transformation $\mathscr{C}$, we obtain

$$\begin{aligned}|\Psi_2)_\mathrm{AB} &= \mathscr{C}\,|\rho_2)_\mathrm{A}^{\otimes N} \\ &= p^N \mathscr{C}\,|\rho_1)_\mathrm{A}^{\otimes N} + (1-p)^N \mathscr{C}\,|\rho_3)_\mathrm{A}^{\otimes N} + |\rho_{rest})_\mathrm{AB} \\ &= p^N\,|\Psi_1)_\mathrm{AB} + (1-p)^N\,|\Psi_3)_\mathrm{AB} + |\rho_{rest})_\mathrm{AB},\end{aligned} \tag{79}$$

where ρ_{rest} is a suitable non-normalized state. This is clearly absurd, since we obtained a non-trivial convex decomposition of the pure state Ψ_2. ■

If Ψ is a purification of ρ and $\mathscr{U}_\mathrm{B}$ is a reversible transformation on the purifying system, then also $|\Psi')_\mathrm{AB} = \mathscr{U}_B\,|\Psi)_\mathrm{AB}$ is a new purification of ρ. Indeed, $\mathscr{U}_B\,|\Psi)_\mathrm{AB}$ must be pure, otherwise by inverting $\mathscr{U}_B$ on $\mathscr{U}_B\,|\Psi)_\mathrm{AB}$ by linearity one would find that $|\Psi)_\mathrm{AB}$ is mixed. In the following Postulate we impose that all purifications are of this form:

Postulate 1 (Purification) *Every state has a purification, unique up to reversible transformations on the purifying system: if $\Psi, \Psi' \in \mathfrak{S}_1(\mathrm{AB})$ are two purifications of the same state, then they are connected by a reversible transformation $\mathscr{U} \in \mathfrak{T}(\mathrm{B})$, namely*



(80)

Remark (Uniqueness of the complementary state) Note that uniqueness of the purification assumed in the purification postulate is equivalent to the uniqueness (up to reversible transformations) of the complementary state defined in Def. 36.

We now show some simple consequences of the purification postulate. First, it implies that all pure states of a system are connected by reversible transformations:

Lemma 20 (Transitivity of the group of reversible transformations on the set of pure states) *For any couple of pure states $\psi, \psi' \in \mathfrak{S}_1(\mathrm{A})$ there is a reversible transformation $\mathscr{U} \in \mathfrak{T}(\mathrm{A})$ such that $\psi' = \mathscr{U}\psi$.*

Proof. Every system is a purifying system for the trivial system. Then just apply Eq. (80) with $\mathrm{A} \equiv \mathrm{I}$. ■

An obvious consequence of the purification postulate is that in a theory with purification there are *entangled states*, according to the usual definition:

Definition 37 (Separable states/entangled states) *A bipartite state $\sigma \in \mathfrak{S}_1(\mathrm{AB})$ is* separable *if it can be written as a convex combination of product states, that is, as $|\sigma)_\mathrm{AB} = \sum_i p_i\,|\phi_i)_\mathrm{A}\,|\psi_i)_\mathrm{B}$ with $p_i \geq 0, \sum_i p_i = 1$. A bipartite state is* entangled *if it is not separable.*

As already anticipated, one has the following (trivial) Corollary:

Corollary 6 (Existence of entangled states) *If $\Psi_\rho \in \mathfrak{S}_1(\mathrm{AB})$ is a purification of $\rho \in \mathfrak{S}_1(\mathrm{A})$ and ρ is mixed, then Ψ_ρ is entangled.*

Proof. By contradiction, suppose that Ψ_ρ is separable. Because it is pure, it must be of the form $|\Psi_\rho)_\mathrm{AB} =$

$|\varphi)_{\mathrm{A}}\,|\psi)_{\mathrm{B}}$ with $|\varphi)_{\mathrm{A}}$ and $|\psi)_{\mathrm{B}}$ pure. Then the marginal $|\rho)_{\mathrm{A}} = (e|_{\mathrm{B}}\,|\Psi_\rho)_{\mathrm{AB}} = |\varphi)_{\mathrm{A}}$ is pure, in contradiction with the hypothesis. ■

Remark (Purification and classical theories). Clearly, Corollary 6 shows that the purification postulate rules out classical probability theory. In fact, there is only one possibility for a causal theory to satisfy the purification postulate without having entangled states: the theory must not contain mixed states. This necessarily implies that the theory is deterministic, that is, that the probabilities of outcomes in any test are either 0 or 1 (if the theory were not deterministic one could construct mixed states by randomization). In particular, this also implies that in such a theory the pure states of an arbitrary system A are perfectly distinguishable. In conclusion, the only causal theories that satisfy the purification postulate and have no entanglement are classical deterministic theories.

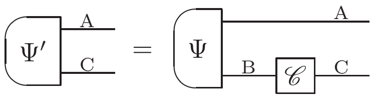
Another elementary consequence of the purification postulate is that "purity implies independence from the rest of the world":

Corollary 7 (Purity implies independence) *If* $\psi \in \mathfrak{S}_1(\mathrm{A})$ *is pure and* $\rho \in \mathfrak{S}_1(\mathrm{AB})$ *is an extension of* ψ*, namely* $|\psi)_{\mathrm{A}} = (e|_{\mathrm{B}}\,|\rho)_{\mathrm{AB}}$*, then* $\rho = \psi \otimes \sigma$*, for some state* $\sigma \in \mathfrak{S}_1(\mathrm{B})$*.*

Proof. Let $\Psi \in \mathfrak{S}_1(\mathrm{ABC})$ be a purification of ρ. Since Ψ is also a purification of ψ, by the Lemma 18 we have $|\Psi)_{\mathrm{ABC}} = |\psi)_{\mathrm{A}}\,|\eta)_{\mathrm{BC}}$, for some pure state $\eta \in \mathfrak{S}_1(\mathrm{BC})$. But since Ψ is a purification of ρ we have $|\rho) = (e|_{\mathrm{C}}\,|\Psi)_{\mathrm{ABC}} = |\psi)_{\mathrm{A}}\,|\sigma)_{\mathrm{B}}$, with $|\sigma)_{\mathrm{B}} := (e|_{\mathrm{C}}\,|\eta)_{\mathrm{BC}}$. ■

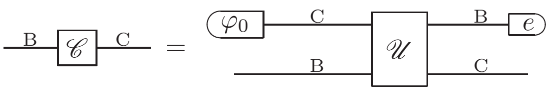
We conclude this subsection with an important Lemma that extends the uniqueness of purification to the case of purifications with different purifying systems:

Lemma 21 (Uniqueness of the purification up to channels on the purifying systems) *Let* $\Psi \in \mathfrak{S}_1(\mathrm{AB})$ *and* $\Psi' \in \mathfrak{S}_1(\mathrm{AC})$ *be two purifications of* $\rho \in \mathfrak{S}_1(\mathrm{A})$*. Then there exists a channel* $\mathscr{C} \in \mathfrak{T}(\mathrm{B},\mathrm{C})$ *such that*



(81)

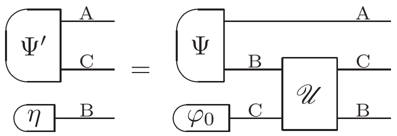
Moreover, channel $\mathscr{C}$ *has the form*



(82)

for some pure state $\varphi_0 \in \mathfrak{S}_1(\mathrm{C})$ *and some reversible channel* $\mathscr{U} \in \mathbf{G}_{\mathrm{BC}}$*.*

Proof. Let $|\eta)_{\mathrm{B}}$ and $|\varphi_0)_{\mathrm{C}}$ be an arbitrary pure state of B and C, respectively. Then $|\Psi')_{\mathrm{AC}}\,|\eta)_{\mathrm{B}}$ and $|\Psi)_{\mathrm{AB}}\,|\varphi_0)_{\mathrm{C}}$ are two purifications of ρ with the same purifying system BC. Due to Eq. (80), we have



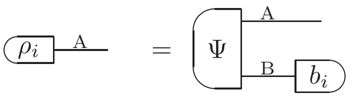
(83)

Applying the deterministic effect e on system B we obtain Eq. (81), with $\mathscr{C} := (e|_{\mathrm{B}}\,\mathscr{U}\,|\varphi_0)$. ■

B. Purification of preparation-tests

We now show that the purification of normalized states implies the purification of preparation-tests.

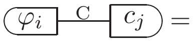
Theorem 6 (Purification of preparation-tests) *Let* $\{\rho_i\}_{i\in\mathrm{X}}$ *be a preparation-test for system* A*, and let* $\Psi \in \mathfrak{S}_1(\mathrm{AB})$ *be a purification of the coarse-grained state* $\rho := \sum_{i\in\mathrm{X}} \rho_i$*. Then there exists an observation-test* $\{b_i\}_{i\in\mathrm{X}}$ *on system* B *such that*



(84)

for any outcome $i \in \mathrm{X}$*. By suitably choosing the purifying system* B*, the observation-test* $\{b_i\}_{i\in\mathrm{X}}$ *can be taken to be discriminating (Definition 26).*

Proof. Take a set of $|\mathrm{X}|$ perfectly distinguishable states $\{\varphi_i\}_{i\in\mathrm{X}} \subset \mathfrak{S}_1(\mathrm{C})$ for some system C. By definition of perfect distinguishability, there exists a discriminating test $\{c_i\}_{i\in X}$ such that

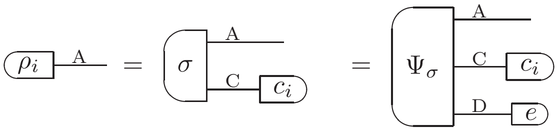


$= \delta_{ij}$ (85)

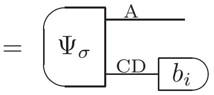
for all $i, j \in \mathrm{X}$. Now consider the state

$$\sigma := \sum_{i\in\mathrm{X}} (\rho_i \otimes \varphi_i) \in \mathfrak{S}_1(\mathrm{AC}), \tag{86}$$

which is clearly an extension of ρ, namely $|\rho)_A = (e|_{\mathrm{C}}\,|\sigma)_{\mathrm{AC}}$. Let $\Psi_\sigma \in \mathfrak{S}_1(\mathrm{ACD})$ be a purification of σ. By definition, Ψ is also a purification of ρ. Using Eq.(85) we obtain for every outcome $i \in \mathrm{X}$



(87)

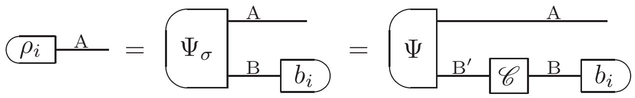


(88)

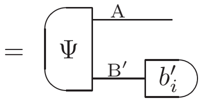
having defined the discriminating test $(b_i|_{\mathrm{CD}} := (c_i|_{\mathrm{C}}\,(e|_{\mathrm{D}}$. This proves that there exists a purification of

ρ with purifying system $\mathrm{B} := \mathrm{CD}$, and a discriminating test $\{b_i\}_{i\in\mathrm{X}}$ on B such that the thesis holds.

Finally, if $\Psi \in \mathfrak{S}_1(\mathrm{AB}')$ is any other purification of ρ, using Lemma 21 we have



(89)



(90)

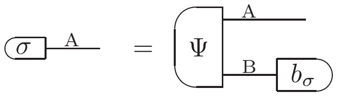
where $\{b'_i\}_{i\in\mathrm{X}}$ is the observation-test on B' defined by $(b'_i| := (b_i|\mathscr{C}$. ■

The property stated by Theorem 6 is sometimes called *steering* in quantum theory, with a terminology that dates back to Schrödinger [37] (see also Ref. [38], for a very recent discussion in the general probabilistic framework): one says that a bipartite state $|\sigma)_{\mathrm{AB}}$ steers its marginal $|\rho)_{\mathrm{A}} = (e|_{\mathrm{B}}\, |\sigma)_{\mathrm{AB}}$ on system A, if every convex decomposition $|\rho)_{\mathrm{A}} = \sum_{i\in\mathrm{X}} p_i\, |\rho_i)_{\mathrm{A}}$ is induced by a suitable observation-test on system B. Using the notion of steering, we may state the following:

Corollary 8 (Pure bipartite states are steering for their marginals) *In a theory with purification any pure state* $|\Psi) \in \mathfrak{S}_1(\mathrm{AB})$ *steers its marginal states* $|\rho)_{\mathrm{A}} = (e|_{\mathrm{B}}\, |\Psi)_{\mathrm{AB}}$ *and* $|\tilde\rho)_{\mathrm{A}} = (e|_{\mathrm{A}}\, |\Psi)_{\mathrm{AB}}$.

We now present a few other corollaries of the purification of preparation-tests stated by Theorem 6.

Corollary 9 *Let* $\Psi \in \mathfrak{S}_1(\mathrm{AB})$ *be a purification of* ρ. *Then, a state* σ *is in the refinement set* D_ρ *if and only if there is an effect* $b_\sigma \in \mathfrak{E}(\mathrm{B})$ *such that*



(91)

Proof. The "if" part is trivial. Conversely, if σ is in D_ρ, by definition there exists a preparation-test $\{\rho_i\}_{i\in\mathrm{X}}$ and an outcome i_0 such that $\rho_{i_0} = \sigma$. Using Theorem 6 and taking the effect $b_\sigma := b_{i_0}$ one proves the thesis. ■

Corollary 10 (Bound on dimensions) *Let* $\Psi \in \mathfrak{S}_1(\mathrm{AB})$ *be a purification of* $\rho \in \mathfrak{S}_1(\mathrm{A})$. *Then, one has the bound*

$$\dim \mathfrak{S}_{\mathbb{R}}(\mathrm{B}) \geq \dim \mathsf{Span}(D_\rho). \tag{92}$$

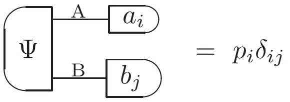
In particular, if ρ *is an internal state, one has*

$$\dim \mathfrak{S}_{\mathbb{R}}(\mathrm{B})) \geq \dim \mathfrak{S}_{\mathbb{R}}(\mathrm{A}). \tag{93}$$

Proof. Consider the map $\hat\omega : \mathfrak{E}_{\mathbb{R}}(\mathrm{B}) \to \mathfrak{S}_{\mathbb{R}}(\mathrm{A})$ defined by $b \mapsto |\hat\omega_b)_{\mathrm{A}} := (b|\,|\Psi)_{\mathrm{AB}}$. By the previous corollary, the range of $\hat\omega$ contains D_ρ. Since $\hat\omega$ is linear, this implies $\dim \mathfrak{E}_{\mathbb{R}}(\mathrm{B}) \geq \dim \mathsf{Span}(D_\rho)$. On the other hand, since states and effects span dual vector spaces, one has $\dim \mathfrak{E}_{\mathbb{R}}(\mathrm{B}) \equiv \dim \mathfrak{S}_{\mathbb{R}}(\mathrm{B})$, thus proving Eq. (92). ■

Theorem 6 implies the existence of pure bipartite states exhibiting perfect correlations in the statistics of independent observations:

Corollary 11 (Pure states with perfect correlations) *Let* $\rho = \sum_{i\in\mathrm{X}} p_i\varphi_i$ *be a mixture of perfectly distinguishable states* $\{\varphi_i\} \subset \mathfrak{S}_1(\mathrm{A})$, *and let* $\Psi \in \mathfrak{S}_1(\mathrm{AB})$ *be a purification of* ρ. *Then there exist two observation-tests* $\{a_i\}_{i\in\mathrm{X}}$ *and* $\{b_j\}_{j\in\mathrm{X}}$ *on systems* A *and* B, *respectively, such that*



(94)

Proof. Consider the preparation-test $\{\rho_i\}_{i\in\mathrm{X}}$ with $\rho_i = p_i\varphi_i$. Since its coarse-grained state is ρ, by Theorem 6 there exists an observation-test $\{b_i\}$ such that $|\rho_i)_{\mathrm{A}} = (b_i|_{\mathrm{B}}\, |\Psi)_{\mathrm{AB}}$. On the other hand, the states $\{\varphi_i\}$ are perfectly distinguishable with a test $\{a_i\}_{i\in\mathrm{X}}$. Hence, we have $(a_i|_{\mathrm{A}}\, (b_j|_{\mathrm{B}}\, |\Psi)_{\mathrm{AB}} = (a_i|_{\mathrm{A}}\, |\rho_j) = p_i\delta_{ij}$. ■

This directly implies the following property

Corollary 12 *Let* $\rho = \sum_{i\in\mathrm{X}} p_i\varphi_i \in \mathfrak{S}_1(\mathrm{A})$ *be a mixture of perfectly distinguishable states,* $\Psi \in \mathfrak{S}_1(\mathrm{AB})$ *be a purification of* ρ, *and* $\tilde\rho = (e|_{\mathrm{A}}\, |\Psi)_{\mathrm{AB}}$ *be the complementary state of* ρ. *Then, one has*

$$\tilde\rho = \sum_{i\in\mathrm{X}} p_i\tilde\varphi_i, \tag{95}$$

where $\{\tilde\varphi\}_{i\in\mathrm{X}}$ *are perfectly distinguishable states of* B.

We conclude this subsection with a crucial consequence of the purification of preparation-test stated by Theorem 6, namely that if two transformations coincide on a purification of ρ, they also coincide *upon input of* ρ, according to the following definition:

Definition 38 (Equality upon input of ρ) *Two transformations* $\mathscr{A}, \mathscr{A}' \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ *are* equal upon input of ρ, *denoted by* $\mathscr{A} =_\rho \mathscr{A}'$, *if one has*

$$\mathscr{A}\, |\sigma)_{\mathrm{A}} = \mathscr{A}'\, |\sigma)_{\mathrm{A}} \qquad \forall\, |\sigma)_{\mathrm{A}} \in D_\rho. \tag{96}$$

In quantum theory two quantum operations $\mathscr{A}, \mathscr{A}'$ are equal upon input of ρ if and only of one has $\mathscr{A}(\sigma) = \mathscr{A}'(\sigma)$ for every density matrix σ whose support is contained in the support of ρ.

We then have the following:

Theorem 7 (Equality upon input of ρ vs equality on purifications) *Let* $\Psi \in \mathfrak{S}_1(\mathrm{AC})$ *be a purification of* $\rho \in \mathfrak{S}_1(\mathrm{A})$, *and let* $\mathscr{A}, \mathscr{A}' \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ *be two transformations. Then one has*

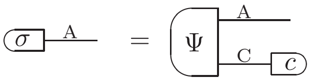
$$\mathscr{A}\, |\Psi)_{\mathrm{AC}} = \mathscr{A}'\, |\Psi)_{\mathrm{AC}} \qquad \Longrightarrow \qquad \mathscr{A} =_\rho \mathscr{A}'\ . \tag{97}$$

If local discriminability holds, one has the equivalence

$$\mathscr{A}\,|\Psi)_{\mathrm{AC}} = \mathscr{A}'\,|\Psi)_{\mathrm{AC}} \qquad \Longleftrightarrow \qquad \mathscr{A} =_\rho \mathscr{A}' \ . \qquad (98)$$

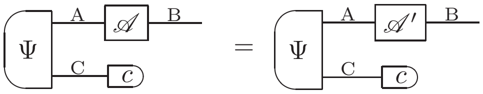
If one of the two transformations is proportional to a reversible transformation the equivalence of Eq. (98) holds under the weaker assumption of local discriminability on pure states.

Proof. By definition, a state σ is in the refinement set D_ρ iff there exists a preparation-test $\{\rho_i\}_{i\in\mathrm{X}}$ and an outcome i_0 such that $\rho_{i_0} = \sigma$. Using Corollary 9, we have that σ is in D_ρ iff there exist an effect c on C such that



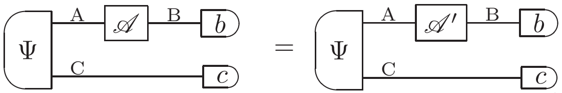
(99)

Therefore, we have that $\mathscr{A} =_\rho \mathscr{A}'$ if and only if



(100)

that is, if and only if the states $\mathscr{A}\,|\Psi)_{\mathrm{AC}}$ and $\mathscr{A}'\,|\Psi)_{\mathrm{AC}}$ cannot be distinguished by local tests, that is, if and only if



(101)

for every product effect $(b|_{\mathrm{B}}\,(c|_{\mathrm{C}}$. Clearly, if $\mathscr{A}\,|\Psi)_{\mathrm{AC}} = \mathscr{A}'\,|\Psi)_{\mathrm{AC}}$, this condition is verified: this proves Eq. (97). When local discriminability holds, equality on local tests implies equality on global tests, hence Eq. (98). Finally, if $\mathscr{A}' = \lambda\mathscr{U}$ with $\mathscr{U}$ reversible, then the state $\mathscr{A}'\,|\Psi)_{\mathrm{AC}} = \lambda\mathscr{U}\,|\Psi)_{\mathrm{AC}}$ is pure, and, by local discriminability of pure states, equality on local tests implies equality. ■

C. Dynamically faithful pure states

We show now an important feature of theories with purification: the possibility of imprinting physical transformations into states in an injective way (that is, if two transformations differ, then the corresponding states are differ). This feature reduces the tomography of a physical process to the tomography of the corresponding state. Technically speaking, we call *dynamically faithful* any state that allows for the tomography of physical processes.

Definition 39 (Dynamically faithful state) *We say that a state $\sigma \in \mathfrak{S}(\mathrm{AC})$ is* dynamically faithful *for system* A *if for any couple of transformations $\mathscr{A}, \mathscr{A}' \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ on has*

$$\mathscr{A}\,|\sigma)_{\mathrm{AC}} = \mathscr{A}'\,|\sigma)_{\mathrm{AC}} \qquad \Longrightarrow \qquad \mathscr{A} = \mathscr{A}'. \qquad (102)$$

The existence of dynamically faithful mixed states is a quite generic fact: for example, in any theory with local discriminability if one takes a basis $\{\rho_i\} \subset \mathfrak{S}_1(\mathrm{A})$ for $\mathfrak{S}_{\mathbb{R}}(\mathrm{A})$ and a set $\{\varphi_i\} \subset \mathfrak{S}_1(\mathrm{C})$ of perfectly distinguishable states of some system C, than any mixture $\sigma = \sum_i p_i \rho_i \otimes \varphi_i$ is dynamically faithful. The remarkable fact in a theory with purification is that there exist dynamically faithful states, which, in addition, are pure.

Theorem 8 (Existence of dynamically faithful pure states) *Let $\omega \in \mathfrak{S}_1(\mathrm{A})$ be an internal state, and let $\Psi_\omega \in \mathfrak{S}_1(\mathrm{AC})$ be a purification of ω. Then Ψ_ω is dynamically faithful for system* A.

Proof. Suppose that $\mathscr{A}\,|\Psi_\omega)_{\mathrm{AC}} = \mathscr{A}'\,|\Psi_\omega)_{\mathrm{AC}}$. Then take an arbitrary system D, an internal state $\sigma \in \mathfrak{S}_1(\mathrm{D})$, and a purification of σ, say $\Psi_\sigma \in \mathfrak{S}_1(\mathrm{DE})$. Clearly, we have $\mathscr{A}\,|\Psi_\omega)_{\mathrm{AC}} \otimes |\Psi_\sigma)_{\mathrm{DE}} = \mathscr{A}'\,|\Psi_\omega)_{\mathrm{AC}} \otimes |\Psi_\sigma)_{\mathrm{DE}}$. According to Theorem 7, this implies that $\mathscr{A}$ and $\mathscr{A}'$ coincide upon input of $\omega \otimes \sigma$. Since ω and σ are internal in $\mathfrak{S}(\mathrm{A})$ and $\mathfrak{S}(\mathrm{D})$, respectively, by Theorem 2 $\omega \otimes \sigma$ is internal in $\mathfrak{S}(\mathrm{AD})$, that is, the refinement set $D_{\omega\otimes\sigma}$ is a spanning set for $\mathfrak{S}_{\mathbb{R}}(\mathrm{AD})$. Now, $\mathscr{A}$ and $\mathscr{A}'$ coincide on a spanning set, and, therefore, they coincide on every state of $\mathfrak{S}(\mathrm{AD})$. Since the ancillary system D is arbitrary, this implies $\mathscr{A} = \mathscr{A}'$. ■

The converse of the previous Theorem 8 also holds:

Theorem 9 (Characterization of dynamically faithful pure states) *A pure state $\Psi \in \mathfrak{S}_1(\mathrm{AC})$ is dynamically faithful for system* A *if and only if the marginal state $|\omega)_{\mathrm{A}} := (e|_{\mathrm{C}}\,|\Psi)_{\mathrm{AC}}$ is internal.*

Proof. The "if" part has been just shown in Theorem 8. To prove the "only if" part, let a, a' be two distinct effects for system A. Since Ψ is dynamically faithful one has $(a|_{\mathrm{A}}\,|\Psi)_{\mathrm{AC}} \neq (a|'_{\mathrm{A}}\,|\Psi)_{\mathrm{AC}}$. This means that there exists an effect $(c|_{\mathrm{C}}$ such that $(a|_{\mathrm{A}}\,(c|_{\mathrm{C}}\,|\Psi)_{\mathrm{AC}} \neq (a|'_{\mathrm{A}}\,(c|_{\mathrm{C}}\,|\Psi)_{\mathrm{AC}}$. Defining the state $|\omega_c)_{\mathrm{A}} := (c|_{\mathrm{C}}\,|\Psi)_{\mathrm{AC}}$, this implies $(a|\omega_c)_{\mathrm{A}} \neq (a'|\omega_c)_{\mathrm{A}}$. Since ω_c is in the refinement set of ω, such a refinement set is separating for $\mathfrak{E}_{\mathbb{R}}(\mathrm{A})$. But a separating set for $\mathfrak{E}_{\mathbb{R}}(\mathrm{A})$ must be a spanning set for the dual vector space $\mathfrak{S}_{\mathbb{R}}(\mathrm{A})$. Hence, ω is internal. ■

Using this characterization it is immediate to show that the product of dynamically faithful pure states is dynamically faithful:

Corollary 13 (Product of dynamically faithful states is dynamically faithful) *Let $\Psi^{(\mathrm{A})} \in \mathfrak{S}_1(\mathrm{AC})$ and $\Psi^{(\mathrm{B})} \in \mathfrak{S}_1(\mathrm{BD})$ be dynamically faithful for systems* A *and* B*, respectively. Then $\Psi^{(\mathrm{A})} \otimes \Psi^{(\mathrm{B})}$ is dynamically faithful for the compound system* AB.

Proof. Since the product of two internal states is internal (Theorem 2), the thesis trivially follows from the previous Theorem. ■

The existence of dynamically faithful pure states has remarkable consequences, among which the "no-information without disturbance" and the "no-cloning" Theorems, that will be analyzed in the following Subsections.

D. No information without disturbance

Definition 40 (Non-disturbing tests) *We say that a test* $\{\mathscr{A}_i\}_{i\in\mathrm{X}}$ *on system* A *is* non-disturbing *upon input of* $\rho \in \mathfrak{S}(\mathrm{A})$ *if*

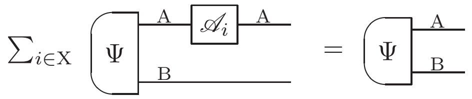
$$\sum_{i\in\mathrm{X}} \mathscr{A}_i \, |\sigma)_\mathrm{A} = |\sigma)_\mathrm{A} \qquad \forall \sigma \in D_\rho, \tag{103}$$

or, equivalently, if $\sum_{i\in\mathrm{X}} \mathscr{A}_i =_\rho \mathscr{I}_\mathrm{A}$. *If* ρ *is an internal state, we say that the test is* non-disturbing, *because in this case one has*

$$\sum_{i\in\mathrm{X}} \mathscr{A}_i \, |\sigma)_\mathrm{A} = |\sigma)_\mathrm{A} \qquad \forall \sigma \in \mathfrak{S}(\mathrm{A}). \tag{104}$$

Theorem 10 (No information without disturbance) *In a theory with purification, a test* $\{\mathscr{A}_i\}$ *on system* A *is non-disturbing upon input of* ρ, *if and only if each transformation* $\mathscr{A}_i$ *is proportional to the identity upon input of* ρ, *namely* $\mathscr{A}_i =_\rho p_i \mathscr{I}_\mathrm{A}$.

Proof. Let Ψ_AB be a purification of ρ. By Theorem 7, the no-disturbance condition $\sum_{i\in\mathrm{X}} \mathscr{A}_i =_\rho \mathscr{I}_\mathrm{A}$ holds if and only if



(105)

Since Ψ is pure, this implies $\mathscr{A}_i |\Psi)_\mathrm{AB} = p_i |\Psi)_\mathrm{AB} = (p_i \mathscr{I}_\mathrm{A}) |\Psi)_\mathrm{AB}$. Now, since the identity is trivially a reversible transformation, according to Theorem 7 this is equivalent to $\mathscr{A}_i =_\rho p_i \mathscr{I}_\mathrm{A}$. ∎

Theorem 11 (No joint discrimination of a spanning set of states) *In a theory with purification the states in a spanning set cannot be perfectly discriminated in a single observation-test.*

Proof. By contradiction, suppose that a collection of states $\{\rho_i\}_{i\in\mathrm{X}}$ is a spanning set—namely $\mathsf{Span}\{\rho_i\}_{i\in\mathrm{X}} = \mathfrak{S}_\mathbb{R}(\mathrm{A})$—and there exists an observation-test $\{a_i\}_{i\in\mathrm{X}}$ such that $(a_i|\rho_j)_\mathrm{A} = \delta_{ij}$. Then, since perfectly distinguishable states are linearly independent, and they must span a finite dimensional vector space, the number of perfectly distinguishable states must be finite. Now consider the measure-and-prepare test $\{\mathscr{A}_i\}_{i\in\mathrm{X}}$ defined by $\mathscr{A}_i = |\rho_i)_\mathrm{A} (a_i|_\mathrm{A}$. Since the states of the spanning set are perfectly distinguishable, the test $\{\mathscr{A}_i\}$ is non-disturbing. Indeed, expanding an arbitrary state ρ on the spanning set, one has

$$\sum_i \mathscr{A}_i \, |\rho)_A = \sum_i \mathscr{A}_i \left(\sum_j c_j \, |\rho_j)_A \right) = \sum_j c_j \, |\rho_j)_A = |\rho). \tag{106}$$

Since $\mathscr{A}_i \neq p_i \mathscr{I}_\mathrm{A}$, this is in contradiction with the no-information without disturbance Theorem 10. ∎

Corollary 14 (No joint discrimination of pure states) *In a theory with purification for every system the pure states cannot be perfectly discriminated in a single observation-test.*

Proof. Since pure states are a spanning set, they cannot be perfectly discriminated in a single test, according to Theorem (11). ∎

Corollary 14 provides a simple alternative way to see that classical probability theory is excluded by the purification Postulate.

Corollary 15 (Maximum number of perfectly distinguishable states) *For every system* A *the maximum cardinality of a set of perfectly distinguishable states is strictly smaller than* $\dim \mathfrak{S}_\mathbb{R}(\mathrm{A})$.

Proof. Since perfectly distinguishable states are linearly independent, if one could find $\dim \mathfrak{S}_\mathbb{R}(\mathrm{A})$ perfectly distinguishable states, then they would form a spanning set, in contradiction with Theorem 11. ∎

Note that the maximum number of distinguishable states in quantum theory satisfies a much stronger bound: such a number is given by the dimension d_A of the system's Hilbert space, while the dimension of the vector space spanned by the density matrices is $\dim \mathfrak{S}_\mathbb{R}(\mathrm{A}) = d_\mathrm{A}^2$.

Corollary 16 (Non-unique convex decomposition on pure states) *In a theory with purification satisfying the no-restriction hypothesis of Def. 16, for every system* A *there is a mixed state* $\rho \in \mathfrak{S}_1(\mathrm{A})$ *with a non-unique convex decomposition on pure states. In other words, the convex set* $\mathfrak{S}_1(\mathrm{A})$ *cannot be a simplex.*

Proof. By contradiction, suppose that $\mathfrak{S}_1(\mathrm{A})$ is a simplex. Then the pure states $\{\varphi_i\}$ of A are a finite set, and for each of them there is a functional $a_i \in \mathfrak{E}_\mathbb{R}(\mathrm{A})$ such that $(a_i|\varphi_j) = \delta_{ij}$. Clearly, a_i is positive on every state, namely $a_i \in \mathfrak{S}_+(\mathrm{A})^*$. Hence, by the consequence of the no-restriction hypothesis stated by Lemma 11, we have $a_i \in \mathfrak{E}_+(\mathrm{A})$. Moreover, one has $\sum_i (a_i|_\mathrm{A} = (e|_\mathrm{A}$. In Corollary 37 we will show that any such collection $\{a_i\}$ is an observation-test. But this test discriminates all pure states, in contradiction with Corollary 14. This proves that $\mathfrak{S}_1(\mathrm{A})$ cannot be a simplex. ∎

E. No-cloning

Definition 41 (Cloning channels) *Let* A, A' *be two operationally equivalent systems, and let* $\{\rho_i\}_{i\in\mathrm{X}}$ *be a set of states of* A*. A channel* $\mathscr{C}$ *from* A *to* AA' *is a* cloning channel *for the set* $\{\rho_i\}_{i\in\mathrm{X}}$ *if*

$$\mathscr{C}\,|\rho_i)_{\mathrm{A}} = |\rho_i)_{\mathrm{A}}\,|\rho_i)_{\mathrm{A}'}. \tag{107}$$

If there is a cloning channel, we say that the states $\{\rho_i\}_{i\in\mathrm{X}}$ *are* perfectly cloneable.

We now show that a spanning set of states (in particular, the set of pure states) cannot be perfectly cloned. To see this we use the equivalence between perfect cloning and perfect discrimination, which was originally proved in Refs. [6, 7] for causal theories with local discriminability using the tomographic limit. Here we use the stronger result of Ref. [39], which proves the equivalence in any convex theory where all "measure-and-prepare" channels are allowed, without requiring causality and local discriminability, and without resorting to the tomographic limit. For convenience of the reader, the argument of Ref. [39] is reproduced here using the notation of the present paper:

Theorem 12 (Cloning/discrimination equivalence) *In a convex theory where all "measure-and-prepare" channels are allowed, the deterministic states* $\{\rho_i\}_{i\in\mathrm{X}} \subset \mathfrak{S}(\mathrm{A})$ *are perfectly cloneable if and only if they are perfectly distinguishable.*

Proof. Suppose that the states $\{\rho_i\}_{i\in\mathrm{X}}$ can be perfectly cloned and consider the binary discrimination between two states $\rho_i, \rho_j, i \neq j$ with a binary observation-test $\{a_i, a_j\}$. Define the worst-case error probability as

$$p_{wc} := \max\{p(i|j), p(j|i)\} \qquad p(k|l) := (a_k|\rho_l)_{\mathrm{A}}, \tag{108}$$

and take its minimum over all binary tests

$$p_{wc}^{(opt)} := \min_{a_i, a_j} p_{wc}. \tag{109}$$

Now, if a cloning channel exists, we can apply it twice to the unknown state, thus getting three identical copies of it. Performing three times the optimal test, and then using majority voting we obtain the new error probabilities given by

$$p'(i|j) = f(p^{(opt)}(i|j)) \qquad f(x) = x^2(3-2x), \tag{110}$$

where $p^{(opt)}(i|j) := \left(a_i^{(opt)}\middle|\rho_j\right)_{\mathrm{A}}$. Since f is a non-decreasing function for $x \in [0,1]$, we also have $p'_{wc} = f\left(p_{wc}^{(opt)}\right)$, and, since $p_{wc}^{(opt)}$ is the minimum error probability, by definition $p'_{wc} \geq p_{wc}^{opt}$. The only solutions of the inequality $f(x) \geq x$ are $x = 0$ and $x \in [1/2, 1]$, and, since $p_{wc}^{(opt)}$ must be in the interval $[0, 1/2)$ (see Lemma 2), we obtain $p_{wc}^{(opt)} = 0$. This proves that any pair of states from the set $\{\rho_i\}_{i\in\mathrm{X}}$ can be perfectly distinguished. But this implies that using $|\mathrm{X}|-1$ pairwise tests we can perfectly discriminate all the states $\{\rho_i\}_{i\in\mathrm{X}}$. This proves the implication "perfect cloning ⇒ perfect discrimination" in any convex theory. If the theory contains all possible "measure-and-prepare" channels, the converse is obviously true: If the states can be perfectly discriminated by an observation-test $\{a_i\}_{i\in\mathrm{X}}$, then the measure-and-prepare channel $\mathscr{C} := \sum_{i\in\mathrm{X}} |\rho_i)_{\mathrm{A}}\,|\rho_i)_{\mathrm{A}'}\,(a_i|_{\mathrm{A}}$ is a cloning channel. ■

Since measure-and-prepare channels can be obtained by conditioning the choice of a preparation-test on the outcome of an observation-test, any causal theory satisfies the hypotheses of the previous Theorem, which becomes

Corollary 17 (Cloning/discrimination equivalence in causal theories) *In a causal theory the states* $\{\rho_i\}_{i\in\mathrm{X}} \subset \mathfrak{S}_1(\mathrm{A})$ *are perfectly cloneable if and only if they are perfectly distinguishable.*

Remark (Non-causal theories with all measure-and-prepare channels). Note that there are also non-causal theories that contain all measure-and prepare channels. An example can be constructed by starting from a causal theory Θ, and by regarding the set of transformations $\mathfrak{T}(\mathrm{A},\mathrm{B})$ from A to B as the set of "states" $\mathfrak{S}'(\mathrm{A}\to\mathrm{B})$ of the system "$\mathrm{A}\to\mathrm{B}$" in a new *second-order* theory Θ'. Performing an observation-test on a "state" $\mathscr{C} \in \mathfrak{S}'(\mathrm{A}\to\mathrm{B})$ is then interpreted in the underlying causal theory Θ as applying the transformation $\mathscr{C} \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ to an input state $\sigma \in \mathfrak{S}_1(\mathrm{AC})$, and subsequently performing an observation-test $\{b_i\}_{i\in\mathrm{X}}$ on the output state $(\mathscr{C}\otimes\mathscr{I}_{\mathrm{C}})\,|\sigma)_{\mathrm{AC}}$. Of course, since the theory Θ is causal, one can use conditioning and perform a channel $\mathscr{C}_i$ that depends on the outcome i. This provides the realization of an arbitrary measure-and-prepare channel in the non-causal theory Θ'.

Coming back to causal theories with purification, the results proved so far imply the following no-cloning statement:

Corollary 18 (No-cloning of states in a spanning set) *In a theory with purification, a cloning channel for a spanning set of states cannot exist. In particular, pure states cannot be cloned.*

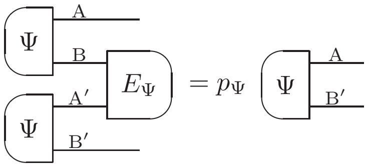
Proof. Immediate consequence of Corollary 17 combined with Theorem 11 and Corollary 14. ■

VIII. PROBABILISTIC TELEPORTATION

A. Entanglement-swapping and teleportation

As we previously showed, in a theory with purification there must be entangled states (according to the usual definition, see Def. 37). We now show the possibility of probabilistic entanglement swapping:

Theorem 13 (Probabilistic entanglement-swapping) *Let* $\Psi \in \mathfrak{S}_1(\mathrm{AB})$ *be a pure state, and let* A' *and* B' *be operationally equivalent to* A *and* B*, respectively. Then there exist an atomic effect* $E_\Psi \in \mathfrak{E}(\mathrm{BA}')$ *(see Def. 22)and a non-zero probability* p_Ψ *such that*



(111)

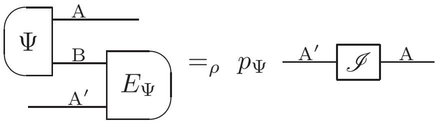
Proof. Let us define the marginal states

$$\begin{aligned} |\rho)_{\mathrm{A}} &:= (e|_{\mathrm{B}}\, |\Psi)_{\mathrm{AB}} \\ |\tilde\rho)_{\mathrm{B}} &:= (e|_{\mathrm{A}}\, |\Psi)_{\mathrm{AB}} \end{aligned} \tag{112}$$

By Theorem 3 we have that there exists a non-zero probability p_Ψ such that $p_\Psi \Psi \in D_{\rho\otimes\tilde\rho}$. Since $|\Psi)_{\mathrm{AB}}\,|\Psi)_{\mathrm{A}'\mathrm{B}'}$ is a purification of $|\rho)_{\mathrm{A}}\,|\tilde\rho)_{\mathrm{B}'}$, using corollary 9 we get the thesis. The effect E_Ψ can be taken to be atomic: indeed, if it were refinable, i.e. $E_\Psi = \sum_i E_i$, since the right hand side of Eq. (111) is a pure state, each effect E_i would achieve entanglement swapping. ■

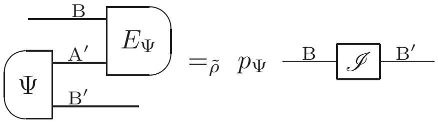
Remark (PR boxes are excluded by the purification Postulate). The possibility of probabilistic entanglement swapping shows that the purification Postulate excludes the theory of Popescu-Rohrlich boxes (see Ref. [8] for the definition of transformations on boxes and states of multipartite boxes). Indeed, Refs. [9, 10] showed that probabilistic entanglement swapping is impossible in this theory.

Corollary 19 (Probabilistic teleportation) *Let* $\Psi \in \mathfrak{S}_1(\mathrm{AB})$ *be a pure state, and let* $\rho \in \mathfrak{S}_1(\mathrm{A})$ *and* $\tilde\rho \in \mathfrak{S}_1(\mathrm{B})$ *be its marginals. Let* A' *and* B' *be operationally equivalent to* A *and* B*, respectively. Then, there exists an atomic effect* $E_\Psi \in \mathfrak{E}(\mathrm{BA}')$ *and a non-zero probability* p_Ψ *such that*



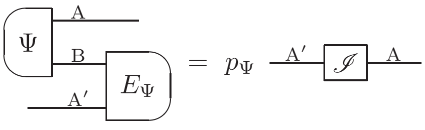
(113)

and



(114)

In particular, if ρ *is an internal state, one has the probabilistic teleportation scheme*



(115)

Proof. Just combine Theorems 13 and 7. ■

The diagram of probabilistic teleportation (115) is one of the main axioms in the categorial approach by Abramsky and Coecke [41]. In the present approach, this property is derived from the purification postulate, rather than being assumed from the start.

For theories with local discriminability the probability of teleportation is related to the dimension of the state space as follows:

Lemma 22 (Maximum teleportation probability) *If local discriminability holds, then the probability of teleportation* p_Ψ *in Eq. (115) satisfies the bound*

$$p_\Psi \le \frac{1}{\dim \mathfrak{S}_{\mathbb{R}}(\mathrm{A})}\ . \tag{116}$$

Proof. Let us choose two bases $\{\rho_i\}$ and $\{\tilde\rho_j\}$ for the vector spaces $\mathfrak{S}_{\mathbb{R}}(\mathrm{A})$ and $\mathfrak{S}_{\mathbb{R}}(\tilde{\mathrm{A}})$, respectively, and write Ψ as $|\Psi)_{\mathrm{A}\tilde{\mathrm{A}}} = \sum_{i,j} A_{ij}\, |\rho_i)_{\mathrm{A}}\, |\tilde\rho_j)_{\tilde{\mathrm{A}}}$. Now take the dual bases $\{\rho_i^*\}$ and $\{\tilde\rho_j^*\}$ for the dual vector spaces $\mathfrak{E}_{\mathbb{R}}(\mathrm{A})$ and $\mathfrak{E}_{\mathbb{R}}(\tilde{\mathrm{A}})$, respectively—so that $(\rho_i^*|\rho_j)_{\mathrm{A}} = \delta_{ij}$ and $(\tilde\rho_k^*|\tilde\rho_l)_{\tilde{\mathrm{A}}} = \delta_{kl}$—, and write E_Ψ as $(E_\Psi|_{\tilde{\mathrm{A}}\mathrm{A}'} = \sum_{k,l} B_{kl}\, (\tilde\rho_k^*|_{\tilde{\mathrm{A}}}\, (\rho_l^*|_{\mathrm{A}'}$. The teleportation diagram (115) is then equivalent to the matrix equality

$$AB = p_\Psi I_{\mathrm{A}}, \tag{117}$$

where I_{A} is the identity matrix of size $\dim(\mathfrak{S}_{\mathbb{R}}(\mathrm{A}))$. Finally, since probabilities are bounded by unit, we obtain

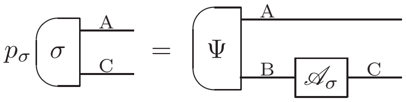
$$1 \ge (E_\Psi|\,\Psi)_{\mathrm{A}\tilde{\mathrm{A}}} = \mathrm{Tr}[AB] = p_\Psi \dim(\mathfrak{S}_{\mathbb{R}}(\mathrm{A})), \tag{118}$$

which is the desired bound. ■

Remark (quantum theory achieves the bound). Note that in quantum theory the teleportation probability achieves the maximum value allowed by the bound of Eq. (116): For a d-dimensional Hilbert space, the real vector space spanned by all density matrices has dimension d^2, which is exactly the maximum probability of conclusive teleportation.

A simple consequence of probabilistic teleportation is the possibility of remotely preparing any bipartite state by acting locally on the purifying system only, according to the following definition

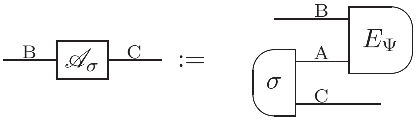
Definition 42 (Preparationally faithful state) *A state* $\Psi \in \mathfrak{S}_1(\mathrm{AB})$ *is* preparationally faithful *for system* B *if for every bipartite state* $\sigma \in \mathfrak{S}_1(\mathrm{AC})$ *there are a transformation* $\mathscr{A}_\sigma \in \mathfrak{T}(\mathrm{B},\mathrm{C})$ *and a non-zero probability* p_σ *such that*



(119)

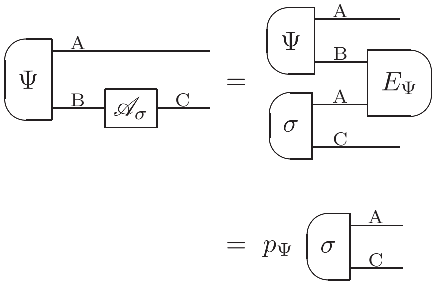
Corollary 20 (Existence of preparationally faithful pure states) *Let* $\Psi \in \mathfrak{S}_1(\mathrm{AB})$ *be the purification of an internal state* $\omega \in \mathfrak{S}_1(\mathrm{A})$*. Then,* Ψ *is preparationally faithful for system* B.

Proof. Let E_Ψ be the teleportation effect for Ψ, as defined in Corollary 19. Define the transformation $\mathscr{A}_\sigma$ as



(120)

Applying $\mathscr{A}_\sigma$ to Ψ and using Eq. (115) with $\mathrm{A}' \equiv \mathrm{A}$ we then obtain



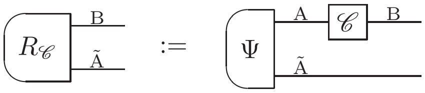
(121)

Hence, the thesis holds with $p_\sigma \equiv p_\Psi$ independently of σ. ■

B. Storing and probabilistic retrieving of transformations

Here we consider the task of storing an unknown transformation in the state of some system. The output state of such a storing protocol then becomes a "program" from which the transformation can be retrieved at later time. The task is achieved probabilistically by a machine that retrieves the transformation from the program and applies it on a new input state.

Corollary 21 (Storing and probabilistic retrieving) *Let $\Psi \in \mathfrak{S}_1(\mathrm{A}\tilde{\mathrm{A}})$ be a pure dynamically faithful state for system* A, *according to Def. 39. The* storing protocol, *consisting in the application of a transformation $\mathscr{C} \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ to the input state Ψ, as in the following diagram*

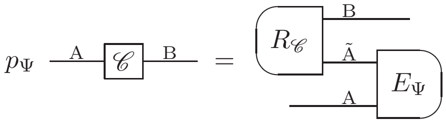


(122)

defines an injective map from transformations $\mathscr{C} \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ to bipartite states $R \in \mathfrak{S}(\mathrm{B}\tilde{\mathrm{A}})$ satisfying the property

$$(e|_{\mathrm{B}}\, |R)_{\mathrm{B}\tilde{\mathrm{A}}} \in D_{\tilde{\omega}}, \tag{123}$$

where $\tilde{\omega}$ is the marginal state $|\tilde{\omega})_{\tilde{\mathrm{A}}} = (e|_{\mathrm{A}}\, |\Psi)_{\mathrm{A}\tilde{\mathrm{A}}}$. The inverse map is given by the probabilistic retrieving protocol



(124)

where E_Ψ is the teleportation effect for state Ψ and p_Ψ is the corresponding teleportation probability, as defined in Eq. (115).

Proof. Since the state Ψ is dynamically faithful, the map $\mathscr{C} \mapsto R_\mathscr{C}$ is injective. Now, any transformation $\mathscr{C}$ is part of a test $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$, namely one has $\mathscr{C} = \mathscr{C}_{i_0}$ for some outcome i_0. Defining the coarse-grained channel $\mathscr{C}_\mathrm{X} := \sum_{i\in\mathrm{X}} \mathscr{C}_i$ we have

$$\begin{aligned}\sum_{i\in\mathrm{X}} (e|_{\mathrm{B}}\, |R_{\mathscr{C}_i})_{\mathrm{B}\tilde{\mathrm{A}}} &= \sum_{i\in\mathrm{X}} (e|_{\mathrm{B}}\, \mathscr{C}_i\, |\Psi)_{\mathrm{A}\tilde{\mathrm{A}}} \\ &= (e|_{\mathrm{B}}\, \mathscr{C}_\mathrm{X}\, |\Psi)_{\mathrm{A}\tilde{\mathrm{A}}} \\ &= (e|_{\mathrm{A}}\, |\Psi)_{\mathrm{A}\tilde{\mathrm{A}}} \\ &= |\tilde{\omega})_{\tilde{\mathrm{A}}}.\end{aligned} \tag{125}$$

having use the normalization condition $(e|_{\mathrm{B}}\, \mathscr{C}_\mathrm{X} = (e|_{\mathrm{A}}$. This implies $(e|_{\mathrm{B}}\, |R_\mathscr{C})_{\mathrm{B}\tilde{\mathrm{A}}}$ is in the refinement set of $\tilde{\omega}$, thus proving Eq. (123). The identity (124) simply follows by writing $p_\Psi \mathscr{C} = \mathscr{C} \circ (p_\Psi \mathscr{I}_\mathrm{A})$ and substituting $p_\Psi \mathscr{I}_\mathrm{A}$ as in Eq. (115). ■

In Section IX we will show that the correspondence $\mathscr{C} \mapsto R_\mathscr{C}$ is also surjective on the set of bipartite states satisfying Eq. (123). This will provide an isomorphism between transformations and bipartite states that enjoys all the structural properties of the Choi-Jamiołkowski isomorphism of quantum theory [42].

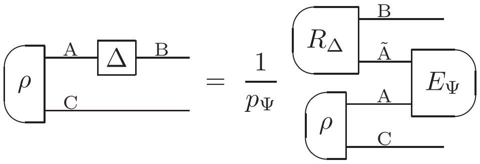
The probabilistic retrieving of Eq. (124) implies a bound on the operational distance between two transformations $\mathscr{A}_0, \mathscr{A}_1$ in terms the operational distance between the corresponding states:

Theorem 14 *Let $\mathscr{A}_0, \mathscr{A}_1 \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ be two transformations and $R_{\mathscr{A}_0}, R_{\mathscr{A}_1} \in \mathfrak{S}(\mathrm{B}\tilde{\mathrm{A}})$ be the corresponding states as in Eq. (122). Then one has the bound*

$$\|\mathscr{A}_1 - \mathscr{A}_0\|_{\mathrm{A},\mathrm{B}} \leq \frac{\|R_{\mathscr{A}_1} - R_{\mathscr{A}_0}\|_{\mathrm{B}\tilde{\mathrm{A}}}}{p_\Psi}, \tag{126}$$

where p_Ψ is the probability of retrieving a transformation from the corresponding state, as defined in Eq. (124).

Proof. Define $\Delta := \mathscr{A}_1 - \mathscr{A}_0$ and $R_\Delta := R_{\mathscr{A}_1} - R_{\mathscr{A}_0}$. Take an ancillary system C and a state $\rho \in \mathfrak{S}_1(\mathrm{AC})$. Then Eq. (124) implies



(127)

Applying a bipartite effect $(a|_{\mathrm{BC}}$ on both sides we then obtain

$$(a|_{\mathrm{BC}}\, \Delta\, |\rho)_{\mathrm{AC}} = \frac{(b_\rho|\, R_\Delta)_{\mathrm{B}\tilde{\mathrm{A}}}}{p_\Psi}, \tag{128}$$

where $(b_\rho|_{\mathrm{B}\tilde{\mathrm{A}}} := [(a|_{\mathrm{BC}} \otimes (E_\Psi|_{\tilde{\mathrm{A}}\mathrm{A}}]\, |\rho)_{\mathrm{AC}}$. Since b_ρ is an effect, the above equality implies the bound

$$\frac{\inf_b (b|\, R_\Delta)_{\mathrm{B}\tilde{\mathrm{A}}}}{p_\Psi} \leq (a|_{\mathrm{BC}}\, \Delta\, |\rho)_{\mathrm{AC}} \leq \frac{\sup_b (b|\, R_\Delta)_{\mathrm{B}\tilde{\mathrm{A}}}}{p_\Psi}. \tag{129}$$

By the definition of operational norm in Eq. 49, this implies

$$\|\Delta\rho\|_{\mathrm{BC}} \leq \|R_\Delta\|_{\mathrm{B\tilde{A}}}/p_\Psi. \tag{130}$$

Finally, taking the supremum over the ancillary system C we get the desired bound. ■

C. Systems of purifications and the link product

For every system A we now fix a dynamically faithful pure state $|\Psi^{(\mathrm{A})})_{\mathrm{A\tilde{A}}}$, where $\tilde{\mathrm{A}}$ is some suitable purifying system. According to the characterization of dynamically faithful pure states given in Theorem 9, the marginal state $|\omega)_{\mathrm{A}} := (e|_{\tilde{\mathrm{A}}} |\Psi^{(\mathrm{A})})_{\mathrm{A\tilde{A}}}$ must be internal. The role of the upper index in $\Psi^{(\mathrm{A})}$ is precisely to recall that the marginal is internal for system A, while it may not be internal for the purifying system $\tilde{\mathrm{A}}$. Moreover, we denote by $(E^{(\mathrm{A})}|_{\tilde{\mathrm{A}}\mathrm{A}}$ and by p_{A} the effect and the probability appearing in the teleportation scheme (115), respectively.

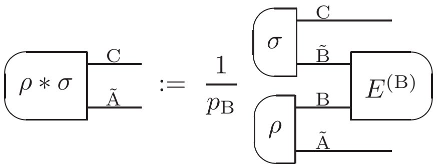
Note that since the product of dynamically faithful pure states is dynamically faithful (Corollary 13), for bipartite systems AB we can choose $|\Psi^{(\mathrm{AB})})_{\mathrm{AB}\widetilde{\mathrm{AB}}} := |\Psi^{(\mathrm{A})})_{\mathrm{A\tilde{A}}} |\Psi^{(\mathrm{B})})_{\mathrm{B\tilde{B}}}$. Likewise, we can choose $(E^{(\mathrm{AB})}|_{\widetilde{\mathrm{AB}}\mathrm{AB}} = (E^{(\mathrm{A})}|_{\tilde{\mathrm{A}}\mathrm{A}} (E^{(\mathrm{B})}|_{\tilde{\mathrm{B}}\mathrm{B}}$, and $p_{\mathrm{AB}} = p_{\mathrm{A}}p_{\mathrm{B}}$. We call a *system of purifications* such a choice of bipartite states and effects:

Definition 43 (System of purifications) *A* system of purification *is a choice of a dynamically faithful pure states* $|\Psi^{(\mathrm{A})})_{\mathrm{A\tilde{A}}}$ *and teleportation effects* $(E^{(\mathrm{A})}|_{\tilde{\mathrm{A}}\mathrm{A}}$ *that satisfies the properties*

$$\begin{aligned} \left|\Psi^{(\mathrm{AB})}\right)_{\mathrm{AB}\widetilde{\mathrm{AB}}} &= \left|\Psi^{(\mathrm{A})}\right)_{\mathrm{A\tilde{A}}} \left|\Psi^{(\mathrm{B})}\right)_{\mathrm{B\tilde{B}}} \\ \left(E^{(\mathrm{AB})}\right|_{\widetilde{\mathrm{AB}}\mathrm{AB}} &= \left(E^{(\mathrm{A})}\right|_{\tilde{\mathrm{A}}\mathrm{A}} \left(E^{(\mathrm{B})}\right|_{\tilde{\mathrm{B}}\mathrm{B}}. \end{aligned} \tag{131}$$

Once a system of purifications has been fixed, one can discuss the composition of transformations in terms of composition of states, generalizing the definitions and the results introduced by Refs. [33, 40] in the quantum setting.

Definition 44 (Link product) *The* link product *of two vectors* $\rho \in \mathfrak{S}_{\mathbb{R}}(\mathrm{B\tilde{A}})$ *and* $\sigma \in \mathfrak{S}_{\mathbb{R}}(\mathrm{C\tilde{B}})$ *is the vector* $\rho * \sigma \in \mathfrak{S}_{\mathbb{R}}(\mathrm{C\tilde{A}})$ *given by*



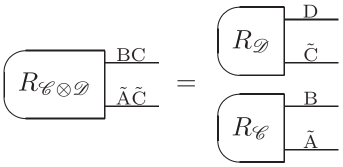
$\rho * \sigma$
C
$\tilde{\mathrm{A}}$
$:=$
$\frac{1}{p_{\mathrm{B}}}$
σ
C
$\tilde{\mathrm{B}}$
ρ
B
$\tilde{\mathrm{A}}$
$E^{(\mathrm{B})}$

(132)

Note that if ρ and σ are proportional to states, then also $\rho * \sigma$ is proportional to a state: one has $\rho * \sigma \in \mathfrak{S}_+(\mathrm{C\tilde{A}})$ for any couple $\rho \in \mathfrak{S}_+(\mathrm{B\tilde{A}}), \sigma \in \mathfrak{S}_+(\mathrm{C\tilde{B}})$.

The product and composition of transformations are then given by the following

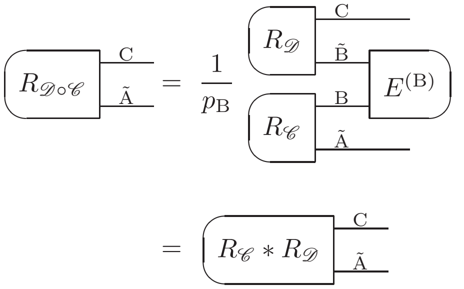
Corollary 22 (Composition of states) *Consider the correspondence given by the storing protocol in Eq. (122). For two transformations* $\mathscr{C} \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ *and* $\mathscr{D} \in \mathfrak{T}(\mathrm{C},\mathrm{D})$ *one has*



$R_{\mathscr{C}\otimes\mathscr{D}}$
BC
$\tilde{\mathrm{A}}\tilde{\mathrm{C}}$
=
$R_{\mathscr{D}}$
D
$\tilde{\mathrm{C}}$
$R_{\mathscr{C}}$
B
$\tilde{\mathrm{A}}$

(133)

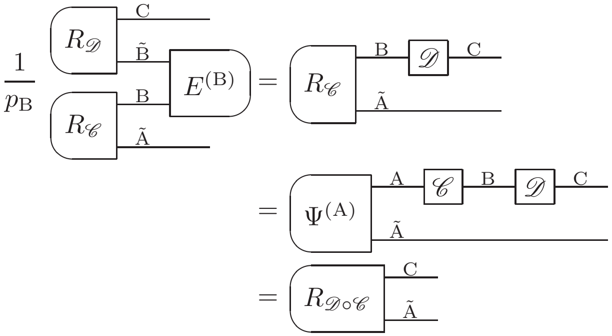
For two transformations $\mathscr{C} \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ *and* $\mathscr{D} \in \mathfrak{T}(\mathrm{B},\mathrm{C})$ *one has*



$R_{\mathscr{D}\circ\mathscr{C}}$
C
$\tilde{\mathrm{A}}$
=
$\frac{1}{p_{\mathrm{B}}}$
$R_{\mathscr{D}}$
C
$\tilde{\mathrm{B}}$
$R_{\mathscr{C}}$
B
$\tilde{\mathrm{A}}$
$E^{(\mathrm{B})}$
=
$R_{\mathscr{C}} * R_{\mathscr{D}}$
C
$\tilde{\mathrm{A}}$

(134)

Proof. The first equation follows from the fact that $|\Psi^{(\mathrm{AC})})_{\mathrm{AC}\widetilde{\mathrm{AC}}} := |\Psi^{(\mathrm{A})})_{\mathrm{A\tilde{A}}} |\Psi^{(\mathrm{C})})_{\mathrm{C\tilde{C}}}$, while the second follows from the probabilistic retrieving of Eq. (124):



$\frac{1}{p_{\mathrm{B}}}$
$R_{\mathscr{D}}$
C
$\tilde{\mathrm{B}}$
$R_{\mathscr{C}}$
B
$\tilde{\mathrm{A}}$
$E^{(\mathrm{B})}$
=
$R_{\mathscr{C}}$
B
$\mathscr{D}$
C
$\tilde{\mathrm{A}}$
=
$\Psi^{(\mathrm{A})}$
A
$\mathscr{C}$
B
$\mathscr{D}$
C
$\tilde{\mathrm{A}}$
=
$R_{\mathscr{D}\circ\mathscr{C}}$
C
$\tilde{\mathrm{A}}$

(135)

■

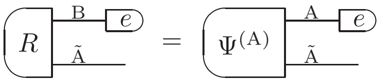
IX. DILATION OF PHYSICAL PROCESSES

In this Section we derive dilation Theorems for channels, observation-test, and general tests. These Theorems extend to all theories with purification the validity of the theorems by Stinespring [43], Naimark [44], and Ozawa [45], originally obtained in the setting of operator algebras.

A. Reversible dilation of channels

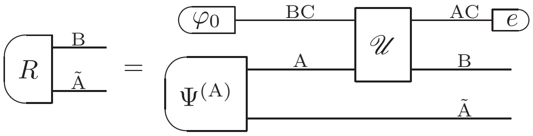
In order to derive the reversible dilation of a channel we need the following lemma:

Lemma 23 *Let $R \in \mathfrak{S}_1(\mathrm{B}\tilde{\mathrm{A}})$ be a state such that*



(136)

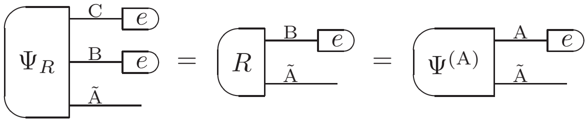
where $\Psi^{(\mathrm{A})}$ is a pure dynamically faithful state for system A. *Then there exist a system* C, *a pure state $\varphi_0 \in \mathfrak{S}_1(\mathrm{BC})$, and a reversible channel $\mathscr{U} \in \mathfrak{T}(\mathrm{ABC})$ such that*



(137)

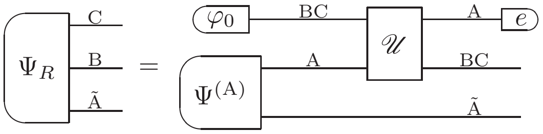
Moreover, the channel $\mathscr{V} \in \mathfrak{T}(\mathrm{A}, \mathrm{ABC})$ defined by $\mathscr{V} := \mathscr{U}\,|\varphi_0)_{\mathrm{BC}}$ is unique up to reversible channels on AC.

Proof. Take a purification of R, say $\Psi_R \in \mathfrak{S}_1(\mathrm{CB}\tilde{\mathrm{A}})$ for some purifying system C. One has



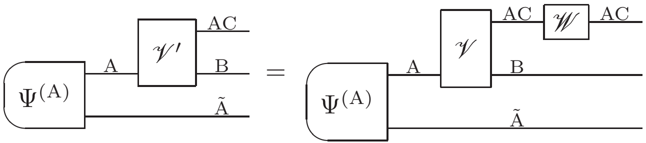
(138)

that is, the pure states Ψ_R and $\Psi^{(\mathrm{A})}$ have the same marginal on system $\tilde{\mathrm{A}}$. Applying the uniqueness of purification as expressed by Lemma 21 one then obtains



(139)

Applying the deterministic effect on system C on both sides, one then proves Eq. (137). Moreover, if $\mathscr{V}' := \mathscr{U}'\,|\varphi_0')_{\mathrm{BC}}$ is channel such that Eq. (137) holds, then the pure states $\mathscr{V}\,\left|\Psi^{(\mathrm{A})}\right)_{\mathrm{A}\tilde{\mathrm{A}}}$ and $\mathscr{V}'\,\left|\Psi^{(\mathrm{A})}\right)_{\mathrm{A}\tilde{\mathrm{A}}}$ have the same marginal on system $\mathrm{B}\tilde{\mathrm{A}}$. Uniqueness of purification then implies

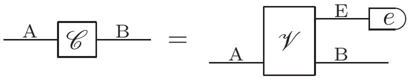


(140)

for some reversible channel $\mathscr{W} \in \mathfrak{T}(\mathrm{AC})$. Since $\Psi^{(\mathrm{A})}$ is dynamically faithful for A, this implies $\mathscr{V}' = \mathscr{W}\mathscr{V}$. ■

We now give the definitions of dilation, environment, and reversible dilation:

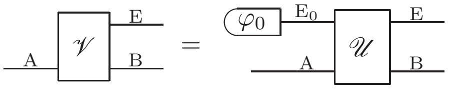
Definition 45 (Dilation of a channel) *A* dilation *of channel $\mathscr{C} \in \mathfrak{T}(\mathrm{A}, \mathrm{B})$ is a channel $\mathscr{V} \in \mathfrak{T}(\mathrm{A}, \mathrm{BE})$ such that*



(141)

We refer to system E *as to the* environment.

Definition 46 (Reversible dilation) *A dilation $\mathscr{V} \in \mathfrak{T}(\mathrm{A}, \mathrm{BE})$ is called* reversible *if there exists a system* E_0 *such that* $\mathrm{AE}_0 \simeq \mathrm{BE}$ *and*

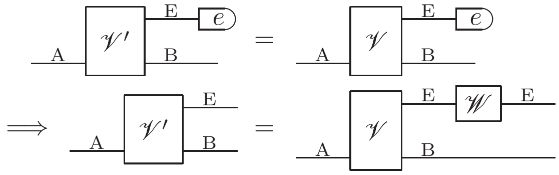


(142)

for some pure state $\varphi_0 \in \mathfrak{S}_1(\mathrm{E}_0)$ and some reversible channel $\mathscr{U} \in \mathfrak{T}(\mathrm{AE}_0, \mathrm{BE})$.

According to the above definitions, we have the following dilation theorem:

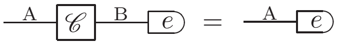
Theorem 15 (Reversible dilation of channels) *Every channel $\mathscr{C} \in \mathfrak{T}(\mathrm{A}, \mathrm{B})$ has a reversible dilation $\mathscr{V} \in \mathfrak{T}(\mathrm{A}, \mathrm{BE})$. If $\mathscr{V}, \mathscr{V}' \in \mathfrak{T}(\mathrm{A}, \mathrm{BE})$ are two reversible dilations of the same channel, then they are connected by a reversible transformation on the environment, namely*



(143)

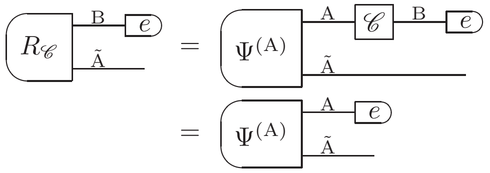
for some reversible channel $\mathscr{W} \in \mathbf{G}_{\mathrm{E}}$.

Proof. Let us store the channel $\mathscr{C}$ in the faithful state $\Psi^{(\mathrm{A})} \in \mathfrak{S}_1(\mathrm{A}\tilde{\mathrm{A}})$, thus getting the state $R_{\mathscr{C}}$, as in Eq. (122). Since $\mathscr{C}$ is a channel, it satisfies the normalization condition



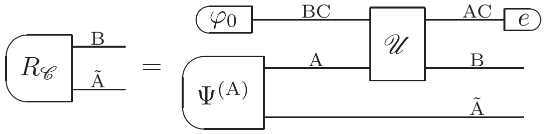
(144)

which implies



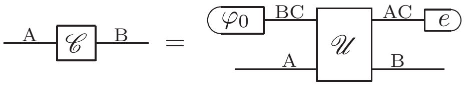
(145)

Now, applying Lemma 23 we obtain



(146)

Since $\Psi^{(\mathrm{A})}$ is dynamically faithful for system A, this implies

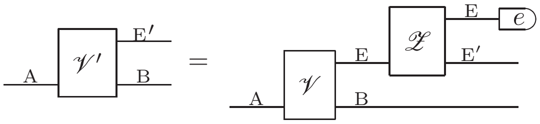


(147)

Therefore, $\mathscr{V} := \mathscr{U}\,|\varphi_0)_{\mathrm{BC}}$ is a reversible dilation of $\mathscr{C}$, with $\mathrm{E}_0 := \mathrm{BC}$ and $\mathrm{E} := \mathrm{AC}$. Finally, the uniqueness clause in Lemma 23 implies uniqueness of the dilation. ■

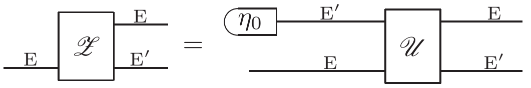
Moreover, two reversible dilations of the same channel with different environments are related as follows:

Lemma 24 *Let $\mathscr{V} \in \mathfrak{T}(\mathrm{A},\mathrm{BE})$ and $\mathscr{V}' \in \mathfrak{T}(\mathrm{A},\mathrm{BE}')$ be two reversible dilations of the same channel $\mathscr{C}$, with generally different environments* E *and* E′. *Then there is a channel $\mathscr{Z}$ from* E *to* EE′ *such that*



(148)

The channel $\mathscr{Z}$ has the form



(149)

for some pure state $\eta_0 \in \mathfrak{S}_1(\mathrm{E}')$ and some reversible transformation $\mathscr{U} \in \mathfrak{T}(\mathrm{EE}')$.

Proof. Apply $\mathscr{V}$ and $\mathscr{V}'$ to the faithful state $\Phi^{(\mathrm{A})}$, and then use the uniqueness of purification stated in Lemma 21. ■

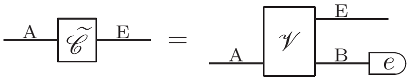
The above results represent the general version—holding in all probabilistic theories with purification—of the dilation scheme implied by Stinespring's Theorem [43] in quantum theory. However, differently from the proof of Stinespring's Theorem, the present proof does not require any C*-algebraic structure, being based just on the purification postulate. In fact, it is easy to see that the purification of states and the reversible dilation of channels are equivalent features, in the following sense:

Corollary 23 (Equivalence between purification and reversible dilation) *Existence and uniqueness (up to reversible channels on the purifying system) of the purification of states is equivalent to existence and uniqueness (up to reversible channels on the environment) of the reversible dilation of channels.*

Proof. The direction "purification ⇒ dilation" has been just proved by the dilation theorem. The converse is obvious, since a normalized state $\rho \in \mathfrak{S}_1(\mathrm{B})$ is a special case of channel from the trivial system I to B, and in this special case purification coincides with dilation. ■

Finally, the reversible dilation of a channel allows one to define the *complementary channel* as follows

Definition 47 (Complementary channel) *Let $\mathscr{V} \in \mathfrak{T}(\mathrm{A},\mathrm{BE})$ be a reversible dilation of channel $\mathscr{C} \in \mathfrak{T}(\mathrm{A},\mathrm{B})$, as in Theorem 15. The* complementary channel *of $\mathscr{C}$ is the channel $\widetilde{\mathscr{C}} \in \mathfrak{T}(\mathrm{A},\mathrm{E})$ defined by*



(150)

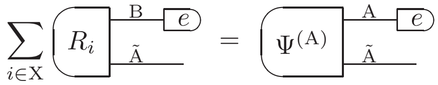
Note that the complementary channel $\widetilde{\mathscr{C}}$ is unique up to reversible transformations on the environment E.

The notion of complementary channel has played a crucial role in the research about capacity of quantum information channels (see e.g. [50–52]) and we expect that having the same definition in general probabilistic theories will be very fruitful (in fact, a number of consequences is already presented in the Section XI).

B. Reversible dilation of tests

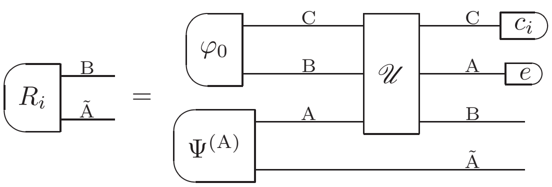
We now generalize the dilation of channels (i.e. single-outcome tests) to the case of arbitrary tests. For this purpose, we need the analogue of Lemma 23:

Lemma 25 *Let $\{R_i\}_{i\in\mathrm{X}}$ be a preparation-test for system $\mathrm{B}\tilde{\mathrm{A}}$ with the property*



(151)

where $\Psi^{(\mathrm{A})}$ is the purification of an internal state of system A. *Then, there exists a system* C, *a pure state $\varphi_0 \in \mathfrak{S}_1(\mathrm{BC})$, a reversible channel $\mathscr{U} \in \mathfrak{T}(\mathrm{ABC})$, and an observation-test $\{c_i\}_{i\in\mathrm{X}}$ on* C *such that*



(152)

for any outcome $i \in \mathrm{X}$. By suitably choosing system C, *the observation-test $\{c_i\}_{i\in\mathrm{X}}$ can be taken to be a discriminating test.*

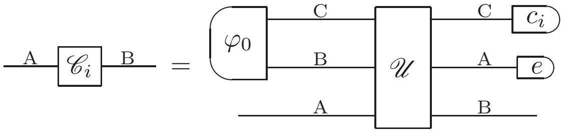
Proof. Take a purification of the coarse-grained state $R = \sum_i R_i$, say $\Psi_R \in \mathfrak{S}_1(\mathrm{CB}\tilde{\mathrm{A}})$ for some purifying system C. According to Theorem 6, there is an observation-test $\{c_i\}_{i\in\mathrm{X}}$ on C such that

$$|R_i)_{\mathrm{B}\tilde{\mathrm{A}}} = (c_i|_{\mathrm{C}}\,|\Psi_R)_{\mathrm{CB}\tilde{\mathrm{A}}} \qquad \forall i \in \mathrm{X}, \tag{153}$$

and, by suitably choosing C, $\{c_i\}$ can be chosen to be a discriminating test. Following the same line of Lemma 23 we then obtain the thesis. ■

Following the proof of the reversible dilation of channels given in Theorem 15 we have the following

Theorem 16 (Reversible dilation of tests) *For every test $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ from system* A *to system* B *there exist a system* C, *a pure state $\varphi_0 \in \mathfrak{S}_1(\mathrm{BC})$, a reversible channel $\mathscr{U} \in \mathfrak{T}(\mathrm{ABC})$, and an observation-test $\{c_i\}_{i\in\mathrm{X}}$ on* C *such that for all outcomes $i \in \mathrm{X}$*



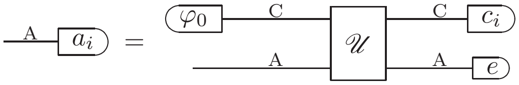
(154)

By suitably choosing system C, *the observation-test $\{c_i\}_{i\in\mathrm{X}}$ can be taken to be a discriminating test.*

In the case we choose the observation-test $\{c_i\}_{i\in\mathrm{X}}$ to be discriminating, the above Theorem yields a (simplified) version of Ozawa's Theorem in quantum theory [45]. Here the simplification comes from the fact that we consider finite dimensional state spaces and tests with finite outcomes, whereas the challenging part of Ozawa's Theorem is the rigorous treatment of infinite dimension and continuous spectrum.

Moreover, we can apply the dilation theorem to tests with trivial output $\mathrm{B}\equiv\mathrm{I}$, thus obtaining the operational version of Naimark's Theorem [44] in the finite-outcome case:

Corollary 24 (Discriminating dilation of observation-tests) *For every observation-test* $\{a_i\}_{i\in\mathrm{X}}$ *on* A *there exists a system* C, *a pure state* $\varphi_0\in\mathfrak{S}_1(\mathrm{C})$, *a reversible channel* $\mathscr{U}\in\mathfrak{T}(\mathrm{AC})$, *and a discriminating test* $\{c_i\}_{i\in\mathrm{X}}$ *on* C *such that*



(155)

for all outcomes $i\in\mathrm{X}$.

Another corollary is the following:

Corollary 25 (Characterization of theories with purification) *In a theory with purification every test can be realized using only pure states, reversible transformations, and discriminating tests.*

In fact, only one pure state for each system is enough, since due to Corollary 20 all pure states can be obtained from a fixed one by acting with reversible channels.

X. STATES-TRANSFORMATIONS ISOMORPHISM

The results of the previous Section allow a complete identification of transformations with bipartite states, thus providing the general version of the Choi-Jamiołkowski isomorphism [46, 47] in quantum theory. The correspondence is summarized in the following

Theorem 17 (States-transformations isomorphism) *The storing map* $\mathscr{C}\mapsto R_{\mathscr{C}}:=\mathscr{C}\left|\Psi^{(\mathrm{A})}\right)_{\mathrm{A}\tilde{\mathrm{A}}}$, *where* $\left|\Psi^{(\mathrm{A})}\right)_{\mathrm{A}\tilde{\mathrm{A}}}$ *is a pure dynamically faithful state for system* A, *has the following properties:*

1. *it defines a bijective correspondence between tests* $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ *from* A *to* B *and preparation-tests* $\{R_i\}_{i\in\mathrm{X}}$ *for* $\mathrm{B}\tilde{\mathrm{A}}$ *satisfying*

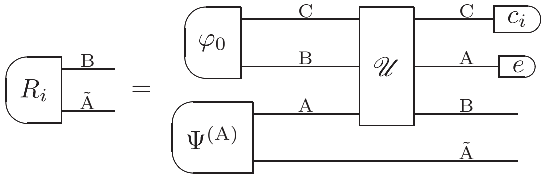
$$\sum_{i\in\mathrm{X}}(e|_{\mathrm{B}}\,|R_i)_{\mathrm{B}\tilde{\mathrm{A}}}=(e|_{\mathrm{A}}\left|\Psi^{(\mathrm{A})}\right)_{\mathrm{A}\tilde{\mathrm{A}}}. \tag{156}$$

2. *a transformation* $\mathscr{C}$ *is atomic (according to Definition 22) if and only if the corresponding state* $R_{\mathscr{C}}$ *is pure.*

3. *in convex theory the map* $\mathscr{C}\mapsto R_{\mathscr{C}}$ *defines a bijective correspondence between transformations* $\mathscr{C}\in\mathfrak{T}(\mathrm{A},\mathrm{B})$ *and bipartite states* $R\in\mathfrak{S}(\mathrm{B}\tilde{\mathrm{A}})$ *satisfying the property*

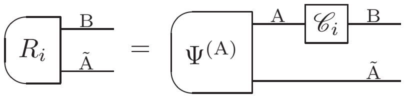
$$(e|_{\mathrm{B}}\,|R)_{\mathrm{B}\tilde{\mathrm{A}}}\in D_{\tilde\omega}\qquad |\tilde\omega)_{\tilde{\mathrm{A}}}=(e|_{\mathrm{A}}\left|\Psi^{(\mathrm{A})}\right)_{\mathrm{A}\tilde{\mathrm{A}}}. \tag{157}$$

Proof. Let us start from the proof of item 1. One direction is obvious: if $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ is a test from A to B, it must satisfy the normalization condition $\sum_{i\in\mathrm{X}}(e|_{\mathrm{B}}\mathscr{C}_i=(e|_{\mathrm{A}}$ (see Eq. (39)). The preparation-test $\{R_{\mathscr{C}_i}\}_{i\in\mathrm{X}}$ defined by $|R_{\mathscr{C}_i})_{\mathrm{B}\tilde{\mathrm{A}}}=\mathscr{C}_i\left|\Psi^{(\mathrm{A})}\right)_{\mathrm{A}\tilde{\mathrm{A}}}$ satisfies the property $\sum_{i\in\mathrm{X}}(e|_{\mathrm{B}}\,|R_{\mathscr{C}_i})_{\mathrm{B}\tilde{\mathrm{A}}}=\sum_{i\in\mathrm{X}}(e|_{\mathrm{B}}\mathscr{C}_i\left|\Phi^{(\mathrm{A})}\right)_{\mathrm{A}\tilde{\mathrm{A}}}=(e|_{\mathrm{A}}\left|\Psi^{(\mathrm{A})}\right)_{\mathrm{A}\tilde{\mathrm{A}}}$, that is, it satisfies Eq. (156). Moreover, if two tests $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ and $\{\mathscr{C}'_i\}_{i\in\mathrm{X}}$ satisfy $R_{\mathscr{C}_i}=R_{\mathscr{C}'_i}$ for all $i\in\mathrm{X}$, then by injectivity of the map $\mathscr{C}\mapsto R_{\mathscr{C}}$ (proved in Corollary 21), one has $\mathscr{C}_i=\mathscr{C}'_i$ for all $i\in\mathrm{X}$. Conversely, suppose that $\{R_i\}_{i\in\mathrm{X}}$ is a preparation-test satisfying Eq. (156). Then, by Lemma 25 there is a a system C, a pure state $\varphi_0\in\mathfrak{S}_1(\mathrm{BC})$, a reversible channel $\mathscr{U}\in\mathfrak{T}(\mathrm{ABC})$, and an observation-test $\{c_i\}_{i\in\mathrm{X}}$ on C such that for every outcome $i\in\mathrm{X}$ one has



(158)

Defining the test $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ by $\mathscr{C}_i:=(c_i|_{\mathrm{C}}\,(e|_{\mathrm{A}}\,\mathscr{U}\,|\varphi_0)_{\mathrm{BC}}$, we then obtain



(159)

This completes the proof of item 1. Item 2 is an immediate consequence of the item 1: If $\mathscr{C}$ is atomic, then $R_{\mathscr{C}}$ must be pure, otherwise we would have a non-trivial decomposition of $\mathscr{C}$. Vice-versa, if $R_{\mathscr{C}}$ is pure, then $\mathscr{C}$ must be atomic, otherwise we would have a non-trivial decomposition of $R_{\mathscr{C}}$. Regarding item 3, injectivity was already established in Corollary 21. To prove surjectivity, suppose that $R\in\mathfrak{S}(\mathrm{B}\tilde{\mathrm{A}})$ is such that $(e|_{\mathrm{B}}\,|R)_{\mathrm{B}\tilde{\mathrm{A}}}$ is in the refinement set of $|\tilde\omega)_{\tilde{\mathrm{A}}}$. This means that there is a preparation-test $\{\tilde\omega_i\}_{i\in\mathrm{X}}$ such that $\tilde\omega=\sum_{i\in\mathrm{X}}\tilde\omega_i$ and $(e|_{\mathrm{B}}\,|R)_{\mathrm{B}\tilde{\mathrm{A}}}=|\tilde\omega_{i_0})_{\tilde{\mathrm{A}}}$ for some outcome i_0. Now choose an arbitrary set of normalized states $\{\rho_i\}_{i\in\mathrm{X}}\subset\mathfrak{S}_1(\mathrm{B})$ and consider the collection of states $\{R_i\}_{i\in\mathrm{X}}$ defined as follows: $R_{i_0}=R$, $R_i=\rho_i\otimes\tilde\omega_i$ for $i\neq i_0$. Because the theory is convex the collection of states $\{R_i\}_{i\in\mathrm{X}}$ is a preparation-test (it can be obtained by randomization of the normalized states $\bar R_i=R_i/\,(e|\,R_i)_{\mathrm{B}\tilde{\mathrm{A}}}$ with probabilities $p_i=(e|\,R_i)_{\mathrm{B}\tilde{\mathrm{A}}})$. Moreover, it clearly satisfies Eq. (156). Therefore, using item 1 we see that there exists a test $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ from A to B such that $R_i=R_{\mathscr{C}_i}$. In particular, $R=R_{i_0}=R_{\mathscr{C}_{i_0}}$, thus proving surjectivity. ■

Clearly, the correspondence $\mathscr{C} \mapsto R_{\mathscr{C}}$ can be extended via linear combinations to an injective linear map between the vector spaces $\mathfrak{T}_{\mathbb{R}}(\mathrm{A},\mathrm{B})$ and $\mathfrak{S}_{\mathbb{R}}(\mathrm{B}\tilde{\mathrm{A}})$.

An immediate consequence of the states-transformations isomorphism is the following

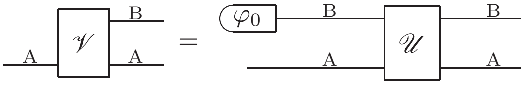
Corollary 26 (Existence of an ultimate refinement) *In a convex theory with purification, every test* $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ *from* A *to* B *admits an ultimate refinement* $\{\mathscr{D}_j\}_{j\in\mathrm{Y}}$ *where every transformation* $\mathscr{D}_j$ *is atomic.*

Proof. Consider the preparation-test $\{R_{\mathscr{C}_i}\}_{i\in\mathrm{X}}$ and take the normalized states $\overline{R}_{\mathscr{C}_i} = R_{\mathscr{C}_i}/\left(e\right|R_{\mathscr{C}_i})_{\mathrm{B}\tilde{\mathrm{A}}}$. Since the states form a finite-dimensional compact convex set, each state $\overline{R}_{\mathscr{C}_i}$ has a convex decomposition on a finite number of pure states. Collecting together all these decompositions yields a preparation-test $\{R_j\}_{j\in\mathrm{Y}}$, containing only pure states, that refines $\{R_{\mathscr{C}_i}\}_{i\in\mathrm{X}}$. By the states-transformations isomorphism, one has $R_j = R_{\mathscr{D}_j}$, for a test $\{\mathscr{D}_j\}_{j\in\mathrm{Y}}$ that refines $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ and contains only atomic transformations. ■

A. First consequences of the isomorphism

Two simple consequences of the states-transformations isomorphism are the following:

Corollary 27 *A channel* $\mathscr{V}$ *from* A *to* AB *is atomic if and only if it is of the form*



(160)

for some pure state $\varphi_0 \in \mathfrak{S}_1(\mathrm{B})$ *and some reversible channel* $\mathscr{U} \in \mathbf{G}_{\mathrm{AB}}$.

Proof. Clearly a channel of the form $\mathscr{V} = \mathscr{U}\,|\varphi_0)_{\mathrm{B}}$ is atomic, since the corresponding state $R_{\mathscr{C}} = \mathscr{U}\,\left|\Psi^{(\mathrm{A})}\right)_{\mathrm{A}\tilde{\mathrm{A}}}|\varphi_0)_{\mathrm{B}}$ is pure. Conversely, if $\mathscr{V}$ is atomic, then $R_{\mathscr{V}}$ is a purification of the state $|\tilde{\omega})_{\tilde{\mathrm{A}}} := (e|_{\mathrm{A}}\left|\Psi^{(A)}\right)_{\mathrm{A}\tilde{\mathrm{A}}}$. Since $R_{\mathscr{V}}$ and $\Psi^{(\mathrm{A})}$ are both purifications of the same state, by the uniqueness of purification stated by Lemma 21 we have $R_{\mathscr{V}} = \mathscr{U}\,\left|\Psi^{(\mathrm{A})}\right)_{\mathrm{A}\tilde{\mathrm{A}}}|\varphi_0)_{\mathrm{B}}$ for some pure state $\varphi_0 \in \mathfrak{S}_1(\mathrm{B})$ and some reversible channel $\mathscr{U} \in \mathbf{G}_{\mathrm{AB}}$. Since $\Psi^{(\mathrm{A})}$ is dynamically faithful for system A, this implies $\mathscr{V} = \mathscr{U}\,|\varphi_0)_{\mathrm{B}}$. ■

When system B is trivial, we have the more specific result:

Corollary 28 *A channel from* A *to* A *is atomic if and only if it is reversible.*

Proof. Special case of Corollary 27 with $\mathrm{B} \equiv \mathrm{I}$. ■

The states-transformations isomorphism also allows one to prove that the sets of transformations, channels, reversible channels, and pure states are compact with respect to the operational norm induced by optimal discrimination:

Corollary 29 *The set of physical transformations* $\mathfrak{T}(\mathrm{A},\mathrm{B})$ *is compact in the operational norm.*

Proof. By Theorem 17, we have $\dim(\mathfrak{T}_{\mathbb{R}}(\mathrm{A},\mathrm{B})) \le \dim(\mathfrak{S}_{\mathbb{R}}(\mathrm{B}\tilde{\mathrm{A}}))$, namely transformations span a finite-dimensional vector space. Since we are in finite dimensions, to prove compactness it is enough to prove that the set of transformations is closed. To see this, suppose that $\{\mathscr{C}_n\}$ is a Cauchy sequence of transformations. By definition, each transformation $\mathscr{C}_n$ arises in some test, which can be taken to be binary without loss of generality. Let $\{\mathscr{C}_n,\mathscr{D}_n\}$ be such a binary test, and let $\{R_{\mathscr{C}_n},R_{\mathscr{D}_n}\}$ be the corresponding preparation-test. Since the set of all states $\mathfrak{S}(\mathrm{B}\tilde{\mathrm{A}})$ is compact (by hypothesis it is finite dimensional and closed), there is a subsequence $\{R_{\mathscr{C}_{n_k}},R_{\mathscr{D}_{n_k}}\}$ converging to a binary preparation-test $\{R_0,R_1\}$. Now, since each test $\{R_{\mathscr{C}_{n_k}},R_{\mathscr{D}_{n_k}}\}$ satisfies Eq. (156), also $\{R_0,R_1\}$ satisfies it. By the states-transformations isomorphism, this implies that there is a binary test $\{\mathscr{C},\mathscr{D}\}$ such that $R_0 = R_{\mathscr{C}}$ and $R_1 = R_{\mathscr{D}}$. Finally, using the bound of Eq. (126) we see that $\mathscr{C}_{n_k}$ (and hence $\mathscr{C}_n$) converges to $\mathscr{C}$ in the operational norm. ■

Corollary 30 *The set of channels from* A *to* B *is compact in the operational norm.*

Proof. Again, since we are in finite dimension, it is enough to prove that the set of channels is closed. Let $\{\mathscr{C}_n\}$ be a Cauchy sequence of channels. Since the set of transformations is closed, the sequence converges to some transformation $\mathscr{C}$. Moreover, $\mathscr{C}$ is a channel. Indeed, since $\mathscr{C}_n$ is a channel we have $(e|_{\mathrm{B}}\mathscr{C}_n = (e|_{\mathrm{A}}$, and, for every state ρ, $(e|_{\mathrm{B}}\mathscr{C}\,|\rho)_{\mathrm{A}} = \lim_{n\to\infty}(e|_{\mathrm{B}}\mathscr{C}_n\,|\rho)_{\mathrm{A}} = (e|\,\rho)_{\mathrm{A}}$, which implies $(e|_{\mathrm{B}}\mathscr{C} = (e|_{\mathrm{A}}$. ■

Corollary 31 *The group* $\mathbf{G}_{\mathrm{A}}$ *of all reversible transformations of system* A *is a compact Lie group.*

Proof. Let $\{\mathscr{U}_n\}$ be a sequence of reversible channels converging to some channel $\mathscr{C}$. We now show that $\mathscr{C}$ is reversible. Indeed, consider the sequence $\{\mathscr{U}_n^{-1}\}$. Since the set of channels is compact, it is possible to choose a subsequence $\{\mathscr{U}_{n_k}^{-1}\}$ that converges to some channel $\mathscr{D}$. But now we have $\mathscr{C}\mathscr{D} = \lim_{k\to\infty}\mathscr{U}_{n_k}\mathscr{U}_{n_k}^{-1} = \mathscr{I}_{\mathrm{A}}$, and, $\mathscr{D}\mathscr{C} = \lim_{k\to\infty}\mathscr{U}_{n_k}^{-1}\mathscr{U}_{n_k} = \mathscr{I}_{\mathrm{A}}$ [48], that is, $\mathscr{C}$ is reversible and $\mathscr{D} = \mathscr{C}^{-1}$. This proves that $\mathbf{G}_{\mathrm{A}}$ is closed, and, therefore, compact. Finally, since $\mathbf{G}_{\mathrm{A}}$ is compact and has a faithful finite-dimensional matrix representation, it is a Lie group (see e.g. Theorem 5.13 of Ref. [49]). ■

Corollary 32 *The set of pure states of system* A *is compact.*

Proof. Let $\{\varphi_n\}$ be a sequence of pure states converging to some state ρ. We now prove that ρ is pure. To see this, let us fix a pure state φ_0. By Lemma 20 for every n there is a reversible channel $\mathscr{U}_n$ such that $\varphi_n = \mathscr{U}_n\varphi_0$. Since the group $\mathbf{G}_\mathrm{A}$ is compact, we can take a subsequence $\{\mathscr{U}_{n_k}\}$ that converges to a reversible channel $\mathscr{U}$. Therefore we have $\rho = \lim_{n\to\infty}\varphi_n = \lim_{k\to\infty}\mathscr{U}_{n_k}\varphi_0 = \mathscr{U}\varphi_0$. Since ρ is connected to a pure state by a reversible channel it must be pure. ■

We conclude this Subsection with two results that will be useful in the construction of deterministic teleportation:

Corollary 33 (Existence of a twirling test) *In a (convex) theory with purification there always exists a twirling test $\{p_i\mathscr{U}_i\}_{i\in\mathrm{X}}$ (according to Def. 15), where $\{p_i\}$ are probabilities and $\{\mathscr{U}_i\}$ are reversible channels. In particular, one of the channels $\{\mathscr{U}_i\}$ can be always chosen to be the identity.*

Proof. Let $\mathrm{d}\mathscr{W}$ be the normalized Haar measure over the compact group $\mathbf{G}_\mathrm{A}$, and define the channel $\mathscr{T} := \int \mathrm{d}\mathscr{W}\ \mathscr{W}$, which is clearly a twirling channel, since by invariance of the Haar measure one has $\mathscr{U}\mathscr{T} = \mathscr{T}$ for every $\mathscr{U} \in \mathbf{G}_\mathrm{A}$. Since the reversible channels span a finite-dimensional space, their convex hull is a finite-dimensional convex set. Then by Caratheodory's theorem the integral can be written as a finite convex combination of reversible transformations, i.e. $\mathscr{T} = \sum_{i\in\mathrm{X}} p_i\mathscr{U}_i$. Since $\mathscr{U}\mathscr{T} = \mathscr{T}$, we can pick an outcome i_0, and apply $\mathscr{U}_{i_0}^{-1}$, thus obtaining a new twirling test where one channel is the identity. ■

Corollary 34 (Uniqueness of the invariant state) *For every system* A, *there is a unique state χ_A invariant under all reversible transformations in $\mathbf{G}_\mathrm{A}$. Moreover, χ_A is internal.*

Proof. Let $\mathscr{T}$ be the twirling channel defined in the previous Corollary. Since for two arbitrary pure states ψ, ψ' there is a reversible channel $\mathscr{U}$ such that $\psi' = \mathscr{U}\psi$ (Lemma 20), this implies

$$\mathscr{T}(\psi') = \int \mathrm{d}\mathscr{W}\ \mathscr{W}\mathscr{U}\psi = \int \mathrm{d}\mathscr{W}'\ \mathscr{W}'\psi = \mathscr{T}(\psi) := \chi, \tag{161}$$

having used the invariance of the Haar measure. Now, since the twirling channel is constant on pure states, it is constant on every state, namely $\mathscr{T}(\rho) = \chi$ for every ρ. In particular, if ρ is an invariant state, then we have $\rho = \mathscr{T}(\rho) = \chi$. This proves that the invariant state is unique. Finally, Corollary 33 implies that the integral $\mathscr{T}(\rho)$ can be written as the sum of the transformations of a twirling test containing the identity, namely

$$\chi = \mathscr{T}(\rho) = \sum_j p_j\mathscr{U}_j\rho = p_{i_0}\rho + \sum_{j\neq i_0}\mathscr{U}_0^{-1}\mathscr{U}_j\rho \tag{162}$$

whence $p_{i_0}\rho$ belongs to the refinement set D_χ of χ for every state ρ. This proves that χ is internal. ■

B. Entanglement breaking channels

An interesting consequence of the states-transformations isomorphism regards the identification of *measure-and-prepare channels* and *entanglement breaking channels* , the latter defined as follows

Definition 48 (Entanglement-breaking channel) *A channel $\mathscr{C} \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ is* entanglement breaking *if the output state $\mathscr{C}\,|\sigma)_\mathrm{AC}$ is separable for every state $\sigma \in \mathfrak{S}_1(\mathrm{AC})$, namely*

$$\mathscr{C}\,|\sigma)_\mathrm{AC} = \sum_{i\in\mathrm{X}} p_i\,|\beta_i)_\mathrm{B}\,|\tilde\rho_i)_\mathrm{C}, \tag{163}$$

for some separable preparation-test $\{p_i\rho_i \otimes \tilde\rho_i\}_{i\in\mathrm{X}}$, $\beta_i \in \mathfrak{S}_1(\mathrm{B}), \tilde\rho_i \in \mathfrak{S}_1(\tilde{\mathrm{A}})$.

The following Theorem extends to arbitrary theories with purification the characterization of entanglement breaking channels presented in quantum theory by Horodecki, Shor, and Ruskai in Ref. [53]:

Corollary 35 (Structure of entanglement-breaking channels) *In a theory with purification, the following are equivalent*

1. *$\mathscr{C}$ is entanglement-breaking*
2. *$R_\mathscr{C}$ is separable*
3. *$\mathscr{C}$ is measure-and-prepare*

Proof. *(1) ⇒ (2)* If $\mathscr{C}$ is entanglement-breaking, then in particular $|R_\mathscr{C})_{\mathrm{B}\tilde{\mathrm{A}}} = \mathscr{C}\left|\Psi^{(\mathrm{A})}\right)_{\mathrm{A}\tilde{\mathrm{A}}}$ is separable. *(2) ⇒ (3)* Suppose that $R_\mathscr{C}$ is separable, namely $R_\mathscr{C} = \sum_{i\in\mathrm{X}} p_i\beta_i \otimes \tilde\rho_i$ for some separable preparation-test $\{p_i\beta_i \otimes \tilde\rho_i\}_{i\in\mathrm{X}}$ (with $\beta_i \in \mathfrak{S}_1(\mathrm{B})$ and $\tilde\rho_i \in \mathfrak{S}_1(\tilde{\mathrm{A}})$). Now, the preparation-test $\{p_i\tilde\rho_i\}_{i\in\mathrm{X}}$ has the property

$$\sum_i p_i\tilde\rho_i = (e|_\mathrm{B}\,|R_\mathscr{C})_{\mathrm{B}\tilde{\mathrm{A}}} = (e|_\mathrm{A}\left|\Psi^{(\mathrm{A})}\right) := |\tilde\chi)_{\tilde{\mathrm{A}}}, \tag{164}$$

having used that $|R_\mathscr{C})_{\mathrm{B}\tilde{\mathrm{A}}} := \mathscr{C}\left|\Psi^{(\mathrm{A})}\right)_{\mathrm{A}\tilde{\mathrm{A}}}$, and the fact that $\mathscr{C}$ is a channel. Applying the first item of Theorem 17 with $\mathrm{B} \equiv \mathrm{I}$, we then deduce that $p_i\tilde\rho_i = R_{a_i}$ for some suitable observation-test $\{a_i\}$ on A. Considering the measure-and-prepare channel $\mathscr{D} := \sum_{i\in\mathrm{X}} |\beta_i)_\mathrm{B}\,(a_i|_\mathrm{A}$ we then obtain $R_\mathscr{D} = R_\mathscr{C}$, which implies $\mathscr{C} = \mathscr{D}$. Hence, $\mathscr{C}$ is measure-and-prepare. *(3) ⇒ (1)* If $\mathscr{C}$ is measure-and-prepare, it is easily seen that it is entanglement-breaking. ■

C. Completeness of theories with purification

As a consequence of the states-transformations isomorphism, in a theory with purification we cannot enlarge the set of transformations without enlarging the set of states. Indeed, we can compare different theories that have the same set of systems in the following way:

Definition 49 (Inclusion of theories) *The theory* Θ' *is* larger *than the theory* Θ *if for every couple of systems* A, B *one has* $\mathfrak{T}(\mathrm{A},\mathrm{B}) \subseteq \mathfrak{T}'(\mathrm{A},\mathrm{B})$*, where* $\mathfrak{T}'(\mathrm{A},\mathrm{B})$ *denotes the set of all transformations from* A *to* B *allowed by* Θ'.

Then we have the following

Lemma 26 (Maximality of theories with purification) *Let* Θ *be a convex theory with purification, and* Θ' *be a convex theory with the same sets of normalized states of* Θ*, i.e.* $\mathfrak{S}_1(\mathrm{A}) = \mathfrak{S}'_1(\mathrm{A})$ *for every* A*. If* Θ' *is larger than* Θ*, then* $\Theta' = \Theta$.

Proof. First of all, note that the deterministic effect, uniquely defined by the condition $(e|\,\rho)_{\mathrm{A}} = 1, \forall \rho \in \mathfrak{S}_1(\mathrm{A})$ is the same in both theories. Now suppose that $\{\mathscr{C}'_i\}_{i\in\mathrm{X}}$ is one of the tests from A to B allowed by theory Θ'. Let $\{R_{\mathscr{C}'_i}\}_{i\in\mathrm{X}}$ be the corresponding preparation-test for system $\mathrm{B}\tilde{\mathrm{A}}$, as defined by the state-transformations isomorphism of Theorem 17. Since the theories Θ' and Θ have the same states, each $R_{\mathscr{C}'_i}$ is also a state in Θ. Now, convexity of the set of states implies that $\{R_{\mathscr{C}'_i}\}_{i\in\mathrm{X}}$ is a legitimate preparation-test in Θ. Moreover, we have $\sum_{i\in\mathrm{X}} (e|_{\mathrm{B}} \left|R_{\mathscr{C}'_i}\right)_{\mathrm{B}\tilde{\mathrm{A}}} = (e|_{\mathrm{B}} \left|\Phi^{(\mathrm{A})}\right)_{\mathrm{B}\tilde{\mathrm{A}}} := |\tilde{\omega})_{\tilde{\mathrm{A}}}$. Then, by Theorem 17 there must be a test $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ from A to B, allowed by theory Θ, such that $R_{\mathscr{C}'_i} = R_{\mathscr{C}_i} := \mathscr{C}_i \left|\Psi^{(\mathrm{A})}\right)_{\mathrm{A}\tilde{\mathrm{A}}}$. Since $\left|\Psi^{(\mathrm{A})}\right)_{\mathrm{A}\tilde{\mathrm{A}}}$ is dynamically faithful for system A, this implies $\mathscr{C}'_i = \mathscr{C}_i$ for every $i \in \mathrm{X}$. Therefore, Θ' and Θ have exactly the same tests. ■

The states-transformations isomorphism has also the very strong consequence that any transformation that is "mathematically admissible" can be actually realized as a test. To make this statement precise, let us give the following definitions:

Definition 50 (Positive transformation) *A transformation* $\mathscr{C} \in \mathfrak{T}_{\mathbb{R}}(\mathrm{A},\mathrm{B})$ *is* positive *if for every* $\rho \in \mathfrak{S}_+(\mathrm{A})$ *one has* $\mathscr{C}\,|\rho)_{\mathrm{A}} \in \mathfrak{S}_+(\mathrm{B})$.

Definition 51 (S-positive transformation) *Given a system* S*, a transformation* $\mathscr{C} \in \mathfrak{T}_{\mathbb{R}}(\mathrm{A},\mathrm{B})$ *is* S-positive *if* $\mathscr{C} \otimes \mathscr{I}_{\mathrm{S}}$ *is positive.*

Definition 52 (Completely positive transformation) *A transformation* $\mathscr{C} \in \mathfrak{T}_{\mathbb{R}}(\mathrm{A},\mathrm{B})$ *is* completely positive *(CP) if it is* S*-positive for every system* S.

Definition 53 (Admissible instrument) *An* admissible instrument *with input* A *and output* B *is a collection of CP transformations* $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ *such that*

$$\sum_{i\in\mathrm{X}} (e|_{\mathrm{B}}\,\mathscr{C}_i = (e|_{\mathrm{A}}\,. \tag{165}$$

The following Theorem establishes that every admissible instrument must be feasible in a convex theory with purification:

Theorem 18 (Completeness of theories with purification) *In a convex theory with purification every admissible instrument from* A *to* B *is a test. In particular, every admissible instrument from* A *to* I *is an observation-test.*

Proof. Call Θ the theory under consideration, and consider the set of all admissible instruments that are conceivable in Θ. This set is closed under parallel/sequential composition and under coarse-graining and conditioning. Therefore this set defines a new theory Θ' that is larger than Θ. Moreover, by construction Θ' and Θ have the same states. By Lemma 26, this implies $\Theta' = \Theta$. ■

Corollary 36 (Characterization of physical transformations) *In a convex theory with purification the following are equivalent*

1. $\mathscr{C}$ *is a physical transformation from* A *to* B
2. $\mathscr{C}$ *is a CP transformation from* A *to* B *and* $(e|_{\mathrm{A}} - (e|_{\mathrm{B}}\mathscr{C}$ *is CP.*

Proof. The direction 1 ⇒ 2 is obvious. Conversely, suppose that condition 2 is satisfied, and define the CP transformations $(a|_{\mathrm{A}} := (e|_{\mathrm{A}} - (e|_{\mathscr{B}}\mathscr{C}$ and $\mathscr{D} := |\beta)_{\mathrm{B}}\,(a|_{\mathrm{A}}$ where $|\beta)_{\mathrm{B}}$ is some normalized state of system B. Then the collection of CP transformations $\{\mathscr{C},\mathscr{D}\}$ is an admissible instrument. By the completeness of Theorem 18 this implies that $\{\mathscr{C},\mathscr{D}\}$ is a test allowed by the theory. Hence, $\mathscr{C}$ is a physical transformation. ■

We are now in position to prove a stronger result than Lemma 26, namely the fact that a theory with purification is completely specified once we have declared the states for every system:

Theorem 19 (States specify the theory) *Let* Θ, Θ' *be two convex theories with purification. If* Θ *and* Θ' *have the same sets of normalized states, then* $\Theta' = \Theta$.

Proof. Given two theories Θ, Θ' with the same set of states we can take the new theory $\Theta \cup \Theta'$ that is generated by Θ and Θ' by taking sequential and parallel composition of the corresponding CP transformations. Since by construction $\Theta \cup \Theta'$ contains Θ and Θ' and has the same sets of states by Lemma 26 we have $\Theta = \Theta \cup \Theta' = \Theta'$. ■

We conclude this Subsection by discussing the implication of the no-restriction hypothesis of Def. 16 and of Lemma 11, which states that every element in the dual cone of states is proportional to a possible effect. In this case, we have the following characterization:

Lemma 27 *In theory satisfying the no-restriction hypothesis of Def. 16 the following are equivalent:*

1. $a \in \mathfrak{T}_{\mathbb{R}}(\mathrm{A},\mathrm{I})$ *is CP*
2. a *is an element of the dual cone* $\mathfrak{S}_+(\mathrm{A})^*$

3. a is an element of the cone $\mathfrak{E}_+(\mathrm{A})$

Proof. $1 \Rightarrow 2$. Any CP transformation $\mathscr{C}$ from A to I defines a unique element a of the dual cone $\mathfrak{S}_+(\mathrm{A})^*$, via the relation $a(\rho) := \mathscr{C}\,|\rho)_\mathrm{A}$. In fact, $\mathscr{C}$ and a are identified: if two CP transformations $\mathscr{C}$ and $\mathscr{C}'$ define the same effect, then we also have $(\mathscr{C} \otimes \mathscr{I}_\mathrm{C})\,|\sigma)_\mathrm{AC} = (\mathscr{C}' \otimes \mathscr{I}_\mathrm{C})\,|\sigma)_\mathrm{AC}$ for every system C and for every state $\rho \in \mathfrak{S}(\mathrm{AC})$. Therefore $\mathscr{C} \equiv \mathscr{C}'$, and we can identify $\mathscr{C}$ with a. $2 \Rightarrow 3$. By the consequence of the no-restriction hypothesis stated by Lemma 11, if a is in the dual $\mathfrak{S}_+(\mathrm{A})^*$ then a is in $\mathfrak{E}_+(\mathrm{A})$. $3 \Rightarrow 1$ By definition, an element of $\mathfrak{E}_+(\mathrm{A})$ is proportional to an effect (with a positive proportionality constant). Now every effect is a physical transformation from A to $\mathscr{I}$, and physical transformations are by definition CP. ■

Definition 54 (Effect-valued measures) *An admissible instrument from* A *to* I *is an* effect-valued measure (EVM)*, that is, a collection of effects $\{a_i\}_{i\in\mathrm{X}}$ such that $\sum_{i\in\mathrm{X}} (a_i|_\mathrm{A} = (e|_\mathrm{A}$.*

The completeness Theorem 18 now implies:

Corollary 37 (Characterization of observation-tests) *In a convex theory with purification every effect-valued measure is an observation-test. If the no-restriction hypothesis of Def. 16 holds, every probability rule (collection of positive functionals that sum to the deterministic effect) is an observation-test.*

Finally, the characterization of Corollary 36 becomes:

Corollary 38 *In a convex theory with purification satisfying the no-restriction hypothesis of Def. 16 the following are equivalent*

1. *$\mathscr{C}$ is a physical transformation from* A *to* B
2. *$\mathscr{C}$ is a CP transformation from* A *to* B *and is* normalization non-increasing*, i.e., $(e|_\mathrm{B} \mathscr{C}\,|\rho)_\mathrm{A} \le (e|\,\rho)_\mathrm{A}$ for every $\rho \in \mathfrak{S}(\mathrm{A})$.*

XI. ERROR CORRECTION

A. Basic definitions

Here we give some basic definitions that will be used in the next Subsections.

Definition 55 (Correctable channels) *A channel $\mathscr{C} \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ is* correctable *upon input of $\rho \in \mathfrak{S}_1(\mathrm{A})$ if there is a* recovery channel *$\mathscr{R} \in \mathfrak{T}(\mathrm{B},\mathrm{A})$ such that $\mathscr{R} \circ \mathscr{C} =_\rho \mathscr{I}_\mathrm{A}$. If ρ is an internal state, we simply say that $\mathscr{C}$ is* correctable.

Definition 56 (Deletion channels) *A channel $\mathscr{C} \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ is a* deletion channel *upon input of $\rho \in \mathfrak{S}_1(\mathrm{A})$ if there is a fixed state $\sigma \in \mathfrak{S}_1(\mathrm{B})$ such that $\mathscr{C} =_\rho |\sigma)_\mathrm{B}\,(e|_\mathrm{A}$.*

Definition 57 (Purification-preserving channels) *A channel $\mathscr{C} \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ is* purification-preserving *for $\rho \in \mathfrak{S}(\mathrm{A})$ if there is a recovery channel $\mathscr{R} \in \mathfrak{T}(\mathrm{B},\mathrm{A})$ such that $\mathscr{R}\mathscr{C}\,|\Psi_\rho)_\mathrm{AR} = |\Psi_\rho)_\mathrm{AR}$, with $\Psi_\rho \in \mathfrak{S}_1(\mathrm{AR})$ arbitrary purification of ρ.*

In the context of error correction, the purifying system R will be referred to as the *reference.*

Definition 58 (Correlation-erasing channels) *A channel $\mathscr{C} \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ is* correlation-erasing *for $\rho \in \mathfrak{S}(\mathrm{A})$ if there is a state $\sigma \in \mathfrak{S}(\mathrm{B})$ such that $\mathscr{C}\,|\Psi_\rho)_\mathrm{AR} = |\sigma)_\mathrm{B}\,|\tilde\rho)_\mathrm{R}$, where $\Psi_\rho \in \mathfrak{S}_1(\mathrm{AR})$ is an arbitrary purification of ρ, and $\tilde\rho$ is the complementary state $|\tilde\rho)_\mathrm{R} := (e|_\mathrm{A}\,|\Psi_\rho)_\mathrm{AR}$.*

In a theory with purification, the interplay between these four definitions is the basic underlying structure of error correction. The simplest relations can be immediately recovered from Theorem 7, which related the equality upon input of ρ to the equality on a purification of ρ.

Corollary 39 *A channel is correctable upon input of ρ if and only if it is purification-preserving for ρ.*

Corollary 40 *If a channel is correlation-erasing for ρ, then it is a deletion channel upon input of ρ. If local discriminability holds, the converse is also true.*

Another simple fact about error correction, which holds in all theories with purification, is the following

Lemma 28 *If a channel $\mathscr{C} \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ is correctable upon input of $\rho \in \mathfrak{S}_1(\mathrm{A})$ with recovery channel $\mathscr{R}$, and $\mathscr{D} \in D_\mathscr{C}$ is a transformation in the refinement set of $\mathscr{C}$ (Def. 21), then $\mathscr{D}$ is correctable upon input of ρ, with recovery channel $\mathscr{R}$,* i.e. *$\mathscr{R}\mathscr{D} =_\rho p\mathscr{I}_\mathrm{A}$ for some probability $p > 0$.*

Proof. By definition, since $\mathscr{D}$ is in the refinement set of $\mathscr{C}$, there is a test $\{\mathscr{D}_i\}_{i\in\mathrm{X}}$ such that $\mathscr{D} \equiv \mathscr{D}_{i_0}$ and $\mathscr{C} = \sum_{i\in\mathrm{X}} \mathscr{D}_i$. Since $\mathscr{C}$ is correctable with recovery channel $\mathscr{R}$, one has $\mathscr{I}_\mathrm{A} =_\rho \mathscr{R}\mathscr{C} = \sum_{i\in\mathrm{X}} \mathscr{R}\mathscr{D}_i$. This means that the test $\{\mathscr{R}\mathscr{D}_i\}_{i\in\mathrm{X}}$ is non-disturbing upon input of ρ. By the "no-information without disturbance" Theorem 10 one then has $\mathscr{R}\mathscr{D}_i =_\rho p_i\mathscr{I}_\mathrm{A}$ for every $i \in \mathrm{X}$. ■

B. Error correction and the complementarity between correctable and deletion channels

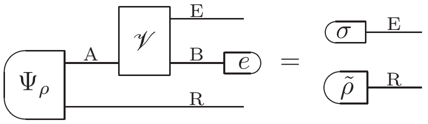
We now discuss some necessary and sufficient conditions for the correctability of channels. The simplest case is that of channels from a system to itself:

Theorem 20 *A channel $\mathscr{C}$ from* A *to* A *is correctable if and only if it is reversible.*

Proof. Clearly, if $\mathscr{C} = \mathscr{U} \in \mathbf{G}_\mathrm{A}$ one can correct $\mathscr{C}$ by applying $\mathscr{U}^{-1}$. Conversely, suppose that $\mathscr{C}$ is correctable with some recovery channel $\mathscr{R}$. Let $\mathscr{C} = \sum_{i\in\mathrm{X}} \mathscr{C}_i$ be a refinement of $\mathscr{C}$ where each $\mathscr{C}_i$ is an atomic transformation. Then, the composition $\{\mathscr{R}\mathscr{C}_i\}_{i\in\mathrm{X}}$ is a non-disturbing test, and Theorem 10 implies $\mathscr{R}\mathscr{C}_i = p_i \mathscr{I}_\mathrm{A}$. Since $\mathscr{R}$ is a channel, applying the deterministic effect we obtain $(e|_\mathrm{A} \mathscr{R}\mathscr{C}_i = (e|_\mathrm{A} \mathscr{C}_i = p_i\, (e|_\mathrm{A}$, that is, $\mathscr{C}_i$ is proportional to an atomic channel $\mathscr{U}_i$. By Corollary 28, an atomic channel from A to A is reversible. Therefore, we have $\mathscr{R}\mathscr{U}_i = \mathscr{I}_\mathrm{A}$, which implies $\mathscr{R} = \mathscr{U}_i^{-1}$ for every i. Hence, all channels $\mathscr{U}_i$ must be equal, and one has $\mathscr{C} = \mathscr{U}$ for some reversible channel $\mathscr{U} \in \mathbf{G}_\mathrm{A}$. ■

We now give necessary and sufficient conditions for error correction in the general case of channels from A to B. The following condition was presented in the quantum case in Refs. [54–56].

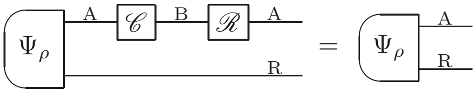
Theorem 21 (Factorization of reference and environment) *A channel $\mathscr{C} \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ is correctable upon input of ρ if and only if there are a reversible dilation $\mathscr{V} \in \mathfrak{T}(\mathrm{A},\mathrm{BE})$ of $\mathscr{C}$ and a purification $|\Psi_\rho)_\mathrm{AR}$ of ρ such that systems* E *and* R *remain uncorrelated. Diagrammatically,*



(166)

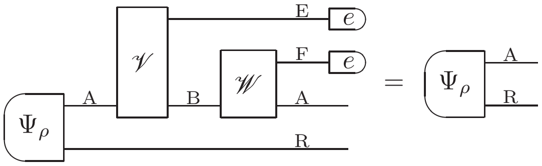
where σ is some state of E *and $\tilde\rho$ is the complementary state of ρ on system* R.

Proof. Suppose that $\mathscr{C}$ is correctable upon input of ρ with some recovery channel $\mathscr{R}$. Then, by Theorem 7 we have



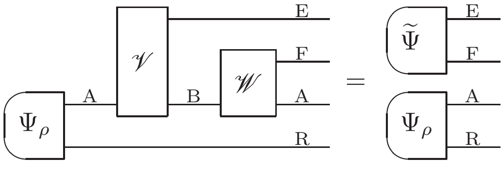
(167)

and, inserting two reversible dilations for $\mathscr{C}$ and $\mathscr{R}$, respectively,



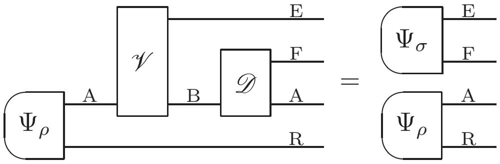
(168)

This means that $\mathscr{W}\mathscr{V}\,|\Psi_\rho)_\mathrm{AR}$ is a purification of $|\Psi_\rho)_\mathrm{AR}$. Then, Lemma 19 ensures that $\mathscr{W}\mathscr{V}\,|\Psi_\rho)_\mathrm{AR}$ is of the form



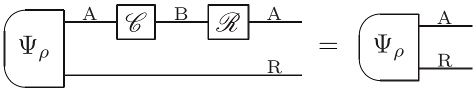
(169)

where $\tilde\Psi$ is some pure state on EF. Applying the deterministic effect on FA and using the fact that $\mathscr{W}$ is a channel, we then obtain Eq. (166). Conversely, suppose that Eq. (166) holds for some dilation $\mathscr{V}$ and some purification $|\Psi_\rho)_\mathrm{AR}$. Then take a purification of σ, say $\Psi_\sigma \in \mathfrak{S}_1(\mathrm{EF})$. Since $\mathscr{V}\,|\Psi_\rho)_\mathrm{AR}$ and $|\Psi_\rho)_\mathrm{AR}\,|\Psi_\sigma)_\mathrm{EF}$ are both purifications of $|\sigma)_\mathrm{E}\,|\tilde\rho)_\mathrm{R}$, by Lemma 21 we have



(170)

for some channel $\mathscr{D} \in \mathfrak{T}(\mathrm{B},\mathrm{FA})$. Applying the deterministic effect on E and F and defining $\mathscr{R} := (e|_\mathrm{F}\, \mathscr{D}$ we then obtain



(171)

By Theorem 7, this implies $\mathscr{R}\circ\mathscr{C} =_\rho \mathscr{I}_\mathrm{A}$, namely $\mathscr{C}$ is correctable upon input of ρ. ■

An immediate consequence of the factorization Theorem 21 is:

Corollary 41 (Complementarity of purification-preserving and correlation-erasing channels) *A channel $\mathscr{C} \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ is purification-preserving for $\rho \in \mathfrak{S}_1(\mathrm{A})$ (according to Def. 57) if and only if its complementary channel $\widetilde{\mathscr{C}} \in \mathfrak{T}(\mathrm{A},\mathrm{E})$ is correlation-erasing for ρ (according to Def. 58).*

Proof. By corollary 39, $\mathscr{C}$ is purification-preserving for ρ iff it is correctable upon input of ρ and, by the previous Theorem, iff Eq. (166) holds. But Eq. (166) is the definition of $\widetilde{\mathscr{C}}$ being a correlation-erasing channel for ρ. ■

In a theory with purification, since the global evolution of system and environment is reversible, it would be natural to expect that if no information goes to the environment, then the whole information about the input state is contained in the system. While this intuition is correct in theories with local discriminability (see Ref. [57] for the quantum case), in general theories this situation is trickier. Indeed, as we will see in the following, in a theory without local discriminability some information can remain "locked" in the global state, in a way that makes it inaccessible both from the system and from the environment separately.

Corollary 42 (Complementarity of correctable and deletion channels) *If a channel $\mathscr{C} \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ is correctable upon input of $\rho \in \mathfrak{S}_1(\mathrm{A})$ (according to Definition 55), then its complementary channel $\widetilde{\mathscr{C}} \in \mathfrak{T}(\mathrm{A},\mathrm{E})$ is a deletion channel upon input of ρ (according to Definition 56). If local discriminability holds, the converse is also true.*

Proof. Direct consequence of corollaries 39, 41, and 40. ■

Counterexample. We show that in a theory without local discriminability the complementarity between correctable and deletion channels does not hold. Consider the case of quantum mechanics on real Hilbert spaces, and consider the isometry V from a real qubit to two real qubits defined by

$$V = |\Phi_+\rangle\langle 0| + |\Psi_-\rangle\langle 1| \tag{172}$$

with $|\Phi_+\rangle := \frac{|0\rangle|0\rangle+|1\rangle|1\rangle}{\sqrt{2}}$, and $|\Psi_-\rangle := \frac{|0\rangle|1\rangle-|1\rangle|0\rangle}{\sqrt{2}}$. In this case the complementary channels $\mathscr{C}(\rho) := \mathrm{Tr}_1[V\rho V^\tau]$ and $\widetilde{\mathscr{C}}(\rho) := \mathrm{Tr}_2[V\rho V^\tau]$ are both deletion channels: indeed, one has

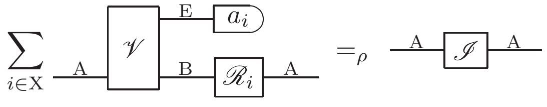
$$\mathscr{C}(\rho) = \frac{I_1}{2} \qquad \widetilde{\mathscr{C}}(\rho) = \frac{I_2}{2}\ , \tag{173}$$

for any real density matrix ρ.

C. Error correction with one-way classical communication from the environment

Here we briefly discuss a more general kind of correction, in which the environment is not completely inaccessible, but rather some operations on it are allowed. Particularly interesting is the case of LOCC operations, which do not require the exchange of systems from the environment, but only communication of outcomes and conditioned operations. In particular, we will focus here on the case of a single round of forward classical communication from the environment to the output system. With the term "classical" we mean that only outcomes are communicated.

Definition 59 (One-way correctable channels) *A channel* $\mathscr{C} \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ *is* one-way correctable *upon input of* ρ *if for every dilation* $\mathscr{V} \in \mathfrak{T}(\mathrm{A},\mathrm{BE})$ *there is an observation-test* $\{a_i\}_{i\in\mathrm{X}}$ *on* E *and a collection of recovery channels* $\{\mathscr{R}_i\}_{i\in\mathrm{X}} \subset \mathfrak{T}(\mathrm{B},\mathrm{A})$ *such that*



(174)

If ρ *is an internal state, we simply say that* $\mathscr{C}$ *is* one-way correctable.

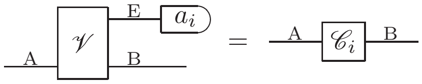
The following theorem states that one-way correctable channels are nothing but randomizations of correctable channels. The quantum version of it was given by Gregoratti and Werner in Ref. [64].

Theorem 22 (Characterization of one-way correctable channels) *A channel* $\mathscr{C} \in \mathfrak{T}(\mathrm{A},\mathrm{B})$ *is one-way correctable upon input of* $\rho \in \mathfrak{S}_1(\mathrm{A})$ *if and only if* $\mathscr{C}$ *is a the coarse-graining of a test* $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ *where each transformation* $\mathscr{C}_i$ *is correctable upon input of* ρ. *In particular, if* ρ *is internal, then* $\mathscr{C}$ *is a randomization of correctable channels.*

Proof. Suppose that $\mathscr{C}$ is one-way correctable upon input of ρ. Defining the test $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ by $\mathscr{C}_i := (a_i|_{\mathrm{E}}\mathscr{V}$, and using Theorem 10, we then obtain $\mathscr{R}_i\mathscr{C}_i =_\rho p_i\mathscr{I}_{\mathrm{A}}$. Therefore, $\mathscr{C}$ is the coarse-graining of a test where each transformation is correctable upon input of ρ. Moreover, if ρ is internal, using the fact that each $\mathscr{R}_i$ is a channel, we obtain

$$(e|_{\mathrm{A}}\mathscr{R}_i\mathscr{C}_i = (e|_{\mathrm{B}}\mathscr{C}_i = p_i\,(e|_{\mathrm{A}}\ , \tag{175}$$

namely each $\mathscr{C}_i$ must be proportional to a channel, say $\mathscr{C}_i = p_i\mathscr{D}_i$, with channel $\mathscr{D}_i$ correctable upon input of ρ. Conversely, suppose that $\mathscr{C} = \sum_{i\in\mathrm{X}}\mathscr{C}_i$ for some test $\{\mathscr{C}_i\}$ where each transformation $\mathscr{C}_i$ is correctable upon input of ρ. Dilating such a test, we then obtain a channel $\mathscr{V} \in \mathfrak{T}(\mathrm{A},\mathrm{BE})$ and an observation-test $\{a_i\}_{i\in\mathrm{X}}$ on E such that



(176)

for every outcome $i \in \mathrm{X}$. Since each $\mathscr{C}_i$ is correctable upon input of ρ, knowing the outcome $i \in \mathrm{X}$, we can perform the recovery channel for $\mathscr{C}_i$, thus correcting channel $\mathscr{C}$. ■

In the case of channels from A to itself, the above theorem takes the simple form

Corollary 43 *A channel* $\mathscr{C} \in \mathfrak{T}(\mathrm{A})$ *is one-way correctable if and only if it is a randomization of reversible channels.*

Proof. Just combine Theorem 22 with the characterization of correctable channels from A to A (Theorem 20). ■

XII. CAUSALLY ORDERED CHANNELS AND CHANNELS WITH MEMORY

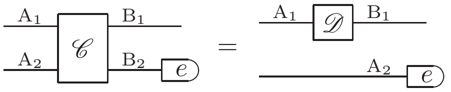
In Ref. [58] Beckman, Gottesmann, Nielsen, and Preskill introduced the notions of *semicausal* and *semilocalizable* quantum channel for the purpose of studying the constraints on quantum dynamics of bipartite systems imposed by relativistic causality. Subsequently, Eggeling, Schlingemann, and Werner [59] proved the equivalence between semicausality and semilocalizability (see also Ref. [60] for an extensive discussion on the topic). The same notions were generalized to the case of multipartite channels by Kretschmann and Werner in Ref. [61]. From different points of view Refs. [33, 61, 62] studied the structure of multipartite causal channels, showing that they can always be realized as sequences of channels with memory. In this Section we show that all these results, originally obtained in quantum mechanics, actually hold in any causal theory with purification.

Unfortunately, the nomenclature used in the literature is not fully consistent if we go from bipartite to multipartite channels [63]. In order to have a consistent nomenclature, instead of "semicausal" and "semilocalizable channel" we use here the plain expressions *causally ordered bipartite channel* and *sequence of two channels with memory*, respectively.

Definition 60 (Causally ordered bipartite channel) *A bipartite channel $\mathscr{C}$ from* $\mathrm{A_1A_2}$ *to* $\mathrm{B_1B_2}$ *is* causally ordered *if there is a channel $\mathscr{D}$ from* $\mathrm{A_1}$ *to* $\mathrm{B_1}$ *such that*

$$(e|_{\mathrm{B_2}}\mathscr{C} = \mathscr{D} \otimes (e|_{\mathrm{A_2}}\,. \tag{177}$$

Diagrammatically,



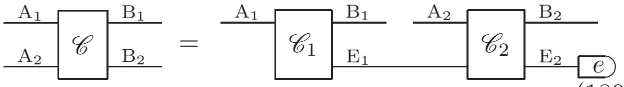
(178)

Eq. (178) means that the channel $\mathscr{C}$ does not allow for signaling from the input system $\mathrm{A_2}$ to the output system $\mathrm{B_1}$. In a relativistic context, this can be interpreted as $\mathrm{B_1}$ being outside the causal future of $\mathrm{A_2}$.

Definition 61 (Sequence of two channels with memory) *A bipartite channel $\mathscr{C}$ from* $\mathrm{A_1A_2}$ *to* $\mathrm{B_1B_2}$ *can be realized as a* sequence of two channels with memory *if there exist two systems* $\mathrm{E_1}, \mathrm{E_2}$, *called* memory systems*, and two channels* $\mathscr{C}_1 \in \mathfrak{T}(\mathrm{A_1}, \mathrm{B_1E_1})$ *and* $\mathscr{C}_2 \in \mathfrak{T}(\mathrm{A_2E_1}, \mathrm{B_2E_2})$ *such that*

$$\mathscr{C} = (e|_{\mathrm{E_2}}(\mathscr{C}_2 \otimes \mathscr{I}_{\mathrm{B_1}})(\mathscr{I}_{\mathrm{A_2}} \otimes \mathscr{C}_1). \tag{179}$$

Diagrammatically,



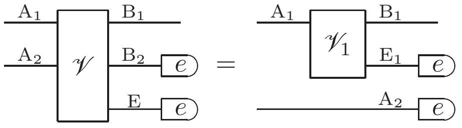
(180)

A. Dilation of causally ordered channels

For causally ordered bipartite channels the dilation theorem implies the following result:

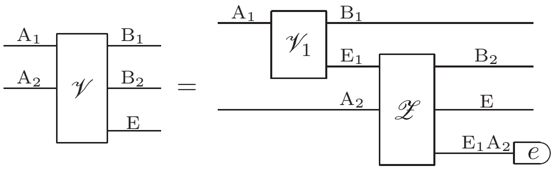
Theorem 23 (Causal ordering is memory) *A bipartite channel $\mathscr{C}$ from* $\mathrm{A_1A_2}$ *to* $\mathrm{B_1B_2}$ *is causally ordered if and only if it can be realized as a sequence of two channels with memory. Moreover, the channels $\mathscr{C}_1, \mathscr{C}_2$ in Eq. (180) can be always chosen such that $\mathscr{C}_2\mathscr{C}_1$ is a reversible dilation of $\mathscr{C}$.*

Proof. If Eq. (180) holds, the channel $\mathscr{C}$ is clearly causally ordered, with the channel $\mathscr{D}$ given by $\mathscr{D} := (e|_{\mathrm{E_1}}\mathscr{C}_1$. Conversely, suppose that $\mathscr{C}$ is causally ordered. Take a reversible dilation of $\mathscr{C}$, say $\mathscr{V} \in \mathfrak{T}(\mathrm{A_1A_2}, \mathrm{B_1B_2E})$, and a reversible dilation of $\mathscr{D}$, say $\mathscr{V}_1 \in \mathfrak{T}(\mathrm{A_1}, \mathrm{B_1E_1})$. Now, by definition of causally ordered channel (Eq. (178))we have



(181)

This means that $\mathscr{V}$ and $\mathscr{V}_1 \otimes \mathscr{I}_{\mathrm{A_2}}$ are two reversible dilations of the same channel. By the uniqueness of the reversible dilation expressed by Lemma 24 we then obtain

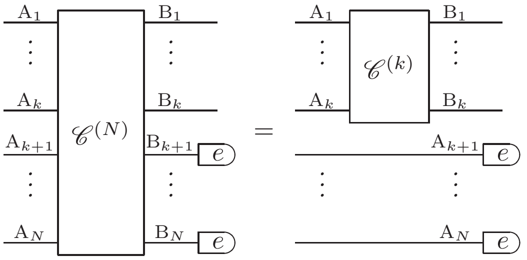


(182)

Once we have defined $\mathrm{E_2} := \mathrm{EE_1A_2}$ it only remains to observe that the above diagram is nothing but the thesis, with $\mathscr{C}_1 = \mathscr{V}_1$ and $\mathscr{C}_2 = \mathscr{Z}$. By construction, $\mathscr{C}_2\mathscr{C}_1$ is a reversible dilation of $\mathscr{C}$. ■

The definition of causally ordered bipartite channel is easily extended to the multipartite case as follows:

Definition 62 (Causally ordered channel) *An N-partite channel $\mathscr{C}^{(N)}$ from* $\mathrm{A_1} \dots \mathrm{A}_N$ *to* $\mathrm{B_1} \dots \mathrm{B}_N$ *is* causally ordered *if for every $k \le N$ there is a channel $\mathscr{C}^{(k)}$ from* $\mathrm{A_1} \dots \mathrm{A}_k$ *to* $\mathrm{B_1} \dots \mathrm{B}_k$ *such that*

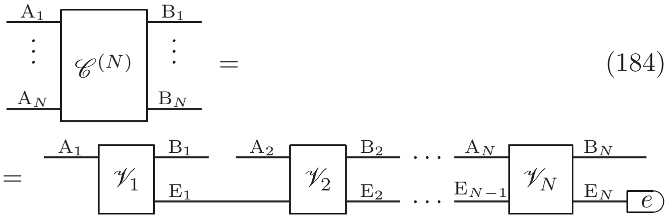


(183)

The definition means that the output systems $\mathrm{B_1} \dots \mathrm{B}_k$ are outside the causal future of any input system A_l with $l > k$.

Causally ordered channels can be characterized as follows:

Theorem 24 (Causal ordering is memory for general N) *An N-partite channel $\mathscr{C}^{(N)}$ from* $\mathrm{A_1} \dots \mathrm{A}_N$ *to* $\mathrm{B_1} \dots \mathrm{B}_N$ *is causally ordered if and only if there exists a sequence of memory systems* $\{\mathrm{E}_k\}_{k=0}^N$ *with* $\mathrm{E_0} = \mathrm{I}$ *and a sequence of channels* $\{\mathscr{V}_k\}_{k=1}^N$*, with* $\mathscr{V}_k \in \mathfrak{T}(\mathrm{A}_k\mathrm{E}_{k-1}, \mathrm{B}_k\mathrm{E}_k)$ *such that*



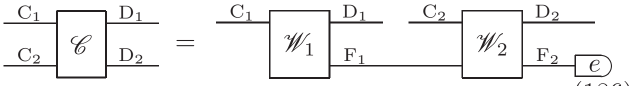
(184)

Moreover, $\mathscr{V}_N \dots \mathscr{V}_1$ is a reversible dilation of $\mathscr{C}$.

Proof. It is trivial to see that if $\mathscr{C}^{(N)}$ is a sequence of channels with memory, it is a causally ordered channel. Here we prove the converse. For $N = 1$ the thesis is just the dilation theorem for channels. We now show that if the thesis holds for N, then it has to hold also for $N+1$. Since $\mathscr{C}^{(N+1)}$ is a causal channel, we have in particular

$$(e|_{\mathrm{B}_{N+1}} \mathscr{C}^{(N+1)} = \mathscr{C}^{(N)} \otimes (e|_{\mathrm{A}_{N+1}} \ . \tag{185}$$

This means that $\mathscr{C}^{(N+1)}$ can be viewed as a bipartite causally ordered channel from $\mathrm{C}_1\mathrm{C}_2$ to $\mathrm{D}_1\mathrm{D}_2$, where $\mathrm{C}_1 := \mathrm{A}_1 \dots \mathrm{A}_N$, $\mathrm{C}_2 := \mathrm{A}_{N+1}$, $\mathrm{D}_1 := \mathrm{B}_1 \dots \mathrm{B}_N$, and $\mathrm{D}_2 := \mathrm{B}_{N+1}$. Then Theorem 23 yields two channels $\mathscr{W}_1 \in \mathfrak{T}(\mathrm{C}_1, \mathrm{D}_1\mathrm{F}_1)$ and $\mathscr{W}_2 \in \mathfrak{T}(\mathrm{C}_2\mathrm{F}_1, \mathrm{D}_2\mathrm{F}_2)$ such that



(186)

Now, applying the deterministic effect on D_2, and using Eq. (185) the above diagram implies also that $\mathscr{W}_1$ is a dilation of $\mathscr{C}^{(N)}$. On the other hand, by the induction hypothesis $\mathscr{C}^{(N)}$ has a reversible dilation $\mathscr{V}^{(N)}$ of the form of Eq. (184), namely

$$\mathscr{V}^{(N)} = \mathscr{T}_N \dots \mathscr{T}_1, \tag{187}$$

for some sequence of channels $(\mathscr{T}_k)_{k=1}^N \in \mathfrak{T}(\mathrm{A}_k\mathrm{G}_{k-1}, \mathrm{B}_k\mathrm{G}_k)$ and some sequence of memory systems $(\mathrm{G}_k)_{k=0}^N$, with $\mathrm{G}_0 = \mathrm{I}$. Since $\mathscr{W}_1$ and $\mathscr{V}^{(N)}$ are reversible dilations of the same channel, the uniqueness of the reversible dilation of Lemma 24 implies $\mathscr{W}_1 = (e|_{\mathrm{G}_N} \mathscr{Z}\mathscr{V}^{(N)}$, with $\mathscr{Z} \in \mathfrak{T}(\mathrm{G}_N, \mathrm{G}_N\mathrm{F}_1)$ of the form of Eq. (149). Then, the thesis follows by defining the memory systems as

$$\mathrm{E}_k := \begin{cases} \mathrm{G}_k & k < N \\ \mathrm{G}_N\mathrm{F}_1 & k = N \\ \mathrm{G}_N\mathrm{F}_2 & k = N+1. \end{cases} \tag{188}$$

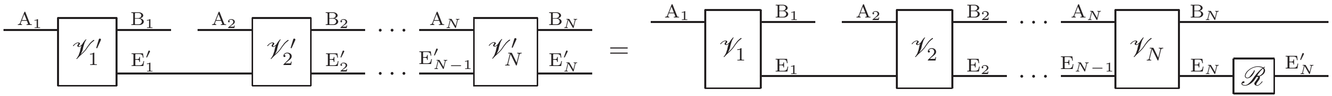
and by defining the channels as

$$\mathscr{V}_k := \begin{cases} \mathscr{T}_k & k < N \\ \mathscr{Z}\mathscr{T}_N & k = N \\ \mathscr{I}_{\mathrm{G}_N} \otimes \mathscr{W}_2 & k = N+1. \end{cases} \tag{189}$$

By construction, the channel $\mathscr{V}_{N+1}\mathscr{V}_N \dots \mathscr{V}_1$ is a reversible dilation of the channel $\mathscr{C}^{(N+1)}$. ■

Moreover, since the realization of the previous Theorem is just the reversible dilation of $\mathscr{C}^{(N)}$, we have the uniqueness result:

Corollary 44 (Uniqueness of the reversible dilation) *Let $\{\mathscr{V}_k\}_{k=1}^N$, $\mathscr{V}_k \in \mathfrak{T}(\mathrm{A}_k\mathrm{E}_{k-1}, \mathrm{B}_k\mathrm{E}_k)$ be a reversible realization of the causally ordered channel $\mathscr{C}^{(N)}$ as a sequence of channels with memory, as in Theorem 24. Suppose that $\{\mathscr{V}'_k\}_{k=1}^N$, $\mathscr{V}'_k \in \mathfrak{T}(\mathrm{A}_k\mathrm{E}'_{k-1}, \mathrm{B}_k\mathrm{E}'_k)$ is another reversible realization of $\mathscr{C}^{(N)}$ as a sequence of channels with memory. Then there exists a channel $\mathscr{R}$ from E_N to E'_N such that*



(190)

Proof. The channels $\mathscr{V} := \mathscr{V}_N \dots \mathscr{V}_1 \in \mathfrak{T}(\mathrm{A}_1 \dots \mathrm{A}_N, \mathrm{B}_1 \dots \mathrm{B}_N\mathrm{E}_N)$ and $\mathscr{V}' := \mathscr{V}'_N \dots \mathscr{V}'_1 \in \mathfrak{T}(\mathrm{A}_1 \dots \mathrm{A}_N, \mathrm{B}_1 \dots \mathrm{B}_N\mathrm{E}'_N)$ are two reversible dilations of the channel $\mathscr{C}^{(N)}$. The statement is the direct application of the uniqueness of the dilation stated by Lemma 24. ■

B. No bit commitment

Sequences of channels with memory can be used to describe sequences of moves of a given party in a cryptographic protocol or in a multiparty game (see Ref. [62] for the case of quantum games). In this scenario, the memory systems are the private systems available to a party, while the other input-output systems are the systems exchanged in the communication with other parties. In this context, the uniqueness of the realization of a causal channel directly implies the impossibility of tasks like unconditionally secure bit commitment (see Refs. [65, 66] and references therein for the definition of the problem). A proof in the general case is given by the following:

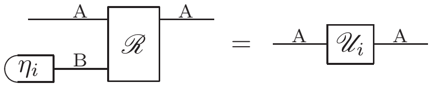
Corollary 45 (No perfectly secure bit commitment) *In a theory with purification, if an N-round protocol is perfectly concealing, then there is a perfect cheating.*

Proof. We first prove the impossibility for protocols that do not involve the exchange of classical information. Let $\mathscr{A}_0, \mathscr{A}_1 \in \mathfrak{T}(\mathrm{A}_1 \dots, \mathrm{A}_N, \mathrm{B}_1 \dots \mathrm{B}_{N-1}\mathrm{B}_N\mathrm{F}_N)$ be two causally ordered N-partite channels (here the last output system of the causally-ordered channels is the bipartite system $\mathrm{B}_N\mathrm{F}_N$), representing Alice's moves to encode the bit value $b = 0, 1$, respectively. The system F_N is the system sent from Alice to Bob at the final phase of the protocol (called the *opening*) in order to unveil the value of the bit. If the protocol is perfectly concealing, then the reduced channels before the opening phase must be indistinguishable, namely $(e|_{\mathrm{F}_N} \mathscr{A}_0 = (e|_{\mathrm{F}_N} \mathscr{A}_1 := \mathscr{C}$. Now, take two reversible dilations $\mathscr{V}_0 \in \mathfrak{T}(\mathrm{A}_1 \ldots, \mathrm{A}_N, \mathrm{B}_1 \ldots \mathrm{B}_N \mathrm{F}_N \mathrm{G}_0)$ and $\mathscr{V}_1 \in \mathfrak{T}(\mathrm{A}_1 \ldots, \mathrm{A}_N, \mathrm{B}_1 \ldots \mathrm{B}_N \mathrm{F}_N \mathrm{G}_1)$ for $\mathscr{A}_0$ and $\mathscr{A}_1$, respectively. Since $\mathscr{V}_0$ and $\mathscr{V}_1$ are also two dilations of the channel $\mathscr{C}$, there is a channel $\mathscr{R}$ from $\mathrm{F}_N\mathrm{G}_0$ to $\mathrm{F}_N\mathrm{G}_1$ such that $\mathscr{V}_1 = \mathscr{R}\mathscr{V}_0$. Applying this channel to her private systems, Alice can switch from $\mathscr{V}_0$ to $\mathscr{V}_1$ just before the opening. Discarding the auxiliary system G_1, this yields channel $\mathscr{A}_1$. The cheating is perfect, since Alice can play the strategy $\mathscr{V}_0$ until the end of the commitment, and decide the bit value before the opening without being detected by Bob. The above reasoning can be extended to N-round protocols involving the exchange of classical information. Indeed, classical messages can be modelled by perfectly distinguishable states, while classical channels can be modelled by measure-and-prepare channels where the observation-test is discriminating, and the prepared states are perfectly distinguishable. The fact that some systems can only be prepared in perfectly distinguishable states will be referred to as the "communication interface" of the protocol [65, 66]. In this case, to construct Alice's cheating strategy we can first take the reversible dilations $\mathscr{V}_0, \mathscr{V}_1$ and the channel $\mathscr{R}$ such that $\mathscr{V}_1 = \mathscr{R}\mathscr{V}_0$. In order to comply with the communication interface of the protocol, one can compose $\mathscr{V}_0$ and $\mathscr{V}_1$ with classical channels on all systems that must be "classical" before the opening, thus obtaining two channels $\mathscr{D}_0$ and $\mathscr{D}_1$ that are no longer reversible, but still satisfy $\mathscr{D}_1 = \mathscr{R}\mathscr{D}_0$. Discarding the auxiliary system G_1 and, if required by the communication interface, applying a classical channel on F_N, Alice then obtains channel $\mathscr{A}_1$. Again, this strategy allows Alice to decide the value of the bit just before the opening without being detected. ■

XIII. DETERMINISTIC PROGRAMMING OF REVERSIBLE TRANSFORMATIONS

In Section VIII we saw that transformations can be stored into states, in such a way that they can be retrieved at later time with non-zero probability of success. This provides an instance of *probabilistic programming*, in which a state plays the role of program for a transformation, and a suitable machine is able to read out the program and to reproduce (with some probability) the correct transformation. Of course, one would like also to have deterministic programmable machines, which correctly retrieve the transformations with unit probability. We now show that such machines are much more demanding in terms of resources: indeed to program a certain number of reversible transformations one needs to have an equal number of perfectly distinguishable program states. This theorem is the general version of the quantum no-programming theorem by Nielsen and Chuang [67].

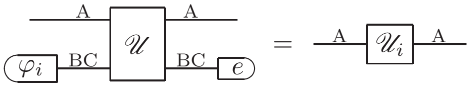
Theorem 25 (No perfect deterministic programming of reversible channels without distinguishable program states) *Let $\{\mathscr{U}_i\}_{i\in\mathrm{X}}$ be a set of reversible channels on* A, *and $\{\eta_i\}_{i\in\mathrm{X}}$ be a set of pure states of* B. *If there exists a channel $\mathscr{R} \in \mathfrak{T}(\mathrm{AB}, \mathrm{A})$ such that*



(191)

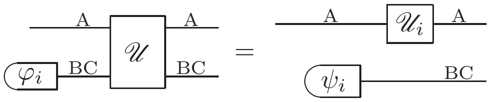
then the states $\{\eta_i\}_{i\in\mathrm{X}}$ are perfectly distinguishable.

Proof. Take a dilation of $\mathscr{R}$, with pure state $\varphi_0 \in \mathfrak{S}_1(\mathrm{C})$ and reversible channel $\mathscr{U} \in \mathfrak{T}(\mathrm{ABC})$. Upon defining the pure states $\varphi_i := \eta_i \otimes \varphi_0$ we have



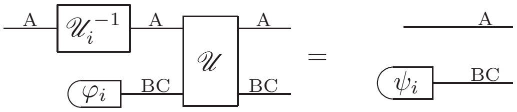
(192)

Since this is a dilation of the reversible transformation $\mathscr{U}_i$, by the uniqueness of the reversible dilation stated by Theorem 15 there must be a pure state $\psi_i \in \mathfrak{S}_1(\mathrm{BC})$ such that



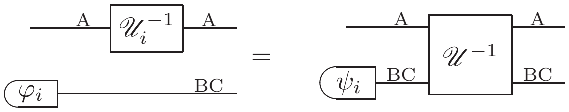
(193)

By applying $\mathscr{U}_i^{-1}$ on both sides of Eq. (193), one has



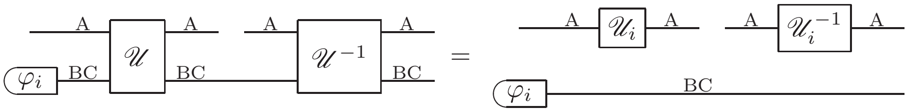
(194)

and, applying $\mathscr{U}^{-1}$,



(195)

Composing Eqs. (193) and (195) we then obtain



(196)

This means that we can obtain an unbounded number of copies of $\mathscr{U}_i$ and $\mathscr{U}_i^{-1}$ by iterating the application of $\mathscr{U}$ and $\mathscr{U}^{-1}$. Now, if $\mathscr{U}_i$ and $\mathscr{U}_j$ are different, the probability of error in discriminating between them using N copies should go to zero as N goes to infinity (this can be seen by repeating N times the optimal test and using majority voting, as in the proof of Theorem 12). Since programming the transformations $\{(\mathscr{U}_i \otimes \mathscr{U}_i^{-1})^{\otimes N}\}$ and discriminating among them is a particular way of discriminating between the program states $\{\varphi_i\}$, the latter must be perfectly distinguishable. Finally, since the states $\varphi_i = \eta_i \otimes \varphi_0$ are perfectly distinguishable, also the program states η_i must be so. ■

Note that trying to use mixed program states $\{\rho_i\}$ cannot help in reducing the number of perfectly distinguishable states needed in the program system B. Indeed, suppose that ρ_i is the following mixture $\rho_i = \sum_j p_j^{(i)} \psi_j^{(i)}$. Since reversible transformations are atomic, this means that each pure state $\psi_j^{(i)}$ must work as a program for $\mathscr{U}_i$. But the above theorem implies that, whichever choice we make, the pure states $\{\varphi_{j_i}^{(i)}\}_{i\in \mathrm{X}}$ must be perfectly distinguishable.

XIV. PURIFICATION WITH CONJUGATE SYSTEMS

A. Conjugate purifying systems

All the results derived so far were consequence of the sole fact that every state has a purification, unique up to reversible transformations of the purifying system. We now add more structure, by introducing the notion of conjugate purifying systems:

Postulate 2 (Conjugate purifying systems) *For every system* A *there exists a* conjugate purifying system $\tilde{\mathrm{A}}$ *such that*

1. *for every state* $\rho \in \mathfrak{S}_1(\mathrm{A})$ *there is a purification* Ψ_ρ *in* $\mathfrak{S}_1(\mathrm{A}\tilde{\mathrm{A}})$ (completeness for purification)
2. $\tilde{\tilde{\mathrm{A}}} = \mathrm{A}$ (symmetry)
3. $\widetilde{\mathrm{AB}} = \tilde{\mathrm{A}}\tilde{\mathrm{B}}$ (regularity under composition)

The above postulate could be derived from more basic assumptions. However, we will not discuss this issue here, and, for the moment, the existence of conjugate systems will be taken as a Postulate.

Conjugate purifying systems have particularly nice properties, some of which are given in the following:

Lemma 29 *Let* $\tilde{\mathrm{A}}$ *be the conjugate system of* A. *Then,* $\dim \mathfrak{S}_{\mathbb{R}}(\tilde{\mathrm{A}}) = \dim \mathfrak{S}_{\mathbb{R}}(\mathrm{A})$.

Proof. Trivial consequence of the bound on dimensions given in Eq. (93) and of the symmetry condition $\tilde{\tilde{\mathrm{A}}} = \mathrm{A}$. ■

In a theory with conjugate purifying systems, the dynamically faithful pure states considered in Subsection VII C enjoy the following symmetry property:

Lemma 30 *If the pure state* $\Psi \in \mathfrak{S}_1(\mathrm{A}\tilde{\mathrm{A}})$ *is dynamically faithful for system* A*, then it is dynamically faithful for system* $\tilde{\mathrm{A}}$.

Proof. Let $\tilde{\omega}$ be the marginal of Ψ on system $\tilde{\mathrm{A}}$, namely $|\tilde{\omega})_{\tilde{\mathrm{A}}} = (e|_{\mathrm{A}} |\Psi)_{\mathrm{A}\tilde{\mathrm{A}}}$. Since Ψ is dynamically faithful for system A, the map $\tau : \mathfrak{E}_{\mathbb{R}}(\mathrm{A}) \to \mathsf{Span}(D_{\tilde{\omega}})$ defined by $(a|_{\mathrm{A}} \mapsto |\tau_a)_{\tilde{\mathrm{A}}} = (a|_{\mathrm{A}} |\Psi)_{\mathrm{A}\tilde{\mathrm{A}}}$ is injective (and surjective, by definition). This implies $\dim \mathsf{Span}(D_{\tilde{\omega}}) = \dim \mathfrak{E}_{\mathbb{R}}(\mathrm{A})$. On the other hand, using the previous Lemma one has $\dim \mathfrak{E}_{\mathbb{R}}(\mathrm{A}) \equiv \dim \mathfrak{S}_{\mathbb{R}}(\mathrm{A}) = \dim \mathfrak{S}_{\mathbb{R}}(\tilde{\mathrm{A}})$. This proves that $\tilde{\omega}$ is internal in $\mathfrak{S}(\tilde{\mathrm{A}})$. Since Ψ is the purification of an internal state, by Theorem 8 it is faithful for system $\tilde{\mathrm{A}}$. ■

Using the previous Lemma it is quite simple to show that conjugate systems are unique up to operational equivalence:

Lemma 31 (Uniqueness of the conjugate system) *For any system* A *the conjugate system* $\tilde{\mathrm{A}}$ *is unique up to operational equivalence (see Def. 5).*

Proof. Suppose that $\tilde{\mathrm{A}}'$ is another conjugate system of A. Then take an internal state $\omega \in \mathfrak{S}_1(\mathrm{A})$ and consider its purifications $\Psi \in \mathfrak{S}_1(\mathrm{A}\tilde{\mathrm{A}})$ and $\Psi' \in \mathfrak{S}_1(\mathrm{A}\tilde{\mathrm{A}}')$. By the uniqueness of purification expressed by Lemma 21, since Ψ and Ψ' are purifications of the same state, there are two channels $\mathscr{C} \in \mathfrak{T}(\tilde{\mathrm{A}}, \tilde{\mathrm{A}}')$ and $\mathscr{D} \in \mathfrak{T}(\tilde{\mathrm{A}}', \tilde{\mathrm{A}})$ such that

$$\begin{aligned} |\Psi')_{\mathrm{A}\tilde{\mathrm{A}}'} &= \mathscr{C} |\Psi)_{\mathrm{A}\tilde{\mathrm{A}}} && (197)\\ |\Psi)_{\mathrm{A}\tilde{\mathrm{A}}} &= \mathscr{D} |\Psi')_{\mathrm{A}\tilde{\mathrm{A}}'}. && (198) \end{aligned}$$

Clearly, this implies that

$$\begin{aligned} |\Psi)_{\mathrm{A}\tilde{\mathrm{A}}} &= \mathscr{D}\mathscr{C} |\Psi)_{\mathrm{A}\tilde{\mathrm{A}}} && (199)\\ |\Psi')_{\mathrm{A}\tilde{\mathrm{A}}'} &= \mathscr{C}\mathscr{D} |\Psi')_{\mathrm{A}\tilde{\mathrm{A}}'}. && (200) \end{aligned}$$

On the other hand, by the previous Lemma the states Ψ and Ψ' are dynamically faithful for systems $\tilde{\mathrm{A}}$ and $\tilde{\mathrm{A}}'$, respectively. Hence, one has $\mathscr{D}\mathscr{C} = \mathscr{I}_{\tilde{\mathrm{A}}}$ and $\mathscr{C}\mathscr{D} = \mathscr{I}_{\tilde{\mathrm{A}}'}$, namely the channels $\mathscr{C}$ and $\mathscr{D}$ are reversible. By Definition 5, this means that A and A$'$ are operationally equivalent. ■

B. States-transformations isomorphism for conjugate purifying systems

If we use conjugate purifying systems to build up dynamically faithful states some of the results derived so far become simpler and more elegant. First of all, according to Lemma 30, if a pure state $\Psi_{\mathrm{A}\tilde{\mathrm{A}}}$ is dynamically faithful for system A, then it is also dynamically faithful for system $\tilde{\mathrm{A}}$. This means that we can simply use the expression "dynamically faithful pure state $|\Psi)_{\mathrm{A}\tilde{\mathrm{A}}}$" without further specifications. Accordingly, we will drop the superscript A in the state $|\Psi^{(\mathrm{A})})$. We now show that we can also drop the condition Eq. (157) in the isomorphism between transformations and bipartite states:

Theorem 26 (Strong version of the states-transformations isomorphism) *The storing map $\mathscr{C} \mapsto |R_{\mathscr{C}})_{\mathrm{B}\tilde{\mathrm{A}}} := \mathscr{C}\,|\Psi)_{\mathrm{A}\tilde{\mathrm{A}}}$, with Ψ dynamically faithful pure state, has the following properties:*

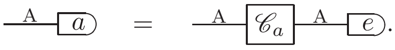
1. *it defines a bijective correspondence between tests $\{\mathscr{C}_i\}_{i\in\mathrm{X}}$ from A to B and preparation-tests $\{R_i\}_{i\in\mathrm{X}}$ for $\mathrm{B}\tilde{\mathrm{A}}$ satisfying*

$$\sum_{i\in\mathrm{X}} (e|_{\mathrm{B}}\,|R_i)_{\mathrm{B}\tilde{\mathrm{A}}} = (e|_{\mathrm{A}}\,|\Psi)_{\mathrm{A}\tilde{\mathrm{A}}}. \tag{201}$$

2. *a transformation $\mathscr{C}$ is atomic (according to Definition 22) if and only if the corresponding state $R_{\mathscr{C}}$ is pure.*
3. *in convex theory the map $\mathscr{C} \mapsto R_{\mathscr{C}}$ defines a bijective correspondence between the cones $\mathfrak{T}_+(\mathrm{A},\mathrm{B})$ and $\mathfrak{S}_+(\mathrm{B}\tilde{\mathrm{A}})$.*

We now have the following remarkable fact:

Theorem 27 *For every effect $a \in \mathfrak{E}(\mathrm{A})$ there is an atomic transformation $\mathscr{C}_a \in \mathfrak{T}(\mathrm{A})$ such that*



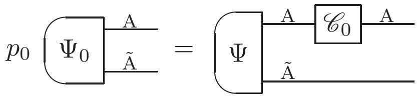
(202)

Moreover, the transformation $\mathscr{C}_a$ is unique up to reversible channels on the output.

Proof. Let p_0 and p_1 be the probabilities defined by $p_0 := (a|_{\mathrm{A}}\,(e|_{\tilde{\mathrm{A}}}\,|\Psi)_{\mathrm{A}\tilde{\mathrm{A}}}$ and $p_1 := (e-a|_{\mathrm{A}}\,(e|_{\tilde{\mathrm{A}}}\,|\Psi)_{\mathrm{A}\tilde{\mathrm{A}}}$. Let $|\Psi_0)_{\mathrm{A}\tilde{\mathrm{A}}}$ and $|\Psi_1)_{\mathrm{A}\tilde{\mathrm{A}}}$ be purifications of the normalized states $|\rho_0)_{\tilde{\mathrm{A}}} := (a|_{\mathrm{A}}\,|\Psi)_{\mathrm{A}\tilde{\mathrm{A}}}/p_0$ and $|\rho_1)_{\tilde{\mathrm{A}}} := (e-a|_{\mathrm{A}}\,|\Psi)_{\mathrm{A}\tilde{\mathrm{A}}}/p_1$, respectively. Now, the collection of states $\{p_0\Psi_0, p_1\Psi_1\}$ is a preparation-test (it can be prepared via randomization). Moreover, such a preparation-test has the property

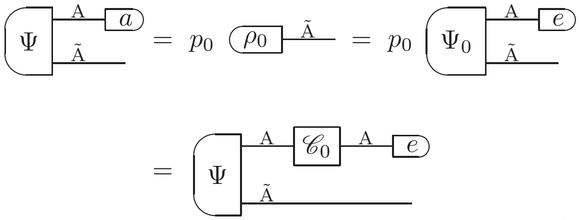
$$p_0\,(e|_{\mathrm{A}}\,|\Psi_0)_{\mathrm{A}\tilde{\mathrm{A}}} + p_1\,(e|_{\mathrm{A}}\,|\Psi_1)_{\mathrm{A}\tilde{\mathrm{A}}} = (e|_{\mathrm{A}}\,|\Psi)_{\mathrm{A}\tilde{\mathrm{A}}}, \tag{203}$$

namely it satisfies Eq. (201). By the states-transformations isomorphism, it must correspond to a test $\{\mathscr{C}_0, \mathscr{C}_1\}$ from A to A: in particular we must have



(204)

Applying the deterministic effect on A we then obtain



(205)

Since Ψ is dynamically faithful, this implies Eq. (202) with $\mathscr{C}_a := \mathscr{C}_0$. Moreover, the states-transformation isomorphism states that $\mathscr{C}_0$ is atomic since $p_0\,|\Psi_0)_{\mathrm{A}\tilde{\mathrm{A}}} = \mathscr{C}_0\,|\Psi)_{\mathrm{A}\tilde{\mathrm{A}}}$ is pure. Finally, suppose that $\mathscr{C}'_0 \in \mathfrak{T}(\mathrm{A})$ is another atomic transformation such that Eq. (202) holds, and define the pure state $|\Psi'_0) := \mathscr{C}'_0\,|\Psi)_{\mathrm{A}\tilde{\mathrm{A}}}/p_0$. Since Ψ_0 and Ψ'_0 are purifications of the same state $|\rho_0)_{\tilde{\mathrm{A}}}$, then they are connected by a reversible channel $\mathscr{U}$ on A. Using the fact that Ψ is dynamically faithful, this implies $\mathscr{C}'_0 = \mathscr{U}\mathscr{C}_0$. ■

Moreover, having conjugate purifying systems allows for a more elegant description of the composition of transformations in terms of composition of states. We recall that to treat the composition of states we need a system of purifications, as defined in Subsect. VIII C. The nice thing now is that we can take the system of purifications to be symmetric:

Definition 63 (Symmetric system of purifications) *A* symmetric system of purification *is a choice of dynamically faithful pure states $|\Psi)_{\mathrm{A}\tilde{\mathrm{A}}}$ and teleportation effects $(E|_{\tilde{\mathrm{A}}\mathrm{A}}$ that satisfies the properties*

$$\begin{aligned} |\Psi)_{\mathrm{A}\mathrm{B}\tilde{\mathrm{A}}\tilde{\mathrm{B}}} &= |\Psi)_{\mathrm{A}\tilde{\mathrm{A}}}\,|\Psi)_{\mathrm{B}\tilde{\mathrm{B}}}\\ (E|_{\tilde{\mathrm{A}}\tilde{\mathrm{B}}\mathrm{A}\mathrm{B}} &= (E|_{\tilde{\mathrm{A}}\mathrm{A}}\,(E|_{\tilde{\mathrm{B}}\mathrm{B}}\,. \end{aligned} \tag{206}$$

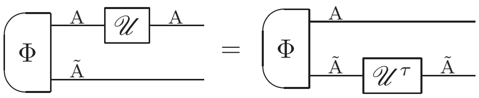
Regarding the probabilities of conclusive teleportation, we now have $p_{\mathrm{A}} = p_{\tilde{\mathrm{A}}}$ (compare Eqs. (113) and (114) in the teleportation protocol of Corollary 19).

In the next Subsection we will see that there is a canonical choice of internal states, namely choosing $|\omega)_{\mathrm{A}} = |\chi)_{\mathrm{A}}$, where $|\chi)_{\mathrm{A}}$ is the unique invariant state of system A (for the uniqueness, see Lemma 34). We will choose a fixed purification of $|\chi)_{\mathrm{A}}$ and refer to it as to the *canonical faithful state*, denoted by $|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}}$. In Corollary 46 we will show that this notation is consistent, since $|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}}$ is also a purification of the unique invariant state of $\tilde{\mathrm{A}}$.

C. Conjugated transformations

The most important consequence of the existence of conjugate purifying systems is the possibility of defining a one-to-one correspondence between the reversible transformations of one system A and the reversible transformations of its conjugate system $\tilde{\mathrm{A}}$. As we will see, this implies in particular the possibility of deterministic teleportation. The correspondence is set by the following Lemma:

Lemma 32 (Transposition of reversible channels) *Let $\Phi \in \mathfrak{S}_1(\mathrm{A}\tilde{\mathrm{A}})$ be a purification of the unique invariant state $\chi \in \mathfrak{S}_1(\mathrm{A})$. Then, for every reversible channel $\mathscr{U} \in \mathbf{G}_{\mathrm{A}}$ there exists a unique reversible channel $\mathscr{U}^\tau \in \mathbf{G}_{\tilde{\mathrm{A}}}$, here called the* transpose *of $\mathscr{U}$ with respect to Φ, such that*



(207)

Transposition is an injective map satisfying the properties

$$\mathscr{I}_{\tilde{\mathrm{A}}}^{\tau} = \mathscr{I}_{\tilde{\mathrm{A}}} \qquad (208)$$

$$(\mathscr{U}_1\mathscr{U}_2)^{\tau} = \mathscr{U}_2^{\tau}\mathscr{U}_1^{\tau}. \qquad (209)$$

Proof. Since $|\chi)_{\mathrm{A}}$ is invariant, the states $|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}}$ and $\mathscr{U}\,|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}}$ are both purifications of it. Then, there must be a reversible transformation $\mathscr{U}^{\tau} \in \mathbf{G}_{\tilde{\mathrm{A}}}$ such that Eq. (207) holds. Moreover, since the invariant state $|\chi)_{\mathrm{A}}$ is internal, its purification Φ is dynamically faithful, both for system A and for system Ã. Dynamical faithfulness on system Ã implies that the transformation $\mathscr{U}^{\tau}$ is uniquely defined, while dynamical faithfulness on system A implies that transposition is injective. Finally, Eq. (208) is obvious, while Eq. (209) is easily proved by repeated application of Eq. (207):

$$\begin{aligned}(\mathscr{I}_{\mathrm{A}} \otimes (\mathscr{U}_1\mathscr{U}_2)^{\tau})\,|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}} &= (\mathscr{U}_1\mathscr{U}_2 \otimes \mathscr{I}_{\tilde{\mathrm{A}}})\,|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}}\\ &= (\mathscr{U}_1 \otimes \mathscr{U}_2^{\tau})\,|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}} \qquad (210)\\ &= (\mathscr{I}_{\mathrm{A}} \otimes \mathscr{U}_2^{\tau}\mathscr{U}_1^{\tau})\,|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}},\end{aligned}$$

using the fact that Φ is dynamically faithful for system Ã. ■

Lemma 33 (Continuity of transposition) *Transposition is continuous with respect to the operational norm. Moreover, if $C \subseteq \mathbf{G}_{\mathrm{A}}$ is closed, then $\tau(C) \subseteq \mathbf{G}_{\tilde{\mathrm{A}}}$ is closed.*

Proof. Let p_{A} be the probability of teleportation for the canonical faithful state $|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}}$. Define $|R_{\mathscr{U}})_{\mathrm{A}\tilde{\mathrm{A}}} := (\mathscr{U} \otimes \mathscr{I}_{\tilde{\mathrm{A}}})\,|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}}$. For every $\epsilon > 0$, if $\mathscr{U}, \mathscr{V} \in \mathbf{G}_{\mathrm{A}}$ are such that $\|\mathscr{U} - \mathscr{V}\|_{\mathrm{A},\mathrm{A}} < \epsilon$, then using Eq. (126) one has $\|\mathscr{U}^{\tau} - \mathscr{V}^{\tau}\|_{\tilde{\mathrm{A}},\tilde{\mathrm{A}}} \leq \|R_{\mathscr{U}} - R_{\mathscr{V}}\|_{\mathrm{A}\tilde{\mathrm{A}}}/p_{\mathrm{A}} < \epsilon/p_{\mathrm{A}}$. This proves continuity. Now, suppose that $C \subseteq \mathbf{G}_{\mathrm{A}}$ is a closed set, and suppose that $\{\mathscr{U}_n^{\tau}\}$ is a sequence in $\tau(C)$ converging to some reversible transformation $\mathscr{V} \in \mathbf{G}_{\tilde{\mathrm{A}}}$. It is easy to see that $\mathscr{V}$ must be in $\tau(C)$. Indeed, consider the sequence $\{\mathscr{U}_n\} \subset \mathbf{G}_{\mathrm{A}}$. Since $\mathbf{G}_{\mathrm{A}}$ is compact, there must be a subsequence $\mathscr{U}_{n_k}$ such that $\mathscr{U}_{n_k} \to \mathscr{U}$ for some $\mathscr{U} \in \mathbf{G}_{\mathrm{A}}$. Moreover, since C is closed, one has $\mathscr{U} \in C$. Now, using continuity we obtain $\mathscr{U}_{n_k}^{\tau} \to \mathscr{U}^{\tau}$. This implies that $\mathscr{V} = \lim_{n\to\infty} \mathscr{U}_n^{\tau} = \mathscr{U}^{\tau}$, that is, the limit point is in $\tau(C)$. Hence, $\tau(C)$ is closed. ■

Lemma 34 *The transposition map $\tau : \mathscr{U} \mapsto \mathscr{U}^{\tau}$ defined in Eq. (207) is surjective on $\mathbf{G}_{\tilde{\mathrm{A}}}$.*

Proof. Take the invariant state $|\chi)_{\tilde{\mathrm{A}}}$, a purification of it, say $\left|\Phi^{(\tilde{\mathrm{A}})}\right)_{\mathrm{A}\tilde{\mathrm{A}}}$, and define the transpose $\tilde{\tau}$ with respect to $\Phi^{(\tilde{\mathrm{A}})}$. Since τ and $\tilde{\tau}$ are both injective transformations, their composition $\iota := \tau\tilde{\tau} : \mathbf{G}_{\tilde{\mathrm{A}}} \to \mathbf{G}_{\tilde{\mathrm{A}}}$ is injective too. Moreover, ι is a homomorphism, since $\iota(\mathscr{I}_{\tilde{\mathrm{A}}}) = \mathscr{I}_{\tilde{\mathrm{A}}}$ and $\iota(\mathscr{V}\mathscr{W}) = \iota(\mathscr{V})\iota(\mathscr{W})$ for every $\mathscr{V}, \mathscr{W}$ in $\mathbf{G}_{\tilde{\mathrm{A}}}$. We now claim that ι is surjective. Of course, since $\iota := \tau\tilde{\tau}$, this will also prove that τ is surjective. Consider the sequence $\{\mathbf{H}_n\}$ defined by $\mathbf{H}_n := \iota^n(\mathbf{G}_{\tilde{\mathrm{A}}})$. By the previous Lemma 33, each $\mathbf{H}_n$ is a closed subgroup of $\mathbf{G}_{\tilde{\mathrm{A}}}$, and one has

$$\mathbf{G}_{\mathrm{A}} := \mathbf{H}_0 \supseteq \mathbf{H}_1 \supseteq \cdots \supseteq \mathbf{H}_n \supseteq \mathbf{H}_{n+1}, \qquad (211)$$

namely $\{\mathbf{H}_n\}$ is a descending chain of subgroups of $\mathbf{G}_{\tilde{\mathrm{A}}}$. Since $\mathbf{G}_{\tilde{\mathrm{A}}}$ is a compact Lie group, every descending chain of closed subgroups must be eventually constant (see e.g. p. 136 of [68]), i.e. there exists a finite $\bar{n}$ such that

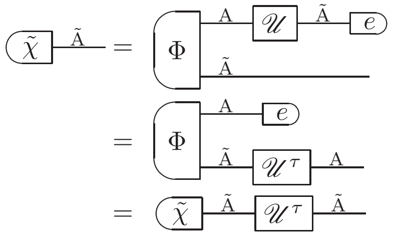
$$\mathbf{H}_n = \mathbf{H}_{n+1} \qquad n \geq \bar{n}. \qquad (212)$$

Applying ι^{-n} on both sides, this implies $\mathbf{H}_0 = \mathbf{H}_1$, namely $\mathbf{G}_{\tilde{\mathrm{A}}} = \iota(\mathbf{G}_{\tilde{\mathrm{A}}})$. Therefore, ι is surjective. ■

The first consequences of the properties of transposition are given by the following corollary

Corollary 46 *Let $\Phi \in \mathfrak{S}_1(\mathrm{A}\tilde{\mathrm{A}})$ be a purification of the unique invariant state $\chi_{\mathrm{A}} \in \mathfrak{S}_1(\mathrm{A})$. Then the complementary state $|\tilde{\chi})_{\tilde{\mathrm{A}}} := (e|_{\mathrm{A}}\,|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}}$ is the unique invariant state of Ã.*

Proof. For every $\mathscr{U} \in \mathbf{G}_{\mathrm{A}}$ we have

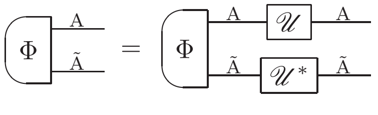


(213)

Since τ is surjective, $\mathscr{U}^{\tau}$ is an arbitrary element of $\mathbf{G}_{\tilde{\mathrm{A}}}$, hence $\tilde{\chi}$ is invariant. ■

Definition 64 (Conjugate of a reversible channel) *The* conjugate *of the reversible channel $\mathscr{U} \in \mathfrak{T}(\mathrm{A})$ with respect to the state $\Phi \in \mathfrak{S}(\mathrm{A}\tilde{\mathrm{A}})$ is the reversible channel $\mathscr{U}^{*} \in \mathfrak{T}(\tilde{\mathrm{A}})$ defined by $\mathscr{U}^{*} := (\mathscr{U}^{\tau})^{-1}$, where the transpose is defined with respect to Φ.*

Note that with this definition the canonical faithful state $|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}}$ is *isotropic*, i.e. it is invariant under combined reversible channels on the conjugate systems A and Ã:



$$\forall \mathscr{U} \in \mathbf{G}_{\mathrm{A}}. \qquad (214)$$

Moreover, we have also the converse:

Corollary 47 (Isotropic states) *A pure state $\Psi \in \mathfrak{S}_1(\mathrm{A}\tilde{\mathrm{A}})$ is isotropic if and only if $|\Psi)_{\mathrm{A}\tilde{\mathrm{A}}} = (\mathscr{V} \otimes \mathscr{I}_{\tilde{\mathrm{A}}})\,|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}}$ for some reversible $\mathscr{V} \in \mathbf{G}_{\mathrm{A}}$ such that*

$$\mathscr{U}\mathscr{V} = \mathscr{V}\mathscr{U} \qquad \forall \mathscr{U} \in \mathbf{G}_{\mathrm{A}}. \qquad (215)$$

Proof. Clearly, a state of the above form is isotropic. Conversely, if Ψ is isotropic, it satisfies Eq. (214), and, therefore, its marginal state on system $\tilde{\mathrm{A}}$ is the invariant state $|\chi)_{\tilde{\mathrm{A}}}$. Since Ψ and Φ are purifications of the same state, there must exist a reversible channel $\mathscr{V} \in \mathbf{G}_{\mathrm{A}}$ such that $|\Psi)_{\mathrm{A}\tilde{\mathrm{A}}} = \mathscr{V}\,|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}}$. The isotropy condition then gives

$$\begin{aligned}(\mathscr{V}\otimes\mathscr{I}_{\tilde{\mathrm{A}}})\,|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}} &= |\Psi)_{\mathrm{A}\tilde{\mathrm{A}}}\\ &= (\mathscr{U}\otimes\mathscr{U}^*)\,|\Psi)_{\mathrm{A}\tilde{\mathrm{A}}}\\ &= (\mathscr{U}\mathscr{V}\otimes\mathscr{U}^*)\,|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}}\\ &= (\mathscr{U}\mathscr{V}\mathscr{U}^{-1}\otimes\mathscr{I}_{\tilde{\mathrm{A}}})\,|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}}\end{aligned} \tag{216}$$

Dynamical faithfulness of Φ then implies $\mathscr{V} = \mathscr{U}\mathscr{V}\mathscr{U}^{-1}$, namely Eq. (215). ■

Recalling that the *center* of the group $\mathbf{G}_{\mathrm{A}}$ is the set of all elements $\mathscr{V} \in \mathbf{G}_{\mathrm{A}}$ such that $\mathscr{U}\mathscr{V} = \mathscr{V}\mathscr{U}$ for every $\mathscr{U} \in \mathbf{G}_{\mathrm{A}}$, it is immediate to state the following

Corollary 48 *The canonical faithful state $|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}}$ is the unique isotropic state of system $\mathrm{A}\tilde{\mathrm{A}}$ if and only if the compact Lie group $\mathbf{G}_A$ has trivial center.*

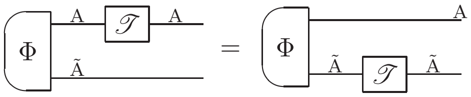
The conclusion of this Subsection is summarized by the following theorem:

Theorem 28 (Isomorphism of groups) *The reversible channels on* A *and* $\tilde{\mathrm{A}}$ *form two isomorphic Lie groups, with the isomorphism given by the conjugation map* $* : \mathbf{G}_{\mathrm{A}} \to \mathbf{G}_{\tilde{\mathrm{A}}}, \mathscr{U} \mapsto \mathscr{U}^*$.

Proof. Clearly, $*$ is a homomorphism, namely $\mathscr{I}^*_{\mathrm{A}} = \mathscr{I}_{\tilde{\mathrm{A}}}$ and $(\mathscr{U}_1\mathscr{U}_2)^* = ((\mathscr{U}_1\mathscr{U}_2)^\tau)^{-1} = \mathscr{U}_1^*\mathscr{U}_2^*$. Moreover, $*$ is injective and surjective, since it is the composition of two injective and surjective maps, namely transposition and inversion. ■

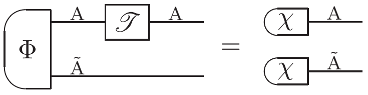
D. Deterministic teleportation

Lemma 35 *Let $\mathscr{T}_{\mathrm{A}}$ and $\mathscr{T}_{\tilde{\mathrm{A}}}$ be the twirling channels on* A *and* $\tilde{\mathrm{A}}$*, respectively, and let $\Phi \in \mathfrak{S}_1(\mathrm{A}\tilde{\mathrm{A}})$ be the canonical faithful state. Then, one has*



(217)

and



(218)

Proof. We have

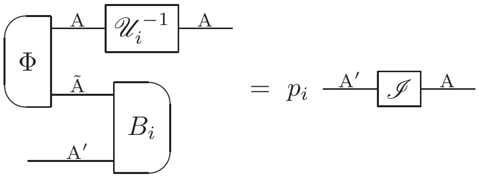
$$\begin{aligned}\mathscr{T}_{\mathrm{A}}\,|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}} &= \int_{\mathbf{G}_{\mathrm{A}}} \mathrm{d}\mathscr{U}\ \mathscr{U}\,|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}}\\ &= \int_{\mathbf{G}_{\mathrm{A}}} \mathrm{d}\mathscr{U}\ (\mathscr{U}^*)^{-1}\,|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}}\\ &= \mathscr{T}_{\tilde{\mathrm{A}}}\,|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}},\end{aligned} \tag{219}$$

having used the fact that $\mathbf{G}_{\mathrm{A}}$ and $\mathbf{G}_{\tilde{\mathrm{A}}}$ are isomorphic, and, therefore, have the same Haar measure. Moreover, since the output of the twirling channel is an invariant state, we have that $|\sigma)_{\mathrm{A}\tilde{\mathrm{A}}} := \mathscr{T}_{\mathrm{A}}\,|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}}$ is invariant under local reversible transformations, i.e.

$$|\sigma)_{\mathrm{A}\tilde{\mathrm{A}}} = (\mathscr{U}\otimes\mathscr{V})\,|\sigma)_{\mathrm{A}\tilde{\mathrm{A}}} \qquad \forall \mathscr{U} \in \mathbf{G}_{\mathrm{A}}, \forall \mathscr{V} \in \mathbf{G}_{\tilde{\mathrm{A}}}. \tag{220}$$

Finally, we invoke Theorem 4, which states that the unique state invariant under local reversible transformations is $\chi_{\mathrm{A}}\otimes\chi_{\tilde{\mathrm{A}}}$. ■

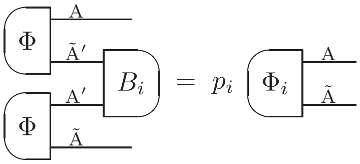
Theorem 29 (Deterministic teleportation) *Let* A *and* A' *be two operationally equivalent systems, and let $\{p_i\mathscr{U}_i\}_{i\in\mathrm{X}}$ be a twirling test, where each $\mathscr{U}_i$ is a reversible channel on* A*. Then there exists an observation-test $\{B_i\}_{i\in\mathrm{X}}$ on $\tilde{\mathrm{A}}\mathrm{A}'$ such that for every outcome i one has*



(221)

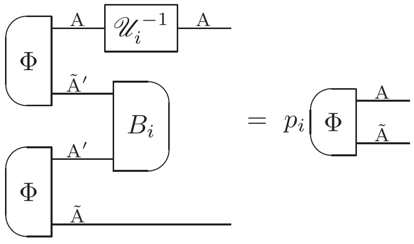
Moreover, each effect B_i must be atomic.

Proof. Define the preparation-test $\{p_i\Phi_i\}_{i\in\mathrm{X}}$ with $|\Phi_i)_{\mathrm{A}\tilde{\mathrm{A}}} := \mathscr{U}_i\,|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}}$. By the previous Lemma, we have $\sum_i p_i\Phi_i = \chi_{\mathrm{A}}\otimes\chi_{\tilde{\mathrm{A}}}$, namely coarse-graining of the preparation-test $\{p_i\Phi_i\}$ yields the invariant state of $\mathrm{A}\tilde{\mathrm{A}}$. By the states-transformations isomorphism of Theorem 26, there exists an observation-test on $\tilde{\mathrm{A}}'\mathrm{A}'$, say $\{B_i\}_{i\in\mathrm{X}}$, such that



(222)

Clearly, the states-transformations isomorphism implies that each effect B_i must be atomic (indeed, the corresponding state is pure). Applying $\mathscr{U}_i^{-1}$ on system A we obtain



(223)

The thesis follows from the fact that Φ is dynamically faithful. ■

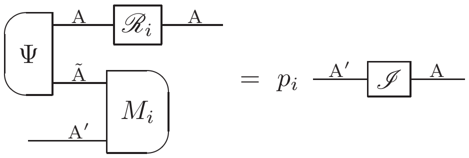
In theories with local discriminability we have the additional result:

Corollary 49 *Let $\{p_i\mathscr{U}_i\}_{i\in\mathrm{X}}$ be a twirling test where each $\mathscr{U}_i$ is a reversible channel. In a theory with local discriminability the number of outcomes $|\mathrm{X}|$ cannot be smaller than* $\dim\mathfrak{S}_{\mathbb{R}}(\mathrm{A})$.

Proof. By Eq. (221) the state $\Psi_i := (\mathscr{U}_i^{-1}\otimes\mathscr{I}_{\mathrm{A}})\Phi$ and the effect B_i achieve teleportation with probability p_i. In a theory with local discriminability the bound of Eq. (116) gives $p_i \leq 1/\dim\mathfrak{S}_{\mathbb{R}}(\mathrm{A})$. We then have $1 = \sum_{i\in\mathrm{X}} p_i \leq |\mathrm{X}|/\dim\mathfrak{S}_{\mathbb{R}}(\mathrm{A})$. ■

If two parties share the pure state $|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}}$, then by the teleportation protocol they can convert it in an arbitrary state $\Psi\in\mathfrak{S}_1(\mathrm{A}\tilde{\mathrm{A}})$ using only local operations and one round of classical communication (one-way LOCC). We now show that the state $|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}}$ is the *maximally entangled* state of $\mathfrak{S}_1(\mathrm{A}\tilde{\mathrm{A}})$, that is, if we can convert another state Ψ to Φ by one-way LOCC, then $\Psi = (\mathscr{U}\otimes\mathscr{I}_{\mathrm{A}})\Phi$ for some local reversible channel $\mathscr{U}\in\mathfrak{T}(\mathrm{A})$. To see that, we show that if Ψ allows for deterministic teleportation, then $\Psi = (\mathscr{U}\otimes\mathscr{I}_{\mathrm{A}})\Phi$.

Theorem 30 (Unique structure of deterministic teleportation) *Let $\Psi\in\mathfrak{S}_1(\mathrm{A}\tilde{\mathrm{A}})$ be a pure state, $\{\mathscr{R}_i\}_{i\in\mathrm{X}}$ be a collection of channels on* A*, $\{p_i\}_{i\in\mathrm{X}}$ a set of probabilities, and $\{M_i\}_{i\in\mathrm{X}}$ be an observation-test on $\tilde{\mathrm{A}}\mathrm{A}'$, with A' and* A *operationally equivalent systems. If for every outcome i one has*

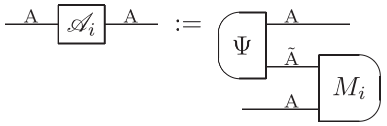


(224)

then

1. *each channel $\mathscr{R}_i$ is reversible, namely $\mathscr{R}_i = \mathscr{U}_i^{-1}$ for some $\mathscr{U}_i\in\mathbf{G}_{\mathrm{A}}$*
2. *there is a reversible channel $\mathscr{U}\in\mathbf{G}_{\mathrm{A}}$ such that $\Psi = \mathscr{U}\Phi$*
3. *each effect M_i has the property $(M_i|_{\tilde{\mathrm{A}}\mathrm{A}'}|\chi)_{\tilde{\mathrm{A}}} = p_i\,(e|_{\mathrm{A}'}$*
4. *$\sum_{i\in\mathrm{X}} p_i\mathscr{U}_i = \mathscr{T}$, where $\mathscr{T}$ is the twirling channel*

Proof. Define the transformation $\mathscr{A}_i$ as

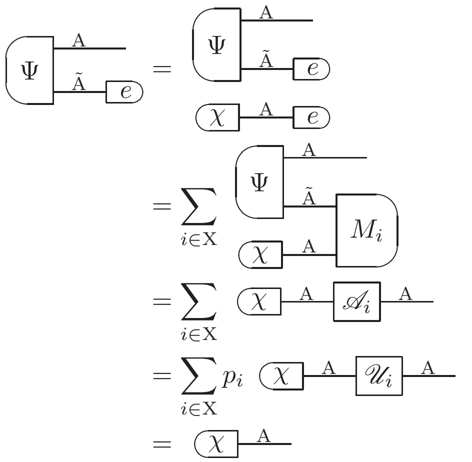


(225)

With this definition we have $\mathscr{R}_i\mathscr{A}_i = p_i\mathscr{I}_{\mathrm{A}}$ for every outcome i. Moreover, applying the deterministic effect on both sides of the equality we obtain

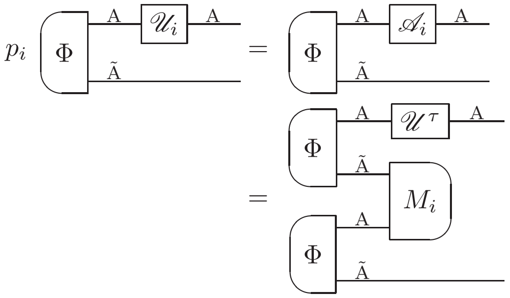
$$(e|_{\mathrm{A}}\,\mathscr{A}_i = (e|_{\mathrm{A}}\,\mathscr{R}_i\mathscr{A}_i = p_i\,(e|_{\mathrm{A}}\,, \qquad (226)$$

that is, each $\mathscr{A}_i$ is proportional to a channel $\mathscr{C}_i$, i.e. $\mathscr{A}_i = p_i\mathscr{C}_i$. We now have $\mathscr{R}_i\circ\mathscr{C}_i = \mathscr{I}_{\mathrm{A}}$, that means that the channel $\mathscr{C}_i$ is invertible. By corollary 27, this implies that $\mathscr{C}_i$ is reversible, namely $\mathscr{C}_i = \mathscr{U}_i$ for some $\mathscr{U}_i\in\mathbf{G}_{\mathrm{A}}$. Clearly, this requires $\mathscr{R}_i = \mathscr{U}_i^{-1}$. Now consider the marginal of Ψ on system A: one has



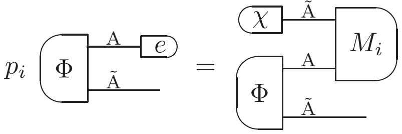
(227)

having used the invariance of χ. But this means that Ψ and Φ have the same marginal on system A, and, therefore, $|\Psi)_{\mathrm{A}\tilde{\mathrm{A}}} = (\mathscr{I}_{\mathrm{A}}\otimes\mathscr{U})\,|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}}$ for some suitable $\mathscr{U}\in\mathbf{G}_{\tilde{\mathrm{A}}}$. Using Lemma 32, we can also transfer $\mathscr{U}$ on system A, getting $|\Psi)_{\mathrm{A}\tilde{\mathrm{A}}} = (\mathscr{U}^{\tau}\otimes\mathscr{I}_{\tilde{\mathrm{A}}})\,|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}}$. Using $\mathscr{A}_i = p_i\mathscr{U}_i$ we then get



(228)

By the states-transformations isomorphism, this means that each M_i is atomic (indeed, the corresponding state is pure). Applying the deterministic effect on system A, the above equation also implies



(229)

which amounts to saying $(M_i|_{\tilde{\mathrm{A}}\mathrm{A}}\,|\chi)_{\tilde{\mathrm{A}}} = p_i\,(e|_{\mathrm{A}}$, because Φ is dynamically faithful. Moreover, summing over the outcomes in Eq. (228) we obtain $\left(\sum_i p_i\mathscr{U}_i\right)|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}} = |\chi)_{\mathrm{A}}\,|\chi)_{\tilde{\mathrm{A}}} = \mathscr{T}\,|\Phi)_{\mathrm{A}\tilde{\mathrm{A}}}$. Again, since Φ is dynamically faithful, this implies $\sum_i p_i\mathscr{U}_i = \mathscr{T}$. ■

In a theory with local discriminability one has also the following result:

Corollary 50 *Let $|\Psi)_{\mathrm{A}\tilde{\mathrm{A}}}$, $\{\mathscr{R}_i\}_{i\in\mathrm{X}}$, $\{p_i\}_{i\in\mathrm{X}}$, and $\{M_i\}_{i\in\mathrm{X}}$ be the state, the recovery channels, the probabilities, and the observation-test in a deterministic teleportation protocol, as in Theorem 30. In a theory with* *local discriminability the number of outcomes satisfies the bound* $|\mathrm{X}| \geq \dim \mathfrak{S}_{\mathbb{R}}(\mathrm{A})$. *The bound is achieved if and only if* $p_i = 1/\dim \mathfrak{S}_{\mathbb{R}}(\mathrm{A})$ *for every* i, *and the states* $|\Psi_i)_{\mathrm{A}\tilde{\mathrm{A}}} := (\mathscr{R}_i \otimes \mathscr{I}_{\tilde{\mathrm{A}}})\,|\Psi)_{\mathrm{A}\tilde{\mathrm{A}}}, \quad i \in \mathrm{X}$ *are perfectly distinguishable with the observation-test* $\{M_i\}$, i.e.

$$(M_i|\,\Psi_j)_{\mathrm{A}\tilde{\mathrm{A}}} = \delta_{ij} \tag{230}$$

Proof. From Eq. (115) we have $p_i \leq 1/\dim \mathfrak{S}_{\mathbb{R}}(\mathrm{A})$ for every i, and, therefore $1 \leq |\mathrm{X}|/\dim \mathfrak{S}_{\mathbb{R}}(\mathrm{A})$. Clearly, the bound is achieved if and only if $p_i = 1/\dim \mathfrak{S}_{\mathbb{R}}(\mathrm{A})$ for every i. In this case, it can be seen from the proof of Eq. (115) that one has $(M_i|\,\Psi_i)_{\mathrm{A}\tilde{\mathrm{A}}} = 1$. Since $\{M_i\}_{i\in\mathrm{X}}$ is an observation-test, and the probabilities of all outcomes must sum up to unit, this implies $(M_j|\,\Psi_i)_{\mathrm{A}\tilde{\mathrm{A}}} = \delta_{ji}$. $\blacksquare$

The above Corollary shows that if teleportation has the minimum possible number of outcomes $|\mathrm{X}| = \dim \mathfrak{S}_{\mathbb{R}}(\mathrm{A})$, then *dense coding* is possible: By acting locally on one side of the state Ψ one can produce $\dim \mathfrak{S}_{\mathbb{R}}(\mathrm{A})$ perfectly distinguishable states. This number exceeds the maximum number of perfectly distinguishable states available in system A, which must be strictly smaller than $\dim \mathfrak{S}_{\mathbb{R}}(\mathrm{A})$ due to Corollary 15. However, we didn't prove here the existence of such a teleportation scheme with $|\mathrm{X}| = \dim \mathfrak{S}_{\mathbb{R}}(\mathrm{A})$. This issue, which is closely related to the topic of discrimination in theories with purification, will be addressed in a future work.

XV. CONCLUSIONS AND PERSPECTIVES ON FUTURE WORK

In this paper we investigated causal probabilistic theories with purification, and derived a surprising wealth of features that are characteristic of quantum theory without resorting to the framework of Hilbert spaces or C*-algebras. Among theories with local discriminability, quantum theory appears as the only known one that satisfies the purification principle. The absence of a counterexample and the amount of quantum features derived suggest that quantum theory could be the only causal theory with purification and local discriminability. However, at the moment we do not have a derivation of quantum theory from the purification principle, and the question whether there are other theories satisfying the above postulates remains open.

Any answer to this question would lead to an interesting scenario: If quantum theory is the only causal theory with purification and local discriminability, then the machinery of Hilbert spaces is a quite redundant way to prove theorems that in fact can be derived directly from basic physical notions. What is more, the general proofs of most theorems are simpler and more intuitive than the original quantum proofs. On the other hand, if quantum theory is not the only theory satisfying our postulates, the existence of more general theories, that share with quantum mechanics the basic structure highlighted in this paper, is also a very fascinating perspective. Moreover, abandoning the standard quantum formalism would be interesting especially in view of a possible reconciliation with general relativity. In this direction, particularly appealing is the possibility of dropping causality from our requirements, and of working with non-unique deterministic effects. The study of non-causal theories with purification is expected to provide new insights toward a formulation of quantum gravity. Such an approach would be related to the informational approaches of Hardy [31] and Lloyd [69]. The study of theories with purification in the non-causal setting will be addressed in a forthcoming paper.

Another direction of further research is the generalization of the notion of subsystem. On the one hand, introducing *classical systems* in the theory and clarifying how they can be viewed as subsystems of the non-classical ones is expected to provide an additional structure that will eventually contribute to the full derivation of quantum mecanics. On the other hand, under suitable assumptions, a face of the convex set of states of a system can be considered as the set of states of some subsystem. Following this observation, we plan to consider information-theoretic tasks like state compression in theories with purification, by analyzing the mechanism that leads the state $\rho^{\otimes N}$ to approach a face corresponding to the state space of $M < N$ systems.

Acknowledgments

We thank the anonymous referees for many suggestions that contributed to improve the original manuscript. GC is grateful to R. Spekkens, B. Coecke, R. Colbeck, S. Facchini, A. Bisio, H. Himai, and A. Doering for useful discussions and suggestions. GMD is grateful to L. Hardy for useful discussions. This work is supported by the Italian Ministry of Education through grant PRIN 2008. 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.

---

[1] S. Popescu and D. Rohrlich, Found. Phys. **24**, 379 (1994).
[2] J. Barrett and S. Pironio, Phys. Rev. Lett. **95**, 140401 (2005).
[3] J. Barrett, L. Hardy, and A. Kent, Phys. Rev. Lett. **95**, 010503 (2005).
[4] A. Acín, N. Gisin, and Ll. Masanes, Phys. Rev. Lett. **97**, 120405 (2006).
[5] M. M. Wolf, D. Perez-Garcia, and C. Fernandez, Phys.

Ref. Lett. **103**, 230402 (2009).

[6] Ll. Masanes, A. Acín, and N. Gisin, Phys. Rev. A **73**, 012112 (2006).

[7] H. Barnum, J. Barrett, M. Leifer, A. Wilce, Phys. Rev. Lett. **99**, 240501 (2007).

[8] J. Barrett, Phys. Rev. A **75**, 032304 (2007).

[9] A. J. Short, S. Popescu, and N. Gisin, Phys. Rev. A **73**, 012101 (2006)

[10] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, arXiv:08053553.

[11] M. Pawlowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Zukowski, Nature **461**, 1101 (2009).

[12] G. Birkhoff and J. von Neumann, Ann. Math. **37**, 743 (1936).

[13] G. W. Mackey, *The mathematical foundations of quantum theory* (W. A. Benjamin Inc, New York, 1963).

[14] J. M. Jauch and C. Piron, Helv. Phys. Acta **36**, 837 (1963); C. Piron, Helv. Phys. Acta **37**, 439 (1964).

[15] G. Ludwig, Commun. Math. Phys. **9**, 1 (1968); G. Ludwig, *Foundations of quantum theory* (Springer-Verlag, New York, 1985).

[16] B. Coecke, D. Moore, and A. Wilce, *Current research in operational quantum logic: algebras, categories, languages* (Fundamental theories of physics series, Kluwer Academic Publishers, 2000).

[17] L. Hardy, quant-ph/0101012.

[18] G. M. D'Ariano, to appear in *Philosophy of Quantum Information and Entanglement*, Eds. A. Bokulich and G. Jaeger (Cambridge University Press, Cambridge UK), see also 0807.4383.

[19] G. M. D'Ariano, AIP Conf. Proc. **810**, 114 (2006).

[20] R. Penrose, p. 221 in D. J. A. Welsh, ed., *Combinatorial Mathematics and its Applications*, Academic Press, New York (1971).

[21] A. Joyal and R. Street, Advances in Mathematics **88**, 55 (1991).

[22] P. Selinger, *A survey of graphical languages for monoidal categories*, available at http://www.mathstat.dal.ca/ selinger/papers/graphical.pdf.

[23] B. Coecke, Advanced Studies in Mathematics and Logic **30**, 45 (2006).

[24] S. Mac Lane, *Categories for the Working Mathematician*, Springer-Verlag, 1971.

[25] In other words, we need to put a probabilistic structure on top of the formal language of circuits. This rule would be given by a function π from $\mathfrak{T}(\mathrm{I},\mathrm{I})$ to the interval $[0,1]$, enjoying the properties $\sum_{i\in\mathrm{X}}\pi(p_i)=1$, and $\pi(p_i\otimes q_j)=\pi(p_i\circ q_j)=\pi(p_i)\pi(q_j)$. In this presentation, however, we will omit the function π, and we will directly identify the event p_i with its probability $\pi(p_i)$.

[26] A. S. Holevo, *Probabilistic and Statistical Aspects of quantum theory*, North-Holland, Amsterdam (1983).

[27] Note that the extension by linearity $\hat{\mathscr{C}}_k\left(\sum_i c_i\rho_i\right) := \sum_i c_i\mathscr{C}_k\,|\rho_i)_{\mathrm{A}}$ is well-defined: Indeed, one has $\sum_i c_i\rho_i=0$ if and only if $\sum_i c_i\,(a|\,\rho_i)_{\mathrm{A}}=0$ for every effect $a\in\mathfrak{S}(\mathrm{A})$. Now, if b is an arbitrary effect in $\mathfrak{E}(\mathrm{B})$, then $(b|_{\mathrm{B}}\mathscr{C}_k$ is an effect in $\mathfrak{E}(\mathrm{A})$, and, therefore, one must have $\sum_i c_i\,(b|_{\mathrm{B}}\mathscr{C}\,|\rho_i)=0$. Since b is arbitrary, this implies $\sum_i c_i\mathscr{C}_k\,|\rho_i)=0$, thus proving that the linear extension is well-defined.

[28] E. B. Davies, J. T. Lewis, Comm. Math. Phys. **17**, 239 (1970).

[29] R. W. Spekkens, Phys. Rev. A **75**, 032110 (2007).

[30] If we toss N times a coin with arbitrary bias q_0, all the $\binom{N}{k}$ sequences with k zeros will have the same probability $q_0^k(1-q_0)^{N-k}$. This means that, conditionally to the fact that there are k zeros, we have $\binom{N}{k}$ equiprobable strings, and, therefore, the probability of the first m_k strings with k zeros is $p(m_k|k)=m_k/\binom{N}{k}$. For a given $\epsilon>0$, if the number $\binom{N}{k}$ is sufficiently large, choosing a suitable m_k we can approximate any given probability $p\in[0,1]$ with $p(m_k|k)$, in such a way that $|p-p(m_k|k)|\le\epsilon/2$. Moreover, by choosing a sufficiently large N one can guarantee that the set S_{ok} of values of k that allow for this approximation has total probability $p_{ok}=\sum_{k\in S_{ok}}p_k=\sum_{k\in S_{ok}}\binom{N}{k}q_0^k(1-q_0)^{N-k}\ge 1-\epsilon/2$. If for every k we associate the outcome 0 to the first m_k strings, and the outcome 1 to the remaining ones, this procedure defines a new coin with bias $p_N=\sum_k p_k p(m_k|k)$ satisfying the bound $|p-p_N|\le\epsilon$.

[31] L. Hardy, J. Phys. A **40**, 3081 (2007).

[32] G. Chiribella, G. M. D'Ariano, and P. Perinotti, EPL **83**, 30004 (2008).

[33] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. A **80**, 022339 (2009).

[34] D. Aharonov, A. Kitaev, and N. Nisan. Quantum Circuits with Mixed States. In Proceedings of the 30th Annual ACM Symposium on Theory of Computing (STOC). ACM, (1998).

[35] V. I. Paulsen, *Completely bounded maps and dilations*, Longman Scientific and Technical (1986).

[36] M. Kleinmann, H. Kampermann, T. Meyer, and D. Bruß, Phys. Rev. A **73**, 062309 (2006).

[37] E. Schrödinger, Proceedings of the Cambridge Philosophical Society **32**, 446 (1936).

[38] H. Barnum, C. P. Gaebler, and A. Wilce, arXiv:0912.5532.

[39] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. Lett. **101**, 180504 (2008).

[40] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. Lett. **101**, 060401 (2008).

[41] S. Abramsky and B. Coecke, Proc. of the 19th IEEE conference on Logic in Computer Science (LiCS'04), IEEE Computer Science Press (2004)

[42] The general version Choi-Jamiołkowski isomorphism considered in this paper is an isomorphism between transformations and bipartite states. This is not the same as the mathematical "CJ isomorphism" of Ref. [18], which was instead the isomorphism between the cone of transformations $\mathfrak{T}_+(\mathrm{A},\mathrm{A})$ and the cone of positive bilinear forms on the complex vector space $\mathfrak{E}_{\mathbb{C}}(\mathrm{A})=\mathsf{Span}_{\mathbb{C}}\{\mathfrak{E}(\mathrm{A})\}$.

[43] W. F. Stinespring, Proc. Amer. Math. Soc. **6**, 211 (1955).

[44] M. A. Naimark, Iza. Akad. Nauk USSR, Ser. Mat. **4**, 277 (1940).

[45] M. Ozawa, J. Math. Phys. **25**, 79 (1984).

[46] M.-D. Choi, Lin. Alg. Appl. **10**, 285 (1975).

[47] A. Jamiołkowski, Rep. Math. Phys. **3**, 275 (1972).

[48] Indeed, it is immediate to see that if two sequences of channels $\{\mathscr{C}_m\}$ and $\{\mathscr{D}_n\}$ converge to two channels $\mathscr{C}$ and $\mathscr{D}$, respectively, then the sequence $\{\mathscr{C}_n\mathscr{D}_n\}$ converges to $\mathscr{C}\mathscr{D}$: $\|\mathscr{C}_n\mathscr{D}_n-\mathscr{C}\mathscr{D}\|_{\mathrm{A},\mathrm{A}}\le\|(\mathscr{C}_n-\mathscr{C})\mathscr{D}_n\|_{\mathrm{A},\mathrm{A}}+\|\mathscr{C}(\mathscr{D}_n-\mathscr{D})\|_{\mathrm{A},\mathrm{A}}\le\|\mathscr{C}_n-\mathscr{C}\|_{\mathrm{A},\mathrm{A}}+\|\mathscr{D}_n-\mathscr{D}\|_{\mathrm{A},\mathrm{A}}\xrightarrow{n\to\infty}0$, having used the triangle inequality and the monotonicity property of Lemma 9.

[49] G. B. Folland, *A course in abstract harmonic analysis*, CRC Press (1995).

[50] I. Devetak and P. Shor, Comm. Math. Phys. **256**, 287 (2005).

[51] A. S. Holevo, Probab. Theory and Appl. **51**, 133 (2006).

[52] C. King, K. Matsumoto, M. Nathanson, and M. B. Ruskai, Markov Process and Related Fields **13**, 391 (2007).

[53] M. Horodecki, P. W. Shor, and M. B. Ruskai, Rev. Math. Phys. **15**, 629 (2003).

[54] B. Schumacher, Phys. Rev. A **54**, 2614 (1996).

[55] B. Schumacher and M. A. Nielsen, Phys. Rev. A **54**, 2629 (1996).

[56] H. Barnum, M. A. Nielsen, and B. Schumacher, Phys. Rev. A **57**, 4153 (1998).

[57] D. Kretschmann, D. W. Kribs, and R. W. Spekkens, Phys. Rev. A **78**, 032330 (2008).

[58] D. Beckman, D. Gottesman, M. A. Nielsen, and J. Preskill, Phys. Rev. A **64**, 052309 (2001).

[59] T. Eggeling, D. Schlingemann, and R.F. Werner, Europhys. Lett. **57**, 782-788 (2002).

[60] M. Piani, M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. A **74**, 012305 (2006).

[61] D. Kretschmann and R. F. Werner, Phys. Rev. A **72**, 062323 (2005).

[62] G. Gutoski and J. Watrous, in *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computation (STOC)* (2007), p. 565.

[63] For bipartite channels the term “semicausal” was used to mean “allowing for signalling in one direction at most”, while the term ”causal” was used to mean ”not allowing for signalling (in any direction)”. For an N-partite channel, however, there are $N!$ possible orderings of the input systems, and a causal N-partite channel was then defined as a channel that is “allowing for the propagation of signals along one of these possible orderings”. Of course, for $N = 2$ the two nomenclatures are conflicting.

[64] M. Gregoratti and R. F. Werner, J. Mod. Opt. **50**, 915 (2003).

[65] G. M. D’Ariano, D. Kretschmann, D. Schlingemann, and R. F. Werner, Phys. Rev. A **76**, 032328 (2007).

[66] G. Chiribella, G. M. D’Ariano, P. Perinotti, D. Schlingemann, and R. F. Werner, arXiv:0905.3801.

[67] M.A. Nielsen and I.L. Chuang, Phys. Rev. Lett. **79**, 321 (1997).

[68] T. Bröcker and T. Tom Dieck, *Representations of compact Lie groups*, Springer Verlag (1985).

[69] S. Lloyd, arXiv:quant-ph/0501135