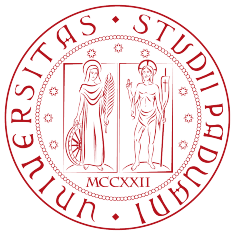



Università degli Studi di Padova

Scuola Galileiana di Studi Superiori

Classe di Scienze Naturali


Tesi di Diploma Galileiano

# A GENERALIZED APPROACH TO RESOURCE THEORIES


**Relatore**
Prof. Pieralberto Marchetti

**Diplomando**
Carlo Maria Scandolo

# Contents

**Introduction** **4**

**1 Introduction to GPTs** **6**
1.1 Basic notions . . . . . 8
1.1.1 Systems and tests . . . . . 8
1.1.2 Sequential and parallel composition . . . . . 10
1.1.3 Operational theories and category theory . . . . . 16
1.1.4 Probabilistic theories . . . . . 20
1.2 Purity and coarse-graining . . . . . 26
1.3 Causality . . . . . 28
1.3.1 Operational norms . . . . . 34

**2 A general framework for resource theories** **39**
2.1 Resource theories . . . . . 40
2.2 A hierarchy among resources . . . . . 47
2.2.1 Some phenomenology . . . . . 51
2.2.2 Resource monotones . . . . . 57
2.3 The general structure of resource theories in GPTs . . . . . 61

**3 Examples of resource theories in quantum mechanics** **64**
3.1 The resource theory of quantum entanglement . . . . . 65
3.1.1 Mixedness relation . . . . . 72
3.1.2 Duality between entanglement and mixedness . . . . . 79
3.1.3 Entanglement monotones . . . . . 84
3.2 The resource theory of purity . . . . . 86
3.2.1 Purity monotones . . . . . 91
3.3 Other examples . . . . . 92
3.3.1 Quantum resource theory of asymmetry . . . . . 92

3.3.2 Quantum resource theory of athermality . . . . . 93

**4 Examples of resource theories in GPTs 96**
4.1 The resource theory of entanglement . . . . . 97
4.2 An operational Lo-Popescu theorem . . . . . 100
4.2.1 Two operational requirements . . . . . 100
4.2.2 Inverting the direction of classical communication . . . 106
4.2.3 Reduction to 1-way LOCC protocols . . . . . 109
4.3 The resource theory of purity . . . . . 110
4.3.1 A resource theory of dynamical control . . . . . 110
4.3.2 From dynamical control to purity . . . . . 112
4.3.3 Maximally mixed states . . . . . 115
4.4 Entanglement-thermodynamics duality . . . . . 117
4.4.1 Purification . . . . . 118
4.4.2 One-way LOCC protocols transforming pure states into pure states . . . . . 120
4.4.3 The more entangled a pure state, the more mixed its marginals . . . . . 123
4.4.4 The more mixed a state, the more entangled its purification . . . . . 124
4.4.5 The duality . . . . . 126
4.5 Consequences of the duality . . . . . 127
4.6 Entanglement and purity monotones . . . . . 129
4.6.1 Information erasure and entanglement generation . . . 131
4.7 Symmetric purification . . . . . 136

**Conclusions 140**

**Acknowledgements 142**

**A Some mathematical results 143**
A.1 A proposition . . . . . 143

# Introduction

In physical theories where the (human) observer plays a crucial role, i.e. theories where physical systems do not arise as external and objective entities, a sensible question we can ask ourselves is up to what extent the observer can control a physical system. Related to this, we have the notion of resource, of which the observer can take advantage to perform some tasks. This is particularly important in a scientific theory having some technological impact. In fact, we can adopt a perspective focused on resources to address a broad class of different scientific theories, encompassing various scientific disciplines.

On of the most paradigmatic examples of a theory of resources is chemistry, especially its industrial branch. Indeed, in this field one wishes to transform "raw" chemical products into useful products (fertilizers, dyes, etc.), which are more valuable resources.

Another theory in which the observer plays a important role is thermodynamics [1, 2]. Here we can distinguish between processes the observer is able to control, which give rise to work, and processes on which the observer has no control, which lead to heat. According to this perspective, the difference between work and heat is solely in the ability one has to control a process. Consequently, work is a more valuable resource than heat in the framework of thermodynamics. Accordingly, heat cannot be completely turned into work (Kelvin's postulate [3]), while the converse can happen. This is because a less valuable resource cannot be transformed for free into a more valuable one. Several approaches to thermodynamics based on resource theories have been proposed so far [4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17], especially from the angle of quantum information, and at the moment this seems to be one of the most promising approaches for an extension of thermodynamic concepts outside the realm of large numbers of particles. In particular, this is significant for a thermodynamic description of systems at the nanoscale,

a topic which has been attracting a lot of attention as a result of recent technological development in the field of nanotechnologies.

However, if there is a theory where the observer plays a crucial role, this is definitely quantum mechanics. This fact has become clear since its foundation and early developments. Therefore, it is interesting to address quantum mechanics from a resource-theoretic viewpoint, and one may expect to find several resource theories related to the quantum world. As an example, consider entanglement, which is a precious resource for communication purposes, in the sense that it enables us to perform communication protocols that would be impossible in classical mechanics (teleportation, dense coding, etc. [18, 19, 20, 21, 22, 23]).

Given that the descriptive power of resource theories encompasses so many different fields, it is a natural question why this happens. There must be an underlying framework common to all resource theories, irrespective to their specific fields. Understanding it is particularly interesting, for it enables one to grasp the similarities between different theories of resources and even to develop new ones.

The aim of this thesis is to understand this common underlying formalism of resource theories and to see how it is realized concretely in some examples taken from quantum information theory. The key methodology of this work is abstraction, and one of the first necessary steps in this direction is to abstract the notion of physical theory itself, in order to understand the common features and the basic structure of physical theories. To this end, we will move to physical theories more general than quantum or classical mechanics, known as general probabilistic theories [24, 25, 26, 27, 28, 29, 30, 31, 32]. By working on such a framework, it will be easier to proceed in our search for a common structure of resource theories, which is given by category theory [33].

Therefore, after presenting the basic notions of the formalism related to general probabilistic theories in chapter 1, we move directly to analyse the general structure of the theories of resources in chapter 2, and we will see that this basic structure arises in a great number of instances. To better understand the abstract formalism for resource theories, we study some interesting examples taken from quantum mechanics in chapter 3. Following our generalization scheme, in chapter 4 we extend these examples to the framework of general probabilistic theories. This extension will enable us to gain a deeper understanding of the structure of the corresponding resource theories in the quantum case.

# Chapter 1

# Introduction to general probabilistic theories

We begin this thesis with an introduction to general probabilistic theories (GPTs), which are one of the most important approaches to the foundations of quantum mechanics. GPTs are generic physical theories admitting a probabilistic structure, with quantum mechanics and classical mechanics as two instances thereof. However, one can devise other examples of such theories, which are neither classical nor quantum, such as the box world theory [25, 34], arising as a generalization of the setting of Bell inequalities (see example 4.2.4 for further details in this respect). GPTs describe what sets of experiments one can do with physical devices, and how to assign probabilities to their outcomes. Therefore, they are a useful framework in which to address a broad class of possible physical theories.

In the past, foundational questions were related essentially to the measurement problem, giving rise to different interpretations of quantum mechanics [35]. However, recently the fields of quantum foundations and quantum information have started to merge, positively influencing each other [36, 37], and GPTs have proved to be a response to the new tendency to focus on the informational content of a theory.

Clearly, one might wonder why we study probabilistic theories more general than quantum mechanics if quantum theory is the theory describing Nature. We can identify some reasons [38].

1. We want to understand quantum mechanics better.
   Indeed, what are the features that single out quantum mechanics among

all the other possible probabilistic theories?

2. We want to study extensions of quantum mechanics.
   Suppose that some day quantum mechanics or some of its axioms will be proved to be wrong. An analysis of more general theories will show how we can modify quantum mechanical axioms to make the theory fit the experiments. Proposals to modify quantum theory have already been presented in the field of quantum gravity [39].

3. We want to study restrictions of quantum mechanics.
   Suppose we are not able to prepare all the states allowed by quantum mechanics. Then, what is our theory like?

In our treatment of general probabilistic theories, we will use a high-level language, borrowed from category theory. The same formalism will be used also in the following chapters, because it is particularly suitable to capture the operational background of a theory, namely, loosely speaking, the way "information is processed" [40]. In this vein, we will carry out our analysis in an abstract way, without resorting to the specific formalism of a certain theory.

Nevertheless, one should not think that our high-level language is completely unrelated to experiments. In fact, it is even closer to an experimental set-up in a laboratory, as we will see presently.

Suppose we have an experimenter in a laboratory. She can build up experiments by connecting devices, and this can be done either sequentially or in parallel. Every device has an input and an output system and possibly some (classical) outcomes that can be read by the experimenter. Each outcome identifies a process which occurred between the input and the output systems when a particular device was applied. In some cases, the experimenter has no control on the outcomes: this means that particular device implements a random process. Some devices have no input: they simply prepare a state. Other devices have no output: they are measurements.

This very simple experimental situation can be treated formally by using a graphical language, in which each device is represented as a box.

Many works have been done on this subject (see for instance [24, 25, 26, 27, 28, 29, 30, 31, 32]); in the present treatment, we will follow the line of reasoning of [28, 29, 41], and of the related works [31, 42]. We present the basic formalism in section 1.1, and in section 1.2 we concentrate on the amount

of information carried by operations performed by an experimenter, defining pure transformations as operations where the information is maximal.

In the study of GPTs is customary to introduce some axioms to restrict our attention only to theories enjoying some reasonable properties. An almost undeniable axiom is Causality, which sets an arrow of time, meaning that operations performed in the "future" cannot have effects on the outcomes of present experiments. Imposing Causality has some interesting consequences for the structure of a probabilistic theory, which are explored in section 1.3.

## 1.1 Basic notions

A GPT has two main parts: the *operational* part and the *probabilistic* part. In this section we introduce first the basic elements of *operational* formalism and their graphical representation, and then we examine the additional features that arise when we introduce the tool of probability theory.

### 1.1.1 Systems and tests

In an operational theory, there are two primitive notions: *systems* and *tests*. We can have an intuition about their meaning by thinking of a concrete experimental situation. A *test* represents a physical device (beam-splitter, polarimeter, Stern-Gerlach magnet, etc.). Every device has an input and an output, which will be called *input* and *output system* respectively. In this way, somehow systems play the role of labels attached to input and output ports of a device.

We denote systems by capital letters in Roman character: A, B, etc. There is also a particular system, the *trivial system*, that simply means "nothing", and we will denote it by letter I. A device with trivial system as input is simply a device with *no* input, and a device with trivial system as output is simply a device with *no* output.

The application of some physical devices can yield various outcomes, where each outcome corresponds to a particular event which occurred in the laboratory and which can be identified by the experimenter. Therefore, we can give the following characterization of tests.

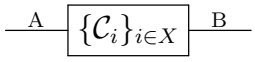
**Definition 1.1.1.** A *test* with input system A and output system B is a collection of *events* $\{\mathcal{C}_i\}_{i\in X}$, labelled by outcome $i$ in some set $X$.

$X$ is called *outcome set*.

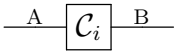
We will often say that $\{\mathcal{C}_i\}_{i\in X}$ is a test *from* system A *to* system B; if A and B coincide, we say that $\{\mathcal{C}_i\}_{i\in X}$ is a test *on* system A.

To clarify the role of outcome $i$ better, we can regard it as what the experimenter actually sees when she performs her experiment (a sequence of digits, a spot in a photographic plate, etc.). The outcome set $X$ is the set containing all the possible outcomes for a given test.

We can represent a test graphically as a box with incoming and outgoing wires representing input and output systems respectively.



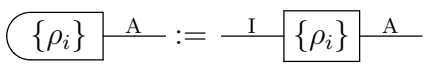

When there is no ambiguity, we will omit the outcome set $X$. If we want to express that the specific event $\mathcal{C}_i$ has occurred, we will write



,

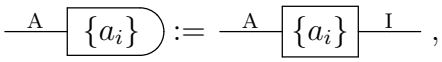
without braces.

Whenever the trivial system I is involved, we omit the corresponding wire and letter. Specifically, when we have no physical input[1] for our device, we have a *preparation-test*, which we represent as



,

namely with a rounded box on its *left* side. Intuitively, preparation-tests prepare a system in a particular "state", although we will clarify this statement later. We will sometimes use the Dirac-like notation $|\rho_i)_{\mathrm{A}}$ for the preparation-event $\rho_i$. The subscript A is intended to highlight the fact that $\rho_i$ is related to system A. Here we use a round bracket to stress the fact that this definition is different and more general than the ket notion in quantum mechanics. Similarly, when we have no physical output[2] for our device, we have an *observation-test*, which we represent as



,

namely with a rounded box on its *right* side. Intuitively, observation-tests destroy a system while acquiring some information from it, so they are related

[1]Recall that *no* physical input means the trivial system I as input.

[2]Again, *no* physical output means the trivial system as output.

to demolition measurements. We will sometimes use the Dirac-like notation $(a_i|_{\mathrm{A}}$ for the observation-event $a_i$.

Finally, if we have a test $\{p_i\}_{i\in X}$ from the trivial system to itself, we omit both the wires and the box.

$$p_i := \ \overset{\mathrm{I}}{\text{—}}\boxed{p_i}\overset{\mathrm{I}}{\text{—}}$$

**Definition 1.1.2.** We say that a test is *deterministic* if its outcome set has one element.

If a test is deterministic, we omit the braces and simply write $\mathcal{C}$ instead of $\{\mathcal{C}\}$. In a non-deterministic test, we cannot predict which particular outcome we will obtain. On the contrary, the outcome of a deterministic test is completely determined. Since we are not able to predict the outcome of non-deterministic tests, we would like to set up a probabilistic structure which enables us at least to define probabilities for the various outcomes. We will address this issue soon, but first some other notions are needed.

### 1.1.2 Sequential and parallel composition

Since we are implementing a graphical language which has a direct link to experimental apparatuses, the next step is to describe how to connect devices. Devices can be connected sequentially or in parallel. Let us start from sequential composition. Intuitively, two devices can be connected sequentially, i.e. one after another, if the output system of the former is the input system of the latter.

**Definition 1.1.3.** If $\{\mathcal{C}_i\}_{i\in X}$ is a test from A to B with outcome set $X$, and $\{\mathcal{D}_j\}_{j\in Y}$ is a test from B to C with outcome set $Y$, we can consider the *sequential composition* $\{\mathcal{D}_j \circ \mathcal{C}_i\}_{(i,j)\in X\times Y}$ , which is a test from A to C and has outcome set $X \times Y$.

Note that sequential composition of tests works exactly as composition of functions: the test $\{\mathcal{D}_j\}_{j\in Y}$ follows the test $\{\mathcal{C}_i\}_{i\in X}$, therefore $\mathcal{D}_j$ is written first.

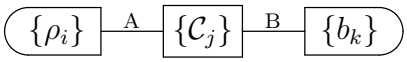
The graphical representation is quite intuitive: suppose we want to compose the event $\mathcal{D}_j$ after the event $\mathcal{C}_i$; we simply write

$$\overset{\mathrm{A}}{\text{—}}\boxed{\mathcal{D}_j \circ \mathcal{C}_i}\overset{\mathrm{C}}{\text{—}} := \overset{\mathrm{A}}{\text{—}}\boxed{\mathcal{C}_i}\overset{\mathrm{B}}{\text{—}}\boxed{\mathcal{D}_j}\overset{\mathrm{C}}{\text{—}} .$$

In this way, there is a natural ordering on tests, given by sequential composition. Indeed, some tests are performed first and other later. In graphical language this ordering goes from left to right: every box follows all the others on its left. However, we must not confuse this ordering with temporal ordering. We will come back to this point in section 1.3.

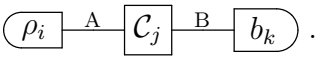
Now let us see an example of sequential composition of tests.

**Example 1.1.4.** Consider the diagram




It gives instructions on how to build up the experiment: first, we initialize system A with the preparation-test $\{\rho_i\}$, then we perform the test $\{\mathcal{C}_j\}$ from A to B and finally we acquire some information from B by destroying it with the observation-test $\{b_k\}$.

If we wish to express which events actually occurred, we write

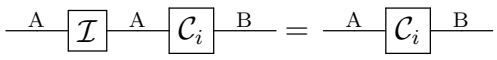



This means that the preparation-event $\rho_i$, the event $\mathcal{C}_j$ and the observation-event $b_k$ occurred. We can represent the whole sequence in Dirac-like notation as $(b_k|\,\mathcal{C}_j\,|\rho_i)$.

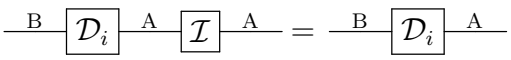
Let us now define the identity test.

**Definition 1.1.5.** The *identity test* for system A is a deterministic test $\mathcal{I}_\mathrm{A}$ on A such that $\mathcal{C}_i \circ \mathcal{I}_\mathrm{A} = \mathcal{C}_i$ for every event $\mathcal{C}_i$ from A to B, and $\mathcal{I}_\mathrm{A} \circ \mathcal{D}_i = \mathcal{D}_i$ for every event $\mathcal{D}_i$ from B to A.

Graphically, we have




for every $\mathcal{C}_i$, and

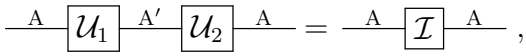



for every $\mathcal{D}_i$. According to this definition, it is clear that for every system A the identity test $\mathcal{I}_\mathrm{A}$ is unique.

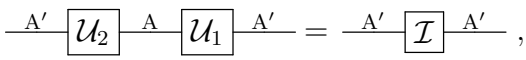
Applying the identity test is just like doing nothing. For this reason we will often omit the box for the identity test.

We sometimes want to "identify" similar system, namely systems that behave exactly in the same way from an operational point of view, yet they are distinct. In quantum mechanics, we can consider the polarization of a photon and the spin of an electron. Although they are completely different physical systems, they are described by the same Hilbert space[3].

**Definition 1.1.6.** We say that system A and system A′ are *operationally equivalent* (and we write $\mathrm{A} \cong \mathrm{A}'$) if there is a deterministic test $\mathcal{U}_1$ from A to A′ and a deterministic test $\mathcal{U}_2$ from A′ to A, such that

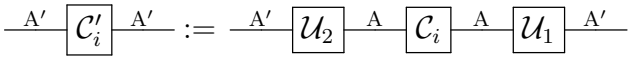



where $\mathcal{I}_{\mathrm{A}}$ is the identity test on A, and




where $\mathcal{I}_{\mathrm{A}'}$ is the identity test on A′.

If $\mathrm{A} \cong \mathrm{A}'$, we can transform tests on system A into tests on system A′ by taking the sequential composition with the intertwining tests $\mathcal{U}_1$ and $\mathcal{U}_2$. Indeed, if $\mathcal{C}_i$ is an event on system A, the corresponding event $\mathcal{C}_i'$ on system A′ is




(1.1.1)

Now we move to the other type of composition: parallel composition. If we have two systems A and B, we can consider them together, forming the composite system $\mathrm{A} \otimes \mathrm{B}$.

**Definition 1.1.7.** If A and B are two systems, the corresponding *composite system* is $\mathrm{A} \otimes \mathrm{B}$. Moreover, system composition has the following properties.

1. $\mathrm{A} \otimes \mathrm{I} = \mathrm{I} \otimes \mathrm{A} = \mathrm{A}$ for every system A;
2. $\mathrm{A} \otimes \mathrm{B} \cong \mathrm{B} \otimes \mathrm{A}$ for all systems $\mathrm{A}, \mathrm{B}$;
3. $\mathrm{A} \otimes (\mathrm{B} \otimes \mathrm{C}) = (\mathrm{A} \otimes \mathrm{B}) \otimes \mathrm{C}$ for all systems $\mathrm{A}, \mathrm{B}, \mathrm{C}$.

These properties have a fairly intuitive meaning.

1. When we combine a system with "nothing", we still have the original system.

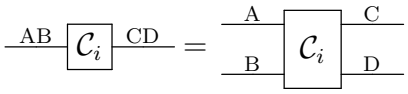
[3]Or by isomorphic Hilbert spaces, to be precise

2. The composition of systems does not depend on the order we compose them.

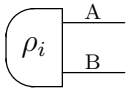
3. This particular form of "associativity" allows us to write simply A⊗B⊗C, without parentheses. Again, the order of composition is irrelevant.

We represent composite systems diagrammatically as a collection of wires one under another. We will typically omit the wire for the trivial system. For the sake of brevity, especially when subscripts are concerned, we will write just AB in place of A ⊗ B.

We can represent an event $\mathcal{C}_i$ from system A ⊗ B to system C ⊗ D as a box with multiple wires, one for each system.

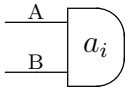



By property 2, it is completely irrelevant to write A rather than B on the upper input wire, and the same holds for every wire. For composite systems we depict preparation-events as




, (1.1.2)

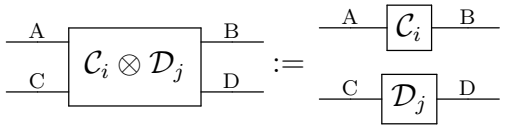
and observation-events as




. (1.1.3)

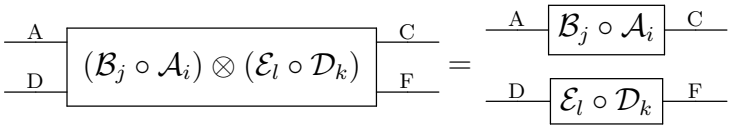
Now we can define the parallel composition of tests.

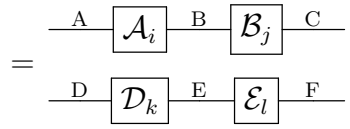
**Definition 1.1.8.** Let $\{\mathcal{C}_i\}_{i\in X}$ be a test from A to B, and let $\{\mathcal{D}_j\}_{j\in Y}$ be a test from C to D. The *parallel composition* $\{\mathcal{C}_i \otimes \mathcal{D}_j\}_{(i,j)\in X\times Y}$ is a test from A⊗C to B⊗D with outcome set $X\times Y$, and it is represented diagrammatically as




.

We can combine parallel and sequential composition in a straightforward way. Suppose $\mathcal{A}_i$ is an event from A to B, $\mathcal{B}_j$ is an event from B to C; $\mathcal{D}_k$ is an event from D to E and $\mathcal{E}_l$ is an event from E to F. Then we have



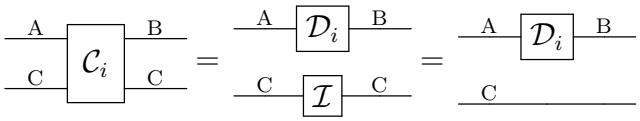



(1.1.4)

If we parallel-compose a test from A to B with the identity test $\mathcal{I}_{\mathrm{C}}$ on another system C, we have a test from $\mathrm{A} \otimes \mathrm{C}$ to $\mathrm{B} \otimes \mathrm{C}$ that in fact acts only on A.

**Definition 1.1.9.** Consider a test $\{\mathcal{C}_i\}_{i \in X}$ from the composite system $\mathrm{A} \otimes \mathrm{C}$ to $\mathrm{B} \otimes \mathrm{C}$. If $\{\mathcal{C}_i\}_{i \in X}$ acts only on A (from A to B), we say that it is a *local test* from A to B.

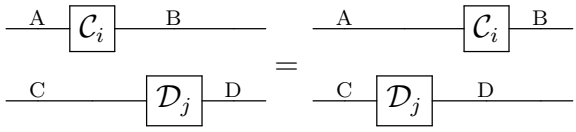
In other words a local test $\{\mathcal{C}_i\}_{i \in X}$ from $\mathrm{A} \otimes \mathrm{C}$ to $\mathrm{B} \otimes \mathrm{C}$ is such that $\mathcal{C}_i = \mathcal{D}_i \otimes \mathcal{I}_{\mathrm{C}}$, for some test $\{\mathcal{D}_i\}_{i \in X}$ from system A to system B.



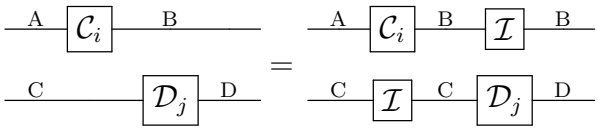

We will write simply $\mathcal{D}_i$ in formulas in place of $\mathcal{D}_i \otimes \mathcal{I}_{\mathrm{C}}$, for example we will write $\mathcal{D}_i \rho_{\mathrm{AC}}$ instead of $(\mathcal{D}_i \otimes \mathcal{I}_{\mathrm{C}})\, \rho_{\mathrm{AC}}$.

We can prove that local tests on different systems commute, like in quantum mechanics,

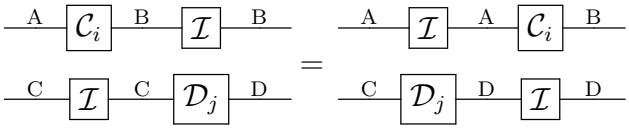
**Proposition 1.1.10.** *Let* $\{\mathcal{C}_i\}_{i \in X}$ *be a test from system* A *to system* B*, and let* $\{\mathcal{D}_j\}_{j \in Y}$ *be a test from system* C *to system* D*. Then we have*



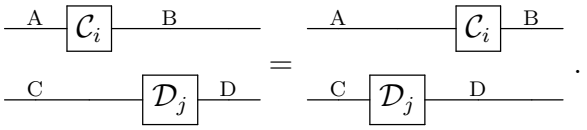

*Proof.* Recall that we can insert the identity test whenever we have a wire. In this way,




Recall that every event commutes with the identity test.



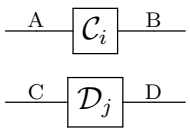
or, in other terms,



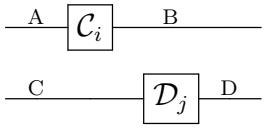

.

□

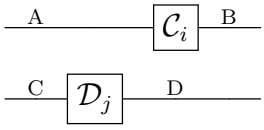
We are then entitled to write




in place of




or

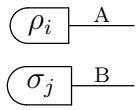



,

since the order with which the two events take place is irrelevant. This also shows that the parallel composition of two tests can be seen as a sequential composition of two local tests on different systems.

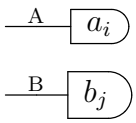
Note that we can compose preparation-tests only in parallel; the same holds for observation-tests, so we will sometimes write simply $|\rho_i)_{\mathrm{A}} |\sigma_j)_{\mathrm{B}}$ in place of $\rho_{i,\mathrm{A}} \otimes \sigma_{j,\mathrm{B}}$; and $(a_i|_{\mathrm{A}} (b_j|_{\mathrm{B}}$ in place of $a_{i,\mathrm{A}} \otimes b_{j,\mathrm{B}}$. Diagrammatically,

$$|\rho_i)_{\mathrm{A}} |\sigma_j)_{\mathrm{B}} = \tag{1.1.5}$$




and

$$(a_i|_{\mathrm{A}} (b_j|_{\mathrm{B}} = \tag{1.1.6}$$




*Remark* 1.1.11. When there is no ambiguity in what kind of composition we are considering, we will write it simply as a product. For instance, if $\mathcal{C}_i$ is an event from A to B and $\mathcal{D}_j$ is an event from B to C, we will write $\mathcal{D}_j \circ \mathcal{C}_i$ simply as $\mathcal{D}_j\mathcal{C}_i$.

Now we can define operational theories.

**Definition 1.1.12.** An *operational theory* is given by a collection of systems, closed under composition, and a collection of tests, closed under sequential and parallel composition.

### 1.1.3 Operational theories and category theory

Although our graphical language can seem naive and not so sound, it has very strong foundations in category theory [43, 44, 45, 40, 46, 47]. Therefore, we are entitled to use graphical language to prove theorems in abstract scenarios for operational theories.

At this point, although we do not wish to enter too much into mathematical details of category theory, it is anyway worthwhile to give some basic definitions [46, 48], which will turn out to be useful for the next chapters as well. We will follow the approach of ref. [46].

**Definition 1.1.13.** A *category* $(\mathbf{C}, \circ)$ is given by

1. a *class*[4] $|\mathbf{C}|$ of *objects*;

2. for any two objects $\mathrm{A}, \mathrm{B} \in |\mathbf{C}|$, a set $\mathbf{C}(\mathrm{A}, \mathrm{B})$ of maps from A to B, called *morphisms*. Such maps can be composed, i.e. if $f$ is a map from A to B ($f \in \mathbf{C}(\mathrm{A}, \mathrm{B})$), and $g$ is a map from B to C ($g \in \mathbf{C}(\mathrm{B}, \mathrm{C})$), there is a map $g \circ f \in \mathbf{C}(\mathrm{A}, \mathrm{C})$ from A to C. Such composition is *associative* and has an *identity*.

Requiring that $\circ : \mathbf{C}(\mathrm{A}, \mathrm{B}) \times \mathbf{C}(\mathrm{B}, \mathrm{C}) \longrightarrow \mathbf{C}(\mathrm{A}, \mathrm{C})$ is associative means that for all $\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D} \in |\mathbf{C}|$ and for all $f \in \mathbf{C}(\mathrm{A}, \mathrm{B})$, $g \in \mathbf{C}(\mathrm{B}, \mathrm{C})$ and $h \in \mathbf{C}(\mathrm{C}, \mathrm{D})$, one has $(h \circ g) \circ f = h \circ (g \circ f)$. If such a composition has an identity, for every object A there exists a morphism $1_{\mathrm{A}} \in \mathbf{C}(\mathrm{A}, \mathrm{A})$ such that if $f \in \mathbf{C}(\mathrm{A}, \mathrm{B})$, then $f = f \circ 1_{\mathrm{A}} = 1_{\mathrm{B}} \circ f$.

We can see that an operational theory enjoys the properties in definition 1.1.13: physical systems are *objects* and the events from a system to another are the *morphisms* between the two systems. Tests are nothing but collections of morphisms labelled by the outcomes.

[4] A *class* is not exactly a *set*, but for all practical purposes one can think of $|\mathbf{C}|$ as a set.

Note that events can be composed in sequence and there is the identity event $\mathcal{I}$ for such a composition. Therefore, a category captures the concepts of physical systems, of tests performed between them and of sequential composition of tests.

What about the other types of composition? We are still missing a mathematical description of the composition of systems and of the parallel composition of tests. Besides, the role of the trivial system is not completely clear from a mathematical point of view. In other terms, if we know that the trivial system I has the special meaning of "nothing" from a physical point of view, so far we do not know what mathematical features make the trivial system "special" among all the other systems. The answer to these questions comes from the following definition.

**Definition 1.1.14.** A *strict monoidal category* $(\mathbf{C}, \circ, \otimes, \mathrm{I})$ is a category such that

1. there is an operation $\otimes$ on the class of objects $|\mathbf{C}|$ such that $(|\mathbf{C}|, \otimes, \mathrm{I})$ is a *monoid* with *identity* $\mathrm{I} \in |\mathbf{C}|$.

2. for all objects $\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D} \in |\mathbf{C}|$, there is an operation $\otimes : \mathbf{C}(\mathrm{A}, \mathrm{B}) \times \mathbf{C}(\mathrm{C}, \mathrm{D}) \longrightarrow \mathbf{C}(\mathrm{A} \otimes \mathrm{C}, \mathrm{B} \otimes \mathrm{D})$ on the morphisms between these objects such that $(f, g) \mapsto f \otimes g$, where $f$ is a map from A to B and $g$ is a map from C to D. Moreover, $\otimes$ is associative and has $1_{\mathrm{I}}$, the identity over I, as its identity.

3. for all objects $\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}, \mathrm{E}, \mathrm{F} \in |\mathbf{C}|$, if $f \in \mathbf{C}(\mathrm{A}, \mathrm{B})$, $g \in \mathbf{C}(\mathrm{B}, \mathrm{C})$, $h \in \mathbf{C}(\mathrm{D}, \mathrm{E})$, $k \in \mathbf{C}(\mathrm{E}, \mathrm{F})$, we have $(g \circ f) \otimes (k \circ h) = (g \otimes k) \circ (f \otimes h)$.

4. for all objects $\mathrm{A}, \mathrm{B} \in |\mathbf{C}|$, one has $1_{\mathrm{A} \otimes \mathrm{B}} = 1_{\mathrm{A}} \otimes 1_{\mathrm{B}}$, where $1_{\mathrm{A} \otimes \mathrm{B}}$ is the identity on $\mathrm{A} \otimes \mathrm{B}$ and $1_{\mathrm{A}}$ and $1_{\mathrm{B}}$ are the identities on A and B respectively.

Let us better clarify the various points of the definition.

1. The fact that $(|\mathbf{C}|, \mathrm{I}, \otimes)$ is a monoid means that, for all objects $\mathrm{A}, \mathrm{B}, \mathrm{C} \in |\mathbf{C}|$, we have $(\mathrm{A} \otimes \mathrm{B}) \otimes \mathrm{C} = \mathrm{A} \otimes (\mathrm{B} \otimes \mathrm{C})$, and $\mathrm{A} \otimes \mathrm{I} = \mathrm{I} \otimes \mathrm{A} = \mathrm{I}$. Sometimes $\otimes$ is called "*tensor*".
   Recalling that objects are physical systems, this is nothing but items 1 and 3 of definition 1.1.7, and I is the trivial system.

2. If the tensor operation $\otimes$ between morphisms is associative, it means that $(f \otimes g) \otimes h = f \otimes (g \otimes h)$ for all $\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}, \mathrm{E}, \mathrm{F}$, and for all $f \in \mathbf{C}(\mathrm{A}, \mathrm{B})$, $g \in \mathbf{C}(\mathrm{C}, \mathrm{D})$, $h \in \mathbf{C}(\mathrm{E}, \mathrm{F})$. Moreover if $1_{\mathrm{I}}$ is the identity, then $f = f \otimes 1_{\mathrm{I}} = 1_{\mathrm{I}} \otimes f$.
   Since in operational theories events are morphisms, this is a property of the parallel composition of events (and consequently of tests). Associativity and the existence of the identity are trivially satisfied in the setting we have presented.

3. This means that "sequential" ($\circ$) and "parallel" ($\otimes$) compositions of morphisms "commute". In our presentation of operational theories, this is nothing but eq. (1.1.4).

4. It is trivial in our setting: if we do nothing on both systems A and B, we do nothing also on the two systems taken together, namely on $\mathrm{A} \otimes \mathrm{B}$.

Now we see that we have almost completed our mathematical picture of operational theories. But the last piece is still missing: the fact that $\mathrm{A} \otimes \mathrm{B} \cong \mathrm{B} \otimes \mathrm{A}$. A couple of definitions will enable us to close the circle.

**Definition 1.1.15.** Let $(\mathbf{C}, \circ)$ be a category. Two objects $\mathrm{A}, \mathrm{B} \in |\mathbf{C}|$ are said to be *isomorphic* if there is an invertible morphism $f$ between A and B. $f$ is called *isomorphism*.

An isomorphism corresponds to an invertible event, and such a notion captures the previously explained idea of operationally equivalent systems. Therefore, we are looking for isomorphisms to describe the equivalence $\mathrm{A} \otimes \mathrm{B} \cong \mathrm{B} \otimes \mathrm{A}$.

**Definition 1.1.16.** A *strict symmetric monoidal category* is a strict monoidal category where for every pair of objects $\mathrm{A}, \mathrm{B} \in |\mathbf{C}|$ there exists an isomorphism $\sigma_{\mathrm{A},\mathrm{B}} : \mathrm{A} \otimes \mathrm{B} \longrightarrow \mathrm{B} \otimes \mathrm{A}$ such that

1. $\sigma_{\mathrm{A},\mathrm{B}}^{-1} = \sigma_{\mathrm{B},\mathrm{A}}$

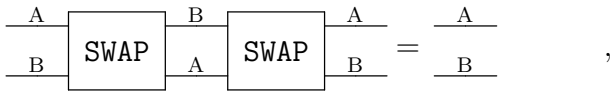
2. for any $\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}$ and for any $f \in \mathbf{C}(\mathrm{A}, \mathrm{B})$ and $g \in \mathbf{C}(\mathrm{C}, \mathrm{D})$, one has

$$\sigma_{\mathrm{B},\mathrm{D}} \circ (f \otimes g) = (g \otimes f) \circ \sigma_{\mathrm{A},\mathrm{C}}.$$

This definition is better understood by making a reference to the operational setting in a laboratory. We can think of $\sigma_{\mathrm{A,B}}$ as a device implementing the swapping between two physical systems. We represent it via the "universal" swapping operation `SWAP`.

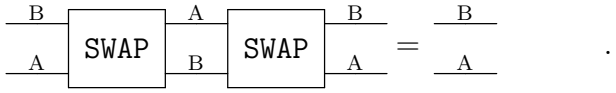

`SWAP` takes $\mathrm{A}\otimes\mathrm{B}$ as input and has $\mathrm{B}\otimes\mathrm{A}$ as output. Let us see that the items in definition 1.1.16 give us reasonable properties for the swapping operation.

1. $\sigma_{\mathrm{A,B}}^{-1} = \sigma_{\mathrm{B,A}}$ means that $\sigma_{\mathrm{B,A}} \circ \sigma_{\mathrm{A,B}} = 1_{\mathrm{A}\otimes\mathrm{B}}$ and $\sigma_{\mathrm{A,B}} \circ \sigma_{\mathrm{B,A}} = 1_{\mathrm{B}\otimes\mathrm{A}}$. In other words, by swapping two systems twice, we get the systems in the original order. It is like applying the identity test.

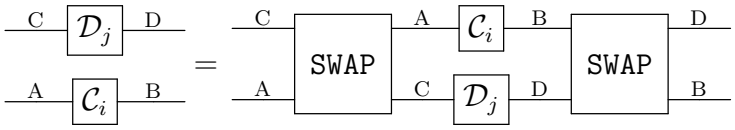
   ,

   and

   .

2. We can rewrite $\sigma_{\mathrm{B,D}} \circ (f \otimes g) = (g \otimes f) \circ \sigma_{\mathrm{A,C}}$ as $g \otimes f = \sigma_{\mathrm{B,D}} \circ (f \otimes g) \circ \sigma_{\mathrm{C,A}}$, recalling that $\sigma_{\mathrm{C,A}} = \sigma_{\mathrm{A,C}}^{-1}$. Thinking of $f$ and $g$ as events $\mathcal{C}_i$ and $\mathcal{D}_j$ respectively, and using diagrams, one has

   .

   This is nothing but eq. (1.1.1), where in place of $\mathcal{U}_1$ and $\mathcal{U}_2$ we have `SWAP`.

Now we are done with the formal definitions about categories. We see that an operational theory is a strict symmetric monoidal category. The strong point about diagrams is that they encode all the properties of strict symmetric monoidal categories in a most apparent way, so that one does not have to remember them, but can reconstruct them from the rules of diagrams. Therefore, we are entitled to consider diagrams as a mathematical language for all strict symmetric monoidal categories, abstracted from the physical interpretation as operational theories in a laboratory.

### 1.1.4 Probabilistic theories

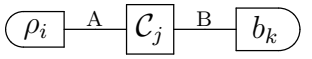
Now we can add the probabilistic ingredient to our theory: basically, we want to assign a number in the interval $[0, 1]$ to every test from the trivial system to itself.

**Definition 1.1.17.** An *operational-probabilistic theory* (general probabilistic theory (GPT) for short) is an operational theory where for every test $\{p_i\}_{i\in X}$ from the trivial system I to itself one has $p_i \in [0, 1]$ and $\sum_{i\in X} p_i = 1$.

Moreover, the sequential and parallel compositions of two events from the trivial system to itself are given by the product of probabilities: $p_i \circ p_j = p_i \otimes p_j = p_i p_j$.

This definition says that every event from I to itself can be interpreted as a probability. As a consequence, we can associate a probability with every diagram with no external lines.

**Example 1.1.18.** Let us consider again



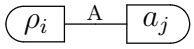

.

This is a diagram without external lines; indeed the sequential composition of the three events is an event from the trivial system I to itself (no input and no output). So we have $p_{ijk} := (b_k|\,\mathcal{C}_j\,|\rho_i)$, that is the *joint probability* of having the preparation-event $\rho_i$, the event $\mathcal{C}_j$ and the observation-event $b_k$.

Henceforth we will focus only on GPTs, namely on operational theories with a probabilistic structure.

Sometimes it happens that we can obtain the same physical configuration with different experimental procedures. For instance, in quantum mechanics, we can consider the mixed state $\rho = \frac{1}{2}\mathbf{1}$ of a q-bit. This state can be prepared either by completely ignoring the state of the system, or by taking the partial trace of one of the Bell states. The issue is now how to distinguish different situations or find out they are equivalent.

Let us consider, for instance, preparation-events. If we compose a preparation-event with an observation-event, we get a probability. Indeed, suppose we have




.

Then we have $p_{ij} = (a_j|\rho_i)$, which is the joint probability of having the preparation-event $\rho_i$ and the observation-event $a_j$.

*Remark* 1.1.19. $p_{ij}$ should not be confused with a conditional probability, namely $p_{ij}$ is *not* the probability of having the observation-event $a_j$ *if* the preparation-event is $\rho_i$. Indeed, assuming this conditional interpretation would imply that information flows from the preparation-event to the observation-event. This assumption is known as *Causality*, to which we will come soon (section 1.3). In general, in a non-causal theory, the observation-event can influence the preparation-event, so, in principle, we are not allowed to say which event influenced the other.

If we have a preparation-event $\rho_i$ on A, we can associate a real-valued function $\widehat{\rho}_i$ with it. This function acts on observation-events $a_j$ on A and yields the joint probability $p_{ij}$.

$$\widehat{\rho}_i : (a_j| \longmapsto (a_j|\rho_i) = p_{ij}$$

Similarly, if we have an observation-test $a_j$ on A, we can associate a real-valued function $\widehat{a}_j$ with it. This function acts on preparation-events $\rho_i$ on A and yields the joint probability $p_{ij}$.

$$\widehat{a}_j : |\rho_i) \longmapsto (a_j|\rho_i) = p_{ij}$$

From a probabilistic point of view, we cannot distinguish two preparations of the system if they yield the same probabilities for all the observation-tests, even if the preparations were obtained operatively in completely different ways. If we consider an experimenter, she can distinguish two unknown preparations of the system by examining the statistics she gets from performing, in principle, all the possible measurements on the system. If she finds any difference in the statistics, then she concludes the preparations were different. A very similar argument holds for observation-events.

In this vein, we can introduce an equivalence relation between preparation-events (and similarly between observation-events). If $\rho_i$ and $\sigma_j$ are two preparation-events on system A, we say that $\rho_i \sim \sigma_j$ if $\widehat{\rho}_i = \widehat{\sigma}_j$, namely if for every observation-event $a_k$ on A we have $(a_k|\rho_i) = (a_k|\sigma_j)$. Similarly, if $a_i$ and $b_j$ are two observation-events on A, we say $a_i \sim b_j$ if $\widehat{a}_i = \widehat{b}_j$, namely if for every preparation-event $\rho_k$ on A we have $(a_i|\rho_k) = (b_j|\rho_k)$.

**Definition 1.1.20.** Equivalence classes of indistinguishable preparation-events are called *states*. The set of states of system A is denoted as $\mathsf{St}(\mathrm{A})$.

Equivalence classes of indistinguishable observation-events are called *effects*. The set of effects of system A is denoted as $\mathsf{Eff}(\mathrm{A})$.

In this way, states and effects are identified with the corresponding functions that yield probabilities, similar to the setting of Gleason's theorem in quantum mechanics [49]. Therefore, two states $\rho_0$ and $\rho_1$ of system A are equal if and only if $(a|\rho_0) = (a|\rho_1)$ for every effect $a \in \mathsf{Eff}(\mathrm{A})$. Similarly, two effects $a_0$ and $a_1$ of system A are equal if and only if $(a_0|\rho) = (a_1|\rho)$ for every state $\rho \in \mathsf{St}(\mathrm{A})$.

We can assume that equivalence classes were taken from the very beginning, so from now on we will say that a preparation-test is made of states and that an observation-test is made of effects. Specifically, when we have a deterministic preparation-test, we will call it *deterministic state*; and when we have a deterministic observation-test, we will call it *deterministic effect*.

**Example 1.1.21.** The trivial system has a unique deterministic state and a unique deterministic effect: it is number 1.

Let us introduce some more terminology about states and effects.

**Definition 1.1.22.** A state of a composite system $\mathrm{A} \otimes \mathrm{B}$ is called *bipartite state*.

An effect of a composite system $\mathrm{A} \otimes \mathrm{B}$ is called *bipartite effect*.

A bipartite state (resp. effect) is called *product state* (resp. *effect*) if it is obtained by parallel composition of states (resp. effects) of A and B.

Bipartite states are depicted as in (1.1.2), bipartite effects are depicted as in (1.1.3). Product states are represented diagrammatically in (1.1.5), product effects are represented diagrammatically in (1.1.6).

Let us see what states and effects are in quantum mechanics.

**Example 1.1.23.** In quantum mechanics, we can associate a Hilbert space $\mathcal{H}_\mathrm{A}$ with every system A. Deterministic states are density operators, which means trace-class positive operators with trace equal to 1. A non-deterministic preparation-test is sometimes called *quantum information source*: it is a collection of trace-class positive operators $\rho_i$, with $\mathrm{tr}\,\rho_i \leq 1$. This is essentially a random preparation: a state $\rho_i$ is prepared with a probability given by $\mathrm{tr}\,\rho_i$. Therefore in quantum mechanics $\mathsf{St}(\mathrm{A})$ is the set of trace-class positive operators with trace less than or equal to one.

An effect is, instead, represented by a positive operator $P$, with $P \leq \mathbf{1}$, where $\mathbf{1}$ is the identity operator. Observation-tests are then POVMs. The pairing between states and effect is given by trace: $(P|\rho) = \mathrm{tr}P\rho$. In quantum mechanics there is only one deterministic effect: the identity $\mathbf{1}$. This is not a coincidence, but it follows from Causality (see section 1.3).

According to definition 1.1.20, states and effects are in fact real-valued functions; as a consequence we can take linear combinations of them with real coefficients; in other words they span real vector spaces. Let $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$ be the vector space spanned by states and let $\mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$ be the vector space spanned by effects. These vector spaces can be finite- or infinite-dimensional. In our treatment, to avoid mathematical subtleties, we will assume that these vector spaces are finite-dimensional. Clearly, $\mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$ is the dual vector space of $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$ and $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$ is the dual vector space of $\mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$. For finite-dimensional vector spaces, we have $\dim \mathsf{St}_{\mathbb{R}}(\mathrm{A}) = \dim \mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$.

**Example 1.1.24.** Let us see what $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$ and $\mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$ are in finite-dimensional quantum theory, namely when the Hilbert space is finite-dimensional ($\mathcal{H} \approx \mathbb{C}^n$, for $n \geq 2$). $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$ is the vector space of hermitian matrices of order $n$. It is a real vector space with dimension $n^2$. $\mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$ is again the vector space of hermitian matrices of order $n$.

*Remark* 1.1.25. In general, $\mathsf{St}(\mathrm{A})$ and $\mathsf{Eff}(\mathrm{A})$ are *not* vector spaces. Indeed, a state is a function which takes values in $[0,1]$ interval according to our probabilistic interpretation. Clearly, a general linear combination of $[0,1]$-valued functions does not take values in $[0,1]$. Instead, if we take a convex combination[5] of $[0,1]$-valued functions, we get another $[0,1]$-valued function. This is the first hint to $\mathsf{St}(\mathrm{A})$ and $\mathsf{Eff}(\mathrm{A})$ being in fact convex sets.

Now we can define the equivalence classes of indistinguishable events for general tests, namely for tests from system A to system B.

First of all, note that every event $\mathcal{C}_i$ from A to B induces a linear operator $\widehat{\mathcal{C}_i}$ from $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$ to $\mathsf{St}_{\mathbb{R}}(\mathrm{B})$. We define $\widehat{\mathcal{C}_i}$ via its action on the spanning set of states $\mathsf{St}(\mathrm{A})$, as follows:

$$\widehat{\mathcal{C}_i} : \rho_{\mathrm{A}} \mapsto \mathcal{C}_i \rho_{\mathrm{A}}, \tag{1.1.7}$$

for every $\rho_{\mathrm{A}} \in \mathsf{St}(\mathrm{A})$. Note that $\mathcal{C}_i \rho_{\mathrm{A}}$ is a state of B. We want to check whether the linear extension of (1.1.7) is well defined. Now, we know how $\widehat{\mathcal{C}_i}$ acts on states, namely on the spanning set $\mathsf{St}(\mathrm{A})$. How can we define its action on all $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$? If $v \in \mathsf{St}_{\mathbb{R}}(\mathrm{A})$, we can express it as a linear combination of states, $v = \sum_j \alpha_j \rho_j$, where $\alpha_j \in \mathbb{R}$ for every $j$. The most obvious linear extension of (1.1.7) is $\widehat{\mathcal{C}_i} v := \sum_j \alpha_j \widehat{\mathcal{C}_i} \rho_j$. The problem is that, in general, $v$ does not have a unique expression in terms of states. Suppose that $v =$

[5] Recall that a convex combinations of the points $x_i$'s is defined as $\sum_i \lambda_i x_i$, where $\lambda_i \geq 0$ for every $i$ and $\sum_i \lambda_i = 1$.

$\sum_j \alpha_j \rho_j$ and $v = \sum_j \beta_j \sigma_j$, where $\beta_j \in \mathbb{R}$ for every $j$. Our extension $\widehat{\mathcal{C}}_i$ is well-defined if and only if $\sum_j \alpha_j \widehat{\mathcal{C}}_i \rho_j = \sum_j \beta_j \widehat{\mathcal{C}}_i \sigma_j$ whenever $\sum_j \alpha_j \rho_j = \sum_j \beta_j \sigma_j$. Using linearity of summations, this problem is equivalent to check if $\sum_j \alpha_j \widehat{\mathcal{C}}_i \rho_j = 0$ whenever $\sum_j \alpha_j \rho_j = 0$.

By definition of effects, we have $\sum_j \alpha_j \rho_j = 0$ if and only if $\sum_j \alpha_j (a|\rho_j) = 0$ for every effect $a \in \mathsf{Eff}(\mathrm{A})$. Let $b$ be an arbitrary effect on B. Then $(b|\widehat{\mathcal{C}}_i$ is an effect on A, therefore $\sum_j \alpha_j (b|\widehat{\mathcal{C}}_i|\rho_j) = 0$. Since $b$ is arbitrary, this implies that $\sum_j \alpha_j \widehat{\mathcal{C}}_i \rho_j = 0$. This proves that the linear extension is well-defined.

Our construction, and (1.1.7) in particular, basically say that events are characterized by their action on states.

Likewise, for every system C, the event $\mathcal{C}_i \otimes \mathcal{I}_\mathrm{C}$ from $\mathrm{A} \otimes \mathrm{C}$ to $\mathrm{B} \otimes \mathrm{C}$ will induce a linear operator from $\mathsf{St}_\mathbb{R}(\mathrm{A} \otimes \mathrm{C})$ to $\mathsf{St}_\mathbb{R}(\mathrm{B} \otimes \mathrm{C})$. We then give the following definition.

**Definition 1.1.26.** Two events $\mathcal{C}_i$ and $\mathcal{C}'_i$ from A to B are *indistinguishable* if for all systems C the linear operators associated with $\mathcal{C}_i \otimes \mathcal{I}_\mathrm{C}$ and $\mathcal{C}'_i \otimes \mathcal{I}_\mathrm{C}$ are the same.

This should remind the reader of the definition of complete positivity in quantum theory [50, 51, 52].

Again, we take the quotient set of events modulo the indistinguishability relation.

**Definition 1.1.27.** Equivalence classes of indistinguishable events from A to B are called *transformations* from A to B.

The set of transformations from A to B is denoted by $\mathsf{Transf}(\mathrm{A}, \mathrm{B})$. The set of transformations from A to itself is denoted simply by $\mathsf{Transf}(\mathrm{A})$.

*Remark* 1.1.28. One may wonder why we have given such a definition of indistinguishable events, involving an ancillary system C. The most obvious way of defining indistinguishability would have been to say that $\mathcal{C}_i$ and $\mathcal{C}'_i$ are indistinguishable if $\mathcal{C}_i \rho = \mathcal{C}'_i \rho$ for every $\rho \in \mathsf{St}(\mathrm{A})$. Actually, this is not enough for general probabilistic theories. Indeed, Wootters provided a counterexample concerning quantum mechanics with real Hilbert space [53]. It can be shown that there exist two transformations that are locally indistinguishable, but if we add an ancillary system, they produce orthogonal output states.

The condition $\mathcal{C}_i \rho = \mathcal{C}'_i \rho$ for every $\rho \in \mathsf{St}(\mathrm{A})$ is sufficient for indistinguishability if the theory satisfies *Local Tomography* (see [28] for further

details). Quantum mechanics with real Hilbert space does not satisfy this property.

We conclude that it is not enough to say that $\mathcal{C}_i$ and $\mathcal{C}'_i$ from A to B are indistinguishable if they act in the same way on every state of system A.

Again, we will assume that equivalence classes have been taken from the very beginning, so we will consider tests as collections of transformations.

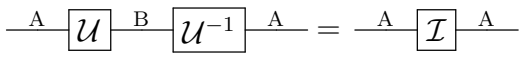
**Definition 1.1.29.** A deterministic transformation $\mathcal{C} \in \mathsf{Transf}(\mathrm{A}, \mathrm{B})$ is called *channel*.

Channels deterministically transform states of system A into states of system B.

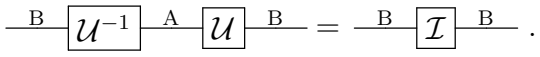
Among all possible channels, reversible channels are particularly important.

**Definition 1.1.30.** A channel $\mathcal{U} \in \mathsf{Transf}(\mathrm{A}, \mathrm{B})$ is said *reversible* if it is invertible, namely if there is another channel $\mathcal{U}^{-1} \in \mathsf{Transf}(\mathrm{B}, \mathrm{A})$, called the *inverse*, such that $\mathcal{U}^{-1} \circ \mathcal{U} = \mathcal{I}_{\mathrm{A}}$ and $\mathcal{U} \circ \mathcal{U}^{-1} = \mathcal{I}_{\mathrm{B}}$.

Using diagrams, we have




and




.

Clearly, reversible channels on A form a group, called $\mathbf{G}_{\mathrm{A}}$.

Now, we can rephrase the definition of operationally equivalent systems: two systems A and A′ are operationally equivalent if there exists a reversible channel from A to A′.

Before moving on, let us see what transformations, channels and reversible channels are in quantum mechanics.

**Example 1.1.31.** A test in quantum mechanics from $\mathcal{H}_{\mathrm{A}}$ to $\mathcal{H}_{\mathrm{B}}$ is a collection of completely positive, trace non-increasing linear maps $\{\mathcal{C}_k\}$, called quantum operations [50], such that $\sum_k \mathcal{C}_k$ is a trace-preserving map. Each quantum operation maps linear operators on $\mathcal{H}_{\mathrm{A}}$ into linear operators on $\mathcal{H}_{\mathrm{B}}$. A test is a quantum instrument, namely a collection of quantum operations [50].

A channel is a completely positive trace-preserving map from linear operators on n $\mathcal{H}_{\mathrm{A}}$ to linear operators on $\mathcal{H}_{\mathrm{B}}$.

Finally, reversible channels are unitary channels. They act on A as $\mathcal{U}(\rho) = U\rho U^\dagger$, where $U$ is a unitary operator. It follows that two systems are operationally equivalent if and only if their Hilbert spaces have the same dimension, otherwise it would not be possible to define unitary operators from one space to the other.

## 1.2 Purity and coarse-graining

Even in an abstract probabilistic theory, it makes sense to define pure and mixed states, or, more generally, about pure and non-pure transformations. The idea behind that is *coarse-graining*. Let us clarify this idea with the example of the roll of a die [54]. In this random experiment, there are some atomic events, which cannot be decomposed further: they are the numbers from 1 to 6. So, an atomic event is, for example, "the outcome of the roll is 2". However, we can consider the event "the outcome of the roll is odd". This event is the union of the atomic events relative to 1, 3, 5. We did a coarse-graining: we joined together some outcomes, neglecting some information. Indeed, if we know only that the outcome was "odd", we cannot retrieve any information about which number actually came out. In this vein, we give the following definition.

**Definition 1.2.1.** A test $\{\mathcal{C}_i\}_{i\in X}$ is a *coarse-graining* of the test $\{\mathcal{D}_j\}_{j\in Y}$ if there is a partition[6] $\{Y_i\}$ of $Y$ such that $\mathcal{C}_i = \sum_{j\in Y_i} \mathcal{D}_j$. In this case, we say that $\{\mathcal{D}_j\}_{j\in Y}$ is a *refinement* of $\{\mathcal{C}_i\}_{i\in X}$.

As we can see, this definition gives a precise characterization of what we mean by "joining together outcomes". A test that refines another extracts more information than the other. It is clear that if $\{\mathcal{C}_i\}_{i\in X}$ is a coarse-graining of the test $\{\mathcal{D}_j\}_{j\in Y}$, it has fewer outcomes.

By performing a coarse-graining, we can associate a deterministic transformation with every test. Indeed, let us take a test $\{\mathcal{C}_i\}_{i\in X}$ from A to B and let us sum over the outcomes $i$. Then we obtain the channel $\mathcal{C} = \sum_{i\in X} \mathcal{C}_i$ from A to B, which is called the channel associated with the test $\{\mathcal{C}_i\}_{i\in X}$. Similarly, we can obtain a deterministic state by summing all the states in a preparation-test; and we can get a deterministic effect by summing all the effects in an observation-test.

[6]Recall that a partition of a set $Y$ is a collection of subsets $Y_i$ such that they are non-empty, they are pairwise disjoint and their union is $Y$.

We can consider also refinements of single transformations.

**Definition 1.2.2.** Let $\mathcal{C}$ be a transformation from system A to system B. Consider a test $\{\mathcal{D}_i\}_{i\in X}$ from system A to system B and a subset $X_0 \subseteq X$ such that $\mathcal{C} = \sum_{i\in X_0} \mathcal{D}_i$. Each transformation $\mathcal{D}_i$, for $i \in X_0$ is a *refinement* of $\mathcal{C}$.

Some transformations cannot be refined further.

**Definition 1.2.3.** A refinement $\mathcal{C}'$ of a transformation $\mathcal{C}$ is called *trivial* if we have $\mathcal{C}' = \lambda\mathcal{C}$, for some $\lambda \in (0, 1]$.

This type of refinement is called trivial because a refinement of any transformation $\mathcal{C}$ can be always obtained by taking a subset of a test, made of $\{p_i\mathcal{C}\}_{i\in X_0}$, with the property that $p_i \in (0, 1]$ for every $i \in X_0$ and $\sum_{i\in X_0} p_i = 1$. Then it is reasonable to give the following definition.

**Definition 1.2.4.** A transformation $\mathcal{C}$ is *pure* (or *atomic*) if it has only trivial refinements.

In other words, it is not possible to extract further information from a pure transformation.

Clearly, this definition applies also to states, which are particular transformations from the trivial system I to a system A. Thus, we have pure states, which admit only trivial refinements. We will denote the set of pure states of system A as $\mathsf{PurSt}(\mathrm{A})$. The non-pure states are called *mixed*. In this way, a pure state represents maximal knowledge about the preparation of a system, whereas a mixed state expresses some lack of information. Let us see some examples in quantum mechanics.

**Example 1.2.5.** If we diagonalize a density operator $\rho = \sum_j p_j |\psi_j\rangle \langle\psi_j|$, each term $p_j |\psi_j\rangle \langle\psi_j|$ is a refinement of $\rho$. More generally, a refinement of $\rho$ is a state $\sigma$ such that $\sigma \leq \rho$. Indeed, in this way the difference $\rho - \sigma$ is a positive operator and can be interpreted as a state. This means that the support[7] of $\sigma$ is contained in the support of $\rho$ (see appendix A.1 of [55] for a proof). A pure state is a density operator $\rho$ proportional (with non-vanishing proportional coefficient) to a rank-one projector.

In quantum mechanics, we can associate Kraus operators $\{M_k\}$ with every quantum operation $\mathcal{C}$, such that $\mathcal{C}(\rho) = \sum_k M_k \rho M_k^\dagger$, for every state $\rho$ [50]. A quantum operation is pure if and only it has only one Kraus operator.

[7]Recall the support of a matrix is the orthogonal complement of its kernel.

Let us analyse the relationship between pure states and reversible channels.

**Lemma 1.2.6.** *Let $\mathcal{U}$ be a reversible channel from* A *to* B. *Then $\psi \in \mathsf{St}(\mathrm{A})$ is pure if and only if $\mathcal{U}\psi \in \mathsf{St}(\mathrm{B})$ is pure.*

*Proof.* Necessity. Let us write $\mathcal{U}\psi$ as a coarse-graining of other states.

$$\mathcal{U}\psi = \sum_i \rho_i \tag{1.2.1}$$

Let us show that each refinement $\rho_i$ of $\mathcal{U}\psi$ is trivial, that is $\rho_i = p_i \mathcal{U}\psi$, for some $p_i \in (0, 1]$, with $\sum_i p_i = 1$. By applying $\mathcal{U}^{-1}$ to both sides of eq. (1.2.1), we have $\psi = \sum_i \mathcal{U}^{-1}\rho_i$. Since $\psi$ is pure, each refinement $\mathcal{U}^{-1}\rho_i$ is trivial, namely

$$\mathcal{U}^{-1}\rho_i = p_i \psi, \tag{1.2.2}$$

for some $p_i \in (0, 1]$, with $\sum_i p_i = 1$. By applying $\mathcal{U}$ to both sides of eq. (1.2.2), we have $\rho_i = p_i \mathcal{U}\psi$. Since every refinement $\mathcal{U}\psi$ is trivial, $\mathcal{U}\psi$ is pure.

Sufficiency follows from necessity, by applying the reversible channel $\mathcal{U}^{-1}$ to $\mathcal{U}\psi$, which is pure by hypothesis. □

This means that reversible channels do not alter the "purity" of a state: they map pure states into pure states and mixed states into mixed states.

A similar statement holds also for effects.

**Lemma 1.2.7.** *Let $\mathcal{U}$ be a reversible channel from* A *to* B. *Then $b \in \mathsf{Eff}(\mathrm{B})$ is pure if and only if $b\mathcal{U} \in \mathsf{Eff}(\mathrm{A})$ is pure.*

*Proof.* The proof is analogous to that of lemma 1.2.6. □

## 1.3 Causality

In this section we will examine the issue of Causality, namely the "direction" in which information flows in an experimental apparatus or in a diagram. We have already mentioned that the order of sequential composition does not correspond, in general, to temporal ordering, which is the ordering given by information flow. When these two ordering coincide, we say that the theory is causal. Causality is a standard setting for any reasonable physical description of Nature, therefore we will assume it in the present treatment

as part of the framework of GPTs. However, there have been some proposals of abandoning causality to formulate a quantum theory of gravity [56].

Let us begin with the formal definition of causal theory.

**Definition 1.3.1** (Causal theories)**.** A theory is *causal* if for every preparation-test $\{\rho_i\}_{i\in X}$ and every observation-test $\{a_j\}_{j\in Y}$ on system A, the marginal probability $p_i := \sum_{j\in Y} (a_j|\rho_i)_{\mathrm{A}}$ is *independent* of the observation-test $\{a_j\}_{j\in Y}$.

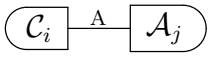
In other words, if $\{a_j\}_{j\in Y}$ and $\{b_k\}_{k\in Z}$ are two observation-tests, we have

$$\sum_{j\in Y} (a_j|\rho_i)_{\mathrm{A}} = \sum_{k\in Z} (b_k|\rho_i)_{\mathrm{A}} \,. \tag{1.3.1}$$

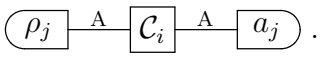
Loosely speaking, the preparation of the system does not depend on the choice of subsequent (or "future") measurements (no-signalling from the future). In this way, the direction in which information flows, which we can identify with temporal ordering, coincides with the ordering given by sequential composition. In general, this is not obvious, as the following example shows [57].

**Example 1.3.2.** Consider a theory where the states of a system are quantum operations on that system. Specifically, deterministic states are quantum channels. Then we can consider the channels of this theory to be quantum "supermaps", which map quantum channels into quantum channels.

Let us consider a preparation of a state $\mathcal{C}_i$ followed by a measurement $\mathcal{A}_j$, which we represent as




Note that the measurement follows the preparation in the composition order. But if we recall that $\mathcal{C}_i$ is a quantum operation, namely a box with an input and an output line, in quantum theory such a diagram will look like




Note that the effect $\mathcal{A}_j$ is split in two parts: one is before the quantum operation and the other is after, otherwise we could not have a diagram with no external wires. Since this diagram is a diagram in quantum theory, which is causal (see below), the order of sequential composition coincides with temporal order. Therefore, in the theory in which states are quantum operations, the preparation of a state is influenced by a subsequent measurement.

We will restrict ourselves only to causal theories. This is essentially the Causality requirement (or axiom).

Now it is possible define conditional probabilities: $p_{ij} = (a_j|\rho_i)$ is the probability of getting outcome $j$ if the prepared state was $i$.

However, definition 1.3.1 is not so practical to work with, although it is operational. We will mostly use the following characterization.

**Proposition 1.3.3.** *A theory is causal if and only if for every system* A *there is a unique deterministic effect* $\mathrm{tr}_\mathrm{A}$.

*Proof.* Necessity. Suppose, by contradiction, that there are two deterministic effects tr and tr$'$ for system A. Deterministic effects are particular examples of observation-tests. Eq. (1.3.1) then states that $\mathrm{tr}\rho_i = \mathrm{tr}'\rho_i$ for every $\rho_i \in \mathsf{St}(\mathrm{A})$. This means that $\mathrm{tr} = \mathrm{tr}'$.

Sufficiency. Suppose there is a unique deterministic effect $\mathrm{tr}_\mathrm{A}$ for system A, and consider the observation-test $\{a_j\}_{j\in Y}$. By doing a coarse-graining over the effects, we obtain the deterministic effect $\mathrm{tr}' = \sum_{j\in Y} a_j$. Since the deterministic effect is unique, it must be $\mathrm{tr}' = \mathrm{tr}$. Hence, for every state $\rho_i$, we have

$$\sum_{j\in Y} (a_j|\rho_i) = \mathrm{tr}\rho_i,$$

and the right-hand side does not depend on the choice of the observation-test. This means that the theory is causal. □

We noticed that if we perform a coarse-graining over the effects in an observation-test, we have a deterministic effect. By uniqueness of the deterministic effect, we have that if $\{a_i\}_{i\in X}$ is an observation-test on system A, then $\sum_{i\in X} a_i = \mathrm{tr}$, where tr is the deterministic effect of A. This is a necessary condition for $\{a_i\}_{i\in X}$ to be an observation-test.

We saw in example 1.1.23 that in quantum mechanics there is only one deterministic effect, the identity operator. Hence quantum mechanics is a causal theory.

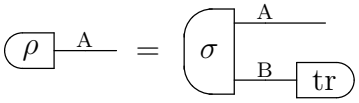
Let us see a straightforward corollary of uniqueness of the deterministic effect.

**Corollary 1.3.4.** *Let* A *and* B *be two systems. In a causal theory, if* $\mathrm{tr}_\mathrm{A}$ *and* $\mathrm{tr}_\mathrm{B}$ *are the deterministic effects of systems* A *and* B *respectively, then the deterministic effect for system* $\mathrm{A}\otimes\mathrm{B}$ *is* $\mathrm{tr}_\mathrm{A}\otimes\mathrm{tr}_\mathrm{B}$.

*Proof.* The parallel composition of two single-outcome tests is clearly a single-outcome test, hence the effect $\mathrm{tr}_\mathrm{A} \otimes \mathrm{tr}_\mathrm{B}$ is deterministic and acts on $\mathrm{A} \otimes \mathrm{B}$. By the uniqueness of the deterministic effect, we conclude that $\mathrm{tr}_{\mathrm{AB}} = \mathrm{tr}_\mathrm{A} \otimes \mathrm{tr}_\mathrm{B}$. □

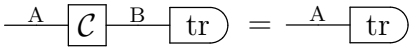
In a causal theory, we can define marginal states. Suppose we have a bipartite state of system $\mathrm{A}\otimes\mathrm{B}$ and we are interested in the state of subsystem A. We want to throw away all the information concerning system B. This operation resembles marginalization in probability theory, whence the name. In quantum mechanics, this operation is simply taking the partial trace over B (whence the choice of the symbol for the deterministic effect).

**Definition 1.3.5.** The *marginal state* (*marginal* for short) $\rho_\mathrm{A}$ on system A of a bipartite state $\sigma_\mathrm{AB}$ is obtained by applying the deterministic effect to B: $\rho_\mathrm{A} = \mathrm{tr}_\mathrm{B}\sigma_\mathrm{AB}$




In a causal theory, we have also useful characterizations of channels and tests.

**Proposition 1.3.6.** *Let* $\mathcal{C} \in \mathsf{Transf}(\mathrm{A}, \mathrm{B})$. *$\mathcal{C}$ is a channel if and only if* $\mathrm{tr}_\mathrm{B}\mathcal{C} = \mathrm{tr}_\mathrm{A}$.




*Proof.* Necessity is straightforward. Since a channel is a deterministic transformation, then $\mathrm{tr}_\mathrm{B}\mathcal{C}$ is a deterministic effect on system A. By uniqueness of the deterministic effect, $\mathrm{tr}_\mathrm{B}\mathcal{C} = \mathrm{tr}_\mathrm{A}$.

Sufficiency. Suppose we have a test $\{\mathcal{C}_i\}_{i\in X}$ from system A to system B such that $\mathcal{C} := \mathcal{C}_{i_0}$ satisfies $\mathrm{tr}_\mathrm{B}\mathcal{C} = \mathrm{tr}_\mathrm{A}$. We want to prove that $\{\mathcal{C}_i\}_{i\in X}$ is a deterministic test. Let us consider the channel $\mathcal{C}'$ associated with the test $\{\mathcal{C}_i\}_{i\in X}$, namely $\mathcal{C}' = \sum_{i\in X}\mathcal{C}_i$. Since $\mathcal{C}'$ is a channel, we have $\mathrm{tr}_\mathrm{B}\mathcal{C}' = \mathrm{tr}_\mathrm{A}$. Recalling the expression of $\mathcal{C}'$, we have

$$\mathrm{tr}_\mathrm{A} = \mathrm{tr}_\mathrm{B}\mathcal{C}' = \mathrm{tr}_\mathrm{B}\mathcal{C}_{i_0} + \mathrm{tr}_\mathrm{B}\sum_{i\neq i_0}\mathcal{C}_i = \mathrm{tr}_\mathrm{A} + \mathrm{tr}_\mathrm{B}\sum_{i\neq i_0}\mathcal{C}_i,$$

because $\mathrm{tr}_\mathrm{B}\mathcal{C}_{i_0} = \mathrm{tr}_\mathrm{A}$. This means $\mathrm{tr}_\mathrm{B}\sum_{i\neq i_0}\mathcal{C}_i = 0$, namely $\sum_{i\neq i_0}\mathcal{C}_i = 0$. Therefore $\mathcal{C} = \mathcal{C}'$, whence the test was in fact deterministic. Thus $\mathcal{C}$ is a channel. □

Note that in quantum theory this is precisely the statement that a quantum operation is a quantum channel if and only if it is trace-preserving.

Specifically, if A is the trivial system, we have that a state $\rho_{\mathrm{B}}$ is deterministic if and only if $\mathrm{tr}\rho = 1$. Moreover, for every test $\{\mathcal{C}_i\}_{i\in X}$ from A to B, we can consider the associated channel $\sum_{i\in X}\mathcal{C}_i$. Therefore we have

$$\sum_{i\in X}\mathrm{tr}_{\mathrm{B}}\mathcal{C}_i = \mathrm{tr}_{\mathrm{A}}.$$

This is a necessary condition. In quantum theory this is the statement that the quantum channel associated with a quantum instrument is trace-preserving.

Suppose we have two parties sharing a bipartite state. In a causal theory it is impossible for a party to send a message to the other by acting locally on her own physical system and relying on correlations she shares with the other party. This form of instantaneous communication is called *signalling*. In more precise terms, in a causal theory it is not possible for a party to communicate the outcome of a local measurement on her system to the other without exchanging physical systems, classical communication included, as it is usually mediated by electromagnetic signals.

**Theorem 1.3.7.** *In a causal theory it is impossible to have signalling without the exchange of physical systems.*

*Proof.* Suppose we have two distant parties, Alice and Bob, that share a bipartite state $\sigma_{\mathrm{AB}}$. Suppose Alice performs a local test $\{\mathcal{A}_i\}_{i\in X}$ on A and Bob performs a local test $\{\mathcal{B}_j\}_{j\in Y}$ on B. Let us define the joint probability $p_{ij} := (\mathrm{tr}|_{\mathrm{AB}}\mathcal{A}_i\otimes\mathcal{B}_j\,|\sigma)_{\mathrm{AB}}$ and the marginal probabilities as $p_i^{(\mathrm{A})} := \sum_{j\in Y}(\mathrm{tr}|_{\mathrm{AB}}\mathcal{A}_i\otimes\mathcal{B}_j\,|\sigma)_{\mathrm{AB}}$ and $p_j^{(\mathrm{B})} := \sum_{i\in X}(\mathrm{tr}|_{\mathrm{AB}}\mathcal{A}_i\otimes\mathcal{B}_j\,|\sigma)_{\mathrm{AB}}$. Each party cannot acquire any information about the outcomes of the other based only on its marginal probability. Indeed, let us examine Alice's marginal probability $p_i^{(\mathrm{A})}$ better. Let $\rho_{\mathrm{A}}$ be the marginal state of $\sigma_{\mathrm{AB}}$ on system A.

$$p_i^{(\mathrm{A})} = \sum_{j\in Y}(\mathrm{tr}|_{\mathrm{AB}}\mathcal{A}_i\otimes\mathcal{B}_j\,|\sigma)_{\mathrm{AB}} = (\mathrm{tr}|_{\mathrm{A}}\,(\mathrm{tr}|_{\mathrm{B}}\mathcal{A}_i\otimes\sum_{j\in Y}\mathcal{B}_j\,|\sigma)_{\mathrm{AB}} =$$

$$= (\mathrm{tr}|_{\mathrm{A}}\mathcal{A}_i\otimes\left((\mathrm{tr}|_{\mathrm{B}}\sum_{j\in Y}\mathcal{B}_j\right)|\sigma)_{\mathrm{AB}} = (\mathrm{tr}|_{\mathrm{A}}\mathcal{A}_i\otimes\mathrm{tr}_{\mathrm{B}}\sigma_{\mathrm{AB}} =$$

$$= (\mathrm{tr}| \mathcal{A}_i |\rho)_{\mathrm{A}}$$

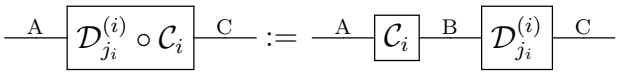
We see that Alice's marginal probability does not depend on the test performed by Bob at all, so she cannot get any information about the outcome of Bob's test based only on her system.

A similar reasoning applies to Bob's party: he cannot gain any information about the outcome of Alice's test. □

Since in a causal theory the order of composition coincides with the order in which information flows, we can choose a later test according to the result of a previous test. Suppose we perform a test $\{\mathcal{C}_i\}_{i\in X}$ from A to B first. Depending on the outcome $i$, we then perform different tests $\left\{\mathcal{D}^{(i)}_{j_i}\right\}_{j_i\in Y_i}$ from B to C. Here the superscript in round brackets is aimed at highlighting the dependence of the test on the outcome of the previous test. Let us make this concept more precise with the following definition.

**Definition 1.3.8.** If $\{\mathcal{C}_i\}_{i\in X}$ is a test from A to B and, for every $i$, $\left\{\mathcal{D}^{(i)}_{j_i}\right\}_{j_i\in Y_i}$ is a test from B to C, then the *conditioned test* is a test from A to C with outcomes $(i, j_i) \in Z := \bigcup_{i\in X} \{i\} \times Y_i$ and events $\left\{\mathcal{D}^{(i)}_{j_i} \circ \mathcal{C}_i\right\}_{(i,j_i)\in Z}$.

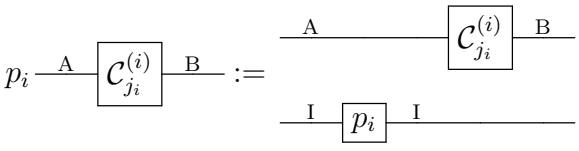
The graphical representation is as usual.




Conditioning expresses the idea of choosing what to do at later steps using classical information about outcomes obtained at previous steps.

A particular case of conditioning is randomization.

**Definition 1.3.9.** If $\{p_i\}_{i\in X}$ is a set of probabilities[8] and, for every $i$, $\left\{\mathcal{C}^{(i)}_{j_i}\right\}_{j_i\in Y_i}$ is a test from A to B, we can construct the *randomized test* $\left\{p_i\mathcal{C}^{(i)}_{j_i}\right\}_{i\in X, j_i\in Y_i}$, which is a test from A to B whose events are defined as




[8]Recall that a set of probabilities can be regarded as a test from the trivial system to itself.

### 1.3.1 Operational norms

In this subsection we want to introduce norms for states, effects and transformations. These norms have a direct relationship with the issue of distinguishing states, effects and transformations, like in quantum theory [50, 52].

**Definition 1.3.10.** Let $\rho \in \mathsf{St}(\mathrm{A})$. We define the *norm* of $\rho$ as

$$\|\rho\| := \mathrm{tr}\,\rho.$$

It can be shown that this function is a norm on $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$. Clearly we have $0 \leq \|\rho\| \leq 1$, because of the probabilistic interpretation of the action of effects on states. We have the following proposition.

**Proposition 1.3.11.** *One has*

$$\|\rho\| = \max_{a \in \mathsf{Eff}(\mathrm{A})} (a|\rho)\,.$$

*Proof.* Let us consider an observation-test $\{a_i\}_{i\in X}$ on A, and let $a := a_{i_0}$. We have $\sum_{i\in X} a_i = \mathrm{tr}$, then

$$\|\rho\| = \mathrm{tr}\,\rho = \sum_{i \neq i_0} (a_i|\rho) + (a|\rho)\,.$$

Since this is a sum of non-negative numbers (each term is a probability), then $(a|\rho) \leq \|\rho\|$. Since $a$ is arbitrary, the thesis follows. □

**Definition 1.3.12.** A state $\rho$ such that $\|\rho\| = 1$ is called *normalized*.

We denote the set of normalized states of system A as $\mathsf{St}_1(\mathrm{A})$, and the set of normalized pure states as $\mathsf{PurSt}_1(\mathrm{A})$.

Normalized states have an operational meaning, expressed by the lemma below.

**Lemma 1.3.13.** *In a causal theory, a state is normalized if and only if it is deterministic.*

*Proof.* It is a trivial corollary of proposition 1.3.6, as we have already noted. □

**Example 1.3.14.** In quantum mechanics, we have

$$\|\rho\| = \mathrm{tr}\, \mathbf{1}\rho = \mathrm{tr}\, \rho.$$

Therefore normalized states are density operators (the trace is equal to 1).

For every state $\rho$ of a causal theory we can consider the normalized state

$$\overline{\rho} = \frac{\rho}{\|\rho\|}.$$

This means that we can perform a rescaled preparation. Suppose we have the preparation-test $\{\rho_i\}$. Clearly $\|\rho_i\| \leq 1$ and one has equality if and only if this is a single-outcome preparation-test. Even in the case of multiple outcomes, if we have the state $\rho_{i_0}$, we can promote it to a normalized state $\overline{\rho}_{i_0}$. This means that in a causal theory, each preparation-event can be promoted to a single-outcome preparation-test, that is a deterministic state. This characterization of causal theories in terms of rescaled preparations is so strong that it is a sufficient condition for causality.

**Lemma 1.3.15.** *A theory where every state is proportional to a deterministic state, which in general depends on the particular state considered, is causal.*

*Proof.* Let $\rho$ be a generic state of system A. Suppose, by contradiction, there are two deterministic effects tr and tr$'$ for system A. By hypothesis, $\rho = k\overline{\rho}$, where $\overline{\rho}$ is a deterministic state and in general it depends on $\rho$. Since $\overline{\rho}$ is deterministic, the composition of $\overline{\rho}$ with tr and tr$'$ is the deterministic effect of the trivial system, which is 1. Then,

$$\mathrm{tr}\, \rho = k\mathrm{tr}\, \overline{\rho} = k = k\mathrm{tr}'\, \overline{\rho} = \mathrm{tr}'\, \rho.$$

Since $\rho$ is arbitrary, tr $=$ tr$'$, hence the theory is causal. □

In a causal theory, every non-normalized state $\rho_i$ can be written as $\rho_i = p_i\overline{\rho}$, where $p_i \in [0,1]$ and $\overline{\rho}$ is a normalized state. Clearly, $p_i = \|\rho_i\|$, and since $p_i \in [0,1]$, we can regard $\rho_i$ as a randomization of the deterministic state $\overline{\rho}$. Indeed, the norm of a state is the probability of preparing that state in a given preparation-test, as in quantum theory. Recall that tr $\rho_i$ gives the conditional probability of tr given $\rho_i$. Since tr is deterministic, the probability comes only from the preparation of $\rho_i$. Therefore states with vanishing norm cannot be prepared, so they are not true states.

The norm of states satisfies the following property.

**Proposition 1.3.16.** *If* $\mathcal{C} \in \mathsf{Transf}(\mathrm{A}, \mathrm{B})$ *is a transformation and* $\rho \in \mathsf{St}(\mathrm{A})$*, then*

$$\|\mathcal{C}\rho\|_{\mathrm{B}} \leq \|\rho\|_{\mathrm{A}},$$

*and one has the equality if and only if* $\mathcal{C}$ *is a channel.*

*Proof.* By definition, $\|\mathcal{C}\rho\|_{\mathrm{B}} = \mathrm{tr}_{\mathrm{B}}\mathcal{C}\rho_{\mathrm{A}}$. Since $\mathrm{tr}_{\mathrm{B}}\mathcal{C}$ is an effect of system A, we have $\mathrm{tr}_{\mathrm{B}}\mathcal{C}\rho_{\mathrm{A}} \leq \|\rho\|_{\mathrm{A}}$ by proposition 1.3.11. Then we have $\|\mathcal{C}\rho\|_{\mathrm{B}} \leq \|\rho\|_{\mathrm{A}}$.

By proposition 1.3.6, $\mathcal{C}$ is a channel if and only if $\mathrm{tr}_{\mathrm{B}}\mathcal{C} = \mathrm{tr}_{\mathrm{A}}$ , then

$$\|\mathcal{C}\rho\|_{\mathrm{B}} = \mathrm{tr}_{\mathrm{B}}\mathcal{C}\rho_{\mathrm{A}} = \mathrm{tr}_{\mathrm{A}}\rho_{\mathrm{A}} = \|\rho\|_{\mathrm{A}}.$$

□

Extending the norm to every element of $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$, we can use it to define a topology on it. Consider the closure of $\mathsf{St}(\mathrm{A})$. It is the set of points of $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$ such that there is a sequence of states converging to them. In other words, every point in the closure of $\mathsf{St}(\mathrm{A})$ can be approximated with arbitrary precision by physical states. It is then sensible to assume that every closure point of $\mathsf{St}(\mathrm{A})$ is a state that can be prepared in a laboratory, therefore $\mathsf{St}(\mathrm{A})$ is closed.

**Assumption 1.3.17.** *For all systems* A *the set* $\mathsf{St}(\mathrm{A})$ *is closed.*

**Lemma 1.3.18.** *If a probabilistic theory is not deterministic, then* $\mathsf{St}(\mathrm{I}) = [0, 1]$.

*Proof.* Let us prove that the closure of $\mathsf{St}(\mathrm{I})$ is $[0, 1]$. If the theory is not deterministic, there is a binary test $\{p_0, p_1\}$ from the trivial system to itself. This test can be thought as a biased coin, and tossing this coin many times, according to the law of large numbers, we can obtain an arbitrary approximation of a coin with any bias $p \in [0, 1]$ (for further details see [28]). This proves that $\mathsf{St}(\mathrm{I})$ is dense in $[0, 1]$. Since $\mathsf{St}(\mathrm{I})$ is closed, then $\mathsf{St}(\mathrm{I}) = [0, 1]$. □

We can define also a norm for effects. The simplest way is the following, close to the statement of proposition 1.3.11.

**Definition 1.3.19.** Let $a \in \mathsf{Eff}(\mathrm{A})$. We define the *norm* of $a$ as

$$\|a\| := \max_{\rho \in \mathsf{St}(\mathrm{A})} (a|\rho).$$

Even in this case $0 \leq \|a\| \leq 1$. Clearly, for the deterministic effect, $\|\mathrm{tr}\| = 1$, because $\mathrm{tr}\,\rho = 1$ if $\rho$ is normalized.

We can also define a norm for general transformations [28].

**Definition 1.3.20.** Let $\mathcal{C} \in \mathsf{Transf}(\mathrm{A}, \mathrm{B})$. We define the *norm* of $\mathcal{C}$ as

$$\|\mathcal{C}\| := \sup_{\mathrm{C}} \sup_{\rho \in \mathsf{St}(\mathrm{A}\otimes\mathrm{C})} \|\mathcal{C}\rho\|_{\mathrm{BC}}\,.$$

We have to add an ancillary system C and to calculate the supremum over the states $\rho$ of $\mathrm{A} \otimes \mathrm{C}$ of the norm of $\mathcal{C}\rho$. Eventually, we take the supremum over all possible ancillary systems.

After defining such norms, it is possible to prove that the sets of states, effects and transformations are convex in a non-deterministic causal theory.

**Proposition 1.3.21.** *If a causal theory is not deterministic, then for all systems* A *and* B*, the sets* $\mathsf{St}(\mathrm{A})$, $\mathsf{Eff}(\mathrm{A})$ *and* $\mathsf{Transf}(\mathrm{A}, \mathrm{B})$ *are convex.*

*Moreover, even* $\mathsf{St}_1(\mathrm{A})$ *is convex.*

*Proof.* Let $p \in [0, 1]$. Since we proved that $\mathsf{St}(\mathrm{I}) = [0, 1]$ for a non-deterministic theory (see lemma 1.3.18), $p \in \mathsf{St}(\mathrm{I})$. Let $\{\mathcal{C}_i\}_{i\in X}$ and $\{\mathcal{D}_j\}_{j\in Y}$ be tests from A to B. By randomization, we can consider the test $\{p\mathcal{C}_i\}_{i\in X} \cup \{(1-p)\,\mathcal{D}_j\}_{j\in Y}$. By coarse-graining, the convex combination $p\mathcal{C}_i + (1-p)\,\mathcal{D}_j$, is still a transformation from A to B. Taking A or B equal to the trivial system, one has the thesis for states and effects.

Let us prove that any convex combination of normalized states is a normalized state. Let $\rho$ and $\sigma$ be two normalized states of system A. Then

$$\|p\rho + (1-p)\,\sigma\| = (\mathrm{tr}|p\rho + (1-p)\,\sigma) = p\mathrm{tr}\,\rho + (1-p)\,\mathrm{tr}\,\sigma =$$

$$= p + 1 - p = 1.$$

□

This proposition shows that convex combinations of (normalized) states, effects and transformations are still (normalized) states, effects and transformations respectively. Clearly, pure states, pure effects and pure transformations are the extreme points of such sets.

Let us focus on the set of normalized states. We want to show that convex combinations of normalized states do not have only a mathematical meaning, but can be also realized operationally. Suppose we have $\rho_p = p\rho_0 + (1-p)\,\rho_1$, where $\rho_0, \rho_1 \in \mathsf{St}_1(\mathrm{A})$. We can prepare $\rho_p$ by using the following procedure.

1. First of all, we perform a binary test in some arbitrary system with outcomes $\{0, 1\}$ and outcome probabilities $p_0 = p$ and $p_1 = 1 - p$.

2. If the outcome is $i$, then we prepare $\rho_i$. In this way, we realize the preparation-test $\{p_0\rho_0, p_1\rho_1\}$. Note that each state $p_i\rho_i$ is not normalized because it is not deterministic: the state $\rho_i$ is prepared with probability $p_i$.

3. Finally, we perform a coarse-graining over the outcomes, getting $\rho_p = p\rho_0 + (1 - p)\,\rho_1$.

In the following, we will mainly focus on normalized states, because every non-normalized state can be reduced to a normalized state. A coarse-graining of normalized states is a non-trivial convex combination of them. Clearly pure states admit only trivial convex decompositions. Every convex decomposition of a state $\rho$ reflects a particular way of preparing $\rho$.

# Chapter 2

# A general framework for resource theories

In this chapter we move to the core topic of the present work, namely how to build a rigorous mathematical description of a resource theory. Resource theories arise in a great number of scientific disciplines, essentially whenever there is something valuable to fulfil some specific task. Clearly, in general, not all the resources are equally valuable with respect to the task one has to perform. Some are very abundant, and therefore not precious, others are the most coveted for their intrinsic value or their ability to facilitate some processes otherwise very hard to realize. The prototypical example of a resource theory is chemistry, where chemical species are the resources. Some of them, "raw" products, are abundant and not so precious, others, instead, are desired because of their industrial relevance as fertilizers, pigments, etc. Then we have chemical reactions, which transform chemical species into others. Here we have the basic ingredients of a theory of resources: resources and their transformations. In this scheme, it is important to quantify the value of resources or, at least, to establish a hierarchy among them.

Given the importance of resources, we would like to find a framework that captures the common features of different theories of resources, even in most different fields. Having such a general mathematical framework will enable us to apply a resource theory even outside the framework in which it was developed originally, and a natural way to do that is to resort to GPTs.

In the present treatment, we will mainly follow the approach of [33]. In section 2.1 we give the general definition of a theory of resource. The strong point about this definition is that it can be exported to several different fields

and in this way some theories come up as resource theories quite unexpectedly. One of the key aspects is that of the cost of resources: some of them are free, others are costly, and these are usually the most coveted. Therefore it is particularly significant to try to establish a hierarchy among resources according to their usefulness, as noted above. This is done in section 2.2. One of the related topics is to find functions which measure the "value" of a resource directly in such a way that it is compatible with their hierarchy.

Finally, in section 2.3 we see how the paradigm of resource theories can be exported to the class of GPTs in which the resources are states.

## 2.1 Resource theories

The key concept of a resource theory, as the name says, is resource. Loosely speaking, a resource is something precious to perform some specific task, such as a chemical reaction, a thermodynamic procedure, a communication or computational task. The definitions of resources in different theories or fields may be very different. To make this apparent, let us consider two basic examples, one taken from chemistry and another taken from computation.

**Example 2.1.1.** Consider the synthesis process of ammonia, according to the following chemical reaction.

$$N_2 + 3H_2 \longrightarrow 2NH_3$$

If our task is to synthesise ammonia, then nitrogen and hydrogen are resources, because once we have them, we are able to fulfil our task. Regarding this situation in the light of GPTs, we can think of nitrogen and hydrogen as "states", for they are related to the preparation of a system. Indeed, if we want the above reaction to occur, we should *prepare* a flask with nitrogen and hydrogen inside.

**Example 2.1.2.** Consider the case of classical computation. Here logical gates are a resource, but, in the language of GPTs, gates are transformations rather than states.

We have just seen that resources can be states or transformations, thus the landscape is varied. Therefore, in order to accomplish a general description of the theories of resources, one must identify the basic and common features resources possess.

A typical example of resource theory is chemistry, to which we will resort for an intuitive understanding of the features of a generic theory of resources.

In general, in a resource theory we have different *resources*, say A, B, C, etc., which we can think of as chemical species. Clearly, if A and B are resources when taken singularly, they are also a resource when taken together, forming a so-called *composite resource*, which we will denote as $\mathrm{A} \otimes \mathrm{B}$. Clearly, the order in which we consider A and B does not matter.

*Remark* 2.1.3. Let us better clarify the statement about the fact that the order in which we consider A and B does not matter. Is it really true that $\mathrm{A} \otimes \mathrm{B} = \mathrm{B} \otimes \mathrm{A}$? The example of chemistry will help to understand the issue better. Suppose we have a flask divided into two parts, where each part contains a different gas (hydrogen and nitrogen in the ammonia example). $\mathrm{A} \otimes \mathrm{B}$ may stand for "gas A is in part 1, and gas B is in part 2"; therefore $\mathrm{B} \otimes \mathrm{A}$ stands for "gas B is in part 1, and gas A is in part 2". Although these two settings are equivalent from the point of view of the resources involved, we cannot say they are exactly the same setting. Therefore, we prefer to say that they are *equivalent*, so that we write $\mathrm{A} \otimes \mathrm{B} \cong \mathrm{B} \otimes \mathrm{A}$.

It is useful to assume that "nothing" is a resource, called the *void resource* I. Clearly, after adding the void resource I to a resource A, we still have A.

In a resource theory, there are also some *processes* or *transformations* which transform a resource into another, say A into B. We can think of chemical reactions as an example of such processes. In general, there may be different ways of transforming A into B (e.g. different chemical reactions from the chemical species A to the chemical species B), therefore it is necessary to label processes with the same input and output resources to be able to distinguish them. For example, we will write $f : \mathrm{A} \longrightarrow \mathrm{B}$ and $g : \mathrm{A} \longrightarrow \mathrm{B}$ to distinguish two processes $f$ and $g$ with the same input and output resources. The fundamental requirement of processes converting resources is that they must be *freely implementable*, or, in other words, "easy" to realize from an operational point of view. At least in principle, there must be no "cost" in implementing an arbitrary number of such operations. We will come back to this point in section 2.3.

Clearly, if we can transform a resource A into a resource B and B into C, we can also transform A into C. In other words, if $f : \mathrm{A} \longrightarrow \mathrm{B}$ and $g : \mathrm{B} \longrightarrow \mathrm{C}$ are the two processes involved, there is another process $g \circ f$ transforming A into C. We say that we have performed the two processes $f$ and $g$ *in sequence*.

Similarly, processes can be composed *in parallel* when we have a composite resource. Indeed, suppose we have again a flask with a partition, where each part contains a different gas. Thanks to the partition, we can have two different chemical reactions occurring at the same time in the two parts. In this way, each of the two resources is transformed independently of the other.

Let us sum up the properties of resources and of their (free) transformations. As far as resources are concerned, one has:

- there is a void resource I which means "nothing";
- resources can be composed;
- $A \otimes I = I \otimes A = A$ for every A;
- $A \otimes B \cong B \otimes A$ for all A and B.

As far as transformations are concerned, one has:

- transformations can be composed in sequence;
- transformations can be composed in parallel;

We note a formal analogy between systems in GPTs and resources, and between transformations in GPTs and transformations in resource theories. This means that a resource theory is a strict symmetric monoidal category, with resources as objects and free transformations as morphisms.

**Definition 2.1.4.** A resource theory is a strict symmetric monoidal category $(\mathbf{D}, \circ, \otimes, I)$, where

- the objects are the resources;
- I is the void resource;
- the morphisms are free transformations between resources;
- $\circ$ is the sequential composition of transformations;
- $\otimes$ gives the composition of resources and the parallel composition of transformations.

Notably, the identity is always a free transformation.

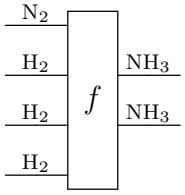
Clearly, strict symmetric monoidal categories are merely the language in which to formulate the structure of resource theories, and they lack all the interpretative part about the content of a given theory of resources. In fact, that part is usually the most important part, for it distinguishes the resource theory of chemistry from, say, the resource theory of thermodynamics. From the categorical perspective, indeed, the resources of different theories are not different in their operational structure, namely when we focus on how to compose and manipulate them.

Since resource theories are strict symmetric monoidal categories, we can apply the same graphical language we used for GPTs. Therefore, we represent resources in the same way as systems (i.e. as wires) and processes in the same way as transformations (i.e. as boxes).

**Example 2.1.5.** Consider again the chemical reaction of example 2.1.1, which we will label as $f$.

$$f : \mathrm{N_2} + 3\mathrm{H_2} \longrightarrow 2\mathrm{NH_3}$$

Here we have two basic resources: nitrogen ($\mathrm{N_2}$) and hydrogen ($\mathrm{H_2}$). Let us regard this reaction from an operational point of view. An experimenter must prepare 1 mol of nitrogen and 3 mol of hydrogen. If we think of one mole as a unit resource, then 3 mol of hydrogen is already a composite resource. The experimenter prepares the composite resource $\mathrm{N_2} + 3\mathrm{H_2}$, and then triggers the reaction. Eventually we get 2 mol of ammonia. This is represented graphically as

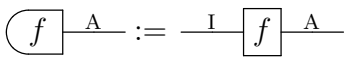



.

Some resources are "free": we may consider them as abundant or easy to get, so there is no cost (whatever it might mean) in producing them. Usually free resources are not very useful and the goal is to produce valuable resources, which are rarer.

The most trivial example of a free resource is the void resource I, because "nothing" is definitely free. Since the processes we are considering are all free, a process from the void resource will have another free resource A as

output. The situation is represented as follows.

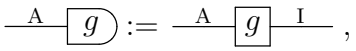



Here A is the free resource, and $f$ is the preparation process for the free resource.

**Definition 2.1.6.** A *free resource* A is an object in $|\mathbf{D}|$ such that there exists a free process from the void resource to A.

A non-free resource is called *costly*.

This definition is particularly important and will play a central role in section 2.3. By a symmetry argument, if a free resource can be prepared from "nothing" via a free process, it can be also destroyed by a free process.

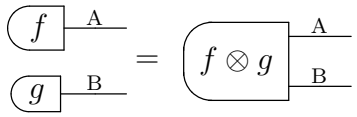



where $g$ is the destruction process of the free resource A. A reason for requiring this symmetry will appear in the next section (remark 2.2.3).

The composition of two free resources is still a free resource, as it is obvious, since the cost of two resources cannot change if we put them together. This is apparent also from the graphical language, which gives a formal proof thereof.

**Proposition 2.1.7.** *Compositions of free resources are still free resources.*

*Proof.* Indeed, if $f$ and $g$ are the preparation processes for resources A and B from the void resource, we have

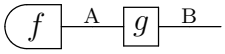



which shows explicitly that $\mathrm{A} \otimes \mathrm{B}$ is still a free resource. □

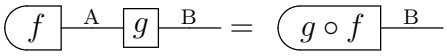
Moreover, the manipulation of free resources through free processes yields free resources; indeed no cost is added in such a manipulation. Again, to have a formal proof, we resort to graphical language.

**Proposition 2.1.8.** *Free processing of free resources yields free resources.*

*Proof.* Suppose $f$ is the preparation process of the free resource A and $g$ is a (free) transformation from A to a resource B.




The sequential composition $g \circ f$ is a transformation with the void resource as input and B as output, thus showing that B is a free resource.




□

We end this section with an example of a non-obvious resource theory, which is important in cryptography.

**Example 2.1.9.** In order to have a secure cryptographic protocol, one must have randomness. This concept can be easily fitted into the framework of resource theories. In such a setting, resources are probability distributions. We can represent them as probability spaces $(\Omega, \mathcal{F}, \mathbb{P})$ [58, 59], where $\Omega$ is the sample space, $\mathcal{F}$ is the $\sigma$-algebra of events and $\mathbb{P}$ is a probability measure on $\mathcal{F}$. Given two resources $\mathrm{A} = (X, \mathcal{F}, \mathbb{P})$ and $\mathrm{B} = (Y, \mathcal{G}, \mathbb{Q})$ we can consider their composition $\mathrm{A} \otimes \mathrm{B}$ as the probability space $(X \times Y, \mathcal{F} \otimes \mathcal{G}, \mathbb{P} \times \mathbb{Q})$. The void resource is the trivial probability space, which has $X = \{x\}$, $\mathcal{F} = \{\varnothing, \{x\}\}$ and $\mathbb{P}$ such that $\mathbb{P}(\{x\}) = 1$, namely $\mathbb{P}$ is the Dirac delta measure[1] at $x$.

A transformation between two probability spaces $(X, \mathcal{F}, \mathbb{P})$ and $(Y, \mathcal{G}, \mathbb{Q})$ is a measurable function $f : X \longrightarrow Y$ such that for every event $E \in \mathcal{G}$, one has $\mathbb{Q}(E) = \mathbb{P}(f^{-1}(E))$. The definition of sequential composition of processes is straightforward: it is just the composition of measurable functions. The parallel composition of two processes $f : X_1 \longrightarrow Y_1$ and $g : X_2 \longrightarrow Y_2$ is

$$f \otimes g : (X_1 \times X_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mathbb{P}_1 \times \mathbb{P}_2) \mapsto (Y_1 \times Y_2, \mathcal{G}_1 \otimes \mathcal{G}_2, \mathbb{Q}_1 \times \mathbb{Q}_2)$$

[1]Let $x \in X$, where $X$ is an arbitrary set. For every $E \subseteq X$, recall we define the Dirac delta measure at $x$ in the following way.

$$\delta_x(E) = \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{if } x \notin E \end{cases}$$

such that for all events $E_1 \in \mathcal{G}_1$ and $E_2 \in \mathcal{G}_2$ one has $\mathbb{Q}_1(E_1) = \mathbb{P}_1(f^{-1}(E_1))$ and $\mathbb{Q}_2(E_2) = \mathbb{P}_2(g^{-1}(E_2))$. In other words, the two measurable functions act on the two probability spaces independently:

$$f \otimes g : \mathrm{A} \otimes \mathrm{B} \longrightarrow \mathrm{A}' \otimes \mathrm{B}',$$

where $f : \mathrm{A} \longrightarrow \mathrm{A}'$ and $g : \mathrm{B} \longrightarrow \mathrm{B}'$.

Now we can try to identify all the free resources. A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a free resource if there is a transformation $f$ from the trivial probability space to $(\Omega, \mathcal{F}, \mathbb{P})$. Any function $f$ from the trivial probability space is measurable. This means that for every event $E \in \mathcal{F}$, one has $\mathbb{P}(E) = \delta_x(f^{-1}(E))$. Recalling the definition of Dirac delta measure, one has

$$\mathbb{P}(E) = \begin{cases} 1 & \text{if } f(x) \in E \\ 0 & \text{if } f(x) \notin E \end{cases} = \delta_{f(x)}(E),$$

whence the probability measure is in fact a Dirac measure, and we can take $\mathcal{F} = \mathcal{P}(X)$, where $\mathcal{P}(X)$ is the power set of $X$. Therefore all the probability spaces having a Dirac measure as their probability measure are free resources. In such spaces there is no randomness, because there is one event having probability 1. Hence, in this resource theory true randomness is a precious resource.

The situation is simpler and maybe clearer if we are dealing with finite sample spaces, $\Omega = \{\omega_1, \ldots, \omega_n\}$. In this case, we can take $\mathcal{F}$ to be the power set of $\Omega$, $\mathcal{F} = \mathcal{P}(\Omega)$, and we can have a complete description of the probability space by defining a probability distribution $p = \{p_1, \ldots, p_n\}$, where $p_i := \mathbb{P}(\{\omega_i\})$, for every $i$, and $\mathbb{P}$ is a discrete probability measure. The void resource is still defined as above: a sample space with one element. The composition of two resources $(X, p)$ and $(Y, q)$ is now defined as the resource $(X \times Y, pq)$, where $pq$ is the product of the two probability distributions.

Now any function is measurable and transformations between different resources can be defined as follows. Let $(X, p)$ and $(Y, q)$ be two finite sets with probability distributions $p$ and $q$ and probability measures $\mathbb{P}$ and $\mathbb{Q}$. Let $f$ be a transformations between them. According to what we said above,

$$q_i := \mathbb{Q}(\{y_i\}) = \mathbb{P}\left(f^{-1}(y_i)\right) = \mathbb{P}\left(\bigcup_{x_j \in f^{-1}(y_i)} \{x_j\}\right),$$

and the union is finite. According to the properties of probability measures, then

$$\mathbb{P}\left(\bigcup_{x_j\in f^{-1}(y_i)}\{x_j\}\right)=\sum_{x_j\in f^{-1}(y_i)}\mathbb{P}\left(\{x_j\}\right)=\sum_{x_j\in f^{-1}(y_i)}p_j,$$

whence $q_i=\sum_{x_j\in f^{-1}(y_i)}p_j$. Sequential and parallel compositions are defined in the same way as above.

Now free resources are represented by probability distributions associated with a Dirac measure, namely extreme probability distributions, where one outcome has probability 1, such as in $\{1,0,\ldots,0\}$.

## 2.2 A hierarchy among resources

Now we wish to enter into more details about the cost of a resource, which is a measure of its value. In the previous section, we divided resources into two classes: free and costly resources. This is a rough classification of resources. Now we will go a step further by trying to establish a hierarchy even among costly resources. Indeed, some of them might be more valuable than others.

As it often happens experimentally, the processing of resources consumes or degrades the resources, even if such a processing is free. In this way, from a precious resource, we end up with resources which are more and more useless, until eventually we have only free resources. Therefore, the most natural way to define an ordering of resources is to take advantage of this idea about resource processing.

**Definition 2.2.1.** In a resource theory, we say that a resource A is *more valuable* than a resource B, and we write $\mathrm{A}\succsim\mathrm{B}$, if there exists a free transformation $f:\mathrm{A}\longrightarrow\mathrm{B}$.

Now the issue of establishing a hierarchy among resources has been turned into an issue of resource convertibility, namely, given A and B, to find whether there exists a free process transforming A into B or B into A. Such a transformation sometimes does not exist, and in this case A and B are "incomparable": we cannot establish which of the two is the more valuable.

Note that we are not interested in analysing the properties of the specific transformation, it is enough to know that a transformation exists. Therefore it is not necessary to use the full categorical structure, but one can only require some weaker axioms to have a theory of resource convertibility, as

shown in [33]. However, in the present treatment, we prefer to stick to the full structure of a resource theory, rather than giving another definition, because the issue of resource convertibility arises naturally in this framework.

Let us study the properties of the relation "to be more valuable than".

**Proposition 2.2.2.** *The relation "to be more valuable than" is a preorder*[2]*.*

*Proof.* The relation "to be more valuable than" is reflexive. Indeed for any resource A, $A \succsim A$ because the identity transforms A into A, and the identity is a free transformation.

The relation is also transitive. Indeed, suppose we have $A \succsim B$ and $B \succsim C$. This means that there exist a transformation $f : A \longrightarrow B$ and a transformation $g : B \longrightarrow C$. Taking the sequential composition, we have the transformation $g \circ f : A \longrightarrow C$, which means that $A \succsim C$. □

In general, however, we *cannot* conclude that, if $A \succsim B$ and $B \succsim A$, then $A = B$. This only means that it is possible to convert A into B through a process $f$ and B into A through a process $g$. Note that this does *not* even mean that $f = g^{-1}$.

However, if $A \succsim B$ and $B \succsim A$, we can think of A and B as *equivalent*, and we say that A is *as valuable as* B. Indeed, from the theory of preorders, it is known that we can define an equivalence relation $\sim$, where $A \sim B$ if $A \succsim B$ and $B \succsim A$, for all A and B. Taking the quotient set of the set of resources $|\mathbf{D}|$ modulo $\sim$, the preorder $\succsim$ turns into a partial order $\succeq$ between equivalence classes (see appendix A.1).

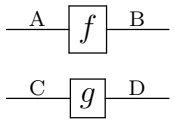
*Remark* 2.2.3. Now we can understand why we required that a free resource is destroyed by a free process. If A is a free resource, there is a free process $f : I \longrightarrow A$ which can prepare it from "nothing". According to our definition of the "more valuable" relation, this means that the void resource I, i.e. "nothing", is more valuable than a real resource! This would be quite absurd if we did not set that there is also a free process destroying A, thus implying also $A \succsim I$. In this way we conclude that $A \sim I$. Note how useless free resources are: they are equivalent to "nothing"!

Let us prove that the preorder we have defined is compatible with the operation of composition of resources. In other words if $A \succsim B$ and $C \succsim D$, then one has $A \otimes C \succsim B \otimes D$.

[2]Recall that a preorder is a relation which is reflexive and transitive.

**Proposition 2.2.4.** *The preorder "to be more valuable than" is compatible with the composition of resources.*

*Proof.* Let A, B, C, D be resources such that $A \succsim B$ and $C \succsim D$; then there exist a process $f : A \longrightarrow B$ and a process $g : C \longrightarrow D$. From a graphical point of view, when we compose the resources, we have




,

this means that there is the process $f \otimes g$ from $A \otimes C$ to $B \otimes D$, therefore $A \otimes C \succsim B \otimes D$. □

Therefore even the equivalence relation $\sim$ associated with the preorder $\succsim$ is compatible with the composition of resources.

**Proposition 2.2.5.** *Let* A *be a* free *resource. Then* $A \sim A \otimes A$.

*Proof.* We know that for any resource A, $A = A \otimes I$, where I is the void resource. Since A is free, then $I \sim A$. Because $A \sim A$ and $\sim$ is compatible with the composition of resources, one has $A = A \otimes I \sim A \otimes A$. □

The meaning of this proposition is that for free resources, as one expects, there is no cost in preparing an arbitrary number of copies of them and having more copies is just like having a single copy.

Before going further to explore other features related to the hierarchy of resources and the issue of resource convertibility, let us better familiarize with the preorder with a couple of examples [33].

The first example is about the resource theory of food.

**Example 2.2.6.** Food is clearly a resource, therefore it is natural to develop a resource theory of it. As a toy example, yet showing the main features of such a resource theory, consider the case in which there are only apples and bananas. Here a resource is a basket of fruit, with $a$ apples and $b$ bananas. We can represent it as a pair of natural numbers $(a, b)$. The void resource is an empty basket, namely the pair $(0, 0)$. The rule of composition of resources is fairly intuitive: $(a, b) \otimes (a', b') = (a + a', b + b')$. In this way, as it is obvious, when we have a basket with $a$ apples and $b$ bananas, and a basket with $a'$ apples and $b'$ bananas, it is as if we had a basket with $a + a'$ apples and $b + b'$ bananas.

Here the only possible (free) process is eating from a basket of fruit, namely we can go from $(a, b)$ to $(a', b')$ if we eat some or no fruit, namely if $a \geq a'$ *and* $b \geq b'$. Therefore, the only free resource is the void resource (all food is expensive!), because the only possibility of having a free process from the void resource $(0, 0)$ to a resource $(a, b)$ is that $0 \geq a$ and $0 \geq b$, which means $a = b = 0$. Here the sequential composition of eating is straightforward: it means eating in sequence. The parallel composition is also straightforward to define: suppose we have two baskets of fruit, the parallel composition of two eating processes means simply eating from each of the two baskets independently.

According to definition 2.2.1, we define a preorder by setting $(a, b) \succsim (a', b')$ if there is an eating process from $(a, b)$ to $(a', b')$, namely if $a \geq a'$ and $b \geq b'$. This definition is sensible, for a basket of fruit is more valuable than another if it contains more apples and more bananas, or the same number of the two kinds of fruit. Note that in this example the consumption of resources by free processes is particularly evident.

Let us find out when two baskets of fruit are equivalent as resources. Suppose $(a, b) \succsim (a', b')$ and $(a', b') \succsim (a, b)$. This means that $a \geq a'$ and $a' \geq a$, and $b \geq b'$ and $b' \geq b$, which implies $a = a'$ and $b = b'$. Therefore two baskets of fruit are equivalent if they have the same number of apples and bananas. Note that even if they have the same number of the two types of fruit, in general they are not the same basket, and this is why antisymmetry fails and we have a preorder instead of an order.

This is not a total[3] preorder, because $(1, 0)$ (one apple) and $(0, 1)$ (one banana) are not "comparable". Indeed, there is no eating process capable of turning an apple into a banana or vice versa. Mathematically, $1 \geq 0$ but $0 \not\geq 1$.

The second example we wish to consider is the resource theory of knowledge.

**Example 2.2.7.** Knowledge a person has is definitely a resource. Again, let us concentrate on a toy model, where we have only two subjects in which to be proficient, say algebra and biology. Let us quantify the level of knowledge a person has of a subject by a natural number (e.g. the grade got in some exam); therefore we represent her proficiency as a pair of natural numbers

[3]Recall that a relation $\mathcal{R}$ on a set $X$ is called total if, for every $x_1, x_2 \in X$, $x_1 \mathcal{R} x_2$ or $x_2 \mathcal{R} x_1$ or both.

$(a, b)$, $a$ for algebra and $b$ for biology. The void resource is represented by a person who knows neither algebra nor biology, so we represent it as the pair $(0, 0)$.

How do we combine the proficiencies of two people to represent the knowledge of the pair? Clearly knowledge is not additive, for if we have two equally proficient people, the level of knowledge of the pair is the same as that of the individuals. It is more reasonable to take the proficiency level of the pair in each of the two subjects as the maximum of the proficiency level of the two people. In other words, the more expert determines the level of proficiency of the group. In mathematical terms, $(a, b) \otimes (a', b') = (\max\{a, a'\}, \max\{b, b'\})$.

Here we consider forgetting or losing proficiency as the only free operation (learning comes at a cost!). Again, we can go from $(a, b)$ to $(a', b')$ if $a \geq a'$ and $b \geq b'$. As in the previous example, the only free resource is the void resource, a person who has knowledge of neither of the disciplines.

The preorder is exactly the same as in the previous example: a person is more valuable than another if she is more proficient in both algebra and biology, and two people are equivalent if their knowledge of algebra and biology is the same. Again, this does not imply at all that the two people are the same! Moreover, once more the preorder is not total.

Although these two example might appear as very similar in the structure of the preorder, they give rise to quite different resource theories. The origin of the difference is their monoidal structure, namely the way in which resources are composed. The resource theory of food (example 2.2.6) is quantitative, meaning that resources compose by addition, whereas the resource theory of knowledge (example 2.2.7) is qualitative. We will enter into further details soon.

### 2.2.1 Some phenomenology

In the light of the hierarchy of resources we have just introduced, we wish to study some of the features a resource theory can have. The first issue is catalysis, like in chemistry. Catalysts facilitate chemical reactions, i.e. processes between resources. In a similar fashion, but in a more general context, sometimes it is not possible to have a conversion of a resource A into a resource B, but, if we add a resource C to A, then the conversion of A into B becomes possible and in the end we recover the resource C, as if it were not consumed during the process. Therefore C acts as a catalyst for

the resource conversion.

**Definition 2.2.8.** A resource C is a *catalyst* for the conversion of resource A into resource B if $\mathrm{A} \not\succsim \mathrm{B}$ but $\mathrm{A} \otimes \mathrm{C} \succsim \mathrm{B} \otimes \mathrm{C}$.

Let us see if we are able to find catalysts for the resource theory of knowledge.

**Example 2.2.9.** Consider the resource theory of knowledge (example 2.2.7). Suppose $\mathrm{A} = (a, b)$ and $\mathrm{B} = (a', b')$. Here $\mathrm{A} \not\succsim \mathrm{B}$ if $a < a'$ or $b < b'$. Now we want to find a catalyst $\mathrm{C} = \left(\widetilde{a}, \widetilde{b}\right)$ such that $\mathrm{A} \otimes \mathrm{C} \succsim \mathrm{B} \otimes \mathrm{C}$. Now,

$$\mathrm{A} \otimes \mathrm{C} = (a, b) \otimes \left(\widetilde{a}, \widetilde{b}\right) = \left(\max\left\{a, \widetilde{a}\right\}, \max\left\{b, \widetilde{b}\right\}\right),$$

and similarly $\mathrm{B} \otimes \mathrm{C} = \left(\max\left\{a', \widetilde{a}\right\}, \max\left\{b', \widetilde{b}\right\}\right)$. One has $\mathrm{A} \otimes \mathrm{C} \succsim \mathrm{B} \otimes \mathrm{C}$ if and only if $\max\left\{a, \widetilde{a}\right\} \geq \max\left\{a', \widetilde{a}\right\}$ and $\max\left\{b, \widetilde{b}\right\} \geq \max\left\{b', \widetilde{b}\right\}$. Let us distinguish three cases.

1. $a < a'$ but $b \geq b'$. In this case A and B are incomparable. Here the condition $\max\left\{b, \widetilde{b}\right\} \geq \max\left\{b', \widetilde{b}\right\}$ is always satisfied. In order to have $\max\left\{a, \widetilde{a}\right\} \geq \max\left\{a', \widetilde{a}\right\}$, it must be $\widetilde{a} \geq a'$.

2. $a \geq a'$ but $b < b'$. Again A and B are incomparable. This is the symmetric situation of the one above: there are no restriction on $a$, but it must be $\widetilde{b} \geq b'$.

3. $a < a'$ and $b < b'$. In this case we have $\mathrm{B} \succsim \mathrm{A}$. Now it must be $\widetilde{a} \geq a'$ and $\widetilde{b} \geq b'$, namely $\mathrm{C} \succsim \mathrm{B}$.

We were able to find constraints on C in each of the three cases so that C is a catalyst. Let us try to understand these constraints intuitively. When we have $\mathrm{A} \not\succsim \mathrm{B}$, it means that A is less proficient than B in some of the two subjects (possibly both). A catalyst C is then an expert whose knowledge in the field in which A defects is higher than the corresponding knowledge B has. To make a comparison with the three cases above, when $\mathrm{A} \not\succsim \mathrm{B}$, we can have three different situations.

1. A is less proficient than B in algebra but not in biology. In this case the catalyst C is an expert in algebra (more expert than B in algebra), irrespective of her own expertise in biology. In this case A and C together have more knowledge than B and C together.

2. A is less proficient than B in biology but not in algebra. Now C is an expert in biology (more expert than B in biology), irrespective of her own expertise in algebra. In this case A and C together have more knowledge than B and C together.

3. A is less proficient than B both in algebra and biology. C must be an expert in both the subjects (more expert than B), so that A and C together are more proficient than B and C together.

In some resource theories, catalysts may be of no help.

**Definition 2.2.10.** A resource theory is said to be *catalysis-free* if $A \otimes C \succsim B \otimes C$ implies $A \succsim B$, for all resources A, B, C.

In this case, if a process is possible with the aid of catalysts, it is also possible without them. In other words, catalysts do not play any role in allowing processes that were impossible without them. It is particularly important to understand what theories do not allow catalysis, because we should minimize the number of resources involved in a process (recall that resources are expensive). Indeed, there is no point in using a resource C to have a process $A \otimes C \longrightarrow B \otimes C$ if it is possible to have a process $A \longrightarrow B$.

Now we will show a sufficient criterion to establish when a resource theory admits catalysts.

**Definition 2.2.11.** A resource theory is called *non-interacting* if, for all resources A, $B_1$, $B_2$ such that $A \succsim B_1 \otimes B_2$ there exist $A_1$ and $A_2$ such that $A \sim A_1 \otimes A_2$ and $A_1 \succsim B_1$ and $A_2 \succsim B_2$.

In other words, whenever a resource A is more valuable than the composition of two resources $B_1$ and $B_2$, we can find another resource $A'$, equivalent to A, such that $A'$ is the composition of two resources $A_1$ and $A_2$ such that each of them is more valuable than $B_1$ or $B_2$.

**Definition 2.2.12.** A resource theory is *quantitative* if, whenever $A_1 \otimes A_2 \sim B_1 \otimes B_2$ and $A_1 \succsim B_1$, then $B_2 \succsim A_2$, for all $A_1$, $A_2$, $B_1$, $B_2$.

Loosely speaking, when the value of two composite resources is the same ($\mathrm{A}_1 \otimes \mathrm{A}_2 \sim \mathrm{B}_1 \otimes \mathrm{B}_2$), but one of the constituent resources of the one is more valuable than one of the other ($\mathrm{A}_1 \succsim \mathrm{B}_1$), than the other constituents have to compensate for this ($\mathrm{B}_2 \succsim \mathrm{A}_2$). In this case the resources must have an extensive or quantitative flavour, because composition behaves more or less like an addition of non-negative numbers. Indeed, by replacing $\otimes$ with $+$, $\sim$ with $=$, and $\succsim$ with $\geq$, we find a similar property of addition: if $\mathrm{A}_1 + \mathrm{A}_2 = \mathrm{B}_1 + \mathrm{B}_2$, and $\mathrm{A}_1 \geq \mathrm{B}_1$, then $\mathrm{B}_2 \geq \mathrm{A}_2$.

**Example 2.2.13.** The resource theory of food (see example 2.2.6) is quantitative. Indeed, take $\mathrm{A}_1 = (a, b)$, $\mathrm{A}_2 = (a', b')$, $\mathrm{B}_1 = \left(\widetilde{a}, \widetilde{b}\right)$ and $\mathrm{B}_2 = \left(\widetilde{a}', \widetilde{b}'\right)$. We have $\mathrm{A}_1 \otimes \mathrm{A}_2 \sim \mathrm{B}_1 \otimes \mathrm{B}_2$ if and only if $(a + a', b + b') \sim \left(\widetilde{a} + \widetilde{a}', \widetilde{b} + \widetilde{b}'\right)$, namely if and only $a + a' = \widetilde{a} + \widetilde{a}'$ and $b + b' = \widetilde{b} + \widetilde{b}'$. Now suppose $\mathrm{A}_1 \succsim \mathrm{B}_1$, which means $a \geq \widetilde{a}$ and $b \geq \widetilde{b}$. In order to have the equalities $a + a' = \widetilde{a} + \widetilde{a}'$ and $b + b' = \widetilde{b} + \widetilde{b}'$, one must have $\widetilde{a}' \geq a'$ and $\widetilde{b}' \geq b'$, which means $\mathrm{B}_2 \succsim \mathrm{A}_2$.

The quantitative flavour of resources in a quantitative resource theory is well captured by the following proposition.

**Proposition 2.2.14.** *In a quantitative resource theory the following are equivalent.*

1. $\mathrm{A}$ *is a free resource*
2. $\mathrm{A} \sim \mathrm{A} \otimes \mathrm{A}$
3. $\mathrm{A} \succsim \mathrm{A} \otimes \mathrm{A}$

*Proof.* Let us prove the various implications.

1 implies 2. This holds in any resource theory, by proposition 2.2.5.

2 implies 3. This is trivial and follows from the definition of $\sim$.

3 implies 1. Clearly, we have $\mathrm{A} \otimes \mathrm{A} \sim \mathrm{A} \otimes \mathrm{A}$, which we rewrite as $(\mathrm{A} \otimes \mathrm{I}) \otimes \mathrm{A} \sim (\mathrm{A} \otimes \mathrm{A}) \otimes \mathrm{I}$. We know that $\mathrm{A} \otimes \mathrm{I} \succsim \mathrm{A} \otimes \mathrm{A}$ because this is nothing but $\mathrm{A} \succsim \mathrm{A} \otimes \mathrm{A}$. Since the resource theory is quantitative, we must have $\mathrm{I} \succsim \mathrm{A}$, which means that $\mathrm{A}$ is a free resource. □

This proposition means that in a quantitative resource theories the only resources for which there is no cost in preparing an arbitrary number of copies of them are the free resources.

Non-interacting quantitative resource theories are catalysis-free.

**Proposition 2.2.15.** *If a resource theory is non-interacting and quantitative, it is catalysis-free.*

*Proof.* Suppose we have $(\mathrm{A}\otimes\mathrm{C})\succsim\mathrm{B}\otimes\mathrm{C}$. Since the resource theory is non-interacting, there exist two resources $\mathrm{A}_1$ and $\mathrm{A}_2$ such that $\mathrm{A}\otimes\mathrm{C}\sim\mathrm{A}_1\otimes\mathrm{A}_2$ and $\mathrm{A}_1\succsim\mathrm{B}$ and $\mathrm{A}_2\succsim\mathrm{C}$. Since the resource theory is quantitative, $\mathrm{A}_2\succsim\mathrm{C}$ implies $\mathrm{A}\succsim\mathrm{A}_1$. Recalling that $\mathrm{A}_1\succsim\mathrm{B}$, one concludes that $\mathrm{A}\succsim\mathrm{B}$. □

Let us show that the resource theory of food is catalysis-free.

**Example 2.2.16.** Let us show that the resource theory of food is non-interacting. This means that whenever $\mathrm{A}\succsim\mathrm{B}_1\otimes\mathrm{B}_2$ there exist $\mathrm{A}_1$ and $\mathrm{A}_2$ such that $\mathrm{A}\sim\mathrm{A}_1\otimes\mathrm{A}_2$ and $\mathrm{A}_1\succsim\mathrm{B}_1$ and $\mathrm{A}_2\succsim\mathrm{B}_2$. Let $\mathrm{A}=(a,b)$, $\mathrm{B}_1=(a_1',b_1')$ and $\mathrm{B}_2=(a_2',b_2')$, the goal is to find $\mathrm{A}_1=(a_1,b_1)$ and $\mathrm{A}_2=(a_2,b_2)$. By hypothesis, $a\geq a_1'+a_2'$ and $b\geq b_1'+b_2'$. Since $a\geq a_1'+a_2'$, there exist a natural number $c$ such that $a=a_1'+a_2'+c$, and similarly $b=b_1'+b_2'+d$, for some natural number $d$. Now we can set $a_1:=a_1'+c$ and $a_2:=a_2'$, and, similarly $b_1:=b_1'+d$ and $b_2:=b_2'$. Note that $a_1\geq a_1'$ and $b_1\geq b_1'$, and $a_2\geq a_2'$ and $b_2\geq b_2'$. This means that $\mathrm{A}_1\succsim\mathrm{B}_1$ and $\mathrm{A}_2\succsim\mathrm{B}_2$, thus proving that the resource theory is non-interacting.

Being non-interacting and quantitative, the resource theory of food is catalysis-free. This could also have been shown directly. Indeed, let $\mathrm{A}=(a,b)$, $\mathrm{B}=(a',b')$ and $\mathrm{C}=\left(\widetilde{a},\widetilde{b}\right)$. Suppose we know that $\mathrm{A}\otimes\mathrm{C}\succsim\mathrm{B}\otimes\mathrm{C}$, namely $a+\widetilde{a}\geq a'+\widetilde{a}$ and $b+\widetilde{b}\geq b'+\widetilde{b}$. Here, $\widetilde{a}$ and $\widetilde{b}$ can be cancelled, yielding $a\geq a'$ and $b\geq b'$, namely $\mathrm{A}\succsim\mathrm{B}$.

While in quantitative resource theories only free resources are such that $\mathrm{A}\sim\mathrm{A}\otimes\mathrm{A}$, at the opposite extreme there are theories in which this holds for every resources, even for costly ones. These are qualitative resource theories.

**Definition 2.2.17.** A resource theory is called *qualitative* if for every resource A one has $\mathrm{A}\sim\mathrm{A}\otimes\mathrm{A}$.

Loosely speaking, the number of A resources does not matter, because all resources have an "intrinsic value". Let us see an example.

**Example 2.2.18.** The resource theory of knowledge (example 2.2.7) is qualitative. Indeed, consider $\mathrm{A}=(a,b)$. We have

$$\mathrm{A}\otimes\mathrm{A}=(\max\{a,a\},\max\{b,b\})=(a,b)\sim\mathrm{A}.$$

Qualitative and quantitative resource theories are almost opposite kinds of resource theories: if a theory is both qualitative and quantitative, then it is *trivial*, in the sense that all resources are free. Indeed, if a theory is qualitative, then for every resource $A \sim A \otimes A$. Since the theory is quantitative as well, by proposition 2.2.14, we conclude that A is free.

Now let us see a necessary and sufficient criterion to see when a given resource theory is qualitative.

**Proposition 2.2.19.** *A resource theory is qualitative if and only if the following are equivalent for all resources* A *and* B*:*

1. $A \otimes A \succsim B$
2. $A \succsim B$
3. $A \succsim B \otimes B$.

*Proof.* Necessity. Suppose we have a qualitative resource theory. Let us prove the various implications.
1 implies 2. Since $A \sim A \otimes A$, we have $A \succsim A \otimes A$, hence also $A \succsim B$.
2 implies 3. Since $B \sim B \otimes B$, then $B \succsim B \otimes B$, whence $A \succsim B \otimes B$.
3 implies 1. Now $A \sim A \otimes A$ and $B \sim B \otimes B$, therefore $A \otimes A \succsim A$ and $B \otimes B \succsim B$. Then we conclude that $A \otimes A \succsim B$.

Sufficiency. Let us take $B = A$, then we have $A \otimes A \succsim A$ and $A \succsim A \otimes A$, whence $A \sim A \otimes A$ for every resource A and the theory is qualitative. □

Finally we study the behaviour of resources under disposal, namely under operations that destroy them. In general, it is not always possible to find free processes that make it possible to get rid of a resource. As an example, think of nuclear waste, whose disposal is costly in terms of time and other resources used to store them. However, we can sometimes freely dispose of some resources.

**Definition 2.2.20.** A resource A is *freely disposable* if $A \succsim I$, where I is the void resource.

This means that a resource is freely disposable if there is a free process that destroys it. In this way it is more valuable than "nothing". Note that free resources are always freely disposable.

**Definition 2.2.21.** If every resource of a theory is freely disposable, the theory is called *waste-free*.

**Example 2.2.22.** The resource theories of food and of knowledge (examples 2.2.6 and 2.2.7) are waste-free. Indeed for every resource $\mathrm{A} = (a, b)$ we have $a \geq 0$ and $b \geq 0$, therefore $\mathrm{A} \gtrsim \mathrm{I}$.

In waste-free resource theories, for every resource A, we have $\mathrm{A} \gtrsim \mathrm{I}$, therefore somehow the equivalence class of void resource (i.e. all free resources) is the minimum of the induced order $\succeq$. In such a situation, $\mathrm{A} \otimes \mathrm{B} \gtrsim \mathrm{A}$, because $\mathrm{A} \gtrsim \mathrm{A}$, $\mathrm{B} \gtrsim \mathrm{I}$ and the preorder is compatible with the operation of composing resources. Therefore, having two resources is always more valuable than having only one.

### 2.2.2 Resource monotones

Having established a hierarchy among resources, it is sometimes useful to have a direct way of quantifying the value of a resource by assigning it a real number. We want to translate the (pre)ordering of resources into an ordering of real numbers. To this end, we need real-valued functions that respect the (pre)ordering of resources.

**Definition 2.2.23.** A real-valued function $M : |\mathbf{D}| \longrightarrow \mathbb{R}$, where $|\mathbf{D}|$ is the class of resources, is said to be a *resource monotone* (*monotone* for short) if $\mathrm{A} \gtrsim \mathrm{B}$ implies $M(\mathrm{A}) \geq M(\mathrm{B})$.

Loosely speaking, monotones assign a price to resources, consistent with their value. Specifically, if $\mathrm{A} \sim \mathrm{B}$, then $M(\mathrm{A}) = M(\mathrm{B})$. Indeed, if $\mathrm{A} \sim \mathrm{B}$, then $\mathrm{A} \gtrsim \mathrm{B}$ and $\mathrm{B} \gtrsim \mathrm{A}$, thus $M(\mathrm{A}) \geq M(\mathrm{B})$ and $M(\mathrm{B}) \geq M(\mathrm{A})$, whence one has $M(\mathrm{A}) = M(\mathrm{B})$. Nonetheless, if $M(\mathrm{A}) = M(\mathrm{B})$ we *cannot* conclude that $\mathrm{A} \sim \mathrm{B}$. Indeed, a trivial monotone is a constant function which assigns the same "price" to all the resources, irrespective of their actual value. In this case, both equivalent and inequivalent resources have the same "price".

However, a careful examination of this definition shows that resource monotones have some tricky subtleties. Indeed, it is not possible to translate the hierarchy of resources faithfully into the ordering of real numbers. The main reason for such a difficulty is the fact that we can only establish a *partial preorder* of resources, whereas we have a *total order* of real numbers. Indeed, we can have two incomparable resources A and B, but if $M$ is a monotone, we have either $M(\mathrm{A}) \geq M(\mathrm{B})$ or $M(\mathrm{B}) \geq M(\mathrm{A})$, because two real numbers can always be compared. Hence, it is not possible to completely characterize the hierarchy among resources by means of an order of real numbers.

According to the definition of resource monotones, $\mathrm{A} \succsim \mathrm{B}$ implies $M(\mathrm{A}) \geq M(\mathrm{B})$, but the converse implication in general does not hold! This means that the preorder $\succsim$ is more fundamental than the order induced by monotones. Indeed, if $M(\mathrm{A}) \geq M(\mathrm{B})$ we *cannot* conclude that $\mathrm{A} \succsim \mathrm{B}$. However, resource monotones are useful to detect non-convertibility of resources. Recalling the definition, we have that $M(\mathrm{A}) < M(\mathrm{B})$ means $\mathrm{A} \not\succsim \mathrm{B}$.

However, we can obtain an equivalence between the preordering of resources and the ordering induced by monotones by taking more than one resource monotone. In this respect, a family of monotones $\{M_i\}_{i \in X}$ is said to be *complete* if we have $\mathrm{A} \succsim \mathrm{B}$ if and only if $M_i(\mathrm{A}) \geq M_i(\mathrm{B})$ for every $i \in X$.

**Proposition 2.2.24.** *Every resource theory admits a complete family of resource monotones.*

*Proof.* Take $X$ to be the class of resources $|\mathbf{D}|$; then the index $i \in X$ is nothing but a resource. Then, for every resource A, define $M_i(\mathrm{A})$ as

$$M_i(\mathrm{A}) = \begin{cases} 1 & \text{if } \mathrm{A} \succsim i \\ 0 & \text{if } \mathrm{A} \not\succsim i \end{cases}.$$

Let us show that $M_i$ is a monotone for every $i \in X$. Suppose $\mathrm{A} \succsim \mathrm{B}$.

- If $\mathrm{B} \succsim i$, then by transitivity $\mathrm{A} \succsim i$. In this case we have $M(\mathrm{A}) = M(\mathrm{B}) = 1$, whence $M(\mathrm{A}) \geq M(\mathrm{B})$.
- If $\mathrm{A} \not\succsim i$ and $\mathrm{B} \not\succsim i$, then $M(\mathrm{A}) = M(\mathrm{B}) = 0$, whence $M(\mathrm{A}) \geq M(\mathrm{B})$.
- If $\mathrm{A} \succsim i$ and $\mathrm{B} \not\succsim i$, $M(\mathrm{A}) = 1$ and $M(\mathrm{B}) = 0$, and again $M(\mathrm{A}) \geq M(\mathrm{B})$.

This shows that $\{M_i\}$ is a family of monotones. Let us show that this family is also complete. To do that, we must prove that if $M_i(\mathrm{A}) \geq M_i(\mathrm{B})$ for every $i \in X$, then $\mathrm{A} \succsim \mathrm{B}$. Suppose, by contradiction that $\mathrm{A} \not\succsim \mathrm{B}$. Consider $i = \mathrm{B}$. Then we have $M_{\mathrm{B}}(\mathrm{A}) = 0$ because $\mathrm{A} \not\succsim \mathrm{B}$, but $M_{\mathrm{B}}(\mathrm{B}) = 1$ because $\mathrm{B} \succsim \mathrm{B}$, therefore $M_{\mathrm{B}}(\mathrm{A}) < M_{\mathrm{B}}(\mathrm{B})$, in contradiction with the hypothesis that $M_i(\mathrm{A}) \geq M_i(\mathrm{B})$ for every $i \in X$. □

Although we managed to construct a complete family of resource monotones for every resource theory, such a family is not so practical, for it is

indexed by the resources themselves. Clearly, the most desirable situation is when we have a complete family of monotones indexed by a few real parameters, but this is not always possible.

It is useful to classify resource monotones into two categories according to their behaviour under composition of resources.

**Additive** They are such that $M(\mathrm{A} \otimes \mathrm{B}) = M(\mathrm{A}) + M(\mathrm{B})$, for all resources A and B.

**Supremal** They are such that $M(\mathrm{A} \otimes \mathrm{B}) = \max\{M(\mathrm{A}), M(\mathrm{B})\}$, for all resources A and B.

For additive monotones, one has $M(\mathrm{I}) = 0$, where I is the void resource. Indeed $M(\mathrm{A}) = M(\mathrm{A} \otimes \mathrm{I}) = M(\mathrm{A}) + M(\mathrm{I})$, whence $M(\mathrm{I}) = 0$.

For supremal monotones, one can always take $M(\mathrm{I}) = 0$. If this is not the case, it is enough to consider the monotone $M' = M - M(\mathrm{I})$, which is still supremal. Assuming $M(\mathrm{I}) = 0$, means that $M(\mathrm{A}) \geq 0$ for every resource A. Indeed,

$$M(\mathrm{A}) = M(\mathrm{A} \otimes \mathrm{I}) = \max\{M(\mathrm{A}), M(\mathrm{I})\} = \max\{M(\mathrm{A}), 0\},$$

and one must have $M(\mathrm{A}) \geq 0$ in order to have equality.

**Example 2.2.25.** Let us see an example of an additive monotone for the resource theory of food (cf. example 2.2.6). A monotone assigns a price to every basket of fruit, and a reasonable way to do that is to fix the price of a single apple $M(1,0)$, and of a single banana $M(0,1)$. We will have an additive monotone. Since $(1,0) \succsim (0,0)$ and $(0,1) \succsim (0,0)$, the prices of a single apple and a single banana are non-negative (recall $(0,0)$ is the void resource). Note that[4] $(a,b) = (1,0)^{\otimes a} \otimes (0,1)^{\otimes b}$, whence $M(a,b) = aM(1,0) + bM(0,1)$. This is most reasonable: the price of a basket of fruit is the price of an apple ($M(1,0)$) times the numbers of apples ($a$), plus the price of a banana $M(0,1)$ times the number of bananas ($b$).

In resource theories, activation processes sometimes take place by increasing the number of copies of resources. In other words sometimes[5] $\mathrm{A} \not\succsim \mathrm{B}$ but

[4] We will use the short-hand notation $\mathrm{A}^{\otimes n} = \underbrace{\mathrm{A} \otimes \ldots \otimes \mathrm{A}}_{n \text{ times}}$.

[5] Note that, instead, whenever $\mathrm{A} \succsim \mathrm{B}$, one has $\mathrm{A}^{\otimes n} \succsim \mathrm{B}^{\otimes n}$ for all positive integers $n$, because the preorder is compatible with the composition of resources.

$\mathrm{A}^{\otimes n} \succsim \mathrm{B}^{\otimes n}$, for some $n > 1$. In this case, a larger number of copies of the resources activates a process of resource conversion. In mathematical terms, activation processes are possible because the composition of resources has the algebraic structure of a monoid and not of a group: multiple copies, like in $\mathrm{A}^{\otimes n} \succsim \mathrm{B}^{\otimes n}$, cannot be cancelled out.

Clearly, activation cannot occur in qualitative resource theories (see definition 2.2.17), where $\mathrm{A} \sim \mathrm{A}^{\otimes n}$ and therefore $\mathrm{A}^{\otimes n} \succsim \mathrm{B}^{\otimes n}$ is equivalent to $\mathrm{A} \succsim \mathrm{B}$. Additive monotones play a special role in activation processes, as we will show soon.

Suppose we are tasked to turn many copies, say $n$, of the resource A into many copies, say $m$, of the resource B. A quantity of interest is the number of copies of B one can produce on average starting from a single copy of A.

**Definition 2.2.26.** The *rate* of the conversion $\mathrm{A} \longrightarrow \mathrm{B}$ is defined as

$$R(\mathrm{A} \longrightarrow \mathrm{B}) := \sup_{n,m \in \mathbb{Z}_+} \left\{ \frac{m}{n} : \mathrm{A}^{\otimes n} \succsim \mathrm{B}^{\otimes m} \right\},$$

where $n$ and $m$ are positive integers.

If there exist no $n$ and $m$ such that $\mathrm{A}^{\otimes n} \succsim \mathrm{B}^{\otimes m}$, we say that the conversion rate is 0. In all the other cases the conversion rate is strictly positive (possibly $+\infty$).

**Example 2.2.27.** As a simple example, consider the conversion $\mathrm{A} \longrightarrow \mathrm{A}$. Clearly $\mathrm{A}^{\otimes n} \succsim \mathrm{A}^{\otimes n}$, therefore the rate is at least 1, $R(\mathrm{A} \longrightarrow \mathrm{A}) \geq 1$. Specifically, if A is a free resource, $\mathrm{A} \sim \mathrm{A}^{\otimes n}$ (see proposition 2.2.5) for any positive integer $n$ and therefore the rate is infinite $R(\mathrm{A} \longrightarrow \mathrm{A}) = +\infty$. This captures the idea that free resources can be produced in an extremely large number of copies at no cost.

*Remark* 2.2.28. In definition 2.2.26 we have the supremum because we regard B as a more precious resource than A (indeed the interesting case is when $\mathrm{A} \not\succsim \mathrm{B}$), therefore we would like to have the maximum achievable number of copies of B. If, instead B is not a desirable resource, we should take the infimum.

Now we show the relationship between additive resource monotones and conversion rates.

**Proposition 2.2.29.** *Let $M$ be an additive resource monotone. Then*

$$R\left(\mathrm{A}\longrightarrow\mathrm{B}\right)M\left(\mathrm{B}\right)\leq M\left(\mathrm{A}\right),$$

*for all resources* A *and* B.

*Proof.* Consider two positive integers $n$ and $m$ such that $\mathrm{A}^{\otimes n}\succsim\mathrm{B}^{\otimes m}$. Since $M$ is additive, one has $nM\left(\mathrm{A}\right)\geq mM\left(\mathrm{B}\right)$, namely $\frac{m}{n}M\left(\mathrm{B}\right)\leq M\left(\mathrm{A}\right)$. $M\left(\mathrm{A}\right)$ is an upper bound for the set $\left\{\frac{m}{n}M\left(\mathrm{B}\right)\right\}$, therefore the supremum of such a set is less than or equal to $M\left(\mathrm{A}\right)$, or, in other words,

$$R\left(\mathrm{A}\longrightarrow\mathrm{B}\right)M\left(\mathrm{B}\right)\leq M\left(\mathrm{A}\right).$$

□

If $M\left(\mathrm{B}\right)>0$, interpreting $M\left(\mathrm{A}\right)$ and $M\left(\mathrm{B}\right)$ as prices or currencies, the rate of conversion from A to B is upper-bounded by the ratio of the prices of the two resources.

$$R\left(\mathrm{A}\longrightarrow\mathrm{B}\right)\leq\frac{M\left(\mathrm{A}\right)}{M\left(\mathrm{B}\right)}$$

If B is very precious, then the rate will be rather small.

## 2.3 The general structure of resource theories in GPTs

In the previous sections we analysed the topic of resource theories from a fairly abstract angle, focusing mainly on their mathematical structure. Before moving to discuss some examples in quantum mechanics, and extending them to GPTs, in this section we enter into more details about how the general notions introduced in the previous sections fit into the framework of GPTs (see also ref. [60] for a similar analysis in the framework of quantum mechanics). Specifically, we want to understand how the mathematical structure of a resource theory is put into place from an operational point of view, namely form the perspective of an observer performing some experiments in a laboratory. We can think of a GPT as the underlying physical theory supporting a resource theory, which is developed on the formalism of that GPT.

In other words, what are really resources in a GPT? In most cases, resources will be the states allowed by the theory. For the sake of simplicity, we

will restrict ourselves only to this case, recommending the interested reader to refer to [33] for more general situations in which even transformations are resources.

If states are resources, we can divide them into free and costly ones, as explained in the previous sections. Now the processes which convert resources are operations transforming states into states, therefore in the setting of GPTs, they are transformations (see subsection 1.1.4). However, in general, not all transformations will be free. Therefore one needs to distinguish also between free and costly transformations, where free and costly states emerge as special instances of transformations. If there are only free transformations, we say that the resource theory is *trivial*.

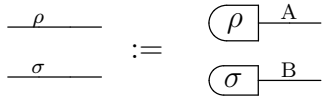
As states are particular kinds of transformations, the notion of free processes will naturally induce the notion of free states.

Since free transformations are closed under sequential and parallel composition, and the parallel composition of systems is a system, a resource theory in a GPT gives rise to a strict symmetric monoidal subcategory of the theory, whose objects are all the systems allowed by the theory, and whose morphisms are free transformations.

One of the most common situations is when one defines a certain set of processes to be free, and then one has to take the closure of this set under sequential and parallel composition in order to have a subcategory. We will encounter this situation in the next chapter, when dealing with the resource theories of purity and athermality.

In section 2.1 we saw that a resource theory is a strict symmetric monoidal category. How do we connect this category to the strict symmetric monoidal category describing a GPT? First of all, one must take the states of the GPT as the objects of the category describing the resource theory, because here the states are the resources. Hence, the parallel composition of resources becomes the parallel composition of states.

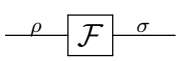
In the following diagrams we will relate the graphical language of the category describing the resource theory to the graphical language of GPTs, introduced in chapter 1. Let us start with the diagram concerning parallel composition of resources.




Here, on the left-hand side, we have the graphical representation of resources as objects of the category describing the resource theory supported by the

GPT. Resources are objects and therefore they are represented as wires. On the right-hand side we represent states (of systems) according to the rules of the graphical language for GPTs.

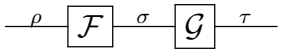
The morphisms in the category of the resource theory are *only* the *free* transformations of the supporting GPT. Therefore we write



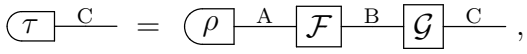

if and only if in the GPT we have (recall $\rho$ and $\sigma$ are states)


, (2.3.1)

for some systems A and B, where $\mathcal{F}$ is a free transformation. The sequential composition of free processes in the resource theory is trivially the sequential composition of free transformations in the GPT. Indeed we write

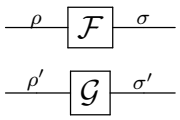



if and only if in the GPT we have

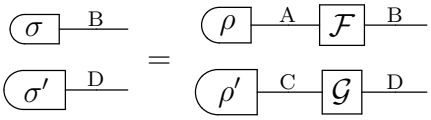



,

for some systems A, B, C, where $\mathcal{F}$ and $\mathcal{G}$ are free transformations and $\sigma$ is defined as in eq. (2.3.1). Similarly one has parallel composition of free processes, defined as the parallel composition of free transformations in the GPT.




if and only if in the GPT




,

for some systems A, B, C, D, where $\mathcal{F}$ and $\mathcal{G}$ are free transformations. We also assume that the swapping of systems (cf. subsection 1.1.3) is a free transformation. In this way we see that we can build a resource theory in the sense of section 2.1 starting from a GPT in which we define a subcategory of free transformations, and where all states are resources.

The next chapters will be devoted to examining some examples first in the familiar context of quantum theory, and then to the more abstract setting of GPTs.

# Chapter 3

# Examples of resource theories in quantum mechanics

Quantum mechanics is the most significant example of a theory where the observer plays a fundamental role in the physical description of reality. As such, the observer's presence has a direct and deep impact on physical systems, and, as a result of such a tight relationship between the observer and the observed, quantum entities emerge naturally as resources available to the observer. Furthermore, after the development of quantum information theory, it has become clear that quantum theory often offers us some tools capable of outperforming some classical protocols in various fields, from communication and cryptography [18, 19, 20, 21, 22, 23], to computation [61, 62, 63, 64]. Therefore, to have improvements on classical protocols, we must search for entirely non-classical features present in quantum mechanics, which will be the most valuable resources in the language of resource theories. The most famous example of a non-classical feature in quantum mechanics is definitely quantum entanglement, and it is also the source of the improved performances of most protocols in quantum information. Therefore, the most natural resource theory to study in the framework of quantum mechanics is the theory of quantum entanglement. Quite surprisingly, this theory is intimately related to another resource theory, the resource theory of purity, developed for thermodynamic reasons. This shows that quantum entanglement can be considered as one of the routes towards a formulation of a theory of quantum thermodynamics, as shown in some works [65, 66, 67, 68, 69, 70, 71].

In this chapter we focus on the resource theory of entanglement and on the related resource theory of purity, and in the next chapter we will try to extend

them to the case of GPTs. In section 3.1 we present the resource theory of entanglement, which is defined using a particular type of free operations, known as $LOCC$[1] *protocols*. Using these protocols, we can order bipartite states according to their degree of entanglement. When the states are pure, this ordering is equivalent to the ordering of their marginals according their degree of mixedness [72, 73, 74, 75, 50]. In section 3.2 we introduce the resource theory of purity, which is defined via a particular class of quantum operations, called *noisy operations*, in which a system is put into contact with an ancillary system in the maximally mixed state. Finally, in section 3.3 we briefly present two other examples of quantum resource theories.

In quantum mechanics, systems are represented by their associated Hilbert space $\mathcal{H}$ and transformations are quantum operations (cf. example 1.1.31), i.e. completely positive trace-non-increasing maps $\mathcal{M} : \mathcal{L}(\mathcal{H}) \longrightarrow \mathcal{L}(\mathcal{H}')$ [50], where $\mathcal{L}(\mathcal{H})$ is the set of bounded linear operators on $\mathcal{H}$ (all the linear operators acting on $\mathcal{H}$ if $\mathcal{H}$ is finite-dimensional). The rule of parallel composition of system is simply the tensor product of the associated Hilbert spaces.

## 3.1 The resource theory of quantum entanglement

Quantum entanglement, besides being a source of foundational puzzles in quantum theory, it is also a very powerful resource for quantum information and communication [18, 19, 20, 21, 22, 23]. It is then natural to develop a resource theory of it. For the sake of simplicity, we will restrict ourselves to the bipartite case, namely to the case when we have a composite system made only of two parts, $\mathcal{H}_{\mathrm{AB}} = \mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{B}}$.

Before entering into details, let us briefly recall the definition of entangled states.

**Definition 3.1.1.** A (bipartite) state $\rho_{\mathrm{AB}}$ is called *separable* if it can be written as $\rho_{\mathrm{AB}} = \sum_j p_j \rho_{j,\mathrm{A}} \otimes \sigma_{j,\mathrm{B}}$, where the $p_j$'s are probabilities and $\rho_{j,\mathrm{A}}$ is a state of system A, and $\sigma_{j,\mathrm{B}}$ is a state of system B, for every $j$.

A state which is not separable is called *entangled*.

Entangled states will be the costly resources, while separable states will be the free resources [76]. To fit everything in the paradigm of resource

[1] Local Operations and Classical Communication

theories developed in the previous chapter, we have to define a class of free transformations (i.e. quantum operations) such that separable states can be thought as particular instances thereof (recall that states in GPTs are a particular kind of transformations). The answer comes from LOCC protocols [77], which are a fully operational notion, without any references to the Hilbert space structure of quantum mechanics. As such, they can be exported easily to GPTs (see chapter 4).

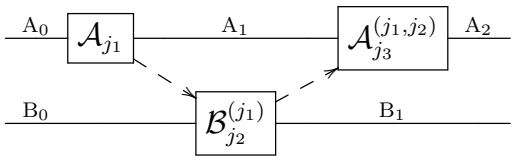
In the LOCC scenario, we have two (distant) parties, Alice and Bob, who perform a sequence of quantum instruments[2] [50], performed by Alice and Bob, with the property that the choice of the quantum instrument at a given step may depend on the outcomes obtained at the previous steps. Alice and Bob are allowed to exchange classical communication to each other, so that even Bob's choice of quantum instrument may depend on Alice's previous choices, and vice versa. LOCC protocols are classified according to how many rounds of classical communications are involved: a $n$-way LOCC protocols is an LOCC protocols where there are $n$ rounds of classical communication.

Now, let us be more concrete and consider a 2-way protocol where

1. Alice performs a quantum instrument $\{\mathcal{A}_{j_1}\}$ and communicates her outcome to Bob;
2. Bob performs a quantum instrument $\left\{\mathcal{B}_{j_2}^{(j_1)}\right\}$ and communicates his outcome to Alice;
3. Alice performs a quantum instrument $\left\{\mathcal{A}_{j_3}^{(j_1,j_2)}\right\}$.

As explained in section 1.3, superscripts in round brackets want to highlight the dependence of a test on the outcome of a previous one. Here the pictorial representation will prove to be rather useful to understand the setting.

An instance of the protocol is identified by the sequence of outcomes $(j_1, j_2, j_3)$ and can be represented by a diagram of the form




, (3.1.1)

[2]Recall that a quantum instrument $\{\mathcal{M}_a\}$ is a collection of quantum operations $\mathcal{M}_a$ such that $\sum_a \mathcal{M}_a$ is trace-preserving.

where the dashed arrows represent classical communication. By coarse-graining over all possible outcomes, one obtains a quantum channel, called *LOCC channel*, given by

$$\mathcal{C} = \sum_{j_1,j_2,j_3} \left[ \mathcal{A}_{j_3}^{(j_1,j_2)} \mathcal{A}_{j_1} \otimes \mathcal{B}_{j_2}^{(j_1)} \right].$$

We declare LOCC protocols to be our free processes.

Let us see if separable states fit into this framework as free resources and special cases of LOCC (free) operations. Consider the generic separable state $\rho_{\mathrm{AB}} = \sum_j p_j \rho_{j,\mathrm{A}} \otimes \sigma_{j,\mathrm{B}}$. It can be prepared via an LOCC protocol according to the following procedure. Suppose Alice has an initial ensemble of states $\{\rho_{j,\mathrm{A}}\}$, where each $\rho_{j,\mathrm{A}}$ is prepared with probability $p_j$.

1. Alice prepares the state $\rho_{j,\mathrm{A}}$ with probability $p_j$.
2. Alice calls Bob to inform him of the outcome of her preparation.
3. Correspondingly, Bob prepares the state $\sigma_{j,\mathrm{B}}$ depending on Alice's outcome $j$.
4. By doing a coarse-graining over $j$, we obtain the desired state.

Therefore, separable states are really free states. Are there any other states which are free but entangled? The answer is negative: a state is free if and only if it is separable. Indeed, consider the most general LOCC preparation protocol. Alice has an ensemble $\{(\rho_{j,\mathrm{A}}, p_j)\}$, where $p_j$ is the probability of preparing the state $\rho_{j,\mathrm{A}}$, and Bob has a family of ensembles from which he can choose according to the outcome of Alice's preparation, namely $\left\{ \left( \sigma_{k,\mathrm{B}}^{(j)}, q_k^{(j)} \right) \right\}$, where $q_k^{(j)}$ is the probability of preparing the state $\sigma_{k,\mathrm{B}}^{(j)}$.

1. Alice prepares the state $\rho_{j,\mathrm{A}}$ with probability $p_j$.
2. She communicates Bob her outcome $j$.
3. Bob prepares the state $\sigma_{k,\mathrm{B}}^{(j)}$ with probability $q_k^{(j)}$.
4. After performing a coarse-graining, we get the state

$$\rho_{\mathrm{AB}} = \sum_{j,k} p_j q_k^{(j)} \rho_{j,\mathrm{A}} \otimes \sigma_{k,\mathrm{B}}^{(j)} = \sum_j p_j \rho_{j,\mathrm{A}} \otimes \sum_k q_k^{(j)} \sigma_{k,\mathrm{B}}^{(j)} =$$

$$= \sum_j p_j \rho_{j,\mathrm{A}} \otimes \sigma'_{j,\mathrm{B}},$$

where we have defined $\sigma'_{j,\mathrm{B}} := \sum_k q_k^{(j)} \sigma_{k,\mathrm{B}}^{(j)}$.

This proves that the output of an LOCC preparation process is necessarily a separable state. This also shows that the definition of LOCC protocols as free operations well captures the structure of the resource theory of entanglement.

Now, we know that all entangled states represent costly resources. In other words, entangled states cannot be prepared by LOCC protols.

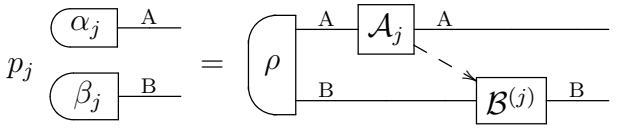
Since LOCC protocols are our free operations, we can use them to establish a hierarchy among bipartite states according to their degree of entanglement.

**Definition 3.1.2.** Given two states $\rho \in \mathcal{L}(\mathcal{H}_{\mathrm{AB}})$ and $\rho' \in \mathcal{L}(\mathcal{H}_{\mathrm{AB}})$, we say that $\rho$ is *more entangled* than $\rho'$, denoted by $\rho \succeq_{\mathrm{ent}} \rho'$, if there exists an LOCC protocol that transforms $\rho$ into $\rho'$, i.e. if $\rho' = \mathcal{C}(\rho)$ for some LOCC channel $\mathcal{C}$.

As one may anticipate, entangled states are more valuable than separable states.

**Proposition 3.1.3.** *Every bipartite state is more entangled than any separable state.*

*Proof.* Consider the separable state $\sigma_{\mathrm{AB}} = \sum_j p_j \alpha_{j,\mathrm{A}} \otimes \beta_{j,\mathrm{B}}$, where the $\alpha_{j,\mathrm{A}}$'s are density operators on the Hilbert space $\mathcal{H}_{\mathrm{A}}$, and the $\beta_j$'s are density operators on the Hilbert space $\mathcal{H}_{\mathrm{B}}$. We can consider the LOCC protocol acting on a generic density operator $\rho$ on the Hilbert space $\mathcal{H}_{\mathrm{AB}}$




where $\mathcal{A}_j$ is a measure-and-prepare quantum operation on A, which prepares $\alpha_{j,\mathrm{A}}$ with probability $p_j$ and $\mathcal{B}^{(j)}$ is a quantum channel on B which prepares $\beta_{j,\mathrm{B}}$, depending on the outcome of the test $\{\mathcal{A}_j\}$. In other words, $\mathcal{A}_j = p_j |\alpha_j\rangle_{\mathrm{A}} \circ \mathrm{tr}_{\mathrm{A}}$ and $\mathcal{B}^{(j)} = |\beta_j\rangle_{\mathrm{B}} \circ \mathrm{tr}_{\mathrm{B}}$. By taking the coarse-graining over the outcome $j$, we get the separable state $\sigma_{\mathrm{AB}} = \sum_j p_j \alpha_{j,\mathrm{A}} \otimes \beta_{j,\mathrm{B}}$. This proves that $\rho_{\mathrm{AB}}$ is more entangled than $\sigma_{\mathrm{AB}}$. □

According to the general framework of resource theories, the relation $\succeq_{\text{ent}}$ is a preorder.

**Definition 3.1.4.** If $\rho \succeq_{\text{ent}} \rho'$ and $\rho' \succeq_{\text{ent}} \rho$, then we say that $\rho$ and $\rho'$ are *equally entangled* (or that $\rho$ is as entangled as $\rho'$), denoted by $\rho \sim_{\text{ent}} \rho'$.

Note that $\rho \sim_{\text{ent}} \rho'$ does not imply that $\rho$ and $\rho'$ are equal: for example, every two separable states are equally (un)entangled, as proposition 4.1.4 shows; indeed it is enough to take $\rho$ to be separable. Specifically, the equivalence class of separable states is the minimum of the entanglement order.

Similarly, two bipartite states that differ by local unitary channels (see example 1.1.31) are equally entangled: $\rho_{\text{AB}} \sim_{\text{ent}} (\mathcal{U}_{\text{A}} \otimes \mathcal{V}_{\text{B}})\, \rho_{\text{AB}}$, indeed the two local unitary channels make up a (reversible) LOCC protocol from $\rho_{\text{AB}}$ to $(\mathcal{U}_{\text{A}} \otimes \mathcal{V}_{\text{B}})\, \rho_{\text{AB}}$.

The entanglement preorder has a special structure in the case of pure bipartite states, which we will explore in the next subsections. Here we note that for the pure case, separable states coincide with (pure) product states, and all the phenomenology is less rich. For instance, while in general the issue of distinguishing entangled from separable states in the mixed case is pretty thorny, in the pure case we have a necessary and sufficient condition for a bipartite state to be entangled.

A similar statement will appear later in the following chapter (proposition 4.4.5). Even though the proposition in the quantum mechanical case could be proved along the same lines as proposition 4.4.5, we prefer to explicitly use the mathematical formalism of quantum theory to better show the difference with the operational approach of chapter 4.

**Proposition 3.1.5.** *Let $|\Psi\rangle_{\text{AB}}$ be a pure bipartite state of system* AB. *The following are equivalent.*

1. $|\Psi\rangle_{\text{AB}}$ is entangled.
2. Its marginal on A is mixed.
3. Its marginal on B is mixed.

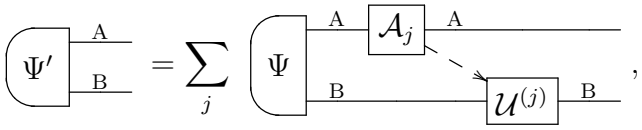
*Proof.* Let $|\Psi\rangle_{\text{AB}} = \sum_{j=1}^{r} \sqrt{p_j}\, |j\rangle_{\text{A}} |j'\rangle_{\text{B}}$ be a Schmidt decomposition of $|\Psi\rangle_{\text{AB}}$, where all the $p_j$'s are non-vanishing for $j = 1, \ldots, r$. Both the marginals of $|\Psi\rangle_{\text{AB}}$ have the $p_j$'s as non-vanishing eigenvalues, therefore 2 and 3 are equivalent. We will prove the equivalence between 1 and 2.

1 implies 2. Suppose, by contradiction that the marginal on A is pure. This means that there is only one non-vanishing $p_j$, say $p_1 = 1$. As a result, we have $|\Psi\rangle_{\mathrm{AB}} = |1\rangle_{\mathrm{A}} |1'\rangle_{\mathrm{B}}$, in contradiction to the hypothesis that $|\Psi\rangle_{\mathrm{AB}}$ is entangled.

2 implies 1. Suppose, by contradiction that $|\Psi\rangle_{\mathrm{AB}}$ is a product state, say $|\Psi\rangle_{\mathrm{AB}} = |\alpha\rangle_{\mathrm{A}} |\beta\rangle_{\mathrm{B}}$, for some $|\alpha\rangle_{\mathrm{A}}$ and $|\beta\rangle_{\mathrm{B}}$. If we take the partial trace over B, we get $\mathrm{tr}_{\mathrm{B}} |\Psi\rangle \langle\Psi|_{\mathrm{AB}} = |\alpha\rangle \langle\alpha|_{\mathrm{A}}$, which contradicts the hypothesis that the marginal on A is mixed. □

This is a first hint of a close relationship between pure-state entanglement and mixed single-system states. This relationship is deeper than this, and will be explored further in subsection 3.1.2.

The entanglement preorder in the pure bipartite case is special because every LOCC protocol between pure states can always be reduced to a 1-way LOCC protocol with a unitary channel performed by the party who receives classical communication. In other words, if $|\Psi\rangle \succeq_{\mathrm{ent}} |\Psi'\rangle$, then there exists a 1-way LOCC protocol




where $\mathcal{A}_j$ is a quantum operation and $\mathcal{U}^{(j)}$ is a unitary channel depending on the outcome $j$. This is the content of Lo-Popescu theorem [78].

**Theorem 3.1.6** (Lo-Popescu)**.** *If $|\Psi\rangle_{\mathrm{AB}}$ can be transformed into $|\Psi'\rangle_{\mathrm{AB}}$ by an LOCC protocol, then it can be transformed into $|\Psi'\rangle_{\mathrm{AB}}$ by a 1-way LOCC protocol, where Alice applies a quantum instrument, she communicates her outcome to Bob, and Bob applies a unitary channel on his system.*

*Proof.* The core of this proof is to show that every quantum operation made by Bob can be "simulated" by one made by Alice, followed by a unitary correction channel on Bob's system. Without loss of generality, we can assume that all quantum operations are pure, i.e. there is only one Kraus operator for each quantum operation, because summations in the non-pure case can always be taken out of the protocol. Let us start from the bipartite pure state $|\Psi\rangle_{\mathrm{AB}}$ and let us consider its Schmidt decomposition[3] $|\Psi\rangle_{\mathrm{AB}} = \sum_j \sqrt{\lambda_j} |j\rangle_{\mathrm{A}} |j\rangle_{\mathrm{B}}$.

[3]Here, for the sake of simplicity, we omit the prime for kets of system B in Schmidt decompositions.

Suppose Bob applies a quantum instrument with Kraus operators $\{M_j\}$, which can be expressed in his Schmidt basis[4] as

$$M_j = \sum_{k,l} M_{j,kl} \left|k\right\rangle_{\mathrm{B}} \left\langle l\right|_{\mathrm{B}} .$$

Suppose Bob gets outcome $j$. The state after his measurement is

$$\left|\Psi_j\right\rangle_{\mathrm{AB}} = \frac{1}{p_j} \left(\mathbf{1}_{\mathrm{A}} \otimes M_j\right) \left|\Psi\right\rangle_{\mathrm{AB}} = \frac{1}{p_j} \sum_{k,l} \sqrt{\lambda_l} M_{j,kl} \left|l\right\rangle_{\mathrm{A}} \left|k\right\rangle_{\mathrm{B}} , \tag{3.1.2}$$

where $p_j$ is the probability of outcome $j$, and it is given by

$$p_j = \left\langle\Psi\right|_{\mathrm{AB}} M_j^\dagger M_j \left|\Psi\right\rangle_{\mathrm{AB}} = \sum_{k,l} \lambda_l \left|M_{j,kl}\right|^2 . \tag{3.1.3}$$

Let us construct a quantum instrument on Alice's system with Kraus operators $\{N_j\}$, defined with respect to Alice's Schmidt basis as

$$N_j = \sum_{k,l} M_{j,kl} \left|k\right\rangle_{\mathrm{A}} \left\langle l\right|_{\mathrm{A}} .$$

In this way, they are perfectly equivalent to Bob's ones, the difference is that now the vectors of Bob's Schmidt basis have become the corresponding vectors of Alice's Schmidt basis. If Alice gets outcome $j$, then the state after her measurement is

$$\left|\Phi_j\right\rangle_{\mathrm{AB}} = \frac{1}{p_j} \left(N_j \otimes \mathbf{1}_{\mathrm{B}}\right) \left|\Psi\right\rangle_{\mathrm{AB}} = \frac{1}{p_j} \sum_{k,l} \sqrt{\lambda_l} M_{j,kl} \left|k\right\rangle_{\mathrm{A}} \left|l\right\rangle_{\mathrm{B}} , \tag{3.1.4}$$

where $p_j$ is still given by eq. (3.1.3). Comparing eq. (3.1.2) with eq. (3.1.4), we see that $\left|\Psi_j\right\rangle_{\mathrm{AB}}$ and $\left|\Phi_j\right\rangle_{\mathrm{AB}}$ are the same state, up to exchanging the role of system A and system B. Therefore they have the same Schmidt coefficients, and they differ by a tensor product of unitary operators, namely $\left|\Psi_j\right\rangle_{\mathrm{AB}} = U_{j,\mathrm{A}} \otimes V_{j,\mathrm{B}} \left|\Phi_j\right\rangle_{\mathrm{AB}}$.
Therefore, when Bob applies a quantum instrument with Kraus operators $\{M_j\}$, this is equivalent to the situation when Alice applies a quantum instrument with Kraus operators $\{U_j N_j\}$ on her system, and then Bob applies the appropriate unitary operator $V_j$ on his system.

[4]Here, we are enlarging the set of Schmidt vectors of system B suitably if it is not an orthonormal basis for $\mathcal{H}_{\mathrm{B}}$ already.

If the original LOCC protocol is multi-way, whenever Bob performs a measurement and communicates his result to Alice, we simulate his measurement by a measurement performed by Alice. In this situation, Alice communicates her outcome to Bob, and he applies a unitary transformation. Taking the sequential composition of all Alice's measurements and of all Bob's unitary channels, we see that this protocol is equivalent to one where there is only one measurement by Alice, followed by classical communication from Alice to Bob and a unitary channel applied by Bob. □

### 3.1.1 Mixedness relation

Quite surprisingly, it turns out that entanglement preorder is intimately related to a relation which orders single-system states according to their degree of mixedness. The way to accomplish such an ordering takes inspiration from classical probability distributions. In classical probability, a state is a probability distribution $\mathbf{p} = \begin{pmatrix} p_1 & \dots & p_n \end{pmatrix}^t$. A state is more mixed than another if in the former we have more uncertainty than in the latter. Intuitively, the uniform distribution is the one with maximum uncertainty, and the closer a distribution is to the uniform distribution, the more mixed the state it represents. The closeness to the uniform is described by how flat a given distribution is, and this concept is captured by the notion of majorization [79, 80]. Given a vector $x \in \mathbb{R}^n$, we define $x_\downarrow$ as the decreasing rearrangement of $x$. We denote the $i$-th component of $x_\downarrow$ as $x_{[i]}$.

**Definition 3.1.7** (Majorization)**.** Let $x, y \in \mathbb{R}^n$. We say that $x$ is *majorized* by $y$ (or that $y$ *majorizes* $x$), and we write $x \preceq y$, if

- $\sum_{i=1}^k x_{[i]} \leq \sum_{i=1}^k y_{[i]}$ for every $k = 1, \dots, n-1$
- $\sum_{i=1}^n x_{[i]} = \sum_{i=1}^n y_{[i]}$

When applied to vectors of probabilities (i.e. probability distributions), the second condition is always met, for $\sum_{i=1}^n x_{[i]} = \sum_{i=1}^n y_{[i]} = 1$.

Equivalently, the conditions for majorization can be given using the increasing arrangement $x^\uparrow$ of a vector. We denote the $i$-th component of $x^\uparrow$ as $x_{(i)}$. In this setting, as it easy to check, $x \preceq y$ if and only if

- $\sum_{i=1}^k x_{(i)} \geq \sum_{i=1}^k y_{(i)}$ for every $k = 1, \dots, n-1$
- $\sum_{i=1}^n x_{(i)} = \sum_{i=1}^n y_{(i)}$ .

Majorization is a preorder. Indeed, if $x \preceq y$ and $y \preceq x$, one has that the decreasing rearrangements of $x$ and $y$ are equal. This means that, in general, $x$ and $y$ are not equal, but they differ by a permutation of their entries.

**Example 3.1.8.** For vectors of probabilities with $n$ entries one has

$$\begin{pmatrix} \frac{1}{n} \\ \vdots \\ \frac{1}{n} \end{pmatrix} \preceq \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix} \preceq \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

as it easy to check. This means that the uniform distribution (the most mixed probability distribution) is the minimum of the majorization, while any pure state (the least mixed distribution) is the maximum of the majorization.

This example leads naturally to the following definition.

**Definition 3.1.9** (Mixedness relation, classical probability theory)**.** We say that a classical probability distribution $\mathbf{p}$ is *more mixed* than another probability distribution $\mathbf{p}'$ if $\mathbf{p} \preceq \mathbf{p}'$.

We say that $\mathbf{p}$ is *as mixed as* $\mathbf{p}'$ (or that $\mathbf{p}$ and $\mathbf{p}'$ are *equally mixed*) if $\mathbf{p} \preceq \mathbf{p}'$ and $\mathbf{p}' \preceq \mathbf{p}$.

As noted above, two equally mixed probability distributions differ by a permutation.

Majorization in general is not a total preorder, as the following counterexample shows.

**Example 3.1.10.** Let us consider two vectors $x$ and $y$ in $\mathbb{R}^3$.

$$x = \begin{pmatrix} \frac{2}{5} \\ \frac{2}{5} \\ \frac{1}{5} \end{pmatrix} \quad y = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{4} \\ \frac{1}{4} \end{pmatrix}$$

In this case we have neither $x \preceq y$, nor $y \preceq x$. Indeed, $\frac{2}{5} \leq \frac{1}{2}$, so $y \not\preceq x$, but $\frac{2}{5} + \frac{2}{5} \geq \frac{1}{2} + \frac{1}{4}$, so $x \not\preceq y$. This shows that the preorder is not total.

However, the preorder is total on vectors of probabilities in $\mathbb{R}^2$. Indeed, in this case we have to compare

$$x = \begin{pmatrix} p \\ 1-p \end{pmatrix} \quad y = \begin{pmatrix} q \\ 1-q \end{pmatrix},$$

and we can always tell if $p \leq q$ or $q \leq p$.

Before moving to the quantum case, let us explore some properties of majorization. One of the key ingredients, as mentioned above, are permutations.

Suppose we want to permute the entries of a vector $x \in \mathbb{R}^n$ by the permutation $\pi \in S_n$, where $S_n$ is the symmetric group over $n$ elements. If $x = \sum_{i=1}^n x_i e_i$, where $\{e_i\}_{i=1}^n$ is the canonical basis for $\mathbb{R}^n$, and we want to move the $i$-th entry to the $\pi(i)$-th entry, then the resulting vector is $x_\pi = \sum_{i=1}^n x_i e_{\pi(i)}$. Therefore we look for a matrix that transforms $e_i$ into $e_{\pi(i)}$. This matrix simply permutes the basis vectors. We will call it permutation matrix.

We can associate a $n \times n$ matrix $\Pi$ with every permutation $\pi \in S_n$. It is the matrix whose $i$-th column is $e_{\pi(i)}$. We sum up all these remarks in the following definition.

**Definition 3.1.11.** A square matrix $\Pi$ of order $n$ is said to be a *permutation matrix* if $\Pi_{ij} = \delta_{i,\pi(j)}$, for some permutation $\pi \in S_n$.

In this way, a permutation matrix can be obtained simply permuting the columns of the identity matrix.

Permutation matrices give a representation of the permutation group $S_n$. Indeed, if $\Pi$ and $\Sigma$ are the matrices associated with the permutations $\pi$ and $\sigma$, then $\Pi\Sigma$ is the matrix associated with $\pi \circ \sigma$. Indeed

$$(\Pi\Sigma)_{ik} = \sum_j \delta_{i,\pi(j)}\delta_{j,\sigma(k)} = \delta_{i,\pi(\sigma(k))} = \delta_{i,\pi\circ\sigma(k)}.$$

The notion directly related to majorization is that of doubly stochastic matrix, which, in turn, is related to permutation matrices.

**Definition 3.1.12.** A square matrix $P$ of order $n$ is called *doubly stochastic* if each entry is non-negative and the sum of all the entries in each row and in each column is 1. In symbols, $P_{ij} \geq 0$, $\sum_j P_{ij} = 1$ (each row sums to 1) and $\sum_i P_{ij} = 1$ (each column sums to 1).

Doubly stochastic matrices can be expressed in terms of permutation matrices.

**Theorem 3.1.13** (Birkhoff [81])**.** *Doubly stochastic matrices are the convex hull of permutation matrices, where permutation matrices are the extreme points of it.*

*Proof.* See ref. [80]. □

Every doubly stochastic matrix can be written as a convex combination of permutation matrices, therefore doubly stochastic matrices can be thought to implement *random permutation*s. Finally, we have the following important theorem [82, 80].

**Theorem 3.1.14.** *Let $x, y \in \mathbb{R}^n$. Then we have $x \preceq y$ if and only if $x = Py$, where $P$ is a doubly stochastic matrix of order $n$.*

*Proof.* See ref. [80]. □

This means that in classical probability theory the source of mixedness are random permutations.

Now we want to turn the machinery of majorization to the quantum case, so as to establish an ordering of quantum states according to their mixedness. The idea will be to apply majorization to the classical probability distribution associated canonically with a density operator, that is the probability distribution of its eigenvalues [72, 73, 74].

**Definition 3.1.15.** Let $\rho$ and $\rho'$ be two quantum states. We say that $\rho$ is *more mixed* than $\rho'$ if $\mathbf{p} \preceq \mathbf{p}'$, where $\mathbf{p}$ is the vector of the eigenvalues of $\rho$ and $\mathbf{p}'$ is the vector of the eigenvalues of $\rho'$.

Recalling example 3.1.8, we can easily establish a hierarchy among quantum states according to their mixedness. Note that mixedness depend only on the eigenvalues of states and not, for instance, on their eigenvectors.

**Example 3.1.16.** From example 3.1.8 we know that

$$\begin{pmatrix} \frac{1}{n} \\ \vdots \\ \frac{1}{n} \end{pmatrix} \preceq \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix} \preceq \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Let $\{|j\rangle\}_{j=1}^n$ be an orthonormal basis of the Hilbert space $\mathcal{H} \approx \mathbb{C}^n$. A state $\chi$ with $n$ equal eigenvalues,

$$\chi = \frac{1}{n} \sum_{j=1}^{n} |j\rangle \langle j| = \frac{1}{n} \mathbf{1}$$

is more mixed than any state. On the other hand, every state is more mixed than any pure state, whose eigenvalues are $\begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}^t$.

Mixedness relation between density operators inherits the property of majorization. Specifically, it is a preorder, which is total only when the Hilbert space has dimension 2. We can define equally mixed states.

**Definition 3.1.17.** Let $\rho$ and $\rho'$ be density operators, and let $\mathbf{p}$ and $\mathbf{p}'$ be the vectors of their eigenvalues respectively. We say that $\rho$ is *as mixed as* $\rho'$ (or $\rho$ and $\rho'$ are *equally mixed*) if $\mathbf{p} \preceq \mathbf{p}'$ and $\mathbf{p}' \preceq \mathbf{p}$.

The conditions for being equally mixed is that two states $\rho$ and $\rho'$ have the same eigenvalues, because if $\mathbf{p} \preceq \mathbf{p}'$ and $\mathbf{p}' \preceq \mathbf{p}$, $\mathbf{p}$ and $\mathbf{p}'$ differ only by a permutation of their entries. Then it is easy to prove the following proposition.

**Proposition 3.1.18.** *$\rho$ is as mixed as $\rho'$ if and only if there exists a unitary operator $U$ such that $\rho' = U\rho U^\dagger$.*

*Proof.* Sufficiency. Given a state $\rho$, $U\rho U^\dagger$, where $U$ is a unitary operator, has exactly the same eigenvalues of $\rho$. Therefore $\rho$ and $\rho'$ have the same vector of eigenvalues and therefore they are equally mixed.

Necessity. Suppose $\rho$ is as mixed as $\rho'$. Then, $\rho$ and $\rho'$ have the same eigenvalues, and they can differ only by the (orthonormal) basis of their eigenvectors. Let $\{|\psi_j\rangle\}_{j=1}^n$ be the basis of eigenvectors of $\rho$, namely $\rho = \sum_{j=1}^n p_j |\psi_j\rangle \langle\psi_j|$, and let $\{|\varphi_j\rangle\}_{j=1}^n$ be the basis of the eigenvectors of $\rho'$, namely $\rho' = \sum_{j=1}^n p_j |\varphi_j\rangle \langle\varphi_j|$. We know that there is a unitary operator $U$ transforming $\{|\psi_j\rangle\}_{j=1}^n$ into $\{|\varphi_j\rangle\}_{j=1}^n$. Then one has

$$U\rho U^\dagger = \sum_{j=1}^n p_j U |\psi_j\rangle \langle\psi_j| U^\dagger = \sum_{j=1}^n p_j |\varphi_j\rangle \langle\varphi_j| = \rho'.$$

□

Unitary operators play an important part not only for characterizing equally mixed states, but also for giving an alternative and operational definition of the mixedness preorder.

**Theorem 3.1.19.** *$\rho$ is more mixed than $\rho'$ if and only if there exist some unitary operators $U_j$, and some probabilities $\lambda_j$ such that $\rho = \sum_j \lambda_j U_j \rho' U_j^\dagger$.*

*Proof.* Necessity. Let $\mathbf{p} = \begin{pmatrix} p_1 & \dots & p_n \end{pmatrix}^t$ and $\mathbf{p}' = \begin{pmatrix} p'_1 & \dots & p'_n \end{pmatrix}^t$ be the vectors of the eigenvalues of $\rho$ and $\rho'$ respectively. Since $\rho$ is more mixed than $\rho'$, then $\mathbf{p} \preceq \mathbf{p}'$. By theorem 3.1.14 there exists a doubly stochastic matrix $P$ such that $\mathbf{p} = P\mathbf{p}'$. If $\{|\psi_j\rangle\}_{j=1}^n$ is the basis of the eigenvectors of $\rho$, then

$$\rho = \sum_{j=1}^{n} p_j \left|\psi_j\right\rangle \left\langle\psi_j\right| = \sum_{j,k=1}^{n} P_{jk} p'_k \left|\psi_j\right\rangle \left\langle\psi_j\right| .$$

By Birkhoff's theorem (theorem 3.1.13), $P$ can be decomposed as a convex combination of permutation matrices $\Pi_j$, $P = \sum_l \lambda_l \Pi_l$, where the $\lambda_l$'s are probabilities.

$$\rho = \sum_{j,k=1}^{n} \sum_{l} \lambda_l \Pi_{l,jk} p'_k \left|\psi_j\right\rangle \left\langle\psi_j\right| = \sum_{l} \lambda_l \sum_{k=1}^{n} p'_k \sum_{j=1}^{n} \Pi_{l,jk} \left|\psi_j\right\rangle \left\langle\psi_j\right| . \tag{3.1.5}$$

Our aim is to prove that $\rho = \sum_l \lambda_l U_l \rho' U_l^\dagger$, and if $\{|\varphi_k\rangle\}_{k=1}^n$ is the basis of the eigenvectors of $\rho'$, we must prove that

$$\sum_{j=1}^{n} \Pi_{l,jk} \left|\psi_j\right\rangle \left\langle\psi_j\right| = U_l \left|\varphi_k\right\rangle \left\langle\varphi_k\right| U_l^\dagger$$

for some unitary operator $U_l$. Now, the matrix $\Pi_l$ simply permutes the order in which the various terms $|\psi_j\rangle \langle\psi_j|$ appear. Indeed, $\Pi_{l,jk} = \delta_{j,\pi_l(k)}$. Then

$$\sum_{j=1}^{n} \Pi_{l,jk} \left|\psi_j\right\rangle \left\langle\psi_j\right| = \sum_{j=1}^{n} \delta_{j,\pi_l(k)} \left|\psi_j\right\rangle \left\langle\psi_j\right| = \left|\psi_{\pi_l(k)}\right\rangle \left\langle\psi_{\pi_l(k)}\right| .$$

Now eq. (3.1.5) becomes

$$\rho = \sum_{l} \lambda_l \sum_{k=1}^{n} p'_k \left|\psi_{\pi_l(k)}\right\rangle \left\langle\psi_{\pi_l(k)}\right| .$$

We know that for every $l$ there exists a unitary operator $U_l$ transforming the orthonormal basis $\{|\varphi_k\rangle\}_{k=1}^n$ into the orthonormal basis $\left\{\left|\psi_{\pi_l(k)}\right\rangle\right\}_{k=1}^n$. Therefore we finally have

$$\rho = \sum_{l} \lambda_l \sum_{k=1}^{n} p'_k U_l \left|\varphi_k\right\rangle \left\langle\varphi_k\right| U_l^\dagger = \sum_{l} \lambda_l U_l \left( \sum_{k=1}^{n} p'_k \left|\varphi_k\right\rangle \left\langle\varphi_k\right| \right) U_l^\dagger = \sum_{l} \lambda_l U_l \rho' U_l^\dagger ,$$

thus proving necessity.

Sufficiency. Suppose we know that $\rho = \sum_j \lambda_j U_j \rho' U_j^\dagger$. Let $\rho$ be diagonalized as $\rho = \sum_{k=1}^n p_k |\psi_k\rangle \langle\psi_k|$ and let $\rho'$ be diagonalized as $\rho' = \sum_{k=1}^n p'_k |\varphi_k\rangle \langle\varphi_k|$, where $\{|\psi_k\rangle\}_{k=1}^n$ and $\{|\varphi_k\rangle\}_{k=1}^n$ are orthonormal bases. Then, inserting the diagonalizations of the two states into $\rho = \sum_j \lambda_j U_j \rho' U_j^\dagger$, we get

$$\sum_{k=1}^n p_k |\psi_k\rangle \langle\psi_k| = \sum_j \lambda_j \sum_{k=1}^n p'_k U_j |\varphi_k\rangle \langle\varphi_k| U_j^\dagger$$

Now, multiply both sides on the left by $\langle\psi_l|$ and on the right by $|\psi_l\rangle$, getting

$$p_l = \sum_j \lambda_j \sum_{k=1}^n p'_k \langle\psi_l| U_j |\varphi_k\rangle \langle\varphi_k| U_j^\dagger |\psi_l\rangle . \tag{3.1.6}$$

Let us define $M_{j,lk} := \langle\psi_l| U_j |\varphi_k\rangle$. We can rewrite eq. (3.1.6) as

$$p_l = \sum_{k=1}^n p'_k \sum_j \lambda_j |M_{j,lk}|^2 .$$

If we are able to prove that $P_{lk} := \sum_j \lambda_j |M_{j,lk}|^2$ are the entries of a doubly stochastic matrix, we are done, because, by theorem 3.1.14 we conclude that $\mathbf{p} \preceq \mathbf{p}'$, and therefore $\rho$ is more mixed than $\rho'$. To this end, again, if we manage to prove that, for every $j$, $|M_{j,lk}|^2$ are the entries of a doubly stochastic matrix $M_j$, we are done, because doubly stochastic matrices are a convex set (see theorem 3.1.13).
By definition $|M_{j,lk}|^2 \geq 0$. Now let us prove that $\sum_{l=1}^n |M_{j,lk}|^2 = 1$. We have

$$\sum_{l=1}^n |M_{j,lk}|^2 = \sum_{l=1}^n \langle\psi_l| U_j |\varphi_k\rangle \langle\varphi_k| U_j^\dagger |\psi_l\rangle =$$

$$= \langle\varphi_k| U_j^\dagger \left( \sum_{l=1}^n |\psi_l\rangle \langle\psi_l| \right) U_j |\varphi_k\rangle = \langle\varphi_k| U_j^\dagger U_j |\varphi_k\rangle =$$

$$= \langle\varphi_k | \varphi_k\rangle = 1.$$

Similarly one proves also that $\sum_{k=1}^n |M_{j,lk}|^2 = 1$. This means that $|M_{j,lk}|^2$ are the entries of a doubly stochastic matrix and this concludes the proof. □

Using unitary operators and probabilities, we can define a particular type of channels, called *random unitary channels*, which are those appearing in the statement of theorem 3.1.19.

**Definition 3.1.20.** A channel $\mathcal{R}$ on a quantum system is called *Random Unitary* ($RU$) if for every state $\rho$ one has

$$\mathcal{R}(\rho) = \sum_j p_j U_j \rho U_j^\dagger = \sum_j p_j \mathcal{U}_j(\rho),$$

where the $p_j$'s are probabilities, the $U_j$'s are unitary operators and we have used the notation $\mathcal{U}_j(\rho)$ for the unitary channel $U_j \rho U_j^\dagger$.

Note that theorem 3.1.19 is the quantum version of theorem 3.1.14, with RU channels playing the role of doubly stochastic matrices and unitary channels playing the role of permutation matrices (cf. theorem 3.1.13).

We will find something very close to RU channels when studying mixedness in GPTs.

### 3.1.2 Duality between entanglement and mixedness

LOCC protocols provide the foundations for the resource theory of entanglement. However, they are not such a practical tool to work with, for they involve quantum operations performed by two parties and rounds of classical communication. Therefore, when presented with two bipartite states $\rho$ and $\rho'$ it is not always easy to see which is the more entangled, or, even, if they are comparable. Indeed, one should check all possible LOCC protocols to find whether there is one transforming $\rho$ into $\rho'$. If both $\rho$ and $\rho'$ are pure, this check is simplified, for, by Lo-Popescu theorem (theorem 3.1.6), we can restrict ourselves to 1-way LOCC protocols. However, even in this case, checking whether a pure state is more entangled than another is not an easy task.

In this subsection we will show that for pure states there is a duality between entanglement and mixedness: the more entangled a pure state, the more mixed its marginals. This is in fact a necessary and sufficient condition, providing us with a powerful tool to study the entanglement preorder of pure bipartite states [75, 50].

Again, the duality theorem will be proved using the full power of the Hilbert space formalism, leaving the operational proof for the case of GPTs

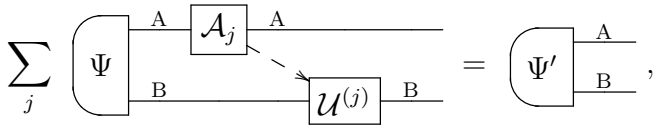
(section 4.4). In this way, the reader will be able to appreciate the difference between these two approaches.

**Theorem 3.1.21.** *Let $|\Psi\rangle$ and $|\Psi'\rangle$ be two pure bipartite states of the quantum system* AB*, and let $\rho$ (resp. $\sigma$) and $\rho'$ (resp. $\sigma'$) be their marginals on system* A *(resp.* B*). The following are equivalent.*

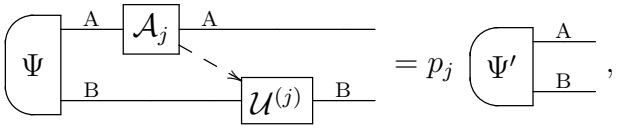
1. *$|\Psi\rangle$ is more entangled than $|\Psi'\rangle$.*
2. *$\rho$ is more mixed than $\rho'$.*
3. *$\sigma$ is more mixed than $\sigma'$.*

*Proof.* Thanks to Schmidt decomposition, the marginals of a pure bipartite state on each of the two subsystems A and B have the same eigenvalues. Therefore 2 and 3 are equivalent. Now let us prove the equivalence between 1 and 2.

1 implies 2. Suppose $|\Psi\rangle$ is more entangled than $|\Psi'\rangle$. Then, by Lo-Popescu theorem, there exists a 1-way LOCC protocol where Alice performs a quantum instrument $\{\mathcal{A}_j\}$, which can be assumed to be pure without loss of generality[5], with Kraus operators $\{M_j\}$, she communicates her outcome to Bob, and Bob applies a unitary channel $\mathcal{U}^{(j)}$.




namely




(3.1.7)

where $p_j$ is the probability of getting outcome $j$. By taking the partial trace over B, we have $\mathcal{A}_j(\rho) = p_j\rho'$, or, using Kraus operators, $M_j\rho M_j^\dagger = p_j\rho'$. Let us consider the polar decomposition[6] of $M_j\sqrt{\rho}$. Then, there exists a unitary operator $U_j$ such that $M_j\sqrt{\rho} = \sqrt{M_j\rho M_j^\dagger}U_j$. Since $M_j\rho M_j^\dagger = p_j\rho'$, we have

$$M_j\sqrt{\rho} = \sqrt{M_j\rho M_j^\dagger}U_j = \sqrt{p_j\rho'}U_j.$$

[5]If it is not pure, we can take the sum outside the protocol.

[6]Recall the polar decomposition of a square complex matrix $A$ means writing $A$ as $A = UP$, where $U$ is a unitary matrix and $P$ is a positive semi-definite matrix. Such $U$ and $P$ always exist, specifically $P$ is given by $P = \sqrt{AA^\dagger}$.

Multiplying on the left by $\sqrt{\rho}M_j^\dagger$, which is the adjoint of $M_j\sqrt{\rho}$, and recalling that, as a consequence, $\sqrt{\rho}M_j^\dagger = U_j^\dagger\sqrt{p_j\rho'}$, we have

$$\sqrt{\rho}M_j^\dagger M_j\sqrt{\rho} = U_j^\dagger\sqrt{p_j\rho'}\sqrt{p_j\rho'}U_j$$

$$\sqrt{\rho}M_j^\dagger M_j\sqrt{\rho} = p_j U_j^\dagger\rho' U_j.$$

Now we sum over $j$, recalling that $\sum_j M_j^\dagger M_j = \mathbf{1}$, and we obtain

$$\rho = \sum_j p_j U_j^\dagger\rho' U_j.$$

By theorem 3.1.19, this means that $\rho$ is more mixed than $\rho'$.

2 implies 1. Suppose $\rho$ is more mixed than $\rho'$ and that $\dim\mathcal{H}_\mathrm{A} = n$, then by theorem 3.1.19 there exist probabilities $\lambda_j$ and unitary matrices $U_j$ such that $\rho = \sum_j \lambda_j U_j\rho U_j^\dagger$. Consider now $|\Psi\rangle$, and define Kraus operators for Alice's quantum instrument $\{\mathcal{A}_j\}$ applied to $|\Psi\rangle$ like in eq. (3.1.7) as

$$M_j\sqrt{\rho} := \sqrt{\lambda_j\rho'}U_j^\dagger. \tag{3.1.8}$$

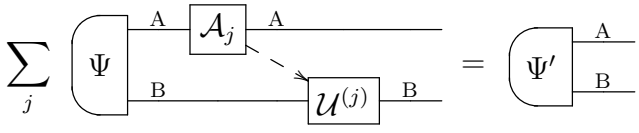
Let us multiply on the left by the adjoint, that is $\sqrt{\rho}M_j^\dagger$,

$$\sqrt{\rho}M_j^\dagger M_j\sqrt{\rho} = \lambda_j U_j\rho' U_j^\dagger$$

and sum over $j$

$$\sqrt{\rho}\left(\sum_j M_j^\dagger M_j\right)\sqrt{\rho} = \sum_j \lambda_j U_j\rho' U_j^\dagger = \rho.$$

First, let us deal with the case when $\rho$ is invertible, namely when it has full rank $n$. In this case, $\sqrt{\rho}$ is invertible too, and therefore we conclude that $\sum_j M_j^\dagger M_j = \mathbf{1}$, and therefore the $M_j$'s are really Kraus operators of a quantum instrument. If Alice obtains outcome $j$, then the state on A becomes $M_j\rho M_j^\dagger = \lambda_j\rho'$ (see eq. (3.1.8)). This means that $\frac{1}{\lambda_j}(\mathcal{A}_j\otimes\mathcal{I})(|\Psi\rangle\langle\Psi|)$ and $|\Psi'\rangle\langle\Psi'|$ have the same marginal on A. Hence, they differ by a unitary channel $\mathcal{U}_j$ on B, yielding



Now consider the case when the rank of $\rho$ is $k < n$, namely $\rho$ has $k$ non-vanishing eigenvalues. Its vector of eigenvalues, arranged in increasing order, can be represented as $\mathbf{p} = \begin{pmatrix} \underbrace{0 \dots 0}_{n-k} & p_1 & \dots & p_k \end{pmatrix}^t$. Let $\mathbf{p}' = \begin{pmatrix} p'_1 & \dots & p'_n \end{pmatrix}^t$ be the vector of the eigenvalues of $\rho'$ in increasing order. By hypothesis, $\mathbf{p} \preceq \mathbf{p}'$, therefore, recalling the majorization conditions for vectors in increasing order, we have

$$0 \geq \sum_{j=1}^{n-k} p'_j,$$

which means that $p'_1 = \ldots = p'_{n-k} = 0$. Therefore $\rho'$ has at most $k$ non-vanishing eigenvalues, and its rank is less than or equal to $k$. As a consequence, if the rank of $\rho$ is $k < n$, we can choose A to be a smaller Hilbert space of dimension $k$, because both $\rho$ and $\rho'$ live there. Then we repeat the same argument as above for the full-rank case, because in this smaller Hilbert space $\rho$ has maximum rank. □

Now the issue of pure state convertibility by LOCC protocol has been turned into checking the majorization conditions on the marginals of the pure states involved, which is definitely an easier task.

The duality enables us to better understand the entanglement preorder of pure bipartite states. Indeed, now we can characterize the equivalence classes of equally entangled pure states.

**Proposition 3.1.22.** *The pure bipartite states $|\Psi\rangle$ and $|\Psi'\rangle$ of system* AB *are equally entangled if and only if there exist two unitary operators $U_{\mathrm{A}}$ on* A *and $V_{\mathrm{B}}$ on* B *such that $|\Psi'\rangle = (U_{\mathrm{A}} \otimes V_{\mathrm{B}}) |\Psi\rangle$.*

*Proof.* Sufficiency was proven on page 69, and it does not require duality.

To prove necessity we must resort to duality. Suppose $|\Psi\rangle$ is as entangled as $|\Psi'\rangle$. Then, if $\rho$ and $\rho'$ are the marginals on A of $|\Psi\rangle$ and $|\Psi'\rangle$ respectively, we know that $\rho$ is as mixed as $\rho'$. By proposition 3.1.18, there exists a unitary operator $U_{\mathrm{A}}$ acting on A such that $\rho' = U_{\mathrm{A}} \rho U_{\mathrm{A}}^\dagger$. As a consequence, the pure states $(U_{\mathrm{A}} \otimes \mathbf{1}) |\Psi\rangle$ and $|\Psi'\rangle$ have the same marginal on A. Therefore they differ by a unitary operator $V_{\mathrm{B}}$ on B, namely $|\Psi'\rangle = (U_{\mathrm{A}} \otimes V_{\mathrm{B}}) |\Psi\rangle$. □

We can also identify the states which are the maximum in the entanglement preorder, not only between pure states, but also between generic bipartite states.

**Proposition 3.1.23.** *The maximum of the entanglement order is given by the equivalence class of the purifications of the maximally mixed state $\chi = \frac{1}{n}\mathbf{1}$, where $n$ is the dimension of the Hilbert space concerned.*

*Proof.* Since $\chi$ is more mixed than any state, by the duality, every purification of it is more entangled than any other *pure* state.

To prove that the states in the equivalence class of the purifications of $\chi$ are more entangled than any (possibly mixed) state, consider a generic bipartite state $\Sigma$. $\Sigma$ can be written as a mixture of pure states, $\Sigma = \sum_j p_j |\Psi_j\rangle \langle\Psi_j|$, where the $p_j$'s are probabilities. Now, for every pure state $|\Psi_j\rangle$ there exists an LOCC channel $\mathcal{C}_j$ such that $|\Psi_j\rangle \langle\Psi_j| = \mathcal{C}_j (|\Phi\rangle \langle\Phi|)$, where $|\Phi\rangle$ is a generic element of the equivalence class of the purifications of $\chi$. Therefore, summing over $j$, one has $\Sigma = \sum_j p_j \mathcal{C}_j (|\Phi\rangle \langle\Phi|)$, and clearly $\sum_j p_j \mathcal{C}_j$ is an LOCC channel. This shows that $|\Phi\rangle \langle\Phi| \succeq_{\text{ent}} \Sigma$ for any bipartite state $\Sigma$. □

Finally, by using the duality, we can show that the resource theory of entanglement admits catalysts [50].

**Example 3.1.24.** Consider the bipartite system AB described by the Hilbert space $\mathcal{H}_{\text{AB}} \approx \mathbb{C}^4 \otimes \mathbb{C}^4$. Let $\{|0\rangle, |1\rangle, |2\rangle, |3\rangle\}$ be an orthonormal basis for each of the two subsystems. Consider the bipartite states

$$|\Psi\rangle = \sqrt{\frac{2}{5}} |00\rangle + \sqrt{\frac{2}{5}} |11\rangle + \frac{1}{\sqrt{10}} |22\rangle + \frac{1}{\sqrt{10}} |33\rangle$$

and

$$|\Psi'\rangle = \frac{1}{\sqrt{2}} |00\rangle + \frac{1}{2} |11\rangle + \frac{1}{2} |22\rangle .$$

Let us compare the degree of entanglement of $|\Psi\rangle$ and $|\Psi'\rangle$. To do that, let us consider the vectors of the eigenvalues of their marginals, They are

$$\mathbf{p} = \begin{pmatrix} \frac{2}{5} \\ \frac{2}{5} \\ \frac{1}{10} \\ \frac{1}{10} \end{pmatrix} \qquad \mathbf{p}' = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{4} \\ \frac{1}{4} \\ 0 \end{pmatrix} .$$

We have $\frac{1}{2} > \frac{2}{5}$, therefore $\mathbf{p}' \not\preceq \mathbf{p}$, but we have $\frac{2}{5} + \frac{2}{5} > \frac{3}{4}$, therefore $\mathbf{p} \not\preceq \mathbf{p}'$. This means that $|\Psi\rangle$ and $|\Psi'\rangle$ are not comparable.

Now consider an additional system C with Hilbert space $\mathcal{H}_{\mathrm{C}} \approx \mathbb{C}^2 \otimes \mathbb{C}^2$, and the pure state

$$|c\rangle = \sqrt{\frac{3}{5}}\,|00\rangle + \sqrt{\frac{2}{5}}\,|11\rangle\,.$$

$|c\rangle$ acts as a catalyst for the transformation of $|\Psi\rangle$ into $|\Psi'\rangle$. We will prove that there exists an LOCC protocol transforming $|\Psi\rangle_{\mathrm{AB}}\,|c\rangle_{\mathrm{C}}$ into $|\Psi'\rangle_{\mathrm{AB}}\,|c\rangle_{\mathrm{C}}$. To do that it is enough to check the majorization condition when we take the tensor product of the marginals of $|\Psi\rangle$ and $|\Psi'\rangle$ with the marginals of $|c\rangle$. Both marginals of $|c\rangle$ have eigenvalues $\mathbf{c} = \left(\begin{array}{cc} \frac{3}{5} & \frac{2}{5} \end{array}\right)^t$. Therefore now we have

$$\mathbf{p} \otimes \mathbf{c} = \left(\begin{array}{cccccccc} \frac{6}{25} & \frac{6}{25} & \frac{4}{25} & \frac{4}{25} & \frac{3}{50} & \frac{3}{50} & \frac{1}{25} & \frac{1}{25} \end{array}\right)^t$$

$$\mathbf{p}' \otimes \mathbf{c} = \left(\begin{array}{cccccccc} \frac{3}{10} & \frac{1}{5} & \frac{3}{20} & \frac{3}{20} & \frac{1}{10} & \frac{1}{10} & 0 & 0 \end{array}\right)^t$$

Now it is easy to check that $\mathbf{p} \otimes \mathbf{c} \preceq \mathbf{p}' \otimes \mathbf{c}$, therefore there exists an LOCC protocol transforming $|\Psi\rangle_{\mathrm{AB}}\,|c\rangle_{\mathrm{C}}$ into $|\Psi'\rangle_{\mathrm{AB}}\,|c\rangle_{\mathrm{C}}$. Hence $|c\rangle$ is a catalyst.

### 3.1.3 Entanglement monotones

Since we have set forth a duality between pure-state entanglement and mixedness, searching for pure-state entanglement monotones is equivalent to searching for measures of mixedness for their marginals. Such functions preserve the mixedness preorder in the following sense.

**Definition 3.1.25.** A real function $f : \mathcal{L}(\mathcal{H}) \longrightarrow \mathbb{R}$ is called a *measure of mixedness* if for every pair of states $\rho, \rho' \in \mathcal{L}(\mathcal{H})$, we have $f(\rho) \geq f(\rho')$ whenever $\rho$ is more mixed than $\rho'$.

As it is known, functions of states are functions of their eigenvalues, and this is a good point, for mixedness depend only on the eigenvalues of a state. According to definition 3.1.15, a measure of mixedness is such that $f(\mathbf{p}) \geq f(\mathbf{p}')$ whenever $\mathbf{p} \preceq \mathbf{p}'$. Such functions are called Schur-concave functions.

**Definition 3.1.26.** A real-valued function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is called *Schur-concave* if $f(x) \geq f(y)$ whenever $x \preceq y$, for every $x, y \in \mathbb{R}^n$.

A real-valued function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is called *Schur-convex* if $f(x) \leq f(y)$ whenever $x \preceq y$, for every $x, y \in \mathbb{R}^n$.

Using doubly stochastic matrices, the definitions of Schur-concave and Schur-convex functions can be rephrased as follows.

**Proposition 3.1.27.** *A function $f$ is Schur-concave if and only if*

$$f(Px) \geq f(x)$$

*for every doubly stochastic matrix $P$ of order $n$ and for every $x \in \mathbb{R}^n$.*

*A function $f$ is Schur-convex if and only if*

$$f(Px) \leq f(x)$$

*for every doubly stochastic matrix $P$ of order $n$ and for every $x \in \mathbb{R}^n$.*

*Proof.* By theorem 3.1.14, $Px \preceq x$, for every doubly stochastic matrix $P$. Therefore $f$ is Schur-concave if and only if $f(Px) \geq f(x)$. Similarly one proves the statement for Schur-convex functions. □

Therefore in our setting measures of mixedness coincide with Schur-concave functions, which are also pure-state entanglement monotones, thanks to the duality (subsection 3.1.2). Let us see some examples of Schur-concave functions.

**Example 3.1.28.** A function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is called *symmetric* if $f(\Pi x) = f(x)$ for any permutation matrix $\Pi$. Every symmetric and concave function is also Schur-concave [80]. As a special case, every concave and separate-variable[7] function is Schur-concave.

One of the most paradigmatic example of Schur-concave function on vectors of probabilities is Shannon entropy [83] $H(\mathbf{p}) = -\sum_{i=1}^{n} p_i \log_a p_i$, where $a > 1$. Note that Shannon entropy is a concave and separate-variable function. In the quantum context, regarded as a function of a density operator, it becomes von Neumann entropy [84]

$$S(\rho) = -\mathrm{tr}\,\rho \log_a \rho = -\sum_{j=1}^{n} p_j \log_a p_j = H(\mathbf{p}),$$

where $\mathbf{p}$ is the vector of the eigenvalues of $\rho$. Therefore von Neumann entropy is an entanglement monotone.

[7]Recall a function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is called a *separate-variable function* if $f(x) = \sum_{i=1}^{n} g(x_i)$, where $g : \mathbb{R} \longrightarrow \mathbb{R}$ is a real function.

## 3.2 The resource theory of purity

Now let us turn to the other important example of resource theory in the quantum case, which has been already considered in the literature as as a starting point towards an informational axiomatization of quantum thermodynamics [85, 86].

The free operations of the resource theory of purity is generated by the following operations:

- adding ancillary systems in the maximally mixed state $\chi = \frac{1}{n}\mathbf{1}$, where $n$ is the dimension of the ancillary system;
- applying unitary channels $\mathcal{U}(\bullet) = U \bullet U^\dagger$;
- taking the partial trace over a system.

According to this definition, the only free state is the maximally mixed state, and this is the reason why we call this theory a resource theory of purity.

*Remark* 3.2.1. Note that the fact that $\chi$ is the only free state is compatible with the requirement that free states have to be closed under parallel composition (see section 2.3). Indeed, in quantum mechanics $\chi_{\mathrm{A}} \otimes \chi_{\mathrm{B}} = \chi_{\mathrm{AB}}$. This is not true in general in GPTs, and this is why we will have to depart from the definition of resource theory of purity presented in this section, when trying to extend it to GPTs.

Combining all the three kinds of generating free operations, it is not hard to see that the most general free transformation on system A is a quantum channel of the form

$$\mathcal{N}(\rho_{\mathrm{A}}) = \mathrm{tr}_{\mathrm{E}}\left[U_{\mathrm{AE}}\left(\rho_{\mathrm{A}} \otimes \frac{1}{d_{\mathrm{E}}}\mathbf{1}_{\mathrm{E}}\right)U_{\mathrm{AE}}^\dagger\right], \tag{3.2.1}$$

where $d_{\mathrm{E}}$ is the dimension of the Hilbert space associated with the ancillary system E, and $U_{\mathrm{AE}}$ is a unitary operator acting on AE. A channel of the form of eq. (3.2.1) is called *noisy operation*. Unfortunately, noisy operations are not a (topologically) closed set, as shown by Shor [87], therefore we choose to consider as a noisy operation any quantum channel which can be arbitrarily well approximated by noisy operations of the form of eq. (3.2.1).

As explained in section 2.2, we can establish a hierarchy among the states of a system by using noisy operations.

**Definition 3.2.2.** We say that a state $\rho$ is *purer* than a state $\rho'$ if there exists a noisy operation $\mathcal{N}$, such that $\rho' = \mathcal{N}(\rho)$. In this case we will write $\rho \succeq_{\text{pur}} \rho'$.

This can be explained intuitively as the fact that during the evolution described by $\mathcal{N}$, the state $\rho$ is put into contact with a maximally mixed state, which is later removed, thereby "absorbing" some mixedness from the ancillary system in the maximally mixed state. Let us show that the maximally mixed state $\chi$ is a minimal element of the purity relation, namely if $\chi \succeq_{\text{pur}} \rho$, then $\rho = \chi$.

**Proposition 3.2.3.** *Consider a Hilbert space $\mathcal{H} \approx \mathbb{C}^n$. For any state $\rho$ on $\mathcal{H}$, if $\chi \succeq_{\text{pur}} \rho$, then $\rho = \chi$, where $\chi = \frac{1}{n}\mathbf{1}$.*

*Proof.* Suppose $\chi$ is purer than $\rho$. Then there exists a noisy transformation $\mathcal{N}$ such that $\mathcal{N}(\chi) = \rho$. Recalling eq. (3.2.1), we get

$$\rho = \mathcal{N}(\chi) = \text{tr}_{\text{E}}\left[U_{\text{AE}}\left(\frac{1}{n}\mathbf{1} \otimes \frac{1}{d_{\text{E}}}\mathbf{1}_{\text{E}}\right)U_{\text{AE}}^{\dagger}\right] =$$

$$= \text{tr}_{\text{E}}\left[U_{\text{AE}}\left(\frac{1}{nd_{\text{E}}}\mathbf{1}_{\text{AE}}\right)U_{\text{AE}}^{\dagger}\right] = \text{tr}_{\text{E}}\left(\frac{1}{nd_{\text{E}}}\mathbf{1}_{\text{AE}}\right) = \frac{1}{n}\mathbf{1} = \chi.$$

This concludes the proof. □

In subsection 3.1.1 we defined a mixedness preorder between single-system states. Mixedness and purity seem to be opposite concepts, therefore we anticipate that the mixedness relation will be related somehow to the present definition of the purity preorder. Is the mixedness relation just the reverse preorder? To answer this question we need to understand the relation between noisy operations and RU channels [86], which are the channels related to the mixedness preorder (see theorem 3.1.19). The proof of the following lemma is an improvement of the proof presented in [86], and fixes some bugs present there.

**Lemma 3.2.4.** *RU channels are a strict subset of noisy operations.*

*Proof.* Let us prove that every RU channel $\mathcal{R} = \sum_{j=1}^{m} p_j \mathcal{U}_j$ on system A, where the $p_j$'s are probabilities and the $\mathcal{U}_j$'s are unitary channels, can be realized as a noisy operation. Consider an ancillary system E with Hilbert space $\mathcal{H}_{\text{E}} \approx \mathbb{C}^N$, initially completely uncorrelated with system A, and in the

maximally mixed state $\chi_{\mathrm{E}}$. For every $j = 1, \ldots, m$ arising in the definition of $\mathcal{R}$, consider a subspace $\mathcal{H}_j$ of $\mathcal{H}_{\mathrm{E}}$, with $\dim \mathcal{H}_j = [p_j N]$, where $[p_j N]$ is the integral part[8] of $p_j N$, such that we can write $\mathcal{H}_{\mathrm{E}}$ as $\mathcal{H}_{\mathrm{E}} = \bigoplus_{j=1}^{m+1} \mathcal{H}_j$, and the subspaces $\mathcal{H}_j$'s are orthogonal to each other for every $j$. Now, define a unitary operator $U_{\mathrm{AE}}$ as follows

$$U_{\mathrm{AE}} : \begin{cases} |\psi\rangle_{\mathrm{A}} |\varphi\rangle_{\mathrm{E}} \mapsto U_j |\psi\rangle_{\mathrm{A}} |\varphi\rangle_{\mathrm{E}} & \text{if } |\varphi\rangle_{\mathrm{E}} \in \mathcal{H}_j, \text{ for all } j = 1, \ldots, m \\ |\psi\rangle_{\mathrm{A}} |\varphi\rangle_{\mathrm{E}} \mapsto |\psi\rangle_{\mathrm{A}} |\varphi\rangle_{\mathrm{E}} & \text{if } |\varphi\rangle_{\mathrm{E}} \in \mathcal{H}_{m+1} \end{cases},$$

and extended by linearity. We end up having a unitary operator of the form[9]

$$U_{\mathrm{AE}} = \sum_{j=1}^{m+1} U_j \otimes P_j, \tag{3.2.2}$$

where we set $U_{m+1} = \mathbf{1}$, and $P_j$ is the orthogonal projector on the subspace $\mathcal{H}_j$ for all $j = 1, \ldots, m+1$. Now let us consider the noisy operation

$$\mathcal{N}(\rho) = \mathrm{tr}_{\mathrm{E}} \left[ U_{\mathrm{AE}} (\rho \otimes \chi_{\mathrm{E}}) U_{\mathrm{AE}}^{\dagger} \right],$$

where $U_{\mathrm{AE}}$ is given by eq. (3.2.2). Inserting the expression of $U_{\mathrm{AE}}$, we get

$$\mathcal{N}(\rho) = \mathrm{tr}_{\mathrm{E}} \left( \sum_{j,k=1}^{m+1} U_j \rho U_k^{\dagger} \otimes P_j \chi P_k \right) = \frac{1}{N} \mathrm{tr}_{\mathrm{E}} \left( \sum_{j,k=1}^{m+1} U_j \rho U_k^{\dagger} \otimes P_j P_k \right)$$

$$= \frac{1}{N} \mathrm{tr}_{\mathrm{E}} \left( \sum_{j,k=1}^{m+1} U_j \rho U_k^{\dagger} \otimes \delta_{jk} P_j \right) = \frac{1}{N} \mathrm{tr}_{\mathrm{E}} \left( \sum_{j=1}^{m+1} U_j \rho U_j^{\dagger} \otimes P_j \right) = \frac{1}{N} \sum_{j=1}^{m+1} (\mathrm{tr}_{\mathrm{E}} P_j) U_j \rho U_j^{\dagger}.$$

Now, the trace of a projector is the dimension of the subspace on which it projects. Hence

$$\mathcal{N}(\rho) = \sum_{j=1}^{m+1} \left( \frac{1}{N} \dim \mathcal{H}_j \right) U_j \rho U_j^{\dagger}. \tag{3.2.3}$$

We have obtained that the noisy operation gives rise to a RU channel $\mathcal{R}'(\rho) = \sum_{j=1}^{m+1} p'_j U_j \rho U_j^{\dagger}$, where $p'_j := \frac{1}{N} \dim \mathcal{H}_j$, and where $U_{m+1} = \mathbf{1}$ (the other

[8] Recall the integral part of a real number is defined as $[x] := \max_{n \in \mathbb{Z}} \{n \leq x\}$.

[9] A unitary operator of this form is usually called *control-unitary*, the system where the projectors $P_j$ act is called *control system*, and the system where the unitary operators $U_j$ act is called *target system*.

unitary operators being the same as in $\mathcal{R}$). This is almost the desired result. Now we want to show that in the limit of large $N$, we get $\mathcal{R}$. This would mean that $\mathcal{R}$ can be arbitrarily well approximated by a noisy operation, therefore it is a noisy operation itself.

The first step is to prove that we can get rid of the term with the identity ($j = m+1$). The dimension of the "residual" subspace $\mathcal{H}_{m+1}$ is $\dim \mathcal{H}_{m+1} = N - \sum_{j=1}^{m} [p_j N]$. Since by definition $[p_j N] \leq p_j N \leq [p_j N] + 1$, then $0 \leq p_j N - [p_j N] \leq 1$, and by summing over $j$, for $j = 1, \ldots, m$, we get $0 \leq \dim \mathcal{H}_{m+1} \leq m$. In other words

$$0 \leq \frac{1}{N} \dim \mathcal{H}_{m+1} \leq \frac{m}{N}$$

and therefore in the limit of $N \to +\infty$, the term $\frac{1}{N} \dim \mathcal{H}_{m+1}$ vanishes: the sum in eq. (3.2.3) has at most $m$ non-vanishing terms, like $\mathcal{R}$.

Now we want to prove that for $j = 1, \ldots, m$, the terms $\frac{1}{N} \dim \mathcal{H}_j = \frac{[p_j N]}{N}$ converge to $p_j$ in the limit of $N \to +\infty$. As $p_j N - 1 \leq [p_j N] \leq p_j N$, we have

$$p_j - \frac{1}{N} \leq \frac{[p_j N]}{N} \leq p_j.$$

This means that when $N \to +\infty$, $\frac{[p_j N]}{N}$ converges to $p_j$. We have shown that every RU channel can be approximated arbitrarily well by noisy operations.

The proof of the strict inclusion is far more technical and we recommend the interested reader to refer to [87]. □

Therefore there is some relation between noisy operations and RU channels. Since noisy operations are more general than RU channels, one may expect that the purity defined given via RU channels (as the reverse mixedness relation) is a restricted version of the purity relation defined via noisy operations. Nonetheless, the relation is tighter than what it seems at first glance, as shown in the following theorem.

**Theorem 3.2.5.** *Let $\rho$ and $\rho'$ be two states. The following are equivalent.*

1. *$\rho' = \mathcal{N}(\rho)$, where $\mathcal{N}$ is a noisy operation.*
2. *$\rho' = \mathcal{R}(\rho)$, where $\mathcal{R}$ is a RU channel.*

*Proof.* By lemma 3.2.4, RU channels are noisy operations, therefore it is immediate that 2 implies 1. The converse implication is non-trivial. Diagonalize $\rho$ and $\rho'$ as $\rho = \sum_{j=1}^{n} p_j \left|\psi_j\right\rangle \left\langle\psi_j\right|$ and $\rho' = \sum_{j=1}^{n} p'_j \left|\varphi_j\right\rangle \left\langle\varphi_j\right|$. Suppose $\rho' = \mathcal{N}(\rho)$, then

$$\sum_{j=1}^{n} p'_j \left|\varphi_j\right\rangle \left\langle\varphi_j\right| = \sum_{j=1}^{n} p_j \mathcal{N}\left(\left|\psi_j\right\rangle \left\langle\psi_j\right|\right).$$

Let us multiply on the left by $\langle\varphi_k|$ and on the right by $|\varphi_k\rangle$. We get

$$p'_k = \sum_{j=1}^{n} p_j \left\langle\varphi_k\right| \mathcal{N}\left(\left|\psi_j\right\rangle \left\langle\psi_j\right|\right) \left|\varphi_k\right\rangle.$$

This expression can be rewritten as $p'_k = \sum_{j=1}^{n} P_{kj} p_j$, where $P_{kj} := \langle\varphi_k| \mathcal{N}(|\psi_j\rangle \langle\psi_j|) |\varphi_k\rangle$. Now we will prove that the $P_{kj}$'s are the entries of a doubly stochastic matrix $P$. Now,

$$\left\langle\varphi_k\right| \mathcal{N}\left(\left|\psi_j\right\rangle \left\langle\psi_j\right|\right) \left|\varphi_k\right\rangle \geq 0,$$

because $\mathcal{N}$ is a (completely) positive map (quantum channel). Then

$$\sum_{k=1}^{n} P_{jk} = \sum_{k=1}^{n} \left\langle\varphi_k\right| \mathcal{N}\left(\left|\psi_j\right\rangle \left\langle\psi_j\right|\right) \left|\varphi_k\right\rangle = \operatorname{tr} \mathcal{N}\left(\left|\psi_j\right\rangle \left\langle\psi_j\right|\right) = 1,$$

because $\{|\varphi_k\rangle\}_{k=1}^{n}$ is an orthonormal basis, and $\mathcal{N}$ is trace-preserving. Finally,

$$\sum_{j=1}^{n} P_{jk} = \sum_{j=1}^{n} \left\langle\varphi_k\right| \mathcal{N}\left(\left|\psi_j\right\rangle \left\langle\psi_j\right|\right) \left|\varphi_k\right\rangle = \left\langle\varphi_k\right| \mathcal{N}(\mathbf{1}) \left|\varphi_k\right\rangle =$$

$$= n \left\langle\varphi_k\right| \mathcal{N}(\chi) \left|\varphi_k\right\rangle,$$

where we used the fact that $\{|\psi_j\rangle\}_{j=1}^{n}$ is an orthonormal basis. By proposition 3.2.3, $\mathcal{N}(\chi) = \chi$. Therefore,

$$\sum_{j=1}^{n} P_{jk} = n \left\langle\varphi_k\right| \frac{1}{n}\mathbf{1} \left|\varphi_k\right\rangle = \left\langle\varphi_k \,\middle|\, \varphi_k\right\rangle = 1.$$

If $\mathbf{p} = \begin{pmatrix} p_1 & \dots & p_n \end{pmatrix}^t$ is the vector of the eigenvalues of $\rho$, and $\mathbf{p}' = \begin{pmatrix} p'_1 & \dots & p'_n \end{pmatrix}^t$ is the vector of the eigenvalues of $\rho'$, we have $\mathbf{p}' = P\mathbf{p}$, and since $P$ is doubly stochastic, this means $\mathbf{p}' \preceq \mathbf{p}$. Hence $\rho'$ is more mixed than $\rho$, and by theorem 3.1.19 there exists a RU channel $\mathcal{R}$ such that $\rho' = \mathcal{R}(\rho)$. □

Therefore, whenever we can transform a state into another by a noisy operation, we can achieve the same result by using a RU channel. As a consequence, noisy operations and RU channels give rise exactly to the same purity preorder. In the case of RU channel, this is defined as the reverse order given by mixedness relation, namely $\rho$ is purer than $\rho'$ if there exists a RU channel such that $\rho' = \mathcal{R}(\rho)$. Therefore the terminology of definition 3.2.2 is well posed. As special cases, we get that the maximally mixed state is not only a minimal element, but also a minimum in the purity relation and that pure states are purer than every state, even according to the definition via noisy operations.

Theorem 3.2.5 is important also because it gives us a practical way to study the convertibility of states under noisy operations, which is again majorization. In this way, we have $\rho \succeq_{\text{pur}} \rho'$ if and only if $\mathbf{p}' \preceq \mathbf{p}$, where $\mathbf{p}$ is the vector of the eigenvalues of $\rho$ and $\mathbf{p}'$ is the vector of the eigenvalues of $\rho'$.

Rephrasing example 3.1.24, we see that the resource theory of purity admits catalysts.

**Example 3.2.6.** Consider the states $\rho$, and $\rho'$ on $\mathcal{H}_{\text{A}} \approx \mathbb{C}^4$, with eigenvalues $\mathbf{p} = \left(\begin{array}{cccc} \frac{2}{5} & \frac{2}{5} & \frac{1}{10} & \frac{1}{10} \end{array}\right)^t$ and $\mathbf{p}' = \left(\begin{array}{cccc} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 \end{array}\right)^t$ respectively. As noted in example 3.1.24, since $\mathbf{p} \not\preceq \mathbf{p}'$, and $\mathbf{p}' \not\preceq \mathbf{p}$, there exist no noisy operations converting $\rho$ into $\rho'$ or vice versa. But any state $\sigma$ on $\mathcal{H}_{\text{C}}$ with eigenvalues $\mathbf{c} = \left(\begin{array}{cc} \frac{3}{5} & \frac{2}{5} \end{array}\right)^t$ acts as a catalysts, for the state $\rho' \otimes \sigma$ can be transformed into $\rho \otimes \sigma$ by a noisy operation (see again example 3.1.24).

### 3.2.1 Purity monotones

Now we can look for purity monotones. Essentially, since the purity preorder induced by noisy operations is just the reverse preorder given by mixedness (see subsection 3.1.1), purity monotones can be generated from measures of mixedness (which are also entanglement monotones) in the following way: if $f : \mathcal{L}(\mathcal{H}) \longrightarrow \mathbb{R}$ is a measure of mixedness, then $g \circ f : \mathcal{L}(\mathcal{H}) \longrightarrow \mathbb{R}$ is a purity monotone, where $g : \mathbb{R} \longrightarrow \mathbb{R}$ is a decreasing function. The simplest choice is to take $g(x) = -x$.

Purity monotones are therefore Schur-convex functions (see definition 3.1.26) of states, indeed the basic requisite is that $f(\rho) \geq f(\rho')$ whenever $\rho$ is purer than $\rho'$, namely $\mathbf{p}' \preceq \mathbf{p}$, where $\mathbf{p}$ and $\mathbf{p}'$ are the vectors of the eigenvalues of $\rho$ and $\rho'$ respectively. Examples of Schur-convex functions can be derived easily from example 3.1.28.

**Example 3.2.7.** Every symmetric and convex function is a Schur-convex function [80], and, as a special case, every convex and separate-variable function is Schur-convex. This class of Schur-convex functions plays a special role, for it is a complete family of purity monotones.

In subsection 2.2.2 we saw that every theory admits a complete family of resource monotones. For the resource theory of purity, it is given by convex and separate-variable functions [82, 80].

**Proposition 3.2.8.** *Let $x, y \in \mathbb{R}^n$. We have $x \preceq y$ if and only if $f(x) \leq f(y)$ for* every *convex and separate-variable function $f$.*

*Proof.* See ref. [80]. □

In this way we have $\rho \succeq_{\text{pur}} \rho'$ if and only if $f(\rho) \geq f(\rho')$ for *every* convex and separate-variable function $f$.

## 3.3 Other examples

In this section we present a couple of further examples of quantum resource theories.

### 3.3.1 Quantum resource theory of asymmetry

We start from the quantum resource theory of *asymmetry* [88, 89, 90]. Symmetries are a powerful tool in theoretical physics to reduce the complexity of a problem, which might not be solved exactly without their aid.

Let us consider a group $G$ acting on the density operators of a Hilbert space $\mathcal{H}$ through a projective unitary representation $\mathcal{U}_g$, $\mathcal{U}_g(\rho) = U_g \rho U_g^\dagger$, where $U_g$ is a projective unitary representation on the rays of $\mathcal{H}$ of $G$. Note that in the parallel composition of two systems one needs to consider the tensor product of the representations associated with the two Hilbert spaces.

Now we must define which quantum states and operations are free.

**Definition 3.3.1.** A quantum operation $\mathcal{M} : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H}')$ is called *covariant* under the action of the group representation if

$$\mathcal{M}\mathcal{U}_g = \mathcal{U}'_g \mathcal{M}, \tag{3.3.1}$$

for all $g \in G$, where $\mathcal{U}_g$ acts on $\mathcal{L}(\mathcal{H})$ and $\mathcal{U}'_g$ on $\mathcal{L}(\mathcal{H}')$.

In this way, $\mathcal{M}$ preserves the symmetry of systems under the action of $G$. We declare that all covariant quantum operations are *free*. This induces also a notion of free states (see section 2.3). Indeed, when we have a state $\rho$, the input representation $\mathcal{U}_g$ is the trivial representation, therefore eq. (3.3.1) becomes, after getting rid of primes,

$$\rho = \mathcal{U}_g(\rho) = U_g \rho U_g^\dagger \tag{3.3.2}$$

for all $g \in G$. Eq. (3.3.2) defines *free states*. In other words, a state is free if and only if it is left invariant by the group representation.

In a resource theory of asymmetry, one can introduce a preorder by setting that a state $\rho$ is *more asymmetric* than $\rho'$ (or that $\rho'$ is *more symmetric* than $\rho$) if there exists a covariant channel $\mathcal{C}$ such that $\rho' = \mathcal{C}(\rho)$. The invariant states are the minimal elements. Indeed, if an invariant state $\rho$ is more asymmetric than $\sigma$, then $\sigma$ is invariant too. In symbols, if $\rho$ is more asymmetric than $\sigma$, then $\sigma = \mathcal{C}(\rho)$. Let us apply a generic $\mathcal{U}'_g$, then

$$\mathcal{U}'_g(\sigma) = \mathcal{U}'_g \mathcal{C}(\rho) = \mathcal{C}\mathcal{U}_g(\rho) = \mathcal{C}(\rho) = \sigma,$$

where we used the fact that the quantum channel $\mathcal{C}$ is covariant and that $\rho$ is invariant. In this way, the application of covariant quantum channels increases the level of symmetry in a system.

Having established a preorder of density operators according to their symmetries, one can define also measures of asymmetry, which have been shown to play a major role in generalizations of Nöther's theorem [91].

Note that all the present discussion can be easily exported to GPTs because it is entirely operational.

### 3.3.2 Quantum resource theory of athermality

The quantum resource theory of *athermality* [4, 12, 10, 86, 17] is a generalization of the quantum resource theory of purity. In such a setting we consider quantum systems equipped with a Hamiltonian operator $H$, and we also *fix* a temperature $T$ throughout. Now, when composing different quantum systems with Hamiltonian operators $H$ and $H'$, the Hamiltonian operator of the composite system will be $H \otimes \mathbf{1} + \mathbf{1} \otimes H'$.

We define free processes starting from a generating set, as done above with the resource theory of purity. Free operations are

- adding ancillary systems in the Boltzmann state $\rho = \frac{1}{Z}\mathrm{e}^{-\beta H}$, where $\beta$ is Boltzmann's factor ($\beta = \frac{1}{kT}$, where $k$ is Boltzmann's constant), $H$ is the Hamiltonian operator, and $Z$ is the canonical partition function ($Z = \operatorname{tr} \mathrm{e}^{-\beta H}$);
- unitary channels $\mathcal{U}(\bullet) = U \bullet U^\dagger$ which preserve energy, namely where the unitary operators $U$ commute with the Hamiltonian operator;
- taking the partial trace over a system.

It can be shown that the most general free process on a system A is a quantum channel of this form:

$$\mathcal{C}(\rho_{\mathrm{A}}) = \operatorname{tr}_{\mathrm{E}}\left[U_{\mathrm{AE}}\left(\rho_{\mathrm{A}} \otimes \frac{1}{Z_{\mathrm{E}}}\mathrm{e}^{-\beta H_{\mathrm{E}}}\right)U_{\mathrm{AE}}^\dagger\right], \tag{3.3.3}$$

where $U_{\mathrm{AE}}$ is a unitary operator that commutes with the Hamiltonian operator of system AE, which is $H_{\mathrm{A}} \otimes \mathbf{1}_{\mathrm{E}} + \mathbf{1}_{\mathrm{A}} \otimes H'_{\mathrm{E}}$. From this result, for a given system at a fixed temperature $T$, there is only one free state: the Boltzmann state at temperature $T$. Any other state is a costly resource, even Boltzmann states at a different temperature $T'$. In this way, free resources are states at thermal equilibrium, because one can think that systems evolve naturally towards thermal equilibrium when they are put into contact with a thermal bath, therefore Boltzmann state are particularly easy to obtain. In this vein, states out of thermal equilibrium represent more valuable (or costly) resources, therefore here we have a resource theory of athermality.

According to section 2.2, if $\rho' = \mathcal{C}(\rho)$, where $\mathcal{C}$ is given by eq. (3.3.3), we say that $\rho$ is *more athermal* than $\rho'$, or, equivalently, that $\rho'$ is *more thermal* than $\rho$. Then reason why we can say so is not only mathematical, but also physical. Indeed, a closer look at eq. (3.3.3) enables one to grasp the physical meaning of such a (free) quantum channel. Indeed, if $\rho$ evolves under $\mathcal{C}$, $\rho$ is put into contact with an environment E in a Boltzmann state at temperature $T$, then it evolves according to an energy-preserving unitary channel (e.g. the time evolution itself, as the system + environment can be regarded as an isolated system) and eventually the environment is removed (discarded by a partial trace). This process puts $\rho$ into thermal contact with a bath at temperature $T$, and this will drive $\rho$ closer to the equilibrium state, that is the Boltzmann state. This is why we can say that $\rho' = \mathcal{C}(\rho)$ is more thermal than $\rho$.

*Remark* 3.3.2. Note that the resource theory of purity is a special case of the resource theory of athermality when the Hamiltonian of all systems is trivial, i.e. a multiple of the identity. In this case we have the maximally mixed state in place of the Boltzmann state.

# Chapter 4

# Examples of resource theories in general probabilistic theories

In this chapter explore the full generalization of the methods presented in chapter 2, by trying to export the resource theories of the previous chapter to the framework of GPTs. The principles of chapter 2 will provide guidance for this task. Besides being a mathematical exercise, this is in fact a most important step one should take to gain a deeper insight on those resource theories. Indeed, if entanglement is so central in quantum mechanics, one would like to understand its power even out of the original framework in which it was developed. Such an extension may potentially bring new ideas about how to harness the resource of entanglement even in the quantum domain itself.

The results of this chapter have already been exposed for the first time in ref. [92], a joint work with G. Chiribella, and partially also in [55]. Here we will stress the resource-theoretic perspective.

In order to have sensible and well-behaved resource theories we need to set some axioms on the framework of *causal* GPTs (see chapter 1). This an interesting point, because we see how desirable resource-theoretic properties can lead us in the search of physical theories supporting them. Indeed, suppose we are looking for an extension or a restriction of quantum mechanics that supports a duality between entanglement and mixedness, in the sense of subsection 3.1.2. By identifying the axioms that allow that, we can rule out some candidate theories from the very beginning.

Like in the quantum case, even in GPTs satisfying some reasonable axioms, the resource theories of entanglement and purity will prove to be in-

timately related to each other, therefore providing a good starting point for an information-theoretic axiomatization of thermodynamics.

We will start by examining the resource theory of entanglement in section 4.1, which will not need any additional axioms besides Causality. In section 4.2 we focus on pure-state entanglement, and specifically on trying to identify the axioms that guarantee the validity of Lo-Popescu theorem even in GPTs. In this way, we are able to prove that every LOCC protocol with a pure state as input can simulated by a 1-way LOCC protocol. The axioms we identify, Purity Preservation and Local Exchangeabiliy, are satisfied by a broad class of causal GPTs. In section 4.3 we define the resource theory of purity, but we will see that we must depart from the definition given in quantum theory (cf. section 3.2). In the GPT case, purity will arise as the ability to control the evolution of a state. In section 4.4 we prove the entanglement-mixedness duality, making use of the previously introduced axioms of Purity Preservation and Local Exchangeability, supplemented by the Purification principle. Specifically, Purification characterizes all the theories admitting a fundamental level where all the processes are pure and reversible. In section 4.5 we examine the consequences of this duality as far as the existence of maximally entangled states is concerned. Then we move to the topic of entanglement and purity monotones in section 4.6. Finally, in section 4.7 we see how resource-theoretic properties pose some constraints on the axiom a theory must satisfy. Specifically, we show that the present set of axioms implies a strengthened version of then Purification principle, called *Symmetric Purification*.

## 4.1 The resource theory of entanglement

Entanglement is not a specifically quantum feature, for it emerges in a number of GPTs [25, 93].

Recall that the quantum resource theory is based on LOCC protocols, which are an operational notion and therefore they can be easily exported to arbitrary theories. Again, consider two parties, Alice and Bob, who perform a sequence of local tests, and exchange classical communication between each other. In this way, the outcome of a test can depend on the outcomes of previous tests. Recall that the ability to perform conditioned tests is guaranteed by causality (see section 1.3). To be more concrete, refer again to diagram (3.1.1). There

1. Alice performs a test $\{\mathcal{A}_{i_1}\}$ and communicates her outcome to Bob;
2. Bob performs a test $\left\{\mathcal{B}_{i_2}^{(i_1)}\right\}$ and communicates his outcome to Alice;
3. Alice performs a test $\left\{\mathcal{A}_{i_3}^{(i_1,i_2)}\right\}$.

Again we can define the associated LOCC channel as[1]

$$\mathcal{L} = \sum_{j_1,j_2,j_3} \left[\mathcal{A}_{j_3}^{(j_1,j_2)} \mathcal{A}_{j_1} \otimes \mathcal{B}_{j_2}^{(j_1)}\right].$$

As explained in section 1.3, superscripts in round brackets want to highlight the dependence of a test on the outcome of a previous one.

Like in the quantum case, we define LOCC protocols to be free operations. Free states are therefore those which can be prepared using an LOCC preparation protocol, which are called *separable states*, in analogy with quantum theory. To characterize them, consider Alice performing a preparation-test $\{\rho_i\}$ and Bob performing a preparation-test $\left\{\sigma_j^{(i)}\right\}$, depending on Alice's outcome $i$, which she communicates to Bob. After coarse-graining, since the two parties act locally, one has that the state is $\rho_{\mathrm{AB}} = \sum_{i,j} \rho_{i,\mathrm{A}} \otimes \sigma_{j,\mathrm{B}}^{(i)}$, which, turning to normalized states, becomes $\rho_{\mathrm{AB}} = \sum_{i,j} p_i \rho_{i,\mathrm{A}} \otimes q_j^{(i)} \sigma_{j,\mathrm{B}}^{(i)}$, where the $p_i := \|\rho_i\|_{\mathrm{A}}$ and $q_j^{(i)} := \left\|\sigma_j^{(i)}\right\|_{\mathrm{B}}$. $\rho_{\mathrm{AB}}$ can be rewritten as

$$\rho_{\mathrm{AB}} = \sum_i p_i \rho_{i,\mathrm{A}} \otimes \sum_j q_j^{(i)} \sigma_{j,\mathrm{B}}^{(i)} = \sum_i p_i \rho_{i,\mathrm{A}} \otimes \sigma'_{i,\mathrm{B}},$$

where $\sigma'_{i,\mathrm{B}} := \sum_j q_j^{(i)} \sigma_{j,\mathrm{B}}^{(i)}$. Now we are entitled to give the following definition.

**Definition 4.1.1.** A (normalized) bipartite state $\rho_{\mathrm{AB}} \in \mathsf{St}_1(\mathrm{A} \otimes \mathrm{B})$ is called *separable* if it can be written as $\rho_{\mathrm{AB}} = \sum_i p_i \rho_{i,\mathrm{A}} \otimes \sigma_{i,\mathrm{B}}$, where the $p_i$'s are probabilities, $\rho_{i,\mathrm{A}}$ is a state of system A, and $\sigma_{i,\mathrm{B}}$ is a state of system B, for every $i$.

Entangled states will be our costly resources; as such, they cannot be generated via LOCC protocols.

[1]Here we can choose letter $\mathcal{L}$, to mean LOCC, since there is no possibility of ambiguity with the same letter used to denote bounded operators on a Hilbert space.

**Definition 4.1.2.** A bipartite state $\rho \in \mathsf{St}(\mathrm{A} \otimes \mathrm{B})$ is called *entangled* if it is not separable.

Not all GPTs admit entangled states, classical theory being a counter-example thereof. Therefore, the following treatment will be relevant only to GPTs admitting entangled states, but, as noted above, this does not mean that such GPTs are quantum.

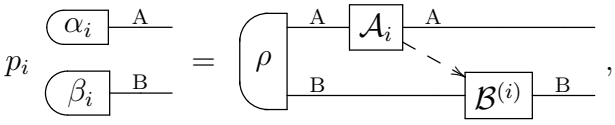
Since LOCC protocols are our free operations, we can use them to establish a hierarchy among bipartite states according to their degree of entanglement, as done in the quantum case.

**Definition 4.1.3.** Given two states $\rho \in \mathsf{St}(\mathrm{A} \otimes \mathrm{B})$ and $\rho' \in \mathsf{St}(\mathrm{A} \otimes \mathrm{B})$, we say that $\rho$ is *more entangled* than $\rho'$, denoted by $\rho \succeq_{\text{ent}} \rho'$, if there exists an LOCC protocol that transforms $\rho$ into $\rho'$, i.e. if $\rho' = \mathcal{L}\rho$ for some LOCC channel $\mathcal{L}$.

We can easily see that every state is more entangled than a separable state.

**Proposition 4.1.4.** *Every bipartite state is more entangled than any separable state.*

*Proof.* The proof is very similar to that of proposition 3.1.3. Consider the separable state $\sigma_{\mathrm{AB}} = \sum_i p_i \alpha_{i,\mathrm{A}} \otimes \beta_{i,\mathrm{B}}$. We can consider the LOCC protocol acting on $\rho \in \mathsf{St}(\mathrm{A} \otimes \mathrm{B})$




where $\mathcal{A}_i$ is a transformation of a test on A, which prepares $\alpha_{i,\mathrm{A}}$ with probability $p_i$ and $\mathcal{B}^{(i)}$ is a channel on B which prepares $\beta_{i,\mathrm{B}}$, depending on the outcome of the test $\{\mathcal{A}_i\}$. In other words, $\mathcal{A}_i = p_i \left|\alpha_i\right\rangle_{\mathrm{A}} \left(\mathrm{tr}\right|_{\mathrm{A}}$ and $\mathcal{B}^{(i)} = \left|\beta_i\right\rangle_{\mathrm{B}} \left(\mathrm{tr}\right|_{\mathrm{B}}$. By taking the coarse-graining over the outcome $i$, we get the separable state $\sigma_{\mathrm{AB}} = \sum_i p_i \alpha_{i,\mathrm{A}} \otimes \beta_{i,\mathrm{B}}$. This proves that $\rho_{\mathrm{AB}}$ is more entangled than $\sigma_{\mathrm{AB}}$. □

Like in the quantum case, and according to the general framework of resource theories, the relation $\succeq_{\text{ent}}$ is a preorder.

**Definition 4.1.5.** If $\rho \succeq_{\text{ent}} \rho'$ and $\rho' \succeq_{\text{ent}} \rho$, then we say that $\rho$ and $\rho'$ are *equally entangled* (or that $\rho$ is as entangled as $\rho'$), denoted by $\rho \sim_{\text{ent}} \rho'$.

Note that $\rho \sim_{\text{ent}} \rho'$ does not imply that $\rho$ and $\rho'$ are equal: for example, every two separable states are equally (un)entangled, as as already noted in the quantum case in section 3.1. Specifically, the equivalence class of separable states is the minimum of the entanglement order.

Similarly, two bipartite states that differ by local reversible channels are equally entangled: $\rho_{\text{AB}} \sim_{\text{ent}} (\mathcal{U}_{\text{A}} \otimes \mathcal{V}_{\text{B}})\, \rho_{\text{AB}}$, for the same reasons as in the quantum case. Is the converse true? Namely, if two bipartite states are equally entangled, can we conclude that they differ by local reversible channels? We will be able to provide an answer to this question later for the *pure* states of a special class of GPTs (see section 4.5).

## 4.2 An operational Lo-Popescu theorem

In this section we want to study under which assumptions we are able to prove the analogue of Lo-Popescu theorem (theorem 3.1.6) for LOCC protocols in GPTs. We would like to show that when we are dealing with pure states, we can always restrict ourselves to 1-way LOCC protocols to study the entanglement relation.

### 4.2.1 Two operational requirements

The first is axiom we are going to set is Purity Preservation.

**Axiom 4.2.1** (Purity Preservation [26, 29, 38])**.** *Sequential and parallel composition of two pure transformations yields a pure transformation.*

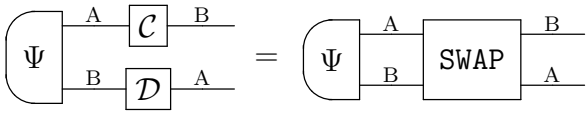
Considering the theory as an algorithm to make deductions about physical processes, Purity Preservation ensures that, when presented with maximal information about two processes, the algorithm outputs maximal information about their composition [38].

Specifically, the product of two pure states is pure, a result that could be proved also using Local Tomography [24, 25, 26, 28, 29], but here we regard Purity Preservation as more fundamental, for Local Tomography is not satisfied by quantum mechanics on a real Hilbert space [94, 53, 95], which instead satisfies Purity Preservation.

Our second requirement imposes a symmetry on pure bipartite states.

**Axiom 4.2.2** (Local Exchangeability)**.** *For every pure bipartite state* $\Psi \in \mathsf{PurSt}(\mathrm{A} \otimes \mathrm{B})$*, there exist two channels* $\mathcal{C} \in \mathsf{Transf}(\mathrm{A}, \mathrm{B})$ *and* $\mathcal{D} \in \mathsf{Transf}(\mathrm{B}, \mathrm{A})$

*such that*




(4.2.1)

*where* SWAP *is the swapping operation, defined in subsection 1.1.3.*

Note that, in general, the two channels $\mathcal{C}$ and $\mathcal{D}$ depend on the specific pure state $\Psi$.

Local Exchangeability is trivially satisfied by classical probability theory, where all pure bipartite states are of the product form. Less trivially, it is satisfied by quantum theory, both on complex and on real Hilbert spaces.

**Example 4.2.3.** Suppose that A and B are quantum systems, and let $\mathcal{H}_{\mathrm{A}}$ and $\mathcal{H}_{\mathrm{B}}$ be their corresponding Hilbert spaces. By Schmidt decomposition, every pure state in the tensor product Hilbert space can be written as

$$|\Psi\rangle_{\mathrm{AB}} = \sum_{j=1}^{r} \sqrt{p_j}\, |\alpha_j\rangle_{\mathrm{A}}\, |\beta_j\rangle_{\mathrm{B}}\,,$$

where $\left\{|\alpha_j\rangle_{\mathrm{A}}\right\}_{j=1}^{r} \subset \mathcal{H}_{\mathrm{A}}$ and $\left\{|\beta_j\rangle_{\mathrm{B}}\right\}_{j=1}^{r} \subset \mathcal{H}_{\mathrm{B}}$ are orthonormal vectors. Then we have

$$\mathtt{SWAP}\, |\Psi\rangle_{\mathrm{AB}} = (C \otimes D)\, |\Psi\rangle_{\mathrm{AB}}$$

where $C := \sum_{j=1}^{r} |\beta_j\rangle_{\mathrm{B}} \langle\alpha_j|_{\mathrm{A}}$ and $D := \sum_{j=1}^{r} |\alpha_j\rangle_{\mathrm{A}} \langle\beta_j|_{\mathrm{B}}$. It is immediate to construct the desired channels $\mathcal{C}$ and $\mathcal{D}$, which can be defined as

$$\mathcal{C}(\rho) := C\rho C^{\dagger} + \sqrt{\mathbf{1}_{\mathrm{A}} - C^{\dagger}C}\,\rho\sqrt{\mathbf{1}_{\mathrm{A}} - C^{\dagger}C},$$

where $\rho$ is a state of system A, and

$$\mathcal{D}(\sigma) := D\sigma D^{\dagger} + \sqrt{\mathbf{1}_{\mathrm{B}} - D^{\dagger}D}\,\sigma\sqrt{\mathbf{1}_{\mathrm{B}} - D^{\dagger}D}.$$

where $\sigma$ is a state of system B. Owing to this definition, one has

$$(\mathcal{C} \otimes \mathcal{D})\left(|\Psi\rangle\langle\Psi|\right) = \mathtt{SWAP}\, |\Psi\rangle\langle\Psi|\, \mathtt{SWAP},$$

which is the Hilbert space version of the Local Exchangeability condition of eq. (4.2.1). Note that everything holds in both the real and the complex case.

There is another example of GPT satisfying Local Exchangeability: it is the theory of *box world* [25, 34], which comes directly from a generalization of the setting of the study of Bell's inequalities [96].

**Example 4.2.4.** To study non-locality and Bell inequalities, the standard approach is to consider two parties, Alice and Bob, who are usually thought to be spacelike separated. Alice and Bob each choose to measure an observable $A$ (resp. $B$) chosen from the set $\{A_x\}_{x\in X}$ (resp. $\{B_y\}_{y\in Y}$), where $x$ (resp. $y$) labels Alice's (resp. Bob's) choice of observable. Suppose the measurement of Alice's observables can yield $a$ outcomes and the measurement of Bob's observables can yield $b$ outcomes. This is the setting in which one obtains CHSH [97] and Cirel'son's [98] inequalities: it is the case when $x$, $y$, $a$, $b$ can take only 2 values, which means Alice and Bob each have a 2-level system (with local hidden variables or genuinely quantum) ($a = b = 2$), and can choose between 2 observables ($A$ and $A'$ for Alice, and $B$ and $B'$ for Bob) to measure. Then they can build the quantity $S = A \otimes B + A' \otimes B + A \otimes B' - A' \otimes B'$ and compute $|\langle S\rangle|$ on the state they share. In a local hidden variable scenario, we have $|\langle S\rangle| \leq 2$ (CHSH inequality), and in the genuinely quantum case $|\langle S\rangle| \leq 2\sqrt{2}$ (Cirel'son's inequality).

Here we take a different approach, for we are interested in the probability $p_{ab|xy}$, which is the probability of getting outcomes $a$ and $b$, given that the choice of observables was $x$ and $y$. In the classical case (or in the case of quantum mechanics with local hidden variables $\lambda$)

$$p_{ab|xy} = \int p_{a|x}(\lambda)\, p_{b|y}(\lambda)\ \mathrm{d}\mathbb{P}_\lambda,$$

where $\mathrm{d}\mathbb{P}_\lambda$ is the probability measure of the hidden variable $\lambda$. Therefore the only influence of Alice to Bob is via the local hidden variable $\lambda$.

In the quantum case, the two parties share a quantum state $\rho$. Let $\{E_a^x\}$ be the spectral projective measurement of Alice's observable $A_x$, and let $\{E_b^y\}$ be the spectral projective measurement of Bob's observable $B_y$. Now, $p_{ab|xy}$ is given by

$$p_{ab|xy} = \mathrm{tr}\left(E_a^x \otimes E_b^y \rho\right).$$

Now we wish to explore the issue in full generality. We want to impose only some constraints because we do not want to violate causality by allowing the case when Alice can communicate Bob a message instantaneously, a fact known as *signalling* from Alice to Bob. Suppose Alice performs her measurement first, and wants to communicate her choice of observable to Bob, namely $x$. Bob's goal is to retrieve Alice's message $x$. To do that, he chooses one of his observables, and he measures it. Only from the statistics of his measurement results would Bob like to recover $x$. We want to rule out this

case, for it would imply the message from Alice is transmitted instantaneously to Bob. Bob's output probability must be independent from Alice's choice of observable irrespective of her measurement result. In symbols,

$$\sum_a p_{ab|xy} = \sum_a p_{ab|x'y} \tag{4.2.2}$$

for all $x$, $x'$, $y$ and $b$. Here we are summing over all possible outcomes of Alice's measurement of her chosen observable. Indeed Bob does not know Alice's outcome of her measurement, nor is he interested in it. He cares only about her choice of observable, not about the outcome of her measurement. Eq. (4.2.2) states that Bob cannot infer anything about $x$ because the $\sum_a p_{ab|xy}$ is in fact independent of $x$, therefore we can rename it as $p_{b|y} := \sum_a p_{ab|xy}$. Clearly, by symmetry, a similar constraint must be imposed also to avoid signalling from Alice to Bob, which gives rise to the condition

$$\sum_b p_{ab|xy} = \sum_b p_{ab|xy'}$$

for all $x$, $y$, $y'$ and $a$. Therefore, the most general constraints on $p_{ab|xy}$ so that it is a physically sensible probability distribution are

1. $p_{ab|xy} \geq 0$ for all $a$, $b$, $x$, $y$;
2. $\sum_{a,b} p_{ab|xy} = 1$ for all $x$, $y$;
3. $\sum_a p_{ab|xy} = \sum_a p_{ab|x'y}$ for all $x$, $x'$, $y$, $b$;
4. $\sum_b p_{ab|xy} = \sum_b p_{ab|xy'}$ for all $x$, $y$, $y'$, $a$.

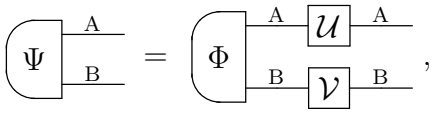
Conditions 1 and 2 simply guarantee that $p_{ab|xy}$ is a probability distribution, and conditions 3 and 4 are termed *no-signalling conditions*, and exclude signalling. Quite surprisingly, these constraints are not enough to single out quantum mechanics or classical theory, but they define a GPT, called *box world*, which possesses stronger non-local correlation than quantum theory. In this GPT, states of a bipartite system $\mathrm{A} \otimes \mathrm{B}$ are given by no-signalling probability distributions $p_{ab|xy}$ [25, 34].

Now we are interested in pure states, because we want to check if box world satisfies Local Exchangeability. Not all possible cases of box world have been completely characterized in the literature, but we will show that Local Exchangeability holds for the cases studied so far.

The first case is when $x, y, a, b \in \{0, 1\}$, which means Alice and Bob each have 2 observables ($x, y \in \{0, 1\}$), each of them having only 2 possible outcomes ($a, b \in \{0, 1\}$). In this case Alice's and Bob's systems are operationally equivalent, with the reversible channel $\mathcal{I} \in \mathsf{Transf}(\mathrm{A}, \mathrm{B})$ implementing the operational equivalence[2]. Pure states were completely characterized in ref. [99], and shown to be equal to the standard state [100]

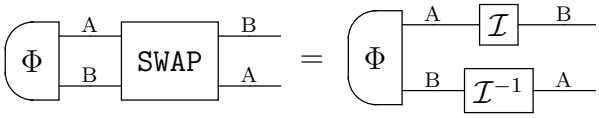
$$p_{ab|xy} = \begin{cases} \frac{1}{2} & a + b \equiv xy \mod 2 \\ 0 & \text{otherwise} \end{cases}, \tag{4.2.3}$$

up to exchange $0 \leftrightarrow 1$ in $x$, $y$, $a$, and $b$. Let $\Phi$ be the standard state; all pure states $\Psi$ can be obtained by performing local reversible channels on $\Phi$, because such exchanges $0 \leftrightarrow 1$ are implemented locally and are clearly reversible. Using diagrams, one has



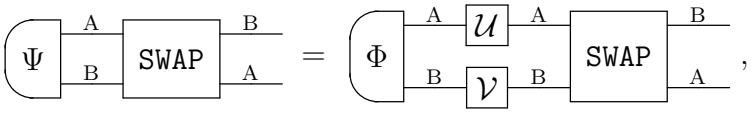

, (4.2.4)

where $\mathcal{U}$ and $\mathcal{V}$ are reversible channels. To see that Local Exchangeability holds, note that swapping systems A and B is equivalent to exchanging $x$ with $y$ and $a$ with $b$. Now, the standard correlation of eq. (4.2.3) is invariant under exchange $x \leftrightarrow y$, $a \leftrightarrow b$, meaning that one has



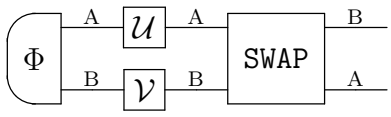

, (4.2.5)

where $\mathcal{I}$ is the reversible channel defined above. Then it is not hard to prove that every pure state of $\mathrm{A} \otimes \mathrm{B}$ can be swapped by local operations. Indeed, one has

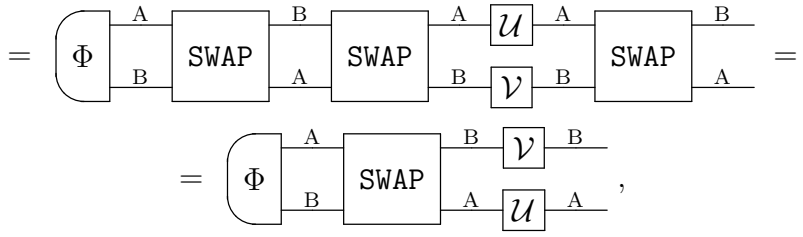



,

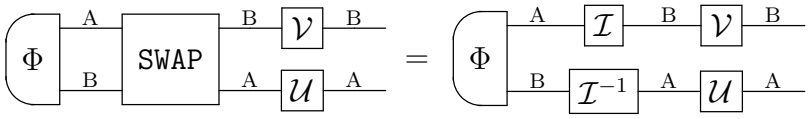
according to eq. (4.2.4). Now,



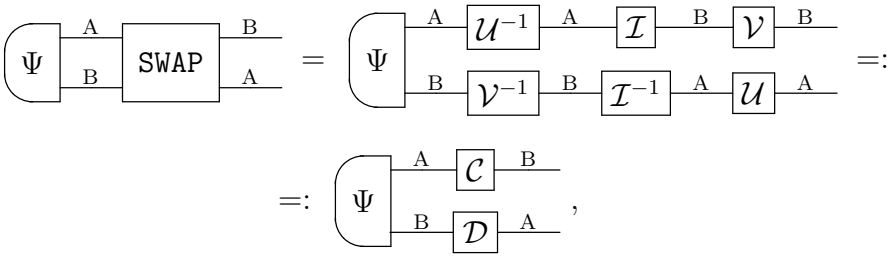

[2]Beware of not confusing $\mathcal{I}$ with the identity channel!




where we made use of the property of the swapping operation (see subsection 1.1.3). Recalling eq. (4.2.5), we have




Again by eq. (4.2.4) we conclude




where $\mathcal{C} := \mathcal{V}\mathcal{I}\mathcal{U}^{-1}$ and $\mathcal{D} := \mathcal{U}\mathcal{I}^{-1}\mathcal{V}^{-1}$. This proves the Local Exchangeability property for all pure bipartite states in the present scenario.

The next case is when we have $x, y \in \{0, 1\}$ (Alice and Bob each have 2 observables), but $a$ and $b$ can take $d_a$ and $d_b$ values respectively (this means that Alice has a $d_a$-level system and Bob has a $d_b$-level system). In this setting pure states have been completely characterized in ref. [101]. Up to local reversible transformations, they are labelled by a parameter $k \in \{2, \ldots, \min\{d_a, d_b\}\}$, and they are such that

$$p_{ab|xy} = \begin{cases} \frac{1}{k} & b - a \equiv xy \mod k \\ 0 & \text{otherwise} \end{cases} . \tag{4.2.6}$$

Thanks to the local equivalence (see the argument above), it is enough to prove the validity of Local Exchangeability for correlations in the standard form of eq. (4.2.6). We distinguish between the two cases $xy = 0$ and $xy = 1$. For $xy = 0$, swapping $x$ with $y$ and $a$ with $b$ has no effect on $p_{ab|xy}$, because

$b-a \equiv 0 \mod k$ if and only if $a-b \equiv 0 \mod k$. For $xy=1$, after swapping $x$ with $y$ and $a$ with $b$, one obtains the probability distribution

$$p'_{ab|xy} = \begin{cases} \frac{1}{k} & a-b \equiv 1 \mod k \\ 0 & \text{otherwise} \end{cases}.$$

This probability distribution can be obtained from the original one by relabelling the outputs as $a' := k-a$ and $b' := k-b$. Such a relabelling corresponds to local reversible operations on A and B. In other words, Local Exchangeability holds.

Finally, the last class of extreme non-local correlations characterized in the literature corresponds to the case of arbitrary number of observables, but Alice and Bob are restricted to 2-level systems. In this case, the extreme correlations were characterized explicitly in ref. [96]. A quick check at the table presented there shows that, up to local reversible transformations, the pure states are invariant under swap. Hence, the same argument used above for the simplest case ($x, y, a, b \in \{0, 1\}$) shows that Local Exchangeability holds. We conclude that for the specific instances of box world described in the literature so far, Local Exchangeability is valid.

### 4.2.2 Inverting the direction of classical communication

Purity Preservation and Local Exchangeability have an important consequence. For 1-way LOCC protocols acting on a pure input state, the direction of classical communication is irrelevant: every 1-way LOCC protocol with communication from Alice to Bob can be replaced by a 1-way LOCC protocol with communication from Bob to Alice. However, first we need an assumption which will prove inessential later, when we introduce the axiom of Purification, which implies it.

**Definition 4.2.5.** A test is said *purely decomposable* if it admits a refinement consisting only of pure transformations.

In principle, not all tests are purely decomposable. Let us see why. If a transformation $\mathcal{A}$ is *not* pure, it can be written as $\mathcal{A} = \sum_i \mathcal{A}_i$, where the $\mathcal{A}_i$'s may not be pure. Those $\mathcal{A}_i$'s which are not pure, can be further decomposed as a sum of other transformations, which, again, may not be pure. In principle, the process of finding refinements can go on indefinitely, but for purely decomposable tests it must come to an end sooner or later.

**Assumption 4.2.6** (Pure Decomposition)**.** *Every test in a theory is purely decomposable.*

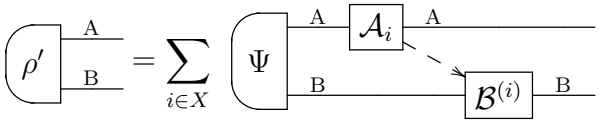
Now we can prove the lemma concerning the inversion of classical communication in a 1-way LOCC protocol.

**Lemma 4.2.7** (Inverting classical communication)**.** *Let* $\Psi$ *be a* pure *state of* $\mathrm{A} \otimes \mathrm{B}$ *and let* $\rho'$ *be a (possibly mixed) state of the same system. Under the validity of Purity Preservation, Local Exchangeabilty and Pure Decomposition, the following are equivalent.*

1. $\Psi$ *can be transformed into* $\rho'$ *by a 1-way LOCC protocol with communication from Alice to Bob.*

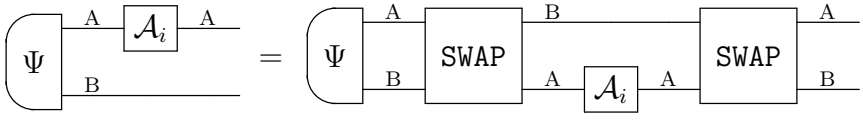
2. $\Psi$ *can be transformed into* $\rho'$ *by a 1-way LOCC protocol with communication from Bob to Alice.*

*Proof.* Clearly it is enough to prove one implication, since the other will be identical. Suppose $\Psi$ can be transformed into $\rho'$ by a 1-way LOCC protocol with communication from Alice to Bob, namely




(4.2.7)

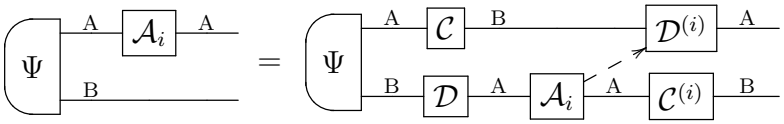
where $\{\mathcal{A}_i\}_{i\in X}$ is a test, and, for every outcome $i \in X$, $\mathcal{B}^{(i)}$ is a channel. Note that one can assume without loss of generality that all transformations $\mathcal{A}_i$'s are pure: if the transformations were not pure, we could refine them by Pure Decomposition, and apply the argument to the refined test consisting of pure transformations. For every fixed $i \in X$, one has



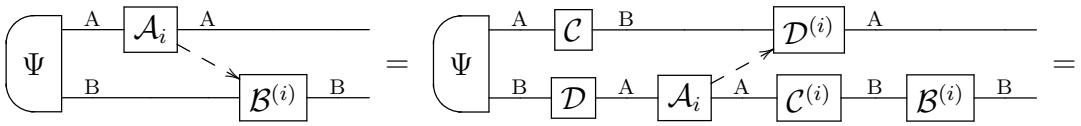

(4.2.8)

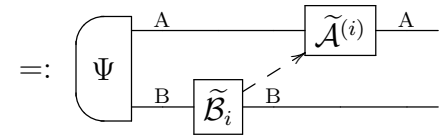
By Local Exchangeability, the first swap can be realized by two local channels $\mathcal{C} \in \mathsf{Transf}(\mathrm{A}, \mathrm{B})$ and $\mathcal{D} \in \mathsf{Transf}(\mathrm{B}, \mathrm{A})$. Moreover, since $\mathcal{A}_i$ is pure, Purity Preservation implies that the (unnormalized) state $(\mathcal{A}_i \otimes \mathcal{I}_\mathrm{B})\,\Psi$ is pure. Hence, also the second swap in eq. (4.2.8) can be realized by two local channels $\mathcal{C}^{(i)} \in \mathsf{Transf}(\mathrm{A}, \mathrm{B})$ and $\mathcal{D}^{(i)} \in \mathsf{Transf}(\mathrm{B}, \mathrm{A})$. Substituting into eq. (4.2.8)

one obtains



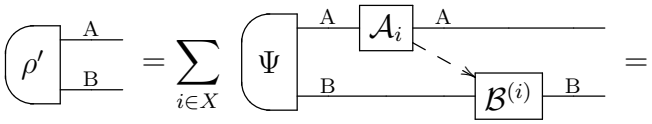

and therefore,

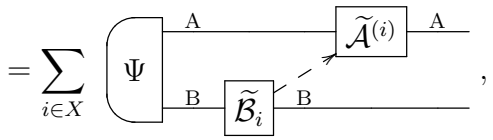



(4.2.9)

having defined $\mathcal{A}^{(i)} := \mathcal{D}^{(i)}\mathcal{C}$ and $\widetilde{\mathcal{B}}_i = \mathcal{B}^{(i)}\mathcal{C}^{(i)}\mathcal{A}_i\mathcal{D}$. By construction $\left\{\widetilde{\mathcal{B}}_i\right\}_{i\in X}$ is a test, because it can be realized by performing the test $\{\mathcal{A}_i\}_{i\in X}$ after the channel $\mathcal{D}$ and subsequently applying the channel $\mathcal{B}^{(i)}\mathcal{C}^{(i)}$, depending on the outcome. On the other hand, $\widetilde{\mathcal{A}}^{(i)}$ is a channel for every $i \in X$. Hence, we have constructed a 1-way LOCC protocol with communication from Bob to Alice. Combining eqs. (4.2.7) and (4.2.9) we obtain




meaning that $\Psi$ can be transformed into $\rho'$ by a 1-way LOCC protocol with communication from Bob to Alice. □

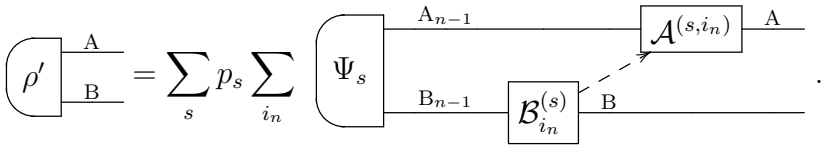
Note that the target state $\rho'$ need not be pure: the fact that the direction of classical communication can be inverted relies only on the purity of the *input* state $\Psi$.

### 4.2.3 Reduction to 1-way LOCC protocols

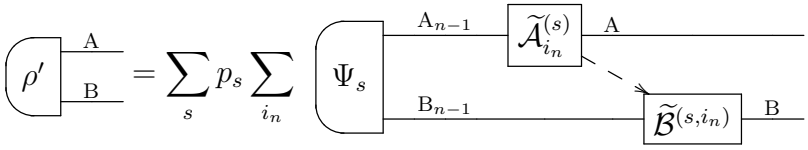
We are now ready to derive the operational version of the Lo-Popescu theorem. Our result shows that the action of an arbitrary LOCC protocol on a pure state can be simulated by a 1-way LOCC protocol.

**Theorem 4.2.8** (Operational Lo-Popescu theorem)**.** *Let $\Psi$ be a pure state of $\mathrm{A}\otimes\mathrm{B}$ and let $\rho'$ be a (possibly mixed) state of the same system. Under the validity of axioms assumed so far, if $\Psi$ can be transformed into $\rho'$ by an LOCC protocol, the same transformation can be achieved by a* 1-way *LOCC protocol.*

*Proof.* Suppose that $\Psi$ can be transformed into $\rho'$ by an LOCC protocol with $n$ rounds of classical communication. We will show by induction over $n$ that every such protocol can be reduced to a 1-way protocol. If $n = 1$ there is nothing to prove. Suppose the thesis holds for $n-1$ and let us prove it holds also for $n$. Thanks to lemma 4.2.7, the party who starts the rounds of classical communication is irrelevant. Therefore, without loss of generality, we assume that Alice starts the protocol and that all transformations occurring in the first $n-1$ rounds are pure. Let $s = (i_1, \dots, i_{N-1})$ be the sequence of all classical outcomes obtained by Alice and Bob up to step $n-1$, let $p_s$ be the probability of getting the sequence $s$, and let $\Psi_s$ be the pure state after step $n-1$, conditional on the occurrence of $s$. For concreteness, suppose that the outcome $i_{n-1}$ has been generated on Alice's side. Then, the rest of the protocol consists in a test $\left\{\mathcal{B}_{i_n}^{(s)}\right\}$, performed on Bob's side, followed by a channel $\mathcal{A}^{(s,i_n)}$ performed on Alice's side. By definition, one has




Now, using lemma 4.2.7 one can invert the direction of the classical communication in the last round, obtaining




for a suitable test $\left\{\widetilde{\mathcal{A}}_{i_n}^{(s)}\right\}$ and suitable channels $\widetilde{\mathcal{B}}^{(s,i_n)}$. Now, since both the $(n-1)$-th and the $n$-th tests are performed by Alice, they can be merged

into a single test, thus reducing the original LOCC protocol to an LOCC protocol with $n-1$ rounds. Since we know that the thesis holds for $n-1$, we conclude the proof. □

## 4.3 The resource theory of purity

In this section we explore one of the possible ways to generalize the resource theory of purity defined in the framework of quantum mechanics. This will turn out to be a non-trivial task. A first attempt to generalize the quantum resource theory of purity is to try to use the same generating set as in the quantum case, which means reversible channels, deterministic effects and something that could be interpreted as a sort of maximally mixed state. In fact, this constitutes the first obstacle. Indeed, first of all it is not even clear what we mean by "maximally mixed state". Since in quantum mechanics the maximally mixed state is also invariant under all unitary channels, we may translate this into operational language by considering an invariant state $\chi$ under all reversible channels. We can prove (see subsection 4.3.3) that such a state exists and it is unique for every system A. However, we encounter an obstacle of mathematical nature. Section 2.3 requires that the set of free states is closed under parallel composition. It is not at all clear whether the product of two invariant states is still an invariant state, namely in general $\chi_{\mathrm{A}} \otimes \chi_{\mathrm{B}} \neq \chi_{\mathrm{AB}}$. Therefore, if we assume $\chi_{\mathrm{A}}$ to be free for every system A, also $\chi_{\mathrm{A}} \otimes \chi_{\mathrm{B}}$ is free, but it will not be, in general, the invariant state of system $\mathrm{A} \otimes \mathrm{B}$. This would generate a lot of free states, therefore we prefer to switch to another approach. Indeed, we proved that in quantum mechanics the preorder given by noisy operations and the preorder given by RU channels are the same. For this reason, we will choose the latter approach to achieve a generalization of the resource theory of purity. The starting point will be the idea of controlling the evolution of states.

### 4.3.1 A resource theory of dynamical control

Consider the situation where a closed system A undergoes a reversible dynamics governed by some parameters under the experimenter's control. For example, system A could be a charged particle in an electric field, which can be controlled in order to obtain a desired trajectory. In general, the experimenter may not have full control, and the actual values of the para-

meters may fluctuate randomly. As a result, the evolution of the system will be described by a *Random Reversible* (*RaRe*) *channel*, that is a channel $\mathcal{R} \in \mathsf{Transf}(\mathrm{A})$ of the form $\mathcal{R} = \sum_{i\in X} p_i \mathcal{U}_i$ where the $\mathcal{U}_i$'s are reversible channels on A, and $\{p_i\}_{i\in X}$ is their probability distribution[3]. Assuming that the system remains closed during the whole evolution, RaRe channels are the most general transformations the experimenter can implement. This will be our free operations. This implies that there are no free preparation processes[4], because the free processes (RaRe channels) having the trivial system as input will have also the trivial system as output. As a consequence, there are *no* free states. Physically, this is in agreement with the fact that the input state in a control problem is not chosen by the experimenter.

According to the general mathematical framework of resource theories, RaRe channels define a preorder on the set of states.

**Definition 4.3.1.** Given two states $\rho$ and $\rho'$ of system A, we say that $\rho$ is *more controllable* than $\rho'$, denoted by $\rho \succeq \rho'$, if $\rho'$ can be obtained from $\rho$ via a RaRe channel.

This definition is sensible, because an important question in all problems of control is whether a given input state can be driven to a target state using the allowed dynamics. With respect to this task, an input state is more valuable than another if the set of target states that can be reached from the former contains the set of target states that can be reached from the latter.

Definition 4.3.1 appeared independently in ref. [102], albeit in a completely different context.

**Definition 4.3.2.** We say that two states $\rho$ and $\rho'$ of system A are equally controllable (or that $\rho$ is *as controllable as* $\rho'$) if $\rho \succeq \rho'$ and $\rho' \succeq \rho$. In this case we write $\rho \sim \rho'$.

The relation $\succeq$ is not an order, for there are different states that are equally controllable: take for instance $\rho$ and $\rho' = \mathcal{U}\rho$, where $\mathcal{U}$ is a reversible

[3]Henceforth, for the convenience of notation, we will sometimes disregard the previous notation in which conditioned transformations are denoted by superscripts in round brackets (cf. definition 1.3.8). Indeed RaRe channels are given by randomization, which is a particular instance of conditioning (cf. definition 1.3.9), over reversible channels, and therefore they should be written as $\mathcal{R} = \sum_{i\in X} p_i \mathcal{U}^{(i)}$. However, not using superscripts will be particularly useful when taking the inverse of reversible channel, so that the notation is not burdened (and maybe even confused) by too many superscripts.

[4]This contrasts with the definition of quantum resource theory of purity in section 3.2, where the maximally mixed state was assumed to be free.

channel different from the identity. In [102] the authors proved by using some results of convex geometry that this is the only possible case: if $\rho \sim \rho'$ there must exist a reversible channel $\mathcal{U}$ such that $\rho' = \mathcal{U}\rho$. In other words, two states that are equally controllable can only differ by a reversible transformation.

### 4.3.2 From dynamical control to purity

There is a close relation between the controllability of a state and its purity. If a state is more controllable than a pure state, it must be pure.

**Proposition 4.3.3.** *If $\alpha \in \mathsf{St}(\mathrm{A})$ is a* pure *state and $\rho$ is more controllable than $\alpha$, then $\rho$ must be pure. Specifically, $\rho = \mathcal{U}\alpha$ for some reversible channel $\mathcal{U}$.*

*Proof.* Since $\alpha$ is pure, the condition $\sum_i p_i \mathcal{U}_i \rho = \alpha$ implies that $\mathcal{U}_i \rho = \alpha$ for every $i$. By lemma 1.2.6, $\rho$ is pure. □

In other words, pure states can be reached only from pure states. A natural question is whether we can reach every state from pure states. The answer is positive in quantum theory and in a large class of theories. Nevertheless, counterexamples exist that prevent an easy identification of the resource theory of dynamical control with a resource theory of purity. This fact is illustrated in the following example.

**Example 4.3.4.** Consider a system with the state space depicted in fig. 4.1a. In this case, there are only two reversible transformations, namely the identity and the reflection around the vertical symmetry axis. As a consequence, there is no way to obtain the mixed states on the two vertical sides by applying a RaRe channel to a pure state. These states represent a valuable resource, even though they are not pure. Since some mixed states are a resource, the resource theory of dynamical control cannot be regarded as a resource theory of purity.

As a second example, consider instead a system whose state space is a half-disk, like in fig. 4.1b. Also in this case there are only two reversible transformations (the identity and the reflection around the vertical axis). However, now every mixed state can be generated from some pure state via a RaRe channel. The state space can be divided into horizontal segments generated by pure states under the action of RaRe channels. As a result,

pure states are the most useful resources and one can interpret the relation $\succeq$ as a way to compare the degree of purity of different states. Nevertheless, pure states on different segments are inequivalent resources, therefore states on different segments cannot be compared. In this case there are several inequivalent classes of pure states: purity is not the only relevant resource into play.

Finally, consider a system with a square state space, like in fig. 4.1c. Now the symmetries are all the transformations in the dihedral group $D_4$. In this case, all the pure states are equivalent under reversible transformations and every mixed state can be obtained by applying a RaRe channel to a fixed pure state. Here, the resource theory of dynamical control becomes a full-fledged resource theory of purity.

The above examples show that not every GPT allows for a sensible resource theory of purity. Motivated by the examples, we give the following definition.

**Definition 4.3.5.** A *resource theory of purity* is a resource theory of dynamical control where every state $\rho$ can be generated from some pure state.

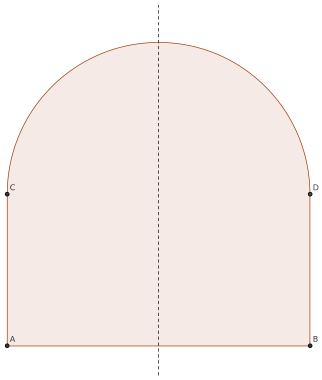
The theory is called *canonical* if every pure state is more controllable than any state.

Henceforth we will mainly focus on canonical theories of purity. As a consequence of the definition, in a canonical theory of purity, all pure states are equivalent to each other as resources. We can give a characterization of canonical theories of purity as follows.

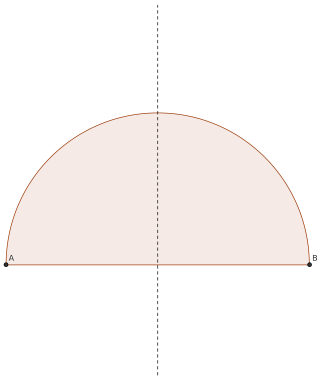
**Proposition 4.3.6.** *The following are equivalent.*

1. *The resource theory is a canonical theory of purity.*
2. *For every pair of* pure *states $\psi$, $\psi'$ there exists a reversible channel $\mathcal{U}$ such that $\psi = \mathcal{U}\psi'$.*
3. *For every system, there exists a state that is more controllable than every state.*

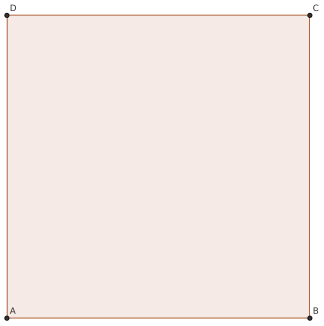
*Proof.* 1 implies 2. If the theory is canonical, the pure state $\psi$ is more controllable than the pure state $\psi'$. By proposition 4.3.3, there is a reversible channel such that $\psi = \mathcal{U}\psi'$.




(a) Example of state space leading to a theory of dynamical control that cannot be interpreted as a resource theory of purity. Owing to the shape of the state space, the only reversible transformations are the identity and the reflection around the vertical axis. The states on the vertical sides cannot be reached from pure states via RaRe channels.




(b) Example of state space leading to a non-canonical theory of purity. In this case, a generic state can be generated from some pure state through a RaRe. However, not all pure states are equally controllable.




(c) Example of state space compatible with a canonical theory of purity. Here the set of maximally controllable states coincides with the set of pure states, and, in addition, all pure states are equivalent resources.

Figure 4.1: Three different examples of theories of dynamical control.

2 implies 3. Every state $\rho$ can be expressed as a convex combination of the form $\rho = \sum_i p_i \varphi_i$, where $\{p_i\}$ is a probability distribution and the $\varphi_i$'s are pure states. Now, consider a pure state $\psi$. For every $i$, by picking a reversible channel $\mathcal{U}_i$ such that $\mathcal{U}_i \psi = \varphi_i$, one obtains the relation $\rho = \sum_i p_i \mathcal{U}_i \psi$, meaning that $\psi$ is more controllable than $\rho$. Since $\rho$ is generic, we conclude that $\psi$ is more controllable than every state.

3 implies 1. Suppose there exists a state $\rho$ that is more controllable than every state. Specifically, $\rho$ must be more controllable than every pure state $\psi$. By proposition 4.3.3, $\rho$ is pure and there exists a reversible channel $\mathcal{U}$ such that $\rho = \mathcal{U}\psi$. This shows that $\psi$ is more controllable than $\rho$, which, in turn, is more controllable than any state. Hence $\psi$ is more controllable than every state. Since $\psi$ is generic, the theory is canonical. □

Statement 2 of proposition 4.3.6, which expresses the transitivity of the action of the group of reversible channels on the set of pure states, has been widely used in axiomatic reconstruction of quantum theory [24, 103, 104, 105, 106]. Now proposition 4.3.6 sheds a new light on the transitivity requirement, identifying it as a necessary and sufficient condition for a canonical resource theory of purity, and, ultimately, for a well-behaved thermodynamics.

In a canonical theory of purity, we are entitled to identify controllability with purity.

**Definition 4.3.7.** In a canonical theory of purity we say that $\rho$ is *purer* than $\rho'$ if $\rho \succeq \rho'$ and we adopt the notation $\rho \succeq_{\text{pur}} \rho'$.

In this case, we also say that $\rho'$ is *more mixed* than $\rho$, denoted by $\rho' \succeq_{\text{mix}} \rho$.

When $\rho \succeq_{\text{mix}} \rho'$ and $\rho' \succeq_{\text{mix}} \rho$ we say that $\rho$ and $\rho'$ are *equally mixed*, denoted by $\rho \sim_{\text{mix}} \rho'$.

Note the mixedness relation is not a preorder arising in the strict sense of resource theories as presented in section 2.2. Indeed, if $\rho$ is more mixed than $\rho'$ the free process (RaRe channel) goes from $\rho'$ to $\rho$, and not from $\rho$ to $\rho'$ as it should if it were a preorder generated by the consumption of resources by free processes. Nevertheless, we want to focus on the mixedness preorder because it will turn out to be intimately related to the entanglement preorder, as we will show in section 4.4.

### 4.3.3 Maximally mixed states

We want to study maximal elements of the mixedness relation.

**Definition 4.3.8.** We say that a state $\chi \in \mathsf{St}(\mathrm{A})$ is *maximally mixed* if for any state $\rho$ such that $\rho \succeq_{\mathrm{mix}} \chi$ we have $\rho = \chi$.

Maximally mixed states can be characterized as the states that are invariant under all reversible channels.

**Proposition 4.3.9.** *A state $\chi \in \mathsf{St}(\mathrm{A})$ is maximally mixed if and only if it is invariant, i.e. if and only if $\chi = \mathcal{U}\chi$ for every reversible channel $\mathcal{U}$ on system* A.

*Proof.* Sufficiency. Suppose $\chi$ is invariant. If $\rho$ is more mixed than $\chi$, then $\rho = \sum_i p_i \mathcal{U}_i \chi$, for some probabilities $p_i$'s and some reversible channels $\mathcal{U}_i$'s. Since $\chi$ is invariant, we have

$$\rho = \sum_i p_i \mathcal{U}_i \chi = \sum_i p_i \chi = \chi.$$

Necessity. Suppose $\chi$ is maximally mixed. Clearly, $\mathcal{U}\chi$ is more mixed than $\chi$ for any reversible channel $\mathcal{U}$. Hence we conclude $\mathcal{U}\chi = \chi$ because $\chi$ is maximally mixed, so $\chi$ is invariant. □

For finite-dimensional canonical theories, the maximally mixed state exists and is unique under the standard assumption of compactness of the state space [28, 104]. In this case, it can be shown that the group of reversible channels is locally compact and therefore it admits a Haar measure [107]. Then the invariant state is defined as

$$\chi = \int_{\mathbf{G}} \mathcal{U}\rho \, \mathrm{d}\mathcal{U}, \tag{4.3.1}$$

where $\mathbf{G}$ is the group of reversible channels, $\mathrm{d}\mathcal{U}$ is the associated normalized Haar measure, and $\rho$ is a generic state. By Carathéodory's theorem [108, 109], eq. (4.3.1) can be rewritten as a finite sum

$$\chi = \sum_i p_i \mathcal{U}_i \rho,$$

thus proving that $\chi$ is more mixed than any state $\rho$. This shows that $\chi$ is not only maximally mixed, but also the maximum of the mixedness relation.

**Example 4.3.10.** In quantum theory, the invariant and maximally mixed state exists only for finite-dimensional systems ($\mathcal{H} \approx \mathbb{C}^n$), and it is $\chi = \frac{1}{n}\mathbf{1}$.

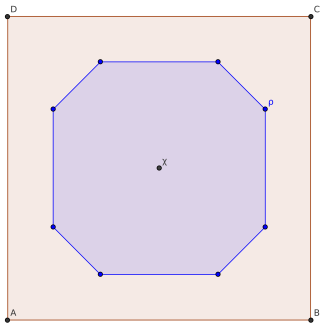



Figure 4.2: Mixedness relation for the state space of a square bit: the vertices of the blue octagon represent the states that can be reached from a given state $\rho$ via reversible transformations. Their convex hull is the set of states that are more mixed than $\rho$. Note that it contains the invariant state $\chi$, which is the maximally mixed state.

Another example is provided by the so-called square bit [25].

**Example 4.3.11.** Consider a system whose state space is a square, as in fig. 4.1c and pick a generic (mixed) state $\rho$. The states that are more mixed than $\rho$ are obtained by applying all possible reversible transformations to $\rho$ (i.e. all the elements of the dihedral group $D_4$) and by taking the convex hull of the orbit. The set of all states that are more mixed than $\rho$ is an octagon, depicted in blue in fig. 4.2. All the vertexes of the octagon are equally mixed. The centre of the square is the maximally mixed state $\chi$, the unique invariant state of the system.

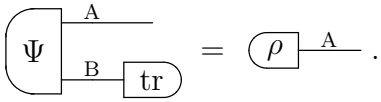
## 4.4 Entanglement-thermodynamics duality

In the previous chapter, we saw that in quantum theory the ordering of pure bipartite states according to the degree of entanglement is equivalent to the ordering of their marginals according to the degree of mixedness. In this section we will prove the validity of this equivalence based only on first principles also in the abstract case of GPTs.

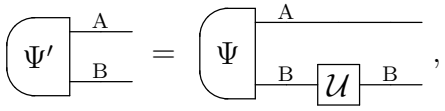
### 4.4.1 Purification

In order to establish the desired duality, we consider theories satisfying Purification [28, 29]. Let us briefly summarize the content of the principle.

**Definition 4.4.1.** We say that a state $\rho \in \mathsf{St}(\mathrm{A})$ has a *purification* if there exists a system B and a pure state $\Psi \in \mathsf{PurSt}(\mathrm{A} \otimes \mathrm{B})$ (the purification) such that




We say that the purification is *essentially unique* if every other purification $\Psi'$ with the same purifying system B satisfies the condition




for some reversible channel $\mathcal{U}$ on the purifying system B.

*Remark* 4.4.2. Note that the concept of purification is stronger than the concept of *extension*. We say that a state $\Sigma \in \mathsf{St}(\mathrm{A} \otimes \mathrm{B})$ is an *extension* of $\rho \in \mathsf{St}(\mathrm{A})$ if $\rho = \mathrm{tr}_{\mathrm{B}}\Sigma$. There is *no* requirement that $\Sigma$ is pure. We can regard a purification as a *pure* extension of a state. Clearly, every state $\rho \in \mathsf{St}(\mathrm{A})$ can be extended; indeed $\rho \otimes \sigma \in \mathsf{St}(\mathrm{A} \otimes \mathrm{B})$, where $\sigma$ is any state of system B, is an extension of $\rho$. Instead, it is not obvious that every state admits a purification. This is precisely the statement of the Purification Principle.

With these definitions, the Purification Principle can be phrased as follows.

**Axiom 4.4.3** (Purification [28, 29])**.** *Every state has a purification, and every purification is essentially unique.*

Purification has a number of important consequences. Here we list only those that will be useful in the following.

First of all, it implies that the group of reversible transformations acts transitively on the set of pure states.

**Proposition 4.4.4.** *For every system* A *and every pair of pure states* $\psi, \psi' \in \mathsf{PurSt}(\mathrm{A})$ *there exists a reversible channel* $\mathcal{U}$ *on* A *such that* $\psi' = \mathcal{U}\psi$.

*Proof.* See lemma 20 of ref. [28]. □

Since all pure states are equivalent under reversible transformations, every theory with purification gives rise to a canonical resource theory of purity (see definition 4.3.5). One may take this fact as a further indication that Purification is a good starting point for a well-behaved thermodynamics.

Another important consequence of Purification is the existence of entangled pure states.

**Proposition 4.4.5.** *Let* $\Psi \in \mathsf{PurSt}(\mathrm{A} \otimes \mathrm{B})$ *be a pure bipartite state. Under the present set of axioms, the following are equivalent.*

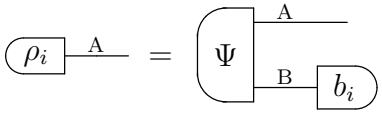
1. $\Psi$ *is entangled.*
2. *Its marginal* $\rho_{\mathrm{A}}$ *on* A *is mixed.*
3. *Its marginal* $\rho_{\mathrm{B}}$ *on* B *is mixed.*

*Proof.* By symmetry, it is enough to prove the equivalence between statement 1 and statement 2. 1 implies 2. Suppose by contradiction that the marginal of $\Psi$ on system A is pure and denote it by $\alpha$. Then, for every pure state $\beta' \in \mathsf{PurSt}(\mathrm{B})$, the product state $\Psi' = \alpha \otimes \beta'$ is pure, thanks to Purity Preservation. Now, $\Psi$ and $\Psi'$ are two purifications of $\alpha$. By the essential uniqueness of purification, one must have $\Psi = (\mathcal{I}_{\mathrm{A}} \otimes \mathcal{U}_{\mathrm{B}})\,\Psi'$ for some reversible transformation $\mathcal{U}_{\mathrm{B}}$ acting on system B. Hence, we have $\Psi = \alpha \otimes \beta$, with $\beta = \mathcal{U}_{\mathrm{B}}\beta'$, in contradiction with the hypothesis that $\Psi$ is entangled.

2 implies 1. Suppose by contradiction that $\Psi$ is separable. Since it is pure, it must be a product of two pure states, say $\Psi = \alpha \otimes \beta$. Clearly, this implies that the marginal on system A is pure, in contradiction with the hypothesis. □

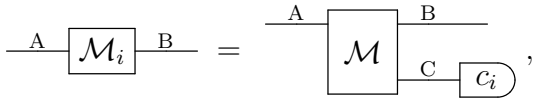
Finally, Purification implies the *steering property* [110, 111], stating that every ensemble decomposition of a given state can be generated by a measurement on the purifying system.

**Theorem 4.4.6** (Steering)**.** *Let* $\rho$ *be a state of system* A *and let* $\Psi \in \mathsf{PurSt}(\mathrm{A} \otimes \mathrm{B})$ *be a purification of* $\rho$*. For every ensemble of states* $\{\rho_i\}_{i\in X}$ *such that* $\sum_i \rho_i = \rho$*, there exists a measurement* $\{b_i\}_{i\in X}$ *on the purifying system* B *such that*



*for all* $i \in X$.

*Proof.* The proof follows the same lines of theorem 6 and corollary 9 in ref. [28], with the only difference that here we do not assume the existence of distinguishable states. In its place, we use the framework assumption called *Physicalization of Readout* [41, 112], which guarantees that the outcome of every test can be read out from a physical system. According to this assumption, for every test $\{\mathcal{M}_i\}_{i\in X}$ from A to B, there exist a system C, a transformation $\mathcal{M} \in \mathsf{Transf}(\mathrm{A}, \mathrm{B}\otimes\mathrm{C})$, and an observation-test $\{c_i\}_{i\in X}$ such that




for all $i \in X$. □

The combination of Purity Preservation and Purification ensures that every test enjoys the Pure Decomposition property (assumption 4.2.6), as shown in corollary 26 of ref. [28]. Furthermore, Purification guarantees that the group of reversible transformations is a compact Lie group [28] and, as a consequence, it guarantees the existence of a unique maximally mixed state (cf. subsection 4.3.3).

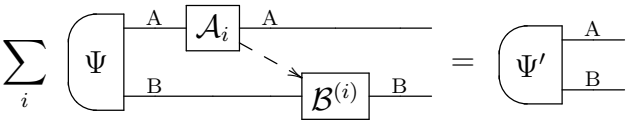
### 4.4.2 One-way LOCC protocols transforming pure states into pure states

The operational Lo-Popescu theorem guarantees that every LOCC protocol acting on a pure bipartite input state can be simulated by a 1-way protocol. Thanks to Purification, we recover the full statement of the theorem: not only is the protocol one-way, but also the conditional channel (i.e. the one applied by the party who receives classical communication) is reversible.

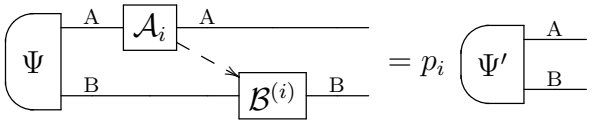
**Lemma 4.4.7.** *Let* $\Psi$ *and* $\Psi'$ *be pure states of* $\mathrm{A}\otimes\mathrm{B}$. *Under the validity of Purification and Purity Preservation, every 1-way LOCC protocol transforming* $\Psi$ *into* $\Psi'$ *can be simulated by a 1-way LOCC protocol where all conditional operations are reversible.*

*Proof.* Suppose that $\Psi$ can be transformed into $\Psi'$ via a 1-way LOCC protocol where Alice performs a test $\{\mathcal{A}_i\}_{i\in X}$ and Bob performs a channel $\mathcal{B}^{(i)}$

conditional on the outcome $i$. By definition, we have

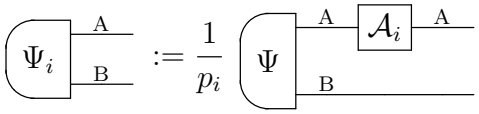



Since $\Psi'$ is pure, this implies that there exists a probability distribution $\{p_i\}$ such that



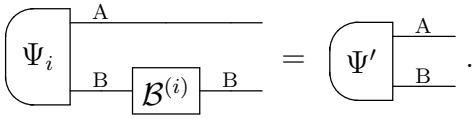

(4.4.1)

for every outcome $i$. Now, without loss of generality each transformation $\mathcal{A}_i$ can be assumed to be pure (if not, one can always decompose it into pure transformations). Then, Purity Preservation guarantees that the normalized state $\Psi_i$ defined as

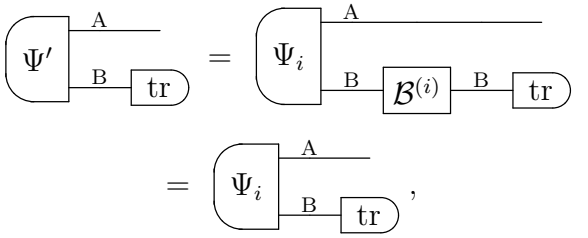



(4.4.2)

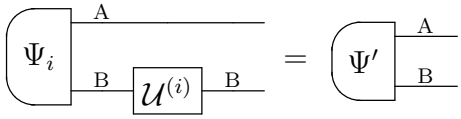
is pure. With this definition, eq. (4.4.1) becomes




Tracing out system B on both sides one obtains

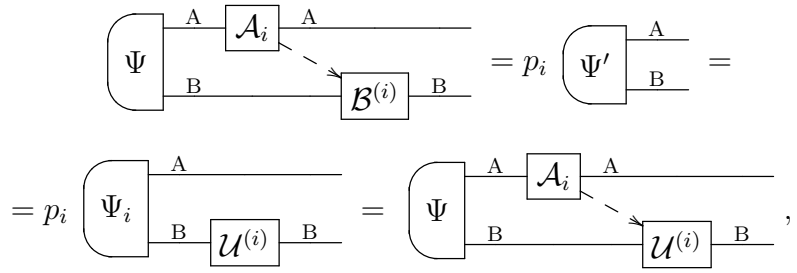



the second equality coming from proposition 1.3.6 applied to channel $\mathcal{B}^{(i)}$. Hence, the pure states $\Psi_i$ and $\Psi'$ have the same marginal on A. By the essential uniqueness of Purification, they must differ by a reversible channel $\mathcal{U}^{(i)}$ on the purifying system B, namely




(4.4.3)

In conclusion, we have obtained




where we have used eqs. (4.4.1), (4.4.3), and (4.4.2). In other words, the initial protocol can be simulated by a protocol where Alice performs the test $\{\mathcal{A}_i\}_{i\in X}$ and Bob performs the reversible transformation $\mathcal{U}^{(i)}$ conditionally on the outcome $i$. □

Note that we must require that $\Psi$ and $\Psi'$ are *both* pure states, otherwise the lemma does not hold.

Combining all the results got so far, we have the reversible version of Lo-Popescu theorem.

**Theorem 4.4.8** (Reversible Lo-Popescu theorem)**.** *Let $\Psi$ and $\Psi'$ be pure states of $\mathrm{A}\otimes\mathrm{B}$. Under the validity of Purity Preservation, Local Exchangeability, and Purification, the following are equivalent.*

1. *There is an LOCC protocol transforming $\Psi$ into $\Psi'$.*
2. *There is a 1-way LOCC protocol from Alice to Bob, with Bob applying a reversible channel.*
3. *There is a 1-way LOCC protocol from Bob to Alice, with Alice applying a reversible channel.*

*Proof.* By theorem 4.2.8, every LOCC from protocol from $\Psi$ to $\Psi'$ can be reduced to a 1-way LOCC protocol (from Alice to Bob or from Bob to Alice). By lemma 4.4.7, the conditional channel can be taken to be reversible. Finally, by lemma 4.2.7, classical communication can be inverted and one has the equivalence between all the statements. □

Our analysis shows that in the quantum mechanical version of Lo-Popescu theorem two assumptions of different nature come into play: leaving Purity Preservation and Causality aside as part of the framework, they are Local

Exchangeability and Purification. We can therefore split the statement of Lo-Popescu theorem into two parts, each of them being a consequence of one of the two axioms.

1. Local Exchangeability guarantees that every LOCC protocol between pure bipartite states of the same system can be turned into a 1-way LOCC protocol.

2. Purification ensures that the conditional channel can be taken to be reversible.

Once more, Purification is related to the reversibility of processes.

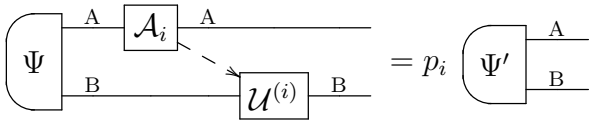
The reduction to 1-way protocols with reversible operations is crucial to connect the resource theory of entanglement with the resource theory of purity. The duality between these two resource theories will be established in the next subsections.

### 4.4.3 The more entangled a pure state, the more mixed its marginals

We start by proving one direction of the entanglement-thermodynamics duality: if a state is more entangled than another, then the marginals of the former are more mixed than the marginals of the latter.

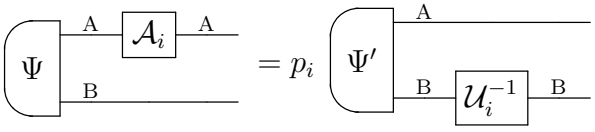
**Lemma 4.4.9.** *Let $\Psi$ and $\Psi'$ be two pure states of system $\mathrm{A}\otimes\mathrm{B}$ and let $\rho$, $\rho'$ and $\sigma$, $\sigma'$ be their marginals on system $\mathrm{A}$ and $\mathrm{B}$ respectively. Under the validity of Purification, Purity Preservation, and Local Exchangeability, if $\Psi$ is more entangled than $\Psi'$, then $\rho$ ($\sigma$) is more mixed than $\rho'$ ($\sigma'$).*

*Proof.* By reversible Lo-Popescu theorem, we know that there exists a 1-way LOCC protocol transforming $\Psi$ into $\Psi'$ with reversible channel (see theorem 4.4.8). Thanks to lemma 4.2.7, without loss of generality, we can choose a protocol with classical communication from Alice to Bob, in which Alice performs the test $\{\mathcal{A}_i\}_{i\in X}$ and Bob applies the reversible channel $\mathcal{U}^{(i)}$, conditional on the outcome $i$. Since $\Psi'$ is pure, we must have

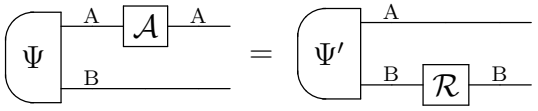



(4.4.4)

for all $i \in X$, where $\{p_i\}_{i\in X}$ is a probability distribution. Let $\mathcal{U}_i^{-1}$ be the inverse of $\mathcal{U}^{(i)}$. Let us apply it to both sides of eq. (4.4.4). We obtain

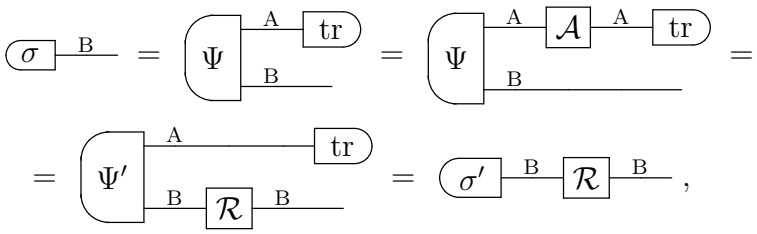



Summing over all outcomes, the equality becomes




(4.4.5)

with $\mathcal{A} := \sum_{i\in X} \mathcal{A}_i$ and $\mathcal{R} = \sum_{i\in X} p_i \mathcal{U}_i^{-1}$. Finally, we obtain




where we used proposition 1.3.6 applied to channel $\mathcal{A}$ in the second equality, and eq. (4.4.5) in the third. Since $\mathcal{R}$ is a RaRe channel by construction, we have proved that $\sigma$ is more mixed than $\sigma'$.

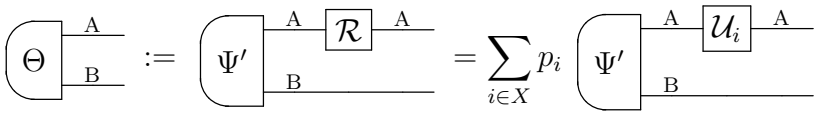
The fact that $\rho$ is more mixed than $\rho'$ can be proved by the same argument, starting from a 1-way protocol with classical communication from Bob to Alice (this is possible by lemma 4.2.7) and with reversible channels on Alice's side. □

### 4.4.4 The more mixed a state, the more entangled its purification

Now we now prove the converse direction of the entanglement-mixedness duality: if a state is more mixed than another, then its purification is more entangled than the other. Remarkably, the proof of this fact requires only the validity of Purification.

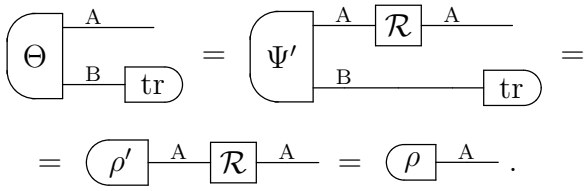
**Lemma 4.4.10.** *Let $\rho$ and $\rho'$ be two states of system* A*, and let $\Psi$ and $\Psi'$ be their purifications respectively, with purifying system* B*. Under the validity of Purification, if $\rho$ is more mixed than $\rho'$, then $\Psi$ is more entangled than $\Psi'$.*

*Proof.* By hypothesis, one has $\rho = \mathcal{R}\rho'$ for some RaRe channel $\mathcal{R} = \sum_{i\in X} p_i \mathcal{U}_i$. Let us define the bipartite state $\Theta$ as

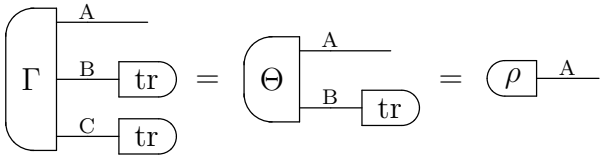



(4.4.6)

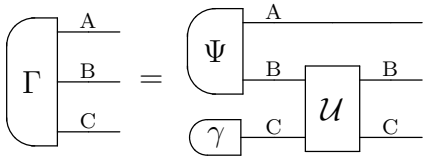
By construction, $\Theta$ is an extension of $\rho$ (see remark 4.4.2 for the definition of extension of a state). Indeed, one has




Let us take a purification of $\Theta$, say $\Gamma \in \mathsf{PurSt}(\mathrm{A}\otimes\mathrm{B}\otimes\mathrm{C})$. Clearly, $\Gamma$ is also a purification of $\rho$, since one has



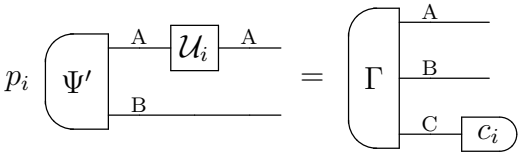

Then, the essential uniqueness of purification implies that $\Gamma$ will be of the form




(4.4.7)

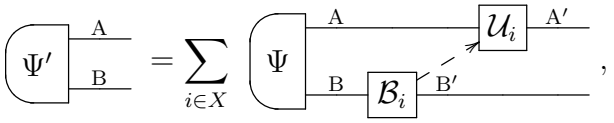
for some reversible channel $\mathcal{U}$ and some pure state $\gamma \in \mathsf{PurSt}(\mathrm{C})$. In other words, $\Psi$ can be transformed into $\Gamma$ by local operations on Bob's side.

Now, eq. (4.4.6) implies that the states $\{p_i (\mathcal{U}_i \otimes \mathcal{I}_{\mathrm{B}})\Psi\}_{i\in X}$ are a decomposition of $\Theta$. Hence, the steering property (theorem 4.4.6) implies that there exists an observation-test $\{c_i\}_{i\in X}$ such that

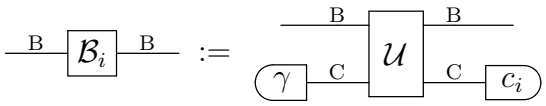



(4.4.8)

for all $i \in X$. Recalling eq. (4.4.7), eq. (4.4.8) gives the desired result.




where $\{\mathcal{B}_i\}_{i\in X}$ is the test defined as




for all $i \in X$. In conclusion, if the marginal state of $\Psi$ is more mixed than the marginal state of $\Psi'$, because there is a (1-way) LOCC protocol converting $\Psi$ into $\Psi'$. □

### 4.4.5 The duality

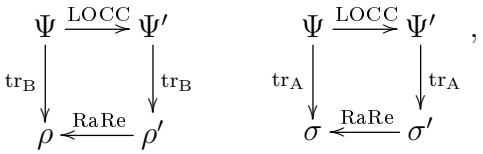
Combining lemmas 4.4.9 and 4.4.10, we identify the degree of entanglement of a pure bipartite state with the degree of mixedness of its marginals.

**Theorem 4.4.11** (Entanglement-mixedness duality)**.** *Let $\Psi$ and $\Psi$ be two pure states of system* A⊗B *and let $\rho$, $\rho'$ and $\sigma$, $\sigma'$ be their marginals on system* A *and* B *respectively. Under the validity of Purification, Purity Preservation, and Local Exchangeability, the following are equivalent.*

1. *$\Psi$ is more entangled than $\Psi'$.*
2. *$\rho$ is more mixed than $\rho'$.*
3. *$\sigma$ is more mixed than $\sigma'$.*

*Proof.* By lemma 4.4.9 we have that 1 implies both 2 and 3. By lemma 4.4.10 we have also the converse implications. □

The duality can be illustrated by the commutative diagrams



and is implemented operationally by discarding one of the component systems. Another illustration of the duality is via the diagram

$$\begin{array}{ccc} \Psi & \xleftarrow{\text{LOCC}} & \Psi' \\ \uparrow & & \uparrow \\ \text{purification} & & \text{purification} \\ \downarrow & & \downarrow \\ \rho & \xrightarrow{\text{RaRe}} & \rho' \end{array} .$$

Here the map implementing the duality is (a choice of) purification. Such a map cannot be realized as a physical operation, as shown in ref. [28], and this is the reason why it has been represented with dashed arrows. Instead, it corresponds to the theoretical operation of modelling mixed states as marginals of pure states.

## 4.5 Consequences of the duality

In this section we discuss the simplest consequences of the entanglement-thermodynamics duality. Specifically, the canonical resource theory of purity, whose structure is simpler because it involves only a single system, helps us understand the resource theory of entanglement better.

The first result is a characterization of the equivalence classes under the relation "to be equally entangled". Recall that in subsection 4.3.1 we noted that two states are equally controllable (which becomes "equally mixed" in the present setting) if and only if they differ by a reversible channel. We will apply this result to the entanglement relation between pure bipartite states.

**Corollary 4.5.1.** *Let* $\Psi$ *and* $\Psi'$ *be two* pure *states of system* $\mathrm{A} \otimes \mathrm{B}$. $\Psi$ *and* $\Psi'$ *are equally entangled if and only if they differ by local reversible channels, namely*

$$\Psi' = (\mathcal{U}_{\mathrm{A}} \otimes \mathcal{V}_{\mathrm{B}})\, \Psi,$$

*where* $\mathcal{U}_{\mathrm{A}}$ *and* $\mathcal{V}_{\mathrm{B}}$ *are reversible channels acting on* A *and* B *respectively.*

*Proof.* Sufficiency. As we noted in section 4.1, if $\Psi$ and $\Psi'$ differ by local reversible transformations, they are equally entangled.

Necessity. Now we must resort to the duality. Owing to it, the marginals of $\Psi$ and $\Psi'$ on system A, denoted as $\rho$ and $\rho'$ respectively, are equally mixed. As a result of ref. [102], $\rho' = \mathcal{U}_{\mathrm{A}}\rho$ for some reversible channel $\mathcal{U}_{\mathrm{A}}$ on A. As a consequence, $\Psi'$ and $(\mathcal{U}_{\mathrm{A}} \otimes \mathcal{I}_{\mathrm{B}})\, \Psi$ are two purifications of $\rho'$. By the essential

uniqueness of purification, we have $\Psi' = (\mathcal{U}_{\mathrm{A}} \otimes \mathcal{V}_{\mathrm{B}})\,\Psi$ for some reversible transformation $\mathcal{V}_{\mathrm{B}}$ on B. □

Now we move to study maximally entangled states. As a consequence of the duality, we anticipate there exists a correspondence between maximally mixed and maximally entangled states. Maximally entangled states can be defined as follows.

**Definition 4.5.2.** A pure state $\Phi$ of system $\mathrm{A} \otimes \mathrm{B}$ is *maximally entangled* if the only states more entangled than $\Phi$ are the states as entangled as $\Phi$.

In other words, if $\Psi \succeq_{\mathrm{ent}} \Phi$, then $\Psi \sim_{\mathrm{ent}} \Phi$, namely, by corollary 4.5.1, $\Psi = (\mathcal{U}_{\mathrm{A}} \otimes \mathcal{V}_{\mathrm{B}})\,\Phi$, where $\mathcal{U}_{\mathrm{A}}$ and $\mathcal{V}_{\mathrm{B}}$ are reversible channels on A and B respectively.

Thanks to the duality we have the following corollary.

**Corollary 4.5.3.** *The purification of a maximally mixed state is maximally entangled.*

*Proof.* Let $\Phi$ be a purification of the maximally mixed state of system A. Suppose that $\Psi \in \mathsf{PurSt}\,(\mathrm{A} \otimes \mathrm{B})$ is more entangled than $\Phi$. By theorem 4.4.11, the marginal of $\Psi$ on system A, denoted by $\rho$, must satisfy $\rho \succeq_{\mathrm{mix}} \chi$. Since $\chi$ is maximally mixed, this implies $\rho = \chi$. The essential uniqueness of purification then implies the condition $\Psi = (\mathcal{I}_{\mathrm{A}} \otimes \mathcal{V}_{\mathrm{B}})\,\Phi$ for some reversible transformation $\mathcal{V}_{\mathrm{B}}$ on B. □

Since the maximally mixed state is the maximum of the mixedness relation (see subsection 4.3.3), all the purifications of it are not only maximal elements of the entanglement relation, but also maxima. This means that, if $\Phi \in \mathsf{PurSt}\,(\mathrm{A} \otimes \mathrm{B})$ is a purification of the maximally mixed state of system A, then $\Phi \succeq_{\mathrm{ent}} \Psi$ for every *pure* state $\Psi \in \mathsf{PurSt}\,(\mathrm{A} \otimes \mathrm{B})$.

Can we conclude that $\Phi$ is more entangled than any bipartite state $\Sigma \in \mathsf{St}_1\,(\mathrm{A} \otimes \mathrm{B})$, even when $\Sigma$ is mixed? The answer is affirmative. Indeed $\Sigma$ can be written as a coarse-graining of pure states, $\Sigma = \sum_{i \in X} p_i \Psi_i$, where $\{p_i\}_{i \in X}$ is a probability distribution and the $\Psi_i$'s are pure states. Now, for every pure state $\Psi_i$ there exists an LOCC channel (see section 4.1) $\mathcal{L}_i$ such that $\Psi_i = \mathcal{L}_i \Phi$. Therefore, summing over $i$, one has $\Sigma = \sum_{i \in X} p_i \mathcal{L}_i \Phi$. We must show that $\sum_{i \in X} p_i \mathcal{L}_i$ is an LOCC channel. Clearly, $\sum_{i \in X} p_i \mathcal{L}_i$ is an LOCC transformation, but it is also a channel thanks to proposition 1.3.6, indeed

$$\mathrm{tr} \left( \sum_{i \in X} p_i \mathcal{L}_i \right) = \sum_{i \in X} p_i \mathrm{tr}\, \mathcal{L}_i = \sum_{i \in X} p_i \mathrm{tr} = \mathrm{tr},$$

where we have used the fact that each $\mathcal{L}_i$ is a channel. Therefore we know that $\Phi \succeq_{\text{ent}} \Sigma$ for any bipartite state $\Sigma$.

## 4.6 Entanglement and purity monotones

According to the general scheme presented in subsection 2.2.2, we introduce resource monotones for the resource theory of entanglement.

**Definition 4.6.1.** An *entanglement monotone* for system $\mathrm{A}\otimes\mathrm{B}$ is a function $E : \mathsf{St}(\mathrm{A}\otimes\mathrm{B}) \longrightarrow \mathbb{R}$ such that $E(\rho) \geq E(\rho')$ whenever $\rho \succeq_{\text{ent}} \rho'$, where $\rho$ and $\rho'$ are states of system $\mathrm{A}\otimes\mathrm{B}$.

Similarly, one can define monotones in the (canonical) resource theory of purity.

**Definition 4.6.2.** A *purity monotone* for system A is a function $P : \mathsf{St}(\mathrm{A}) \longrightarrow \mathbb{R}$ such that $P(\rho) \geq P(\rho')$ whenever $\rho \succeq_{\text{pur}} \rho'$, where $\rho$ and $\rho'$ are states of system A.

These two definitions are clearly independent of the set of axioms we are assuming, therefore they make sense in every causal GPT.

**Example 4.6.3.** Let us see what purity monotones are in classical probability theory. In classical probability theory, states are probability distributions over finite sets and reversible transformations are permutation matrices. According to definition 4.3.1, a state $\mathbf{p} = \begin{pmatrix} p_1 & \dots & p_n \end{pmatrix}$ is purer than another state $\mathbf{p}' = \begin{pmatrix} p'_1 & \dots & p'_n \end{pmatrix}$ if

$$\mathbf{p}' = \sum_i q_i \Pi_i \mathbf{p},$$

where the $q_i$'s are probabilities and the $\Pi_i$'s are permutation matrices. Therefore, the matrix $\sum_i q_i \Pi_i$ is doubly stochastic, and this means that $\mathbf{p}'$ is majorized by $\mathbf{p}$, $\mathbf{p}' \preceq \mathbf{p}$ (see theorem 3.1.14). Hence, a function $P : \mathbb{R}^n \longrightarrow \mathbb{R}$ is a purity monotone if and only if it is a Schur-convex function. Therefore, Schur-concave functions are instead measures of mixedness in classical probability theory.

Constructing purity monotones is fairly easy. For example, every function that is convex and invariant under reversible channels is a purity monotone.

**Proposition 4.6.4.** *Let $P : \mathsf{St}(\mathrm{A}) \longrightarrow \mathbb{R}$ be a function satisfying*

**convexity:** *$P\left(\sum_i p_i \rho_i\right) \leq \sum_i p_i P(\rho_i)$ for every set of states $\{\rho_i\}$ and for every probability distribution $\{p_i\}$;*

**invariance under reversible channels:** *$P(\mathcal{U}\rho) = P(\rho)$ for every state $\rho$ and for every reversible transformation $\mathcal{U}$.*

*Then $P$ is a purity monotone.*

*Proof.* Let $\rho$ be purer than $\rho'$. Then $\rho' = \sum_i p_i \mathcal{U}_i \rho$, for some probabilities $p_i$'s and reversible channels $\mathcal{U}_i$'s. Therefore, since $P$ is convex,

$$P(\rho') = P\left(\sum_i p_i \mathcal{U}_i \rho\right) \leq \sum_i p_i P(\mathcal{U}_i \rho).$$

Since $P$ is also invariant under reversible channels,

$$P(\rho') \leq \sum_i p_i P(\mathcal{U}_i \rho) = \sum_i p_i P(\rho) = P(\rho),$$

where the last equality follows from the fact that the $p_i$'s sum to 1. This shows that $P$ is a purity monotone. □

Let us see the implications of this theorem in the case of classical probability theory.

**Example 4.6.5.** In example 4.6.3 we saw that in classical probability theory purity monotones coincide with Schur-convex functions. Recall that reversible transformations are permutation matrices, and that a function $P : \mathbb{R}^n \longrightarrow \mathbb{R}$ is called symmetric if $P(\Pi x) = P(x)$ for every permutation matrix $\Pi$. Now, proposition 4.6.4 reads that every symmetric convex function is a Schur-convex function, a well-known result in the theory of majorization (cf. example 3.2.7).

Using proposition 4.6.4, one can construct a lot of purity monotones. For every convex function $f : \mathbb{R} \longrightarrow \mathbb{R}$ one can define the *$f$-purity* $P_f : \mathsf{St}(\mathrm{A}) \longrightarrow \mathbb{R}$ as

$$P_f(\rho) := \sup_{\boldsymbol{a} \text{ pure}} \sum_{i \in X} f(p_i),$$

where $p_i := (a_i|\rho)$, and the supremum runs over all pure observation-tests $\boldsymbol{a} = \{a_i\}_{i \in X}$, and over all outcome sets $X$. It is easy to verify that every

$f$-purity is convex (because $f$ is convex) and invariant under reversible transformations (because of the supremum[5]), and therefore is a purity monotone. In the special case of the function $f(x) = x \log_a x$, where $a > 1$, one has $P_f(\rho) = -H(\rho)$, where $H$ is the *measurement entropy* [113, 114, 115], namely the minimum over all pure observation-tests of the Shannon entropy of the probability distribution resulting from the observation-test.

Now, thanks to the duality entanglement monotones for pure states can be obtained from purity monotones: given a purity monotone $P : \mathsf{St}(\mathrm{A}) \longrightarrow \mathbb{R}$, we can define the pure-state entanglement monotone $E : \mathsf{PurSt}(\mathrm{A} \otimes \mathrm{B}) \longrightarrow \mathbb{R}$ as

$$E(\Psi) := g[P(\rho)],$$

where $\rho = \mathrm{tr}_{\mathrm{B}}\Psi$ (or, equivalently, $\rho = \mathrm{tr}_{\mathrm{A}}\Psi$) and $g : \mathbb{R} \longrightarrow \mathbb{R}$ is any decreasing function. Essentially, the decreasing behaviour of $g$ is the quantitative implementation of the reversing of arrows in the duality (cf. the commutative diagrams of subsection 4.4.5). In other words, every measure of mixedness becomes a measure of pure-state entanglement. An easy way to generate entanglement measures is to pick an $f$-purity and take its negative (in this case $g(x) = -x$). For example, the choice $f(x) = x \log_a x$ leads to an generalization of Shannon entropy as a measure of entanglement for pure states.

### 4.6.1 Information erasure and entanglement generation

In 1961 Landauer proposed his principle [116], according to which "information erasure" is a costly operation.

**Definition 4.6.6.** *Information erasure* is the process in which an arbitrary state is transformed into a fixed pure state.

Despite its name, information erasure does not mean a loss of information, but, instead, an information gain, for we go from a mixed state to a pure state. Landauer pointed out that information erasure takes place in (classical) computers as part of their operating principles. Indeed, a computer performs some operations on a (finite) memory made of a string of bits, i.e. a sequence of 0 and 1. We can think of the initial state of a computing process as a sequence of all 0's. During computation, the computer changes the values of the bits in the memory. At some point, since the memory is finite, the

[5]Indeed, thanks to the supremum, we can reabsorb the action of the reversible channel into the observation-test, $(a_i|\mathcal{U}\rho) = (a_i\mathcal{U}|\rho)$, without changing the value of $P_f(\rho)$.

computer has to reset some of the bits of the memory to their original value of 0 (a pure state). Let us concentrate on a single bit. Before restoring its value to 0 (information erasure), the value of the bit is unknown a priori, i.e. it is described by the uniform distribution[6] $\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix}$, where the first element is the probability that the bit value is 0, and the second is the probability that the bit is 1. After the information erasure, the state of the bit is, instead $\begin{pmatrix} 1 & 0 \end{pmatrix}$, because now its value is 0 with certainty. The statistical entropy of the bit before the erasure is

$$S_{\text{in}} = kH \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix} = k \ln 2,$$

where $k$ is Boltzmann's constant and $H$ is Shannon entropy (in the natural basis for the logarithm). The final entropy is $S_{\text{fin}} = 0$ because the final state is pure. Therefore, the balance of entropy is negative, $\Delta S = S_{\text{fin}} - S_{\text{in}} = -k \ln 2$. If the process occurs at constant temperature $T$, the energy of the system has a negative variation $\Delta U = -kT \ln 2$ (there is an energy cost), and according to the first law of thermodynamics, it becomes heat dissipated into the environment.

The information erasure is just a particular instance of the thermodynamic implications of irreversible computation. Landauer explained that every irreversible process in computation gives rise to thermal dissipation. Here irreversible process means an operation which cannot be inverted. Information erasure is an example, indeed once we have erased a bit to the fixed value 0, we cannot recover its value before erasure. Another example of irreversible process is the AND logic operation.

**Example 4.6.7.** The AND logic operation is one of the gates used in computers to do computation. AND is binary operation $\wedge$ on bits: $\wedge : \{0,1\}^2 \longrightarrow \{0,1\}$, defined via the following truth table.

| $a$ | $b$ | $a \wedge b$ |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 1 | 0 |
| 1 | 0 | 0 |
| 1 | 1 | 1 |

We immediately see that the AND operation is irreversible: if the output is 0, we are not able to retrieve the value of the two input bits, indeed they

[6] Recall that in the classical case states are probability distributions (see example 4.6.3).

may be $(0,0)$, $(0,1)$, $(1,0)$, where the first entry represents the value of the bit $a$ and the second entry the value of the bit $b$. Thinking of entropies, at the beginning we have 4 inputs states with equal probability: $(0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$. Therefore the initial entropy is

$$S_{\text{in}} = kH\left(\begin{array}{cccc} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{array}\right) = 2k\ln 2.$$

At the output, we have the state 0 with probability $\frac{3}{4}$ and the state 1 with probability $\frac{1}{4}$. The final entropy is therefore

$$S_{\text{fin}} = kH\left(\begin{array}{cc} \frac{3}{4} & \frac{1}{4} \end{array}\right) = k\left(2\ln 2 - \frac{3}{4}\ln 3\right) < S_{\text{in}}.$$

Again, we have a negative entropy difference $\Delta S = -\frac{3}{4}k\ln 3$, which gives rise to a thermal dissipation of $|Q| = \frac{3}{4}kT\ln 3$ per AND operation performed at temperature $T$.

In this way, we see that irreversible computation contributes to heat generation in computers, thus posing some intrinsic limitations on their size. To overcome this problem the idea of reversible computation was developed [117, 118].

Now we wish to study information erasure in the framework of resource theories in causal GPTs satisfying Purity Preservation, Local Exchangeability and Purification. We have seen that information erasure has an energy cost for the system, therefore we anticipate that it will be described by a costly operation in our resource theory of purity. Indeed, this is what happens, because there is no way to transform a non-pure state into a pure state by using only RaRe channels (cf. proposition 4.3.3). The dual operation in the resource theory of entanglement is the generation of entangled states from product states. By the duality, the impossibility of erasing information by RaRe channels and the impossibility of generating entanglement out of nothing by LOCC are equivalent.

However, in subsection 2.2.1 we saw that catalysts may allow transformations that would be ordinarily forbidden. Is it possible to achieve information erasure for free with the aid of catalysts? In this case, the operation of erasure assisted by the catalyst $\sigma \in \mathsf{St}(\mathrm{C})$ transforms the product state $\rho\otimes\sigma \in \mathsf{St}(\mathrm{A}\otimes\mathrm{C})$ into the state $\alpha\otimes\sigma$ for some fixed pure state $\alpha \in \mathsf{PurSt}(\mathrm{A})$.

By duality, it is immediate to see that catalyst-assisted erasure is equivalent to catalyst-assisted entanglement generation.

**Corollary 4.6.8.** *Let $\Psi$ and $\Gamma$ be two pure states of systems $\mathrm{A}\otimes\mathrm{C}$ and $\mathrm{C}\otimes\mathrm{D}$ respectively, and let $\rho$ and $\sigma$ be their marginals on systems $\mathrm{A}$ and $\mathrm{C}$ respectively. Then, the following are equivalent.*

1. *$\rho$ can be erased by a RaRe channel using $\sigma$ as a catalyst.*
2. *$\Psi$ can be prepared by an LOCC protocol using $\Gamma$ as a catalyst.*

*Proof.* We have that $\rho\in\mathsf{St}(\mathrm{A})$ can be erased by a RaRe channel using $\sigma\in\mathsf{St}(\mathrm{C})$ as a catalyst if and only if $\alpha\otimes\sigma\succeq_{\mathrm{mix}}\rho\otimes\sigma$, where $\alpha$ is a pure state of system A. This happens if and only if the purification of the left-hand side is more entangled than the purification of the right-hand side, namely if and only if

$$\alpha\otimes\beta\otimes\Gamma\succeq_{\mathrm{ent}}\Psi\otimes\Gamma, \tag{4.6.1}$$

where $\beta$ is a pure state of system B. Since $\alpha\otimes\beta$ is a pure state of $\mathrm{A}\otimes\mathrm{B}$, it is a free resource, therefore it is equivalent to the void resource. Therefore (4.6.1) can be rewritten as

$$\Gamma\succeq_{\mathrm{ent}}\Psi\otimes\Gamma. \tag{4.6.2}$$

This proves the corollary. □

Formula (4.6.2) shows the role of the catalyst $\Gamma$ as an "entanglement reservoir", from which entanglement can be extracted indefinitely. Dually $\sigma$ behaves as an infinitely mixed state.

Suppose we have a non-negative and additive entanglement monotone $E$. Formula (4.6.2) implies

$$E(\Gamma)\geq E(\Psi)+E(\Gamma).$$

If $\Psi$ is entangled[7], namely $E(\Psi)>0$, the only possibility is that $E(\Gamma)=+\infty$, showing again the role of $\Gamma$ as an infinitely entangled state.

It is then natural to ask whether the impossibility of infinitely entangled/infinitely mixed states follow from our axioms. The answer is affirmative in the finite-dimensional case, but counterexamples exist in infinite dimension in the quantum case [119].

For the finite-dimensional case (i.e. when the vector spaces spanned by states and effects are finite-dimensional, see subsection 1.1.4), which is the case we are interested in, we have the following proposition.

[7]The case with $\Psi$ product state is not interesting, because in this case catalysts are not necessary.

**Proposition 4.6.9.** *Let* $\mathrm{A} \otimes \mathrm{C}$ *be a finite system. Then, it is impossible to erase a* mixed *state of* A *using a state of system* C *as a catalyst.*

*Proof.* Suppose by contradiction that the *mixed* state $\rho \in \mathsf{St}(\mathrm{A})$ can be erased to a fixed pure state $\alpha \in \mathsf{PurSt}(\mathrm{A})$, with the aid of the catalyst $\sigma \in \mathsf{St}(\mathrm{C})$. This means that $\alpha \otimes \sigma \succeq_{\mathrm{mix}} \rho \otimes \sigma$. Since $\alpha$ is pure, $\rho \succeq_{\mathrm{mix}} \alpha$, which implies $\rho \otimes \sigma \succeq_{\mathrm{mix}} \alpha \otimes \sigma$, hence $\rho \otimes \sigma$ and $\alpha \otimes \sigma$ are equally mixed. We know [102] that there exists a reversible channel $\mathcal{U}$ such that

$$\mathcal{U}(\alpha \otimes \sigma) = \rho \otimes \sigma. \tag{4.6.3}$$

Now, let us choose a basis for $\mathsf{St}_{\mathbb{R}}(\mathrm{A} \otimes \mathrm{C})$, such that the reversible transformations are represented by orthogonal matrices. This is always possible because, by Purification, the group of reversible transformation is a compact Lie group [28]. Let us consider the norm[8] induced by the scalar product associated with (i.e. preserved by) the orthogonal matrices. Take a decomposition of $\rho$ into pure states, $\rho = \sum_i p_i \alpha_i$. By the triangle inequality of the norm

$$\left\| \sum_i p_i (\alpha_i \otimes \sigma) \right\| \leq \sum_i p_i \|\alpha_i \otimes \sigma\|. \tag{4.6.4}$$

The norm on the right-hand side of inequality (4.6.4) is invariant under reversible channels. By applying suitable reversible channels $\mathcal{U}_{i,\mathrm{A}} \otimes \mathcal{I}_{\mathrm{C}}$ to $\alpha_i \otimes \sigma$, every term in the norm becomes $\alpha \otimes \sigma$.

$$\sum_i p_i \|\alpha_i \otimes \sigma\| = \|\alpha \otimes \sigma\|$$

However, eq. (4.6.3) states that

$$\|\alpha \otimes \sigma\| = \|\rho \otimes \sigma\| = \left\| \sum_i p_i (\alpha_i \otimes \sigma) \right\|.$$

This means that in formula (4.6.4) equality holds.

$$\left\| \sum_i p_i (\alpha_i \otimes \sigma) \right\| = \sum_i p_i \|\alpha_i \otimes \sigma\|$$

In order for this to be possible, all the terms $\alpha_i \otimes \sigma$ must be proportional to one another: in other words, $\rho$ must be pure. □

[8] If $v = \begin{pmatrix} v_1 & \dots & v_n \end{pmatrix}^t$ is the vector of the coordinates relative to the fixed basis, the norm of $v$ is defined as $\|v\| = \sqrt{\sum_i v_i^2}$.

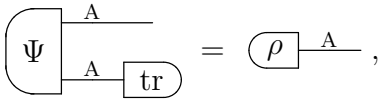
As shown in corollary 4.6.8, this is equivalent to stating that there are no entanglement reservoirs.

## 4.7 Symmetric purification

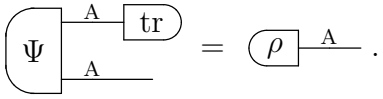
We conclude this work by showing that the study of resource theories can prove valuable also to analyse foundational issues of physical theories. In proposition 4.3.6 we saw how the requirement of a well-behaved resource theory of purity has led naturally to the requirement that reversible channels act transitively on the set of pure states, which is a property concerning the fundamental structure of the theory.

In general, setting the axioms that give a resource theories some sensible and desirable properties helps understand also the basic structure of the physical theory underlying the resource theory under consideration. This is also our case. Indeed, from Purity Preservation, Local Exchangeability and Purification we can derive a strengthened version of the Purification Principle, called Symmetric Purification. In fact, we will be able to prove that Symmetric Purification is equivalent to the set of the other axioms.

**Definition 4.7.1.** Let $\rho$ be a state of system A and let $\Psi$ be a pure state of $\mathrm{A} \otimes \mathrm{A}$. We say that $\Psi$ is a *symmetric purification* of $\rho$ if




and




This definition leads us to an upgraded version of the Purification Principle.

**Axiom 4.7.2** (Symmetric Purification)**.** *Every state has a symmetric purification. Every purification is essentially unique.*

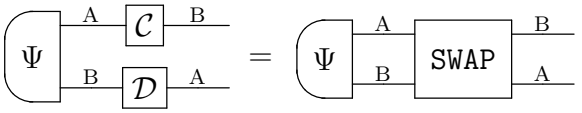
In a symmetric purification, the distinction between the purifying system and the purified system vanishes, essential uniqueness means that two symmetric purification of the same state differ by local reversible channels: if $\mathrm{tr}_{\mathrm{A}}\Psi = \mathrm{tr}_{\mathrm{A}}\Psi'$, where $\Psi, \Psi' \in \mathsf{PurSt}\,(\mathrm{A} \otimes \mathrm{A})$, then $\Psi' = (\mathcal{U} \otimes \mathcal{V})\,\Psi$.

Now we prove the main result.

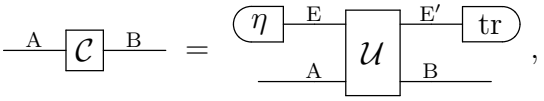
**Theorem 4.7.3.** *In a causal theory satisfying Purity Preservation, the following axioms are equivalent:*

- *Local Exchangeability and Purification*
- *Symmetric Purification.*

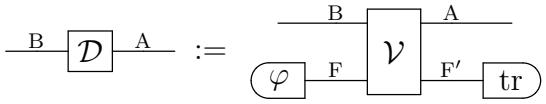
*Proof.* Let us prove that Local Exchangeability and Purification imply Symmetric Purification. Let $\rho$ be a state of system A, and let $\Psi \in \mathsf{PurSt}(\mathrm{A} \otimes \mathrm{B})$ be one of its purifications ($\Psi$ exists by Purification). By Local Exchangeability there exist two channels $\mathcal{C} \in \mathsf{Transf}(\mathrm{A}, \mathrm{B})$ and $\mathcal{D} \in \mathsf{Transf}(\mathrm{B}, \mathrm{A})$ such that




Now, in a theory satisfying Purification, every channel can be realized through a reversible transformation acting on the system and on an environment, initially in a pure state and finally discarded [28]. Specifically, channel $\mathcal{C}$ can be realized as

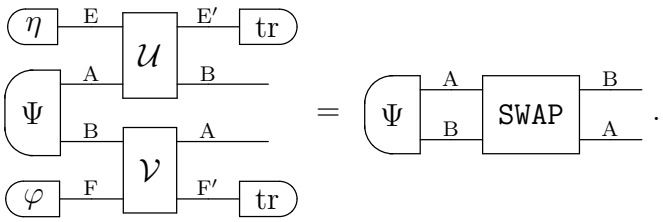



where E and E′ are suitable systems, $\mathcal{U}$ is a reversible channel, and $\eta$ is a pure state. Similarly, channel $\mathcal{D}$ can be realized as

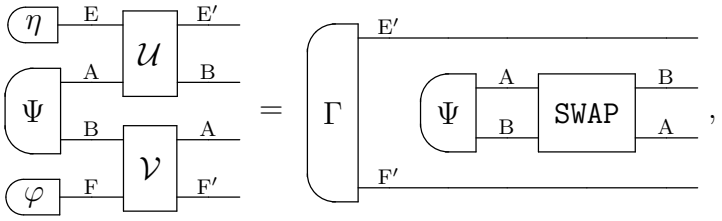



(4.7.1)

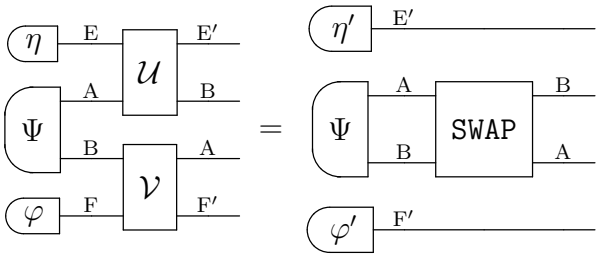
where F and F′ are suitable systems, $\mathcal{V}$ is a reversible channel, and $\varphi$ is a pure state. Inserting the realizations of $\mathcal{C}$ and $\mathcal{D}$ in Local Exchangeability, we obtain



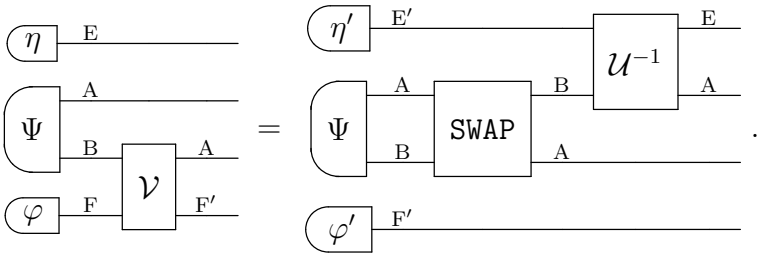
Since the pure state on the left-hand side is the purification of a pure state, by proposition 4.4.5, it must be of the product form



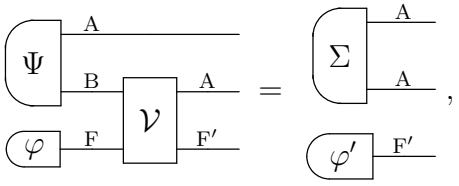

for some pure state $\Gamma$. The above equation shows that the state $\Gamma$ can be generated by a LOCC protocol using $\Psi$ as a catalyst. By proposition 4.6.9, $\Gamma$ cannot be entangled, therefore it is a product state, i.e. $\Gamma = \eta' \otimes \varphi'$ for two pure states $\eta' \in \mathsf{PurSt}(\mathrm{E}')$ and $\varphi' \in \mathsf{PurSt}(\mathrm{F}')$. Hence, Local Exchangeability becomes




or, equivalently,

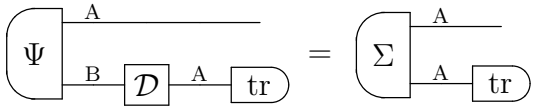



Discarding system E one obtains

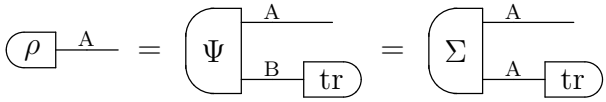



for some suitable state $\Sigma$. Since the left-hand side is a pure state, $\Sigma$ must be a pure state. Now, discarding system $\mathrm{F}'$ and the second copy of system A,

and recalling eq. (4.7.1), we have

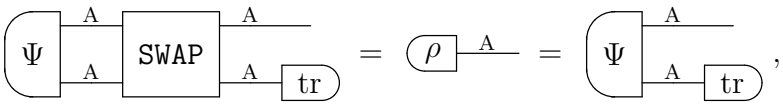



Recalling that $\mathcal{D}$ is a channel, and therefore $\mathrm{tr}_{\mathrm{A}}\mathcal{D} = \mathrm{tr}_{\mathrm{B}}$ (see proposition 1.3.6), we conclude that

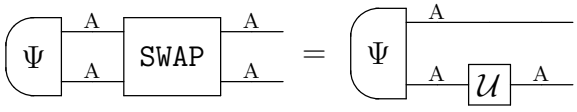



Hence, the marginal of the pure state $\Sigma$ on the first copy of system A is $\rho$. By the same argument, we can prove that the marginal on the second copy of system A is also $\rho$. Hence, $\Sigma$ is a symmetric purification of $\rho$. By the essential uniqueness of the purification, all symmetric purifications of $\rho$ differ by local reversible transformations. Since $\rho$ is arbitrary, we conclude that every state has a symmetric purification, unique up to local reversible transformations.

Conversely, now we show that the Symmetric Purification implies Local Exchangeability and Purification. It is obvious that Symmetric Purification implies Purification, therefore it is enough to prove that Local Exchangeability follows from Symmetric Purification. Symmetric purifications are locally exchangeable: indeed, if $\Psi$ is a symmetric purification one has




and, by the essential uniqueness of symmetric purification,




for some reversible channel $\mathcal{U}$. This shows that the symmetric purification $\Psi$ is locally exchangeable. Since all purifications of $\rho$ are equivalent to $\Psi$ under local (reversible) operations and since $\Psi$ is locally exchangeable, we conclude that every purification of $\rho$ is locally exchangeable (similar to the argument presented in the first of the three examples of non-local boxes in example 4.2.4). This proves Local Exchangeability. □

# Conclusions

In this thesis we examined resource theories, which are theories that describe the processing and the value of some entities called resources. A resource is anything that can be used to accomplish a task, and we saw that examples of resource theories encompass the most diverse fields: from chemistry to food, from knowledge to physics. Specifically, we focused on some examples taken from quantum mechanics and quantum information. Indeed, tasks that can be performed classically, can usually be fulfilled in a more efficient or faster way by using quantum resources that have no classical counterpart, the most paradigmatic example being quantum entanglement. For this reason it is fairly natural to find a great number of resource theories in the framework of quantum mechanics.

The broad spread of resource theories in many disciplines stimulated us to search for a common mathematical background, capable of capturing the significant features of resource theories, common to all of them, irrespective of the content of the theory. This mathematical framework is provided by the theory of strict symmetric monoidal categories, which not only provide a general foundation for all resource theories, but are also a tool to address physical theories in a formalism-independent way. This shows that the basic structure of a resource theory depends ultimately only on the informational structure of the underlying physical theory and not on the content of the theory itself. Essentially, what matters in a resource theory is the value or the price we assign to resources, therefore we can order them according to their degree of usefulness and even establish functions that measure their value in a consistent way.

Once we have such a generalized framework, one can try to extend some resource theories to broader scopes than the original ones. The natural way to do that for quantum resource theories is to export them to general probabilistic theories. This will result in a deeper grasp of their meaning and

features. In the present work, we focused essentially on two quantum resource theories: the resource theory of entanglement and the resource theory of purity. Both are of great importance in quantum mechanics, the former for quantum communication and the latter for the foundations of quantum thermodynamics via the ordering of states according to their degree of purity (or mixedness). The two theories are intimately related by a duality theorem, therefore in quantum mechanics either of them can be taken to be a route towards the foundations of quantum thermodynamics.

In the last part of the thesis, we tried for the first time to extend these two resource theories to GPTs, as done in ref. [92], and to identify what axioms we have to set in order to recover some of the properties of the resource theories considered. Specifically, even though in the quantum case we could provide two different definitions of purity, in the GPTs framework, it turns out that only one of the two approaches is viable. One of the interesting results of this analysis is that requiring some properties for a resource theory has a direct impact on the basic structure of the theory itself.

# Acknowledgements

This thesis was done at IIIS (Institute for Interdisciplinary Information Sciences), Tsinghua University, Beijing. My research at Tsinghua University was funded by the Chinese Government Scholarship for PhD and by a scholarship from "Fondazione Ing. Aldo Gini".

I would like to thank Prof. Pieralberto Marchetti for supervising this thesis with genuine interest and commitment. Many thanks also to Prof. Giulio Chiribella for supervising me in my PhD research, part of which is included in chapter 4 of thesis, and originally published in ref. [92]. The research with Prof. Chiribella was supported by the National Basic Research Program of China (973) 2011CBA00300 (2011CBA00301), by the National Natural Science Foundation of China through Grants 11450110096, 11350110207, 61033001, and 61061130540, by the Foundational Questions Institute through the large grant "The fundamental principles of information dynamics", and by the 1000 Youth Fellowship Program of China. I am grateful to Dr. Shojun Nakayama for useful and interesting discussions about control-unitaries. Thanks also to Prof. Antonio Saggion for valuable discussions and suggestions about the topic of this thesis.

My thanks also to all professors I encountered during my five years at the Galilean School of Higher Education, I have countless reasons for being grateful to them.

# Appendix A

# Some mathematical results

## A.1 A proposition

**Proposition A.1.1.** *Let $(X, \lesssim)$ be a preordered set. If $x, y \in X$, define $x \sim y$ if $x \lesssim y$ and $y \lesssim x$. Then $\sim$ is an equivalence relation.*

*We can define an order $\leq$ on the set $X/\sim$, such that $[x] \leq [y]$ if $x \lesssim y$, where $[x]$ and $[y]$ are the equivalence classes of $x, y \in X$.*

*Proof.* It is easy to prove that $\sim$ is an equivalence relation. Indeed, $x \sim x$ because $x \lesssim x$ for every $x \in X$, since $\lesssim$ is reflexive. In addition, if $x \sim y$, then $y \sim x$, by definition of $\sim$, for every $x, y \in X$. Finally, if $x \sim y$, and $y \sim z$, then $x \sim z$. Indeed, $x \sim y$, means $x \lesssim y$ and $y \lesssim x$; and $y \sim z$ means $y \lesssim z$ and $z \lesssim y$. Since $\lesssim$ is transitive, one has $x \lesssim z$ and $z \lesssim x$, whence $x \sim z$. Therefore, $\sim$ is an equivalence relation.

First of all, let us prove that our definition of $\leq$ is well-posed. If we take $x' \sim x$ and $y' \sim y$, we have that $x' \lesssim x$ (and $x \lesssim x'$), and $y \lesssim y'$ (and $y' \lesssim y$). Thus, if $x \lesssim y$, then $x' \lesssim y'$, hence $\leq$ is well-defined.
Now, let us show that $\leq$ is an order. It is reflexive, for $[x] \leq [x]$ if $x \lesssim x$, and this is true because $\lesssim$ is reflexive. It is transitive, since if $[x] \leq [y]$ and $[y] \leq [z]$, then $x \lesssim y$ and $y \lesssim z$, whence $x \lesssim z$, because $\lesssim$ is transitive. Thus, it follows that $[x] \leq [z]$. Finally, $\leq$ is also antisymmetric. Indeed, if $[x] \leq [y]$ and $[y] \leq [x]$, this means that $x \lesssim y$ and $y \lesssim x$, that is $x \sim y$. Hence $[x] = [y]$. □

# Bibliography


[1] R. Faraldo and A. Saggion. L'osservatore in termodinamica (The observer in thermodynamics). Unpublished research note.

[2] E. A. Guggenheim. *Thermodynamics: An Advanced Treatise for Chemists and Physicists*. North-Holland Publishing Company, 1967.

[3] W. Thompson. On the dynamical theory of heat: with numerical results deduced from Mr. Joule's equivalent of a thermal unit and Messr. Regnault's observation on steam. *Trans. R. Soc. Edinburgh*, 1851.

[4] D. Janzing, P. Wocjan, R. Zeier, R. Geiss, and Th. Beth. Thermodynamic cost of reliability and low temperatures: Tightening Landauer's principle and the second law. *International Journal of Theoretical Physics*, 39(12):2717–2753, 2000.

[5] N. Linden, S. Popescu, and P. Skrzypczyk. How small can thermal machines be? The smallest possible refrigerator. *Phys. Rev. Lett.*, 105:130401, Sep 2010.

[6] O. C. O. Dahlsten, R. Renner, E. Rieper, and V. Vedral. Inadequacy of von Neumann entropy for characterizing extractable work. *New Journal of Physics*, 13(5):053015, 2011.

[7] L. del Rio, J. Åberg, R. Renner, O. C. O. Dahlsten, and V. Vedral. The thermodynamic meaning of negative entropy. *Nature*, 474(7349):61–63, 06 2011.

[8] D. Egloff, O. C. O. Dahlsten, R. Renner, and V. Vedral. Laws of thermodynamics beyond the von neumann regime. *arXiv preprint arXiv:1207.0434*, 2012.

[9] P. Faist, F. Dupuis, J. Oppenheim, and R. Renner. A quantitative landauer's principle. *arXiv preprint arXiv:1211.1037*, 2012.

[10] M. Horodecki and J. Oppenheim. Fundamental limitations for quantum and nanoscale thermodynamics. *Nat Commun*, 4, 06 2013.

[11] J. Åberg. Truly work-like work extraction via a single-shot analysis. *Nat Commun*, 4, 06 2013.

[12] F. G. S. L. Brandão, M. Horodecki, J. Oppenheim, J. M. Renes, and R. W. Spekkens. Resource theory of quantum states out of thermal equilibrium. *Phys. Rev. Lett.*, 111:250404, Dec 2013.

[13] P. Skrzypczyk, A. J. Short, and S. Popescu. Extracting work from quantum systems. *arXiv preprint arXiv:1302.2811*, 2013.

[14] P. Skrzypczyk, A. J. Short, and S. Popescu. Work extraction and thermodynamics for individual quantum systems. *Nat Commun*, 5, 06 2014.

[15] N. Yunger Halpern and J. M. Renes. Beyond heat baths: Generalized resource theories for small-scale thermodynamics. *arXiv preprint arXiv:1409.3998*, 2014.

[16] N. Yunger Halpern. Beyond heat baths II: Framework for generalized thermodynamic resource theories. *arXiv preprint arXiv:1409.7845*, 2014.

[17] F. G. S. L. Brandão, M. Horodecki, N. Ng, J. Oppenheim, and S. Wehner. The second laws of quantum thermodynamics. *Proceedings of the National Academy of Sciences*, 112(11):3275–3279, 2015.

[18] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.*, 70:1895–1899, Mar 1993.

[19] C. H. Bennett and S. J. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Phys. Rev. Lett.*, 69:2881–2884, Nov 1992.

[20] C. H. Bennett and G. Brassard. Public key distribution and coin tossing. In *Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing*, pages 175–179, Bangalore, India, 1984.

[21] A. K. Ekert. Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.*, 67:661–663, Aug 1991.

[22] D. Mayers and A. Yao. Self testing quantum apparatus. *Quantum Info. Comput.*, 4(4):273–286, July 2004.

[23] B. W. Reichardt, F. Unger, and U. Vazirani. A classical leash for a quantum system: Command of quantum systems via rigidity of CHSH games. *arXiv preprint arXiv:1209.0448*, 2012.

[24] L. Hardy. Quantum theory from five reasonable axioms. *arXiv preprint quant-ph/0101012*, 2001.

[25] J. Barrett. Information processing in generalized probabilistic theories. *Phys. Rev. A*, 75:032304, Mar 2007.

[26] G. M. D'Ariano. Probabilistic theories: what is special about quantum mechanics? In A. Bokulich and G. Jaeger, editors, *Philosophy of Quantum Information and Entanglement*, pages 85–126. Cambridge University Press, Cambridge, 2010.

[27] H. Barnum, J. Barrett, M. Leifer, and A. Wilce. Generalized no-broadcasting theorem. *Phys. Rev. Lett.*, 99:240501, Dec 2007.

[28] G. Chiribella, G. M. D'Ariano, and P. Perinotti. Probabilistic theories with purification. *Phys. Rev. A*, 81:062348, Jun 2010.

[29] G. Chiribella, G. M. D'Ariano, and P. Perinotti. Informational derivation of quantum theory. *Phys. Rev. A*, 84:012311, Jul 2011.

[30] L. Hardy. Foliable operational structures for general probabilistic theories. In H. Halvorson, editor, *Deep Beauty: Understanding the Quantum World through Mathematical Innovation*, pages 409–442. Cambridge University Press, Cambridge, 2011.

[31] L. Hardy. Reformulating and reconstructing quantum theory. *arXiv preprint arXiv:1104.2066*, 2011.

[32] H. Barnum and A. Wilce. Information processing in convex operational theories. *Electronic Notes in Theoretical Computer Science*, 270(1):3 – 15, 2011. Proceedings of the Joint 5th International Workshop on Quantum Physics and Logic and 4th Workshop on Developments in Computational Models (QPL/DCM 2008).

[33] B. Coecke, T. Fritz, and R. W. Spekkens. A mathematical theory of resources. *arXiv preprint arXiv:1409.5531*, 2014.

[34] A. J. Short and J. Barrett. Strong nonlocality: a trade-off between states and measurements. *New Journal of Physics*, 12(3):033034, 2010.

[35] G. Auletta. *Foundations and interpretation of quantum mechanics.* World Scientific, Singapore, 2000.

[36] L. Hardy and R. Spekkens. Why physics needs quantum foundations. *Phys. Can.*, 66:73, 2010.

[37] P. Ball. Quantum quest. *Nature*, 501(7466):154–156, 2013.

[38] G. Chiribella and C. M. Scandolo. Conservation of information and the foundations of quantum mechanics. *EPJ Web of Conferences*, 95:03003, 2015.

[39] M. Maggiore. A generalized uncertainty principle in quantum gravity. *Physics Letters B*, 304(1–2):65–69, 1993.

[40] B. Coecke. Quantum picturalism. *Contemporary Physics*, 51:59, 2010.

[41] G. Chiribella. Dilation of states and processes in operational-probabilistic theories. In B. Coecke, I. Hasuo, and P. Panangaden, editors, Proceedings 11th workshop on *Quantum Physics and Logic,* Kyoto, Japan, 4-6th June 2014, volume 172 of *Electronic Proceedings in Theoretical Computer Science*, pages 1–14. Open Publishing Association, 2014.

[42] L. Hardy. Reconstructing quantum theory. *arXiv preprint arXiv:1303.1538*, 2013.

[43] S. Abramsky and B. Coecke. A categorical semantics of quantum protocols. In *Logic in Computer Science, 2004. Proceedings of the 19th Annual IEEE Symposium on*, pages 415–425, July 2004.

[44] B. Coecke. Kindergarten quantum mechanics. *arXiv preprint quant-ph/0510032*, 2005.

[45] S. Abramsky and B. Coecke. Categorical quantum mechanics. In K. Engesser, D. M. Gabbay, and D. Lehmann, editors, *Handbook of Quantum Logic and Quantum Structures: Quantum Logic*, pages 261–324. Elsevier, 2008.

[46] B. Coecke and É. O. Paquette. Categories for the practising physicist. In B. Coecke, editor, *New Structures for Physics*, volume 813 of *Lecture Notes in Physics*, pages 173–286. Springer, Berlin, 2011.

[47] P. Selinger. A survey of graphical languages for monoidal categories. In B. Coecke, editor, *New Structures for Physics*, pages 289–356. Springer, Berlin, 2011.

[48] S. Mac Lane. *Categories for the working mathematician*. Springer Science & Business Media, 1978.

[49] A. Gleason. Measures on the closed subspaces of a Hilbert space. *Indiana Univ. Math. J.*, 6:885–893, 1957.

[50] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, Cambridge, 2010.

[51] J. Preskill. Lecture notes on quantum computation.

[52] M. M. Wilde. *Quantum information theory*. Cambridge University Press, Cambridge, 2013.

[53] W. K Wootters. Local accessibility of quantum states. In W. H. Zurek, editor, *Complexity, entropy and the physics of information*, pages 39–46. Westview Press, 1990.

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

[55] C. M. Scandolo. Entanglement and thermodynamics in general probabilistic theories. Master's thesis, Università degli Studi di Padova, July 2014.

[56] L. Hardy. Towards quantum gravity: a framework for probabilistic theories with non-fixed causal structure. *Journal of Physics A: Mathematical and Theoretical*, 40(12):3081, 2007.

[57] G. Chiribella, G. M. D'Ariano, and P. Perinotti. Transforming quantum operations: Quantum supermaps. *EPL (Europhysics Letters)*, 83(3):30004, 2008.

[58] G. B. Folland. *Real analysis: Modern techniques and their applications*. John Wiley & Sons, 2013.

[59] P. Billingsley. *Probability and measure*. John Wiley & Sons, 2008.

[60] F. G. S. L. Brandão and G. Gour. The general structure of quantum resource theories. *arXiv preprint arXiv:1502.03149*, 2015.

[61] D. Deutsch and R. Jozsa. Rapid solution of problems by quantum computation. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 439(1907):553–558, 1992.

[62] L. K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the 28th Annual ACM Symposium on the Theory of Computing (STOC)*, pages 212–219, May 1996.

[63] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca. Quantum algorithms revisited. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 454(1969):339–354, 1998.

[64] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Review*, 41(2):303–332, 1999.

[65] E. Lubkin and T. Lubkin. Average quantal behavior and thermodynamic isolation. *International Journal of Theoretical Physics*, 32(6):933–943, 1993.

[66] J. Gemmer, A. Otte, and G. Mahler. Quantum approach to a derivation of the second law of thermodynamics. *Phys. Rev. Lett.*, 86:1927–1930, Mar 2001.

[67] S. Popescu, A. J. Short, and A. Winter. Entanglement and the foundations of statistical mechanics. *Nat Phys*, 2(11):754–758, 2006.

[68] S. Goldstein, J. L. Lebowitz, R. Tumulka, and N. Zanghì. Canonical typicality. *Phys. Rev. Lett.*, 96:050403, Feb 2006.

[69] J. Gemmer, M. Michel, and G. Mahler. *Quantum Thermodynamics: Emergence of Thermodynamic Behavior Within Composite Quantum Systems*, volume 784 of *Lecture Notes in Physics*. Springer Verlag, Heidelberg, 2009.

[70] M. P. Müller, D. Gross, and J. Eisert. Concentration of measure for quantum states with a fixed expectation value. *Communications in Mathematical Physics*, 303(3):785–824, 2011.

[71] F. G. S. L. Brandão and M. Cramer. Equivalence of statistical mechanical ensembles for non-critical quantum systems. *arXiv preprint arXiv:1502.03263*, 2015.

[72] A. Uhlmann. Sätze über Dichtematrizen. *Wiss. Z. Karl-Marx-Univ. Leipzig*, 20:633, 1971.

[73] A. Uhlmann. Endlich-dimensionale Dichtematrizen I. *Wiss. Z. Karl-Marx-Univ. Leipzig*, 21:421, 1972.

[74] A. Uhlmann. Endlich-dimensionale Dichtematrizen II. *Wiss. Z. Karl-Marx-Univ. Leipzig*, 22:139, 1973.

[75] M. A. Nielsen. Conditions for a class of entanglement transformations. *Phys. Rev. Lett.*, 83:436–439, Jul 1999.

[76] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. Quantum entanglement. *Rev. Mod. Phys.*, 81:865–942, Jun 2009.

[77] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters. Mixed-state entanglement and quantum error correction. *Phys. Rev. A*, 54:3824–3851, Nov 1996.

[78] H.-K. Lo and S. Popescu. Concentrating entanglement by local actions: Beyond mean values. *Phys. Rev. A*, 63:022301, Jan 2001.

[79] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge University Press, Cambridge, 1952.

[80] A. W. Marshall, I. Olkin, and B. C. Arnold. *Inequalities: Theory of Majorization and Its Applications*. Springer Series in Statistics. Springer, New York, 2011.

[81] G. Birkhoff. Tres observaciones sobre el algebra lineal. *Univ. Nac. Tucumán Rev. Ser. A*, 5:147–151, 1946.

[82] Godfrey H Hardy, John E Littlewood, and Gyorgy Pólya. Some simple inequalities satisfied by convex functions. *Messenger Math*, 58(145–152):310, 1929.

[83] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

[84] J. von Neumann. *Mathematical foundations of quantum mechanics*. Princeton University Press, Princeton, 1955.

[85] M. Horodecki, P. Horodecki, and J. Oppenheim. Reversible transformations from pure to mixed states and the unique measure of information. *Phys. Rev. A*, 67:062104, Jun 2003.

[86] G. Gour, M. P. Müller, V. Narasimhachar, R. W. Spekkens, and N. Yunger Halpern. The resource theory of informational nonequilibrium in thermodynamics. *Physics Reports*, 583(0):1–58, 2015.

[87] P. W. Shor. Structure of unital maps and the asymptotic quantum Birkhoff conjecture. presentation, 2010.

[88] S. D. Bartlett, T. Rudolph, and R. W. Spekkens. Reference frames, superselection rules, and quantum information. *Rev. Mod. Phys.*, 79:555–609, Apr 2007.

[89] G. Gour and R. W. Spekkens. The resource theory of quantum reference frames: manipulations and monotones. *New Journal of Physics*, 10(3):033023, 2008.

[90] I. Marvian and R. W. Spekkens. The theory of manipulations of pure state asymmetry: I. Basic tools, equivalence classes and single copy transformations. *New Journal of Physics*, 15(3):033001, 2013.

[91] I. Marvian and R. W. Spekkens. Extending Noether's theorem by quantifying the asymmetry of quantum states. *Nat Commun*, 5, 05 2014. Article.

[92] G. Chiribella and C. M. Scandolo. Entanglement and thermodynamics in general probabilistic theories. *arXiv preprint arXiv:1504.07045*, 2015.

[93] H. Barnum, J. Barrett, M. Leifer, and A. Wilce. Teleportation in general probabilistic theories. In *Proceedings of Symposia in Applied Mathematics*, volume 71, pages 25–48, 2012.

[94] E. C. G. Stueckelberg. Quantum theory in real Hilbert space. *Helv. Phys. Acta*, 33(8):727–752, 1960.

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

[96] N. S. Jones and L. Masanes. Interconversion of nonlocal correlations. *Phys. Rev. A*, 72:052312, Nov 2005.

[97] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23:880–884, Oct 1969.

[98] B. S. Cirel'son. Quantum generalizations of Bell's inequality. *Letters in Mathematical Physics*, 4(2):93–100, 1980.

[99] B. S. Cirel'son. Some results and problems on quantum Bell-type inequalities. *Hadronic J. Suppl.*, 8:329–345, 1993.

[100] S. Popescu and D. Rohrlich. Quantum nonlocality as an axiom. *Foundations of Physics*, 24(3):379–385, 1994.

[101] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts. Nonlocal correlations as an information-theoretic resource. *Phys. Rev. A*, 71:022101, Feb 2005.

[102] M. P. Müller and L. Masanes. Three-dimensionality of space and the quantum bit: an information-theoretic approach. *New Journal of Physics*, 15(5):053040, 2013.

[103] Borivoje Dakić and Caslav Brukner. Quantum theory and beyond: is entanglement special? In H. Halvorson, editor, *Deep Beauty: Understanding the Quantum World through Mathematical Innovation*, pages 365–392. Cambridge University Press, Cambridge, 2011.

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

[105] M. P. Müller and L. Masanes. Information-theoretic postulates for quantum theory. *arXiv preprint arXiv:1203.4516*, 2012.

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

[107] G. B Folland. *A course in abstract harmonic analysis*. CRC press, 1994.

[108] C. Carathéodory. Über den Variabilitätsbereich der Fourierschen Konstanten von positiven harmonischen Funktionen. *Rendiconti del Circolo Matematico di Palermo*, 32(1):193–217, 1911.

[109] E. Steinitz. Bedingt konvergente Reihen und konvexe Systeme. *J. Reine Angew. Math.*, 143(143):128–175, 1913.

[110] E. Schrödinger. Discussion of probability relations between separated systems. *Mathematical Proceedings of the Cambridge Philosophical Society*, 31:555–563, 10 1935.

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

[112] G. Chiribella and C. M. Scandolo. Operational axioms for state diagonalization. *arXiv preprint arXiv:1506.00380*, 2015.

[113] H. Barnum, J. Barrett, L. Orloff Clark, M. Leifer, R. Spekkens, N. Stepanik, A. Wilce, and R. Wilke. Entropy and information causality in general probabilistic theories. *New Journal of Physics*, 12(3):033024, 2010.

[114] A. J. Short and S. Wehner. Entropy in general physical theories. *New Journal of Physics*, 12(3):033023, 2010.

[115] G. Kimura, K. Nuida, and H. Imai. Distinguishability measures and entropies for general probabilistic theories. *Reports on Mathematical Physics*, 66(2):175 – 206, 2010.

[116] R. Landauer. Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.*, 5(3):183–191, July 1961.

[117] C. H. Bennett. Logical reversibility of computation. *IBM J. Res. Dev.*, 17(6):525–532, November 1973.

[118] E. Fredkin and T. Toffoli. Conservative logic. *International Journal of Theoretical Physics*, 21(3–4):219–253, 1982.

[119] M. Keyl, T. Matsui, D. Schlingemann, and R. F. Werner. Entanglement, Haag-duality and type properties of infinite quantum spin chains. *Reviews in Mathematical Physics*, 18(09):935–970, 2006.