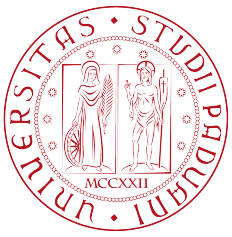


Università degli Studi di Padova

Dipartimento di Fisica e Astronomia

Corso di Laurea Magistrale in Fisica

Tesi di Laurea Magistrale

ENTANGLEMENT AND THERMODYNAMICS IN GENERAL PROBABILISTIC THEORIES

Relatore
Prof. Pieralberto Marchetti

Laureando
Carlo Maria Scandolo

Correlatore
Prof. Giulio Chiribella

Anno Accademico 2013–2014

Abstract

Since the early works of Einstein-Podolsky-Rosen and Schrödinger, entanglement is universally considered one of the most distinctive and puzzling features of quantum mechanics. In traditional introductions to the topics, entanglement is presented as a consequence of the linear structure of the Hilbert space, which imposes that composite systems must have some pure states —the "entangled states"— that are not the product of pure states of the component systems. But is entanglement just a mathematical accident of the linearity of quantum mechanics, or perhaps a more fundamental feature related to the physical content of the theory? This thesis aims at giving a characterization of entanglement and of the transformations of entangled states only in terms of basic information-theoretic principles, without appealing to the specific details of the Hilbert space formalism of quantum mechanics. The principles used in this characterization provide a new angle on the foundations of thermodynamics, on the definition of entropic quantities, and on the relations between thermodynamics and information theory.

Contents

Introduction **5**

1 Operational framework of quantum theory **8**
1.1 Purification 9
1.1.1 Schmidt decomposition 9
1.1.2 Purification 13
1.2 POVMs 15
1.3 Quantum channels 18
1.4 Quantum instruments 22
1.5 Axiomatic approach to quantum operations 28

2 Operational probabilistic theories **34**
2.1 Basic notions 36
2.1.1 Systems and tests 36
2.1.2 Sequential and parallel composition 38
2.1.3 States, effects and transformations 45
2.2 Pure conditioning 51
2.3 Causality 55
2.3.1 Operational norms 60

3 The purification postulate **67**
3.1 The purification postulate 67
3.2 Choi correspondence 76

4 Entanglement **89**
4.1 Entanglement and mixedness 90
4.2 The relation "to be more entangled than" 93
4.2.1 Mathematical properties 96

4.3 The relation "to be more mixed than" 99
4.3.1 Mathematical properties 101
4.4 Equivalence between entanglement and mixedness 103
4.4.1 More mixed implies more entangled 103
4.4.2 More entangled implies more mixed 105
4.5 Lo-Popescu theorem 107

5 Diagonalizing mixed states 118
5.1 Perfect distinguishability 119
5.2 Diagonalizing mixed states 125

6 A two-level system 133
6.1 A two-level system 133
6.2 Majorization and its properties 139
6.2.1 Heuristic introduction 139
6.2.2 Majorization 140
6.2.3 Mathematical properties 143
6.3 Mixedness relation in a two-level system 153
6.4 Schur-concave functions 157
6.4.1 Rényi entropies 163
6.5 Schur-concave functions for a two-level system 170

7 General measures of mixedness 172
7.1 A d-level system 172
7.2 Majorization in a d-level system 174
7.3 Shannon entropy for d-level systems 176
7.4 A derivation of the second law of thermodynamics 184

Conclusions 186

Acknowledgements 188

A Some useful mathematical results 189
A.1 Some theorems 189

Introduction

Thermodynamics has proven to be one of the most successful physical theories, for its applications range from physics, to chemistry, up to biological sciences.

One of the most puzzling aspects of thermodynamics is surely irreversibility and its second law, which inspired many discussions in the community of physicists. Related to this, a paradigmatic issue is the famous Maxwell's demon paradox [1, 2], which is directly related to the notion of information. A further step towards a tight relationship between thermodynamics and information theory came with the work by Szilard [3], but perhaps the most surprising contribution is Landauer's principle [4]. Landauer discovered that irreversible computation in computer feeds entropy to the environment. Therefore, in order to restrain thermal dissipation in computers, the idea of reversible computation was developed [7, 8].

The idea that thermodynamics is related to information theory should not surprise, if one concurs with Wheeler's opinion that each area of physics should be reread from the point of view of information theory [9]. In this vein, this work is aimed at setting an information-theoretic basis for the foundations of thermodynamics. We deal with this issue not in the framework of classical or quantum theory, but in the framework of a generic probabilistic theory. A general probabilistic theory is a physical theory that admits probabilistic processes. Addressing the foundations of thermodynamics for a general probabilistic theory, we will not be bound to the details of a specific physical theory, but we will be able to tackle the issue from a purely operational viewpoint, which means from the way information is processed in a theory.

The central point of our analysis is the purification postulate. Loosely speaking, this means that even when we have partial information about the system we are examining, we can recover a complete picture if we extend our

viewpoint to a larger system including also the environment. This postulate has a strong connection with the customary procedure in thermodynamics of enlarging the system to deal with a larger isolated system. In other words, the purification postulate expresses a sort of information conservation principle: information can never be destroyed, it can only be discarded [10].

In this work, we will follow the route charted by Thirring [11]: he sets the foundations of thermodynamics on measures of mixedness, namely on measures of the quantity of information we have. Therefore, a large part of this work will be devoted to the analysis of mixedness and of its measures. Eventually, it will turn out that a particular measure of mixedness fulfils an inequality that can be interpreted as the second law of thermodynamics.

Thanks to the purification postulate, the subject of mixedness is closely related to entanglement. Entanglement has been proven to be one of the essential ingredients to build the foundations of statistical mechanics [12]: we need no more to resort to the equal a priori probability postulate [14]. According to this postulate, the state of a system is equally likely to be each of the states compatible with its thermodynamic properties. A common argument to justify the equal a priori probability postulate is the ergodic hypothesis. However, using entanglement between a system and the environment, the equal a priori probability postulate is no more a postulate, because it has been proven that almost every state of the system fulfils it.

In this thesis, one of the main original results is a general proof that the connection between entanglement and mixedness is much closer than what we may think at first glance. Indeed, entanglement is a powerful tool for communication purposes, whereas mixedness means lack of information. Yet these two aspects are so related that measures of mixedness are also measures of entanglement. Therefore, it is completely equivalent to build the foundations of thermodynamics starting from entanglement or from mixedness. The other important and original result is an operational procedure of diagonalization of mixed states, without any references to the Hilbert space formalism.

The overview of the present work is as follows. In chapter 1, we will introduce the basic operational formalism for quantum theory. In this way, the reader is introduced to a first example of operational formalism in the familiar context of quantum mechanics. Then we are ready to deal with the operational formalism for general probabilistic theories, which is based on category theory. We will present it in chapter 2, and in the same chapter, we will start setting some reasonable axioms, which might be relevant also for

thermodynamics. Chapter 3 is entirely devoted to the purification postulate, and to explore its consequences for general probabilistic theories. We will see that this postulate is the essence of every admissible quantum theory, because it accounts for the fact that ignorance about a part is always compatible with maximal knowledge about the whole, a feature that Schrödinger thought as the actual essence of quantum theory [15]. In chapter 4 we come to the original part of this work and to the first major new result: we develop some tools to order states according to their entanglement or their mixedness. We will show that these two orderings are perfectly equivalent, establishing a tight relationship between entanglement and mixedness for general probabilistic theories. In chapter 5, adding a new axiom concerning the issue of distinguishing states, we devise a diagonalization procedure for mixed states even in a general probabilistic theory, and this is the second major original result. The remaining two chapters are devoted to exploring methods to measure mixedness based on the abstract version of the eigenvalues of mixed states. In particular, in chapter 6, we present the formalism of majorization, a widely used tool to measure mixedness of probability distributions in statistics. This tool is closely related to measures of mixedness, namely functions that quantify the mixedness of a given state by assigning a number to it. We apply these new tools to arbitrary d-level systems in chapter 7. We will study the properties of a particular measure of mixedness, managing to prove an abstract version of the second law of thermodynamics.

Chapter 1

Operational framework of quantum theory

In this first chapter, we invite the reader to familiarize with the basic operational formalism for a well-known and established theory: quantum theory. This formalism will serve as a guideline to set the operational formalism for general probabilistic theories. In this way, the reader can get sufficient familiarity with operational formalism before coming to the abstract version, which we will use throughout this work and which we will introduce in the next two chapters.

Following the approach of [16], operational formalism for quantum theory originates as a generalization of traditional formalism for isolated systems. We assume the reader to be already familiar with formalism for isolated systems, as well as with mixed states and partial traces.

The simplest way to achieve this generalization is to consider open quantum systems: for these systems, we are forced to abandon traditional formalism in favour of a more general one. The route to this generalization closely follows a method widely used in thermodynamics: whenever we have an open system, we enlarge our viewpoint to include the environment in our treatment. In this way, we end up with an isolated system, where traditional formalism of quantum mechanics still holds. Then, the final step is to restrict our attention to the original open system, by discarding the environment performing a partial trace over it.

Quite surprisingly, quantum theory, unlike classical theory, has an extremely interesting and important feature: each extension of traditional formalism actually originates from formalism for a larger isolated system. This

means that the procedure of considering an open system as a part of a larger isolated system is not a mere matter of convenience, but it is fully justified by the theory itself. This makes quantum theory the natural framework in which to develop a theory of thermodynamics.

In this chapter, some of the proofs are very long and technical, therefore we decided to omit them, for they are beyond the scope of this work. Anyway, references will be given to the benefit of the interested reader.

1.1 Purification

We begin presenting purification in quantum mechanics. This property will play a central role throughout this work, and it is a distinctive feature of quantum theory. However, we must first familiarize with Schmidt decomposition, which is a powerful tool. From now on, we will assume that every state vector is normalized.

1.1.1 Schmidt decomposition

In this section we consider systems made up by two subsystems. The states of such systems are called *bipartite states*. A bipartite pure state can be expressed as a particular linear combination of pure states.

Theorem 1.1.1 (Schmidt decomposition)**.** *Let $\mathcal{H} = \mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{B}}$. Every bipartite pure state $|\psi\rangle_{\mathrm{AB}}$ can be expressed as*

$$|\psi\rangle_{\mathrm{AB}} = \sum_j \sqrt{p_j}\, |j\rangle_{\mathrm{A}}\, |j'\rangle_{\mathrm{B}}\,, \tag{1.1}$$

where p_j are the eigenvalues of the marginal state[1] *ρ_{A} on subsystem* A *and $|j\rangle_{\mathrm{A}}$ and $|j'\rangle_{\mathrm{B}}$ are eigenvectors of ρ_{A} and ρ_{B} respectively, where ρ_{B} is the marginal state of $|\psi\rangle_{\mathrm{AB}}$ on subsystem* B*.*

Proof. Any vector $|\psi\rangle_{\mathrm{AB}}$ in $\mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{B}}$ can be expressed as

$$|\psi\rangle_{\mathrm{AB}} = \sum_{j,k} c_{jk}\, |j\rangle_{\mathrm{A}}\, |k\rangle_{\mathrm{B}}\,, \tag{1.2}$$

[1]Recall the marginal state of $|\psi\rangle_{\mathrm{AB}}$ on $\mathcal{H}_{\mathrm{A}}$ is a density operator defined as $\rho_{\mathrm{A}} = \mathrm{tr}_{\mathrm{B}}\, |\psi\rangle_{\mathrm{AB}} \langle\psi|_{\mathrm{AB}}$.

where $\{|j\rangle_{\rm A}\}$ is an orthonormal basis for $\mathcal{H}_{\rm A}$ and $\{|k\rangle_{\rm B}\}$ is an orthonormal basis for $\mathcal{H}_{\rm B}$. Let us define

$$\left|\widetilde{j}\right\rangle_{\rm B} := \sum_k c_{jk} \, |k\rangle_{\rm B} \, ,$$

then

$$|\psi\rangle_{\rm AB} = \sum_j |j\rangle_{\rm A} \left|\widetilde{j}\right\rangle_{\rm B}$$

Here $\left|\widetilde{j}\right\rangle_{\rm B}$'s need not to be mutually orthogonal or normalized. Some of them can even be the null vector. Suppose we have chosen the basis $\{|j\rangle_{\rm A}\}$ to be a basis of eigenvectors of $\rho_{\rm A}$,

$$\rho_{\rm A} = \sum_j p_j \, |j\rangle_{\rm A} \, \langle j|_{\rm A} \, . \tag{1.3}$$

We can compute $\rho_{\rm A}$ also as a partial trace of $|\psi\rangle_{\rm AB} \, \langle\psi|_{\rm AB}$.

$$\rho_{\rm A} = {\rm tr}_{\rm B} \left[\sum_{j,k} \left(|j\rangle_{\rm A} \, \langle k|_{\rm A} \otimes \left|\widetilde{j}\right\rangle_{\rm B} \left\langle\widetilde{k}\right|_{\rm B} \right) \right] = \sum_{j,k,l} |j\rangle_{\rm A} \, \langle k|_{\rm A} \left\langle l \left| \widetilde{j} \right\rangle_{\rm B} \left\langle \widetilde{k} \right| l \right\rangle_{\rm B} =$$

$$= \sum_{j,k,l} |j\rangle_{\rm A} \, \langle k|_{\rm A} \sum_l \left\langle \widetilde{k} \right| l \right\rangle_{\rm B} \left\langle l \left| \widetilde{j} \right\rangle_{\rm B} = \sum_{j,k} \left\langle \widetilde{k} \left| \widetilde{j} \right\rangle_{\rm B} \left(|j\rangle_{\rm A} \, \langle k|_{\rm A} \right)$$

Comparing this expression of $\rho_{\rm A}$ with the one in eq. (1.3), we find that $\left\langle \widetilde{k} \left| \widetilde{j} \right\rangle_{\rm B} = p_j \delta_{jk}$. This means that the vectors $\left|\widetilde{j}\right\rangle_{\rm B}$ are in fact orthogonal. We can rescale them to make them be an orthonormal set. Let us define

$$|j'\rangle_{\rm B} = \frac{1}{\sqrt{p_j}} \left|\widetilde{j}\right\rangle_{\rm B}$$

if $p_j \neq 0$. We can assume $p_j \neq 0$ because we will use $|j'\rangle_{\rm B}$ only for j appearing in eq. (1.3). Therefore we have

$$|\psi\rangle_{\rm AB} = \sum_j \sqrt{p_j} \, |j\rangle_{\rm A} \, |j'\rangle_{\rm B} \, . \tag{1.4}$$

Note that this expression is well-defined even if some of the p_j's are zero. If we compute the marginal state on system B starting from eq. (1.4), we have

$$\rho_{\rm B} = {\rm tr}_{\rm A} \ |\psi\rangle_{\rm AB} \, \langle\psi|_{\rm AB} = \sum_j p_j \, |j'\rangle_{\rm B} \, \langle j'|_{\rm B} \, .$$

This shows that $|j'\rangle_{\rm B}$'s are eigenvectors of $\rho_{\rm B}$. □

Eq. (1.1) is called *Schmidt decomposition* for $|\psi\rangle_{\mathrm{AB}}$. Note that Schmidt decomposition differs from a generic linear expansion of $|\psi\rangle_{\mathrm{AB}}$, such as eq. (1.2), because there is only one index j, irrespective of the fact that $\mathcal{H}_{\mathrm{A}}$ and $\mathcal{H}_{\mathrm{B}}$ may have different dimensions.

Corollary 1.1.2. *The two marginals of a bipartite pure state have the same non-zero eigenvalues, with the same degeneracy.*

Proof. Immediate from the construction of Schmidt decomposition, for p_j's are the non-zero eigenvalues of both ρ_{A} and ρ_{B}. □

The interesting point is that ρ_{A} and ρ_{B} have the same non-vanishing eigenvalues even if $\mathcal{H}_{\mathrm{A}}$ and $\mathcal{H}_{\mathrm{B}}$ have different dimensions. In that case, the difference of dimension is fully accounted by the presence of additional eigenvectors with zero eigenvalues.

Now we might wonder: given a bipartite pure state, is Schmidt decomposition unique? Clearly Schmidt coefficients cannot vary, the only possibility is that the vectors vary. If ρ_{A} and ρ_{B} have no degenerate eigenvalues, then the corresponding eigenvectors are fixed, and Schmidt decomposition is unique. This gives us also a tool to calculate Schmidt decomposition for states with non-degenerate marginals: we diagonalize ρ_{A} and ρ_{B} and we pair up eigenvectors of ρ_{A} and ρ_{B} with the same eigenvalue.

What if ρ_{A} and ρ_{B} are degenerate? Clearly, in this case, we can associate different bases of eigenvectors to ρ_{A} and ρ_{B}, so Schmidt decomposition is not unique, as the following example shows.

Example 1.1.3. Suppose $\mathcal{H}_{\mathrm{AB}} \approx \mathbb{C}^d \otimes \mathbb{C}^d$. Consider the bipartite pure state

$$|\psi\rangle_{\mathrm{AB}} = \frac{1}{\sqrt{d}} \sum_{j=1}^{d} |j\rangle_{\mathrm{A}} |j'\rangle_{\mathrm{B}} , \tag{1.5}$$

where $\{|j\rangle_{\mathrm{A}} |j'\rangle_{\mathrm{B}}\}$ is an orthonormal basis for $\mathcal{H}_{\mathrm{AB}}$. Eq. (1.5) is already a Schmidt decomposition of $|\psi\rangle_{\mathrm{AB}}$, and we can see that $\rho_{\mathrm{A}} = \rho_{\mathrm{B}} = \frac{1}{d}\mathbf{1}$, which is clearly degenerate. We can obtain another Schmidt decomposition if we consider a $d \times d$ unitary matrix U. We know that $\sum_{k=1}^{d} U_{kj}^* U_{kl} = \delta_{jl}$. Then

$$|\psi\rangle_{\mathrm{AB}} = \frac{1}{\sqrt{d}} \sum_{j,l=1}^{d} \delta_{jl} |j\rangle_{\mathrm{A}} |l'\rangle_{\mathrm{B}} = \frac{1}{\sqrt{d}} \sum_{j,k,l=1}^{d} U_{kj}^* U_{kl} |j\rangle_{\mathrm{A}} |l'\rangle_{\mathrm{B}} =$$

$$= \frac{1}{\sqrt{d}} \sum_{j,k,l=1}^{d} U_{kj}^{*} \left|j\right\rangle_{\mathrm{A}} U_{kl} \left|l'\right\rangle_{\mathrm{B}} .$$

Now define $\left|k\right\rangle_{\mathrm{A}} := \sum_{j=1}^{d} U_{kj}^{*} \left|j\right\rangle_{\mathrm{A}}$ and $\left|k'\right\rangle_{\mathrm{B}} := \sum_{l=1}^{d} U_{kl} \left|l'\right\rangle_{\mathrm{B}}$. In general, $\left|k\right\rangle_{\mathrm{A}} \neq \left|j\right\rangle_{\mathrm{A}}$ and $\left|k'\right\rangle_{\mathrm{B}} \neq \left|j'\right\rangle_{\mathrm{B}}$, but we can write

$$\left|\psi\right\rangle_{\mathrm{AB}} = \frac{1}{\sqrt{d}} \sum_{k=1}^{d} \left|k\right\rangle_{\mathrm{A}} \left|k'\right\rangle_{\mathrm{B}} ,$$

and this is another Schmidt decomposition for $\left|\psi\right\rangle_{\mathrm{AB}}$.

We conclude with the following lemma, which we will use in section 4.5.

Lemma 1.1.4. *Suppose $\mathcal{H}_{\mathrm{AB}} = \mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{B}}$ and let $\left|\psi\right\rangle_{\mathrm{AB}}$ and $\left|\phi\right\rangle_{\mathrm{AB}}$ be two bipartite pure states such that they have the same Schmidt coefficients. Then $\left|\phi\right\rangle_{\mathrm{AB}} = U_{\mathrm{A}} \otimes V_{\mathrm{B}} \left|\psi\right\rangle_{\mathrm{AB}}$.*

Proof. Suppose we have

$$\left|\psi\right\rangle_{\mathrm{AB}} = \sum_{j=1}^{n} \sqrt{p_j} \left|\alpha_j\right\rangle_{\mathrm{A}} \left|\beta_j\right\rangle_{\mathrm{B}}$$

and

$$\left|\phi\right\rangle_{\mathrm{AB}} = \sum_{j=1}^{n} \sqrt{p_j} \left|\alpha'_j\right\rangle_{\mathrm{A}} \left|\beta'_j\right\rangle_{\mathrm{B}}$$

Since $\left\{\left|\alpha_j\right\rangle_{\mathrm{A}}\right\}_{j=1}^{n}$ is an orthonormal set, it can be completed to an orthonormal basis. The same holds for $\left\{\left|\beta_j\right\rangle_{\mathrm{B}}\right\}_{j=1}^{n}$, $\left\{\left|\alpha'_j\right\rangle_{\mathrm{A}}\right\}_{j=1}^{n}$, and $\left\{\left|\beta'_j\right\rangle_{\mathrm{B}}\right\}_{j=1}^{n}$. Let $\left\{\left|\alpha_j\right\rangle_{\mathrm{A}}\right\}_{j=1}^{d_{\mathrm{A}}}$ be the completion of $\left\{\left|\alpha_j\right\rangle_{\mathrm{A}}\right\}_{j=1}^{n}$, and let $\left\{\left|\alpha'_j\right\rangle_{\mathrm{A}}\right\}_{j=1}^{d_{\mathrm{A}}}$ be the completion of $\left\{\left|\alpha'_j\right\rangle_{\mathrm{A}}\right\}_{j=1}^{n}$. Then we know there is a unitary operator U_{A} on A transforming $\left\{\left|\alpha_j\right\rangle_{\mathrm{A}}\right\}_{j=1}^{d_{\mathrm{A}}}$ into $\left\{\left|\alpha'_j\right\rangle_{\mathrm{A}}\right\}_{j=1}^{d_{\mathrm{A}}}$. A similar argument holds also for $\left\{\left|\beta_j\right\rangle_{\mathrm{B}}\right\}_{j=1}^{n}$ and $\left\{\left|\beta'_j\right\rangle_{\mathrm{B}}\right\}_{j=1}^{n}$, eventually yielding

$$U_{\mathrm{A}} \otimes V_{\mathrm{B}} \left|\psi\right\rangle_{\mathrm{AB}} = \sum_{j=1}^{n} \sqrt{p_j} \left|\alpha'_j\right\rangle_{\mathrm{A}} \left|\beta'_j\right\rangle_{\mathrm{B}} = \left|\phi\right\rangle_{\mathrm{AB}} .$$

□

1.1.2 Purification

Purification plays a crucial role in all this work, so it is better we start to familiarize with it in the context of quantum mechanics. It is simply the statement that every mixed state comes from a pure state in a larger system. In quantum theory, purification is a theorem, but in our treatment of general probabilistic theories we will promote it to a postulate.

The concept of purification can be defined as follows.

Definition 1.1.5. Let ρ be a state of a system A with Hilbert space $\mathcal{H}_{\mathrm{A}}$. We call *purification* of ρ a pure state $|\psi\rangle_{\mathrm{AB}}$ of a larger system $\mathcal{H}_{\mathrm{AB}} = \mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{B}}$ such that $\rho = \mathrm{tr}_{\mathrm{B}}\ |\psi\rangle_{\mathrm{AB}} \langle\psi|_{\mathrm{AB}}$.

System B is called *purifying system.*

It is clear that pure states can be purified. Indeed, if $|\alpha\rangle_{\mathrm{A}}$ is a pure state, a purification of $|\alpha\rangle_{\mathrm{A}}$ is, for example, $|\psi\rangle_{\mathrm{AB}} = |\alpha\rangle_{\mathrm{A}} |\beta\rangle_{\mathrm{B}}$, where $|\beta\rangle_{\mathrm{B}}$ is any pure state of any system B.

But the really interesting aspect is whether mixed states can be purified. In classical mechanics this is impossible: a bipartite state is pure if and only if it is the product of pure states, therefore we cannot distil purity from mixed states.

Surprisingly, in quantum theory, every state can be purified.

Theorem 1.1.6. *Every state ρ of system* A *can be purified. Moreover, if ρ has two purifications, they differ by a local unitary operator on the purifying system. In other words, if $|\psi\rangle_{\mathrm{AB}}$ and $|\phi\rangle_{\mathrm{AB}}$ are two purifications of the same state ρ of system* A*, we have*

$$|\phi\rangle_{\mathrm{AB}} = \mathbf{1}_{\mathrm{A}} \otimes U_{\mathrm{B}} |\psi\rangle_{\mathrm{AB}} .$$

Proof. Let us diagonalize ρ, as

$$\rho = \sum_{j=1}^{n} p_j |j\rangle_{\mathrm{A}} \langle j|_{\mathrm{A}} ,$$

where we assume $p_j > 0$, so $n \leq d_{\mathrm{A}}$, and $d_{\mathrm{A}} = \dim \mathcal{H}_{\mathrm{A}}$, being $\mathcal{H}_{\mathrm{A}}$ the Hilbert space associated with system A. Consider another system B with Hilbert space $\mathcal{H}_{\mathrm{B}}$, and the orthonormal set $\{|j'\rangle_{\mathrm{B}}\}_{j=1}^{n}$ for $\mathcal{H}_{\mathrm{B}}$. Recalling the construction of Schmidt decomposition, we have that

$$|\psi\rangle_{\mathrm{AB}} = \sum_{j} \sqrt{p_j} |j\rangle_{\mathrm{A}} |j'\rangle_{\mathrm{B}}$$

is a purification of ρ.

Suppose now we have two purifications of ρ, $|\psi\rangle_{\mathrm{AB}}$ and $|\phi\rangle_{\mathrm{AB}}$. By construction, we have

$$|\psi\rangle_{\mathrm{AB}} = \sum_{j=1}^{n} \sqrt{p_j}\, |j\rangle_{\mathrm{A}}\, |j_1'\rangle_{\mathrm{B}}$$

and

$$|\phi\rangle_{\mathrm{AB}} = \sum_{j=1}^{n} \sqrt{p_j}\, |j\rangle_{\mathrm{A}}\, |j_2'\rangle_{\mathrm{B}}\,.$$

Completing $\{|j_1'\rangle_{\mathrm{B}}\}_{j=1}^{n}$ and $\{|j_2'\rangle_{\mathrm{B}}\}_{j=1}^{n}$ to orthonormal bases of $\mathcal{H}_{\mathrm{B}}$, we find a unitary operator U_{B} on $\mathcal{H}_{\mathrm{B}}$ such that $U_{\mathrm{B}}\, |j_1'\rangle_{\mathrm{B}} = |j_2'\rangle_{\mathrm{B}}$ for every j. Therefore,

$$|\phi\rangle_{\mathrm{AB}} = \sum_{j=1}^{n} \sqrt{p_j}\, |j\rangle_{\mathrm{A}}\, |j_2'\rangle_{\mathrm{B}} = \sum_{j=1}^{n} \sqrt{p_j}\, |j\rangle_{\mathrm{A}}\, U_{\mathrm{B}}\, |j_1'\rangle_{\mathrm{B}} =$$

$$= \mathbf{1}_{\mathrm{A}} \otimes U_{\mathrm{B}}\, |\psi\rangle_{\mathrm{AB}}\,.$$

□

Loosely speaking, this theorem states that in quantum theory, whenever we do not have maximal information, we can "retrieve" it by enlarging the system conveniently. Moreover, this enlargement is not so arbitrary, because purification is essentially unique (up to unitary operators in the purifying system).

Theorem 1.1.6 gives us some information about how large the purifying system must be.

Corollary 1.1.7. *If ρ is a state of $\mathcal{H}_{\mathrm{A}}$, with $n \leq d_{\mathrm{A}}$ non-vanishing eigenvalues, where d_{A} is the dimension of $\mathcal{H}_{\mathrm{A}}$, then the purifying system $\mathcal{H}_{\mathrm{B}}$ is such that* $\dim \mathcal{H}_{\mathrm{B}} \geq n$.

Proof. Immediate from the construction of a purification of ρ. □

Now, let us move to the generalizations of the concept of operation on a quantum system. Loosely speaking, we can consider an operation every map that acts on quantum states. In the traditional formalism, operations are essentially unitary transformations (deterministic transformations) and orthogonal measurements (probabilistic transformations).

1.2 POVMs

We start analysing a first extension of orthogonal measurements. Recall that every projective measurement comprises two ingredients:

1. a rule to assign probabilities of the various outcomes of the measurement;

2. a rule to determine the state immediately after the measurement.

In some applications, we have little interest in knowing the state after the measurement, or we cannot simply have access to it because the measurement has destroyed the system (as when an electron is absorbed by a photographic plate). Therefore, the main concern when performing a measurement is how to assign probabilities to the various outcomes. This question will lead us directly to the formalism of POVMs as a generalization of projective measurements.

Consider an isolated system AB with Hilbert space $\mathcal{H}_{\mathrm{AB}} = \mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{B}}$, which is initially in a product state $\rho_{\mathrm{AB}} = \rho_{\mathrm{A}} \otimes \rho_{\mathrm{B}}$. Suppose we perform a projective measurement on $\mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{B}}$, described by mutually orthogonal projectors $\{E_a\}$. We know that outcome a occurs with probability

$$p_a = \mathrm{tr}_{\mathrm{AB}}\, E_a \left(\rho_{\mathrm{A}} \otimes \rho_{\mathrm{B}}\right).$$

Suppose we are interested only in probabilities, and we want to express p_a using only objects related to $\mathcal{H}_{\mathrm{A}}$. We can write

$$p_a = \mathrm{tr}_{\mathrm{A}} \left[\mathrm{tr}_{\mathrm{B}}\, E_a \left(\rho_{\mathrm{A}} \otimes \rho_{\mathrm{B}}\right)\right].$$

Let us evaluate $\mathrm{tr}_{\mathrm{B}} E_a \left(\rho_{\mathrm{A}} \otimes \rho_{\mathrm{B}}\right)$. Introducing an orthonormal basis $\{|j\rangle_{\mathrm{A}}\}$ for $\mathcal{H}_{\mathrm{A}}$ and an orthonormal basis $\{|\mu\rangle_{\mathrm{B}}\}$ for $\mathcal{H}_{\mathrm{B}}$, we have, using matrix elements,

$$\mathrm{tr}_{\mathrm{B}}\, E_a \left(\rho_{\mathrm{A}} \otimes \rho_{\mathrm{B}}\right) = \sum_{j,k,\mu,\nu} \left(E_a\right)_{j\mu,k\nu} \left(\rho_{\mathrm{A}}\right)_{kj} \left(\rho_{\mathrm{B}}\right)_{\nu\mu} =: \sum_{j,k} \left(F_a\right)_{jk} \left(\rho_{\mathrm{A}}\right)_{kj},$$

where we set

$$\left(F_a\right)_{jk} := \sum_{\mu,\nu} \left(E_a\right)_{j\mu,k\nu} \left(\rho_{\mathrm{B}}\right)_{\nu\mu}.$$

Now, recalling the definition of matrix elements

$$\langle j|_{\mathrm{A}}\, F_a\, |k\rangle_{\mathrm{A}} = \sum_{\mu,\nu} \langle j|_{\mathrm{A}} \langle \mu|_{\mathrm{B}}\, E_a\, |k\rangle_{\mathrm{A}} |\nu\rangle_{\mathrm{B}} \langle \nu|_{\mathrm{B}}\, \rho_{\mathrm{B}}\, |\mu\rangle_{\mathrm{B}},$$

therefore

$$F_a = \sum_{\mu,\nu} \langle\mu|_{\rm B} E_a |\nu\rangle_{\rm B} \langle\nu|_{\rm B} \rho_{\rm B} |\mu\rangle_{\rm B},$$

and recognizing Dirac's resolution of identity, we have

$$F_a = \sum_{\mu} \langle\mu|_{\rm B} E_a \rho_{\rm B} |\mu\rangle_{\rm B} = {\rm tr}_{\rm B}\, E_a \rho_{\rm B},$$

as we might have expected. Hence, we have

$$p_a = {\rm tr}_{\rm A}\, F_a \rho_{\rm A}.$$

Let us see the properties of this new operator F_a on $\mathcal{H}_{\rm A}$.

- It is hermitian, because

$$F_a^\dagger = {\rm tr}_{\rm B}\ (E_a\rho_{\rm B})^\dagger = {\rm tr}_{\rm B}\, \rho_{\rm B}^\dagger E_a^\dagger = {\rm tr}_{\rm B}\, \rho_{\rm B} E_a = F_a.$$

- It is positive. Indeed if $\{|\mu\rangle_{\rm B}\}$ is a basis of eigenvectors of $\rho_{\rm B}$, $\rho_{\rm B} = \sum_\mu p_\mu |\mu\rangle_{\rm B} \langle\mu|_{\rm B}$, then

$$\langle\psi|_{\rm A} F_a |\psi\rangle_{\rm A} = \sum_\mu p_\mu \left(\langle\psi|_{\rm A} \langle\mu|_{\rm B}\right) E_a \left(|\psi\rangle_{\rm A} |\mu\rangle_{\rm B}\right) \geq 0,$$

  because $p_\mu \geq 0$ and E_a is positive. This property is related to the fact that probabilities are non-negative.

- We have $\sum_a F_a = \mathbf{1}_{\rm A}$. Indeed,

$$\sum_a F_a = {\rm tr}_{\rm B} \sum_a E_a \rho_{\rm B} = {\rm tr}_{\rm B}\, \mathbf{1}_{\rm AB} \rho_{\rm B} = \mathbf{1}_{\rm A},$$

  where we used the fact that $\sum_a E_a = \mathbf{1}_{\rm AB}$. This property is related to the fact that probabilities must sum to 1.

An observer who has access only to system A will describe the measurement as a collection of operators $\{F_a\}$ with the properties just shown.

In this vein, we define general measurements as follows.

Definition 1.2.1. A *POVM*[2] on A is a collection of operators $\{F_a\}$ on $\mathcal{H}_{\rm A}$ such that

[2] *Positive-Operator-Valued Measure*

- F_a is hermitian, for every a;
- F_a is positive[3] (and $F_a \leq \mathbf{1}$), for every a;
- $\sum_a F_a = \mathbf{1}$.

Notice that $F_{a_0} \leq \mathbf{1}$ for every a_0 is necessary, otherwise $\sum_{a \neq a_0} F_a$ could not be a positive operator. This means nothing the probability of outcome a is less than or equal to 1.

We therefore set the following axiom.

Axiom 1.2.2. *A (demolition) measurement in quantum theory is described by a POVM $\{F_a\}$. The probability p_a of outcome a, given a state ρ, is*

$$p_a = \operatorname{tr} F_a \rho.$$

Note that POVMs do not give any information about the state after the measurement, either because we are not interested in it, or because the system has been destroyed (whence the term "demolition measurement").

A projective measurement $\{E_a\}$ is clearly a particular case of POVM, with the additional constraint $E_a E_b = \delta_{ab} E_a$.

In our line of reasoning, we derived some POVMs starting from a composite system, performing an orthogonal measurement on it, and then restricting to a subsystem. Does every POVM come from a similar procedure? The answer is affirmative [19].

Theorem 1.2.3 (Naimark)**.** *For every POVM $\{F_a\}$ on $\mathcal{H}_\mathrm{A}$, there exist a Hilbert space $\mathcal{H}_\mathrm{B}$, a density operator ρ_B on $\mathcal{H}_\mathrm{B}$, and a projective measurement $\{E_a\}$ on $\mathcal{H}_\mathrm{A} \otimes \mathcal{H}_\mathrm{B}$ such that*

$$\operatorname{tr}_\mathrm{A} F_a \rho_\mathrm{A} = \operatorname{tr}_\mathrm{AB} E_a \left(\rho_\mathrm{A} \otimes \rho_\mathrm{B}\right)$$

for every state ρ_A of $\mathcal{H}_\mathrm{A}$.

Proof. We omit the proof. The interested reader can see [16]. □

Even in Naimark's theorem we see that the characterizing feature of POVMs is probability: indeed, the extension is defined starting from probability. Therefore, two POVMs are equal if they yield the same probabilities for every state.

[3] We will often write $F_a \geq 0$. Moreover, recall that the writing $A \leq B$, where A and B are two operators, means that $B - A$ is a positive operator, or, in other words, $B - A \geq 0$.

1.3 Quantum channels

Now we start analysing the evolution of a quantum system when some physical operations that do not destroy it are performed. In this section we examine deterministic evolutions, namely when the output of some physical transformation is completely determined by the input. This means that no random or probabilistic processes occur.

The prototypes of all deterministic transformations are unitary transformations, which are the deterministic transformations occurring in an isolated system. We will see that if the system is not isolated, other types of deterministic transformations will appear.

Consider an isolated system AB with Hilbert space $\mathcal{H}_{\mathrm{AB}} = \mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{B}}$, which is initially in a product state $\rho_{\mathrm{AB}} = \rho_{\mathrm{A}} \otimes |0\rangle_{\mathrm{B}} \langle 0|_{\mathrm{B}}$. We are entitled to assume that system B is in a pure state, because, if this is not the case, we can always purify it enlarging system B conveniently.

The composite system undergoes unitary evolution described by unitary operator U_{AB}.

$$\rho_{\mathrm{AB}} \longmapsto U_{\mathrm{AB}} \left(\rho_{\mathrm{A}} \otimes |0\rangle_{\mathrm{B}} \langle 0|_{\mathrm{B}}\right) U_{\mathrm{AB}}^{\dagger}. \tag{1.6}$$

We want to find out how the state of subsystem A evolved. To do so, let us perform a partial trace over $\mathcal{H}_{\mathrm{B}}$.

$$\rho'_{\mathrm{A}} = \mathrm{tr}_{\mathrm{B}} \left[U_{\mathrm{AB}} \left(\rho_{\mathrm{A}} \otimes |0\rangle_{\mathrm{B}} \langle 0|_{\mathrm{B}}\right) U_{\mathrm{AB}}^{\dagger}\right] =$$

$$= \sum_k \langle k|_{\mathrm{B}} U_{\mathrm{AB}} |0\rangle_{\mathrm{B}} \rho_{\mathrm{A}} \langle 0|_{\mathrm{B}} U_{\mathrm{AB}}^{\dagger} |k\rangle_{\mathrm{B}},$$

where $\{|k\rangle_{\mathrm{B}}\}$ is an orthonormal basis for $\mathcal{H}_{\mathrm{B}}$. Here, $M_k := \langle k|_{\mathrm{B}} U_{\mathrm{AB}} |0\rangle_{\mathrm{B}}$ is a linear operator acting on $\mathcal{H}_{\mathrm{A}}$, whose matrix elements are

$$\langle j|_{\mathrm{A}} M_k |l\rangle_{\mathrm{A}} = \langle j|_{\mathrm{A}} \left(\langle k|_{\mathrm{B}} U_{\mathrm{AB}} |0\rangle_{\mathrm{B}}\right) |l\rangle_{\mathrm{A}} =$$

$$= \left(\langle j|_{\mathrm{A}} \langle k|_{\mathrm{B}}\right) U_{\mathrm{AB}} \left(|0\rangle_{\mathrm{B}} |l\rangle_{\mathrm{A}}\right).$$

Therefore we can express ρ'_{A} as

$$\rho'_{\mathrm{A}} = \sum_k M_k \rho_{\mathrm{A}} M_k^{\dagger}.$$

We have that the M_k's satisfy the property $\sum_k M_k^{\dagger} M_k = \mathbf{1}_{\mathrm{A}}$. Indeed,

$$\sum_k M_k^{\dagger} M_k = \sum_k \langle 0|_{\mathrm{B}} U_{\mathrm{AB}}^{\dagger} |k\rangle_{\mathrm{B}} \langle k|_{\mathrm{B}} U_{\mathrm{AB}} |0\rangle_{\mathrm{B}} =$$

$$= \langle 0|_{\mathrm{B}} U_{\mathrm{AB}}^{\dagger} U_{\mathrm{AB}} |0\rangle_{\mathrm{B}} = \langle 0|_{\mathrm{B}} \mathbf{1}_{\mathrm{AB}} |0\rangle_{\mathrm{B}} = \mathbf{1}_{\mathrm{A}}.$$

Let us check if the resulting state ρ'_{A} is still a density operator.

- ρ'_{A} is hermitian. Indeed,

$$(\rho'_{\mathrm{A}})^{\dagger} = \sum_k M_k \rho_{\mathrm{A}}^{\dagger} M_k^{\dagger} = \sum_k M_k \rho_{\mathrm{A}} M_k^{\dagger} = \rho'_{\mathrm{A}}.$$

- ρ'_{A} is positive. Indeed,

$$\langle \psi|_{\mathrm{A}} \rho'_{\mathrm{A}} |\psi\rangle_{\mathrm{A}} = \sum_k \left(\langle \psi|_{\mathrm{A}} M_k\right) \rho_{\mathrm{A}} \left(M_k^{\dagger} |\psi\rangle_{\mathrm{A}}\right) \geq 0,$$

  because ρ_{A} is positive.

- ρ'_{A} has unit trace. Indeed,

$$\mathrm{tr}\ \rho'_{\mathrm{A}} = \sum_k \mathrm{tr}\ M_k \rho_{\mathrm{A}} M_k^{\dagger} = \sum_k \mathrm{tr}\ M_k^{\dagger} M_k \rho_{\mathrm{A}} = \mathrm{tr}\ \rho_{\mathrm{A}} = 1.$$

  Hence the normalization condition $\sum_k M_k^{\dagger} M_k = \mathbf{1}_{\mathrm{A}}$ is related to the fact that the trace is preserved.

Now we can give the following definition of deterministic transformation, called quantum channel.

Definition 1.3.1. A *quantum channel* on $\mathcal{H}_{\mathrm{A}}$ is a deterministic transformation that acts on density operators on $\mathcal{H}_{\mathrm{A}}$ and transforms them into density operators on $\mathcal{H}_{\mathrm{A}}$ according to the rule

$$\rho \longmapsto \sum_k M_k \rho M_k^{\dagger},$$

where M_k's are operators on $\mathcal{H}_{\mathrm{A}}$, called *Kraus operators*, such that $\sum_k M_k^{\dagger} M_k = \mathbf{1}$.

Unitary transformations are special kinds of quantum channels, as shown in the following example.

Example 1.3.2. A unitary transformation on $\mathcal{H}_{\mathrm{A}}$ is a quantum channel. Indeed it acts on states as

$$\rho \longmapsto U\rho U^{\dagger}.$$

We will call it *unitary channel*, and we will write a unitary channel channel as

$$\mathcal{U}(\rho) = U\rho U^{\dagger}.$$

A unitary channel has only one Kraus operator, the unitary operator itself, for $UU^{\dagger} = \mathbf{1}$.

We can generalize our definition of quantum channel allowing an output system different from the input system.

Definition 1.3.3. A *quantum channel* from $\mathcal{H}_{\mathrm{A}}$ to $\mathcal{H}_{\mathrm{B}}$ is a deterministic transformation[4] that acts on density operators on $\mathcal{H}_{\mathrm{A}}$ and transforms them into density operators on $\mathcal{H}_{\mathrm{B}}$ according to the rule

$$\rho \longmapsto \sum_k M_k \rho M_k^{\dagger},$$

where M_k's are operators from $\mathcal{H}_{\mathrm{A}}$ to $\mathcal{H}_{\mathrm{B}}$, called *Kraus operators*, such that $\sum_k M_k^{\dagger} M_k = \mathbf{1}$.

This definition of quantum channel is often called *operator sum representation* of a quantum channel. We will give a more abstract definition in section 1.5.

For the time being, we can set the following axiom.

Axiom 1.3.4. *Every deterministic transformation from $\mathcal{H}_{\mathrm{A}}$ to $\mathcal{H}_{\mathrm{B}}$ is a quantum channel from $\mathcal{H}_{\mathrm{A}}$ to $\mathcal{H}_{\mathrm{B}}$.*

In the following treatment, for the sake of simplicity, we will deal with channel with equal input and output systems, but the generalization is straightforward.

In our derivation of quantum channel, we started from an isolated system, where the evolution is unitary. Is it true that every channel comes from a unitary transformation in a larger system? The answer comes from the following theorem by Stinespring [20].

[4]Recall that a deterministic transformation is a process in which the output state is completely determined by the input state. There is no randomness.

Theorem 1.3.5 (Stinespring)**.** *Every operator sum representation of a quantum channel comes from a unitary transformation in a larger system.*

Proof. Suppose a channel on $\mathcal{H}_{\mathrm{A}}$ has Kraus operators $\{M_k\}_{k=1}^n$. We choose a Hilbert space $\mathcal{H}_{\mathrm{B}}$ with $\dim \mathcal{H}_{\mathrm{B}} \geq n$. If $|\psi\rangle_{\mathrm{A}}$ is any pure state of $\mathcal{H}_{\mathrm{A}}$, $\{|k\rangle_{\mathrm{B}}\}_{k=1}^n$ is an orthonormal set for $\mathcal{H}_{\mathrm{B}}$, and $|0\rangle_{\mathrm{B}}$ is some pure state in $\mathcal{H}_{\mathrm{B}}$, define an operator U_{AB} on $\mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{B}}$ whose action is

$$U_{\mathrm{AB}} |\psi\rangle_{\mathrm{A}} |0\rangle_{\mathrm{B}} := \sum_{k=1}^{n} \left(M_k |\psi\rangle_{\mathrm{A}} \otimes |k\rangle_{\mathrm{B}}\right).$$

This action preserves scalar product, indeed, if we take another pure state $|\psi'\rangle_{\mathrm{A}}$ on $\mathcal{H}_{\mathrm{A}}$, we have

$$\left(\sum_{l=1}^{n} \langle\psi'|_{\mathrm{A}} M_l^\dagger \otimes \langle l|_{\mathrm{B}}\right)\left(\sum_{k=1}^{n} M_k |\psi\rangle_{\mathrm{A}} \otimes |k\rangle_{\mathrm{B}}\right) =$$

$$= \langle\psi'|_{\mathrm{A}} \sum_{k=1}^{n} M_k^\dagger M_k |\psi\rangle_{\mathrm{A}} = \langle\psi' \,|\, \psi\rangle_{\mathrm{A}} = \left(\langle\psi'|_{\mathrm{A}} \langle 0|_{\mathrm{B}}\right)\left(|\psi\rangle_{\mathrm{A}} |0\rangle_{\mathrm{B}}\right).$$

Therefore U_{AB} can be extended to a unitary operator acting on all $\mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{B}}$. Taking the partial trace over B, we find that

$$|\psi\rangle_{\mathrm{A}} \longmapsto \mathrm{tr}_{\mathrm{B}}\left(U_{\mathrm{AB}} |\psi\rangle_{\mathrm{A}} |0\rangle_{\mathrm{B}}\right)\left(\langle\psi|_{\mathrm{A}} \langle 0|_{\mathrm{B}} U_{\mathrm{AB}}^\dagger\right) =$$

$$= \sum_k M_k |\psi\rangle_{\mathrm{A}} \langle\psi|_{\mathrm{A}} M_k^\dagger.$$

Since every mixed state is a convex combination of pure states, we see that U_{AB} gives rise to the quantum channel

$$\rho \longmapsto \sum_k M_k \rho M_k^\dagger$$

in A. □

We see that non-unitary channels have the property of transforming separable pure states of the composite system into entangled pure states. Indeed, the action of U_{AB} on the separable state $|\psi\rangle_{\mathrm{A}} |0\rangle_{\mathrm{B}}$ is

$$|\psi\rangle_{\mathrm{A}} |0\rangle_{\mathrm{B}} \longmapsto \sum_{k=1}^{n} \left(M_k |\psi\rangle_{\mathrm{A}} \otimes |k\rangle_{\mathrm{B}}\right),$$

and the right-hand side is in general an entangled pure state. Therefore, quantum channels in general build up correlations between a system (A) and its environment (B).

Finally, we must note that, in general, quantum channels are not invertible, or as we will often say, are not *reversible*. Indeed, the following proposition holds.

Proposition 1.3.6. *A quantum channel on $\mathcal{H}_{\mathrm{A}}$ is reversible if and only if it is unitary.*

Proof. Omitted. For further details, see [16]. □

1.4 Quantum instruments

In section 1.2, we generalized orthogonal measurements as far as the probabilistic aspect is concerned. Now we want to find a generalization also for non-demolition measurements, that are measurements that do not destroy the system. So, it makes sense to wonder about the state after the measurement.

Recall that a projective measurement $\{E_a\}$ on a state ρ yields outcome a with probability $p_a = \operatorname{tr} E_a\rho$, and in this case the state immediately after the measurement is

$$\rho'_a = \frac{E_a\rho E_a}{\operatorname{tr} E_a\rho}.$$

Note that we must introduce a normalization factor $\operatorname{tr} E_a\rho$, that corresponds to the probability p_a, because the sole action of the projector E_a on ρ gives a state with trace less than or equal to 1, because $\operatorname{tr} E_a\rho E_a = \operatorname{tr} E_a\rho = p_a \leq 1$.

To find out what generalization we need, it is useful to proceed in the customary way of enlarging the system. The idea is to perform an orthogonal measurement on the ancillary system after the compound system has undergone unitary evolution.

Consider a unitary operator U_{AB} on the Hilbert space $\mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{B}}$, and suppose the compound system is initially in a product state $\rho_{\mathrm{AB}} = \rho_{\mathrm{A}} \otimes |0\rangle_{\mathrm{B}} \langle 0|_{\mathrm{B}}$. Let us take an orthonormal basis $\{|a\rangle_{\mathrm{B}}\}$ for $\mathcal{H}_{\mathrm{B}}$. We can describe a measurement that does not destroy system A as a unitary evolution U_{AB}, followed by a measurement on $\mathcal{H}_{\mathrm{B}}$. Because of entanglement, the state of system A will change. Suppose we choose the orthogonal measurement given by $\{|a\rangle_{\mathrm{B}} \langle a|_{\mathrm{B}}\}$. This is clearly an atomic (or pure) measurement, because

it projects on subspaces of dimension 1. This kind of measurement has maximum "resolving power".

After unitary evolution, the state of $\mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{B}}$ is given by eq. (1.6). Then we perform the orthogonal measurement $\{|a\rangle_{\mathrm{B}} \langle a|_{\mathrm{B}}\}$. The probability of outcome a is given by

$$p_a = \mathrm{tr}_{\mathrm{AB}} \left[|a\rangle_{\mathrm{B}} \langle a|_{\mathrm{B}} U_{\mathrm{AB}} \left(\rho_{\mathrm{A}} \otimes |0\rangle_{\mathrm{B}} \langle 0|_{\mathrm{B}} \right) U_{\mathrm{AB}}^{\dagger} \right] =$$

$$= \mathrm{tr}_{\mathrm{A}} \left(\langle a|_{\mathrm{B}} U_{\mathrm{AB}} |0\rangle_{\mathrm{B}} \rho_{\mathrm{A}} \langle 0|_{\mathrm{B}} U_{\mathrm{AB}}^{\dagger} |a\rangle_{\mathrm{B}} \right) =$$

$$= \mathrm{tr}_{\mathrm{A}} M_a \rho_{\mathrm{A}} M_a^{\dagger} = \mathrm{tr}_{\mathrm{A}} M_a^{\dagger} M_a \rho_{\mathrm{A}},$$

where we set $M_a := \langle a|_{\mathrm{B}} U_{\mathrm{AB}} |0\rangle_{\mathrm{B}}$, and we call them Kraus operators, because their definition is analogous to the one given in section 1.3 for quantum channels. Note that this expression for the probability of outcome a resembles the one given for a POVM, in which $p_a = \mathrm{tr}_{\mathrm{A}} F_a \rho_{\mathrm{A}}$. If we set $F_a := M_a^{\dagger} M_a$, we have $\sum_a F_a = \sum_a M_a^{\dagger} M_a = \mathbf{1}_{\mathrm{A}}$, and F_a is clearly positive. So $\{F_a\}$ is a POVM. This should not surprise, because POVMs are naturally associated with probabilities; hence, if a measurement must assign probabilities to the various outcomes, it must comprise a POVM in its formalism. Recall this is true for orthogonal measurements, where the POVM is given by the projectors themselves.

Let us move to examine the state of system A after this non-demolition measurement if the outcome is a. The non-normalized version of the state is

$$\widetilde{\rho}'_{\mathrm{A}} = \mathrm{tr}_{\mathrm{B}} \left[|a\rangle_{\mathrm{B}} \langle a|_{\mathrm{B}} U_{\mathrm{AB}} \left(\rho_{\mathrm{A}} \otimes |0\rangle_{\mathrm{B}} \langle 0|_{\mathrm{B}} \right) U_{\mathrm{AB}}^{\dagger} |a\rangle_{\mathrm{B}} \langle a|_{\mathrm{B}} \right] =$$

$$= \langle a|_{\mathrm{B}} U_{\mathrm{AB}} |0\rangle_{\mathrm{B}} \rho_{\mathrm{A}} \langle 0|_{\mathrm{B}} U_{\mathrm{AB}}^{\dagger} |a\rangle_{\mathrm{B}} = M_a \rho_{\mathrm{A}} M_a^{\dagger}.$$

The normalized state is

$$\rho'_{\mathrm{A}} = \frac{M_a \rho_{\mathrm{A}} M_a^{\dagger}}{\mathrm{tr}_{\mathrm{A}} M_a^{\dagger} M_a \rho_{\mathrm{A}}}.$$

Again, the trace of the non-normalized state is the probability with which the state is prepared. We see that Kraus operators characterize the measurement completely, because they give both the probability and the state after the measurement.

Therefore we can give the following provisional definition for a non-demolition generalization of a projective measurement. Later we will give a more general definition.

Definition 1.4.1 (Provisional definition)**.** A *quantum measurement* on $\mathcal{H}_{\mathrm{A}}$ is a collection of Kraus operators $\{M_a\}$ on $\mathcal{H}_{\mathrm{A}}$ (such that $\sum_a M_a^\dagger M_a = \mathbf{1}$ and $M_a^\dagger M_a \leq \mathbf{1}$), where each operator corresponds to a measurement outcome. Given a state ρ before the measurement, the probability of getting outcome a is

$$p_a = \mathrm{tr}\, M_a^\dagger M_a \rho,$$

and the state immediately after the measurement is

$$\rho'_a = \frac{M_a \rho M_a^\dagger}{\mathrm{tr}\, M_a^\dagger M_a \rho}.$$

Orthogonal measurements are included among quantum measurements, as shown in the following example.

Example 1.4.2. Projective measurements are a special type of quantum measurements. Indeed, if we have an orthogonal measurement $\{E_a\}$, then E_a's are Kraus operators, because $\sum_a E_a^\dagger E_a = \sum_a E_a = \mathbf{1}$. The probability of outcome a is $p_a = \mathrm{tr}\, E_a^\dagger E_a \rho = \mathrm{tr}\, E_a \rho$, and finally the state after a measurement, if the outcome is a, is

$$\rho'_a = \frac{E_a \rho E_a^\dagger}{\mathrm{tr}\, E_a^\dagger E_a \rho} = \frac{E_a \rho E_a}{\mathrm{tr}\, E_a \rho}.$$

Definition 1.4.1 sheds a new light on quantum channels. Suppose we do not know the outcome of a quantum measurement, then the state is the incoherent superposition of all the possible states after the measurement, weighted with their probabilities. In symbols, if we do not know the outcome, the state after the measurement is

$$\rho' = \sum_a p_a \rho'_a = \sum_a \mathrm{tr}\, M_a^\dagger M_a \rho \frac{M_a \rho M_a^\dagger}{\mathrm{tr}\, M_a^\dagger M_a \rho} = \sum_a M_a \rho M_a^\dagger.$$

This is the same expression as a quantum channel. Therefore we can interpret a quantum channel as a coarse-graining over a collection of probabilistic operations. Each of them is described by a Kraus operator M_a and occurs with probability p_a, that depends on the input state. This also means that we can associate a quantum channel with every quantum measurement, a quantum channel with the same Kraus operators as the quantum measurement.

Now we can try to generalize further this notion of quantum measurement. Indeed, in our derivation, we performed an atomic orthogonal measurement

on $\mathcal{H}_{\rm B}$. This accounted for orthogonal measurements on $\mathcal{H}_{\rm A}$, but now we want to enlarge our perspective. A first step is to consider a generic orthogonal measurements $\{E_a\}$ on $\mathcal{H}_{\rm B}$, not only atomic measurements. Recall that every projector of an orthogonal measurement can be decomposed as a sum of atomic projectors, therefore we expect to obtain Kraus operators that are a sum of atomic Kraus operators. Even more generally, we can consider an orthogonal measurement $\{E_a\}$ on $\mathcal{H}_{\rm A}\otimes\mathcal{H}_{\rm B}$ (and not only on $\mathcal{H}_{\rm B}$!), following unitary evolution $U_{\rm AB}$. What is the final state of system A?

Again, we start from a product state $\rho_{\rm AB}=\rho_{\rm A}\otimes|0\rangle_{\rm B}\langle 0|_{\rm B}$. After the projective measurement $\{E_a\}$, if the outcome is a, the (non-normalized) state of AB is

$$\widetilde{\rho}'_{\rm AB}=E_aU_{\rm AB}\left(\rho_{\rm A}\otimes|0\rangle_{\rm B}\langle 0|_{\rm B}\right)U_{\rm AB}^\dagger E_a.$$

If $\{|k\rangle_{\rm B}\}$ is an orthonormal basis for $\mathcal{H}_{\rm B}$, the state of subsystem A is

$$\widetilde{\rho}'_{\rm A}={\rm tr}_{\rm B}\left[E_aU_{\rm AB}\left(\rho_{\rm A}\otimes|0\rangle_{\rm B}\langle 0|_{\rm B}\right)U_{\rm AB}^\dagger E_a\right]=$$

$$=\sum_k\langle k|_{\rm B}E_aU_{\rm AB}|0\rangle_{\rm B}\,\rho_{\rm A}\,\langle 0|_{\rm B}U_{\rm AB}^\dagger E_a|k\rangle_{\rm B}=$$

$$=:\sum_k M_{ak}\rho_{\rm A}M_{ak}^\dagger,$$

where we set $M_{ak}:=\langle k|_{\rm B}E_aU_{\rm AB}|0\rangle_{\rm B}$. Now, these operators satisfy the properties

- $\sum_{a,k}M_{ak}^\dagger M_{ak}=\mathbf{1}$. Indeed,

$$\sum_{a,k}\langle 0|_{\rm B}U_{\rm AB}^\dagger E_a|k\rangle_{\rm B}\langle k|_{\rm B}E_aU_{\rm AB}|0\rangle_{\rm B}=$$

$$=\sum_a\langle 0|_{\rm B}U_{\rm AB}^\dagger E_aU_{\rm AB}|0\rangle_{\rm B}=\langle 0|_{\rm B}U_{\rm AB}^\dagger U_{\rm AB}|0\rangle_{\rm B}=\mathbf{1}_{\rm A}.$$

  where we used the properties of orthogonal measurements and of unitary operators.

- Clearly then $\sum_k M_{ak}^\dagger M_{ak}\leq\mathbf{1}$ for every a. This means that an additional positive operator is needed to obtain the identity as sum.

These two properties mean that $\left\{\sum_k M_{ak}^\dagger M_{ak}\right\}$ is a POVM with outcomes $\{a\}$. $\sum_k M_{ak}^\dagger M_{ak}$ must play a central role in determining probabilities. Indeed, the probability of outcome a is

$$p_a = \mathrm{tr_A}\left\{\mathrm{tr_B}\left[E_a U_{\mathrm{AB}}\left(\rho_{\mathrm{A}}\otimes|0\rangle_{\mathrm{B}}\langle 0|_{\mathrm{B}}\right)U_{\mathrm{AB}}^\dagger E_a\right]\right\} = \mathrm{tr_A}\,\widetilde{\rho}'_{\mathrm{A}}.$$

Even in this case, the probability of outcome a is the trace of the non-normalized state $\widetilde{\rho}'_{\mathrm{A}}$ after the measurement if the outcome is a. Expanding the expression of $\mathrm{tr_A}\,\widetilde{\rho}'_{\mathrm{A}}$ we find

$$\mathrm{tr_A}\,\widetilde{\rho}'_{\mathrm{A}} = \mathrm{tr_A}\left(\sum_k M_{ak}\rho_{\mathrm{A}}M_{ak}^\dagger\right) = \sum_k \mathrm{tr_A}\,M_{ak}\rho_{\mathrm{A}}M_{ak}^\dagger =$$

$$= \mathrm{tr_A}\left(\sum_k M_{ak}^\dagger M_{ak}\rho_{\mathrm{A}}\right),$$

confirming that $\sum_k M_{ak}^\dagger M_{ak} =: F_a$ are elements of a POVM.

We can finally give the following general definition for measurements that have an output state.

Definition 1.4.3. A *quantum instrument* is a collection of operators $\{M_{ak}\}$, called *Kraus operators*, where a labels the various outcomes, such that

- $\sum_k M_{ak}^\dagger M_{ak} \leq \mathbf{1}$
- $\sum_{a,k} M_{ak}^\dagger M_{ak} = \mathbf{1}$.

A quantum instrument performs a probabilistic transformation on states of system A: given an input state ρ, we have the transformation

$$\rho \longmapsto \frac{\sum_k M_{ak}\rho M_{ak}^\dagger}{\mathrm{tr}\left(\sum_k M_{ak}^\dagger M_{ak}\rho\right)}$$

with probability

$$p_a = \mathrm{tr}\left(\sum_k M_{ak}^\dagger M_{ak}\rho\right),$$

for every a.

Each of the transformations $\rho \mapsto \frac{\sum_k M_{ak}\rho M_{ak}^\dagger}{\mathrm{tr}\left(\sum_k M_{ak}^\dagger M_{ak}\rho\right)}$ is called *quantum operation*.

We can associate a POVM and a coarse-graining channel with every quantum instrument. The POVM is needed to assign probabilities and it is

$$\{F_a\} = \left\{ \sum_k M_{ak}^\dagger M_{ak} \right\},$$

as we have seen above. The coarse-graining channel arises if we do not know the outcome, but we know the input state. In this case the output state is

$$\rho' = \sum_a p_a \rho'_a = \sum_{a.k} M_{ak} \rho M_{ak}^\dagger.$$

Therefore the coarse-graining channel is the deterministic transformation

$$\rho \longmapsto \sum_{a,k} M_{ak} \rho M_{ak}^\dagger,$$

whose Kraus operators are M_{ak}'s.

Note that definition is really general. It comprises the previous definition 1.4.1, but also channels.

Example 1.4.4. In definition 1.4.1 we considered only quantum instruments where Kraus operators had only the outcome label (k took only one value, so there was no sum over k). In that case, we find again $M_a^\dagger M_a \leq \mathbf{1}$ and $\sum_a M_a^\dagger M_a = \mathbf{1}$. The associated POVM is $\left\{M_a^\dagger M_a\right\}$, whereas the associated channel is the transformation $\rho \mapsto \sum_a M_a \rho M_a^\dagger$.

Example 1.4.5. Quantum channels are deterministic transformations, and deterministic transformations are a special kind of probabilistic transformations, in which there is only one outcome. The associated POVM is therefore the identity $\mathbf{1}$ and it is made of one element. Since a takes only one value, we have that $\sum_{a,k} M_{ak}^\dagger M_{ak} = \mathbf{1}$ becomes $\sum_k M_k^\dagger M_k = \mathbf{1}$, and this is the normalization condition for Kraus operators of a channel. Therefore we have equality in $\sum_k M_{ak}^\dagger M_{ak} \leq \mathbf{1}$.

In this vein, we can prove the following lemma.

Lemma 1.4.6. *A quantum instrument with Kraus operators $\{M_{ak}\}$ is a quantum channel if and only if $\sum_k M_{ak}^\dagger M_{ak} = \mathbf{1}$ for some a.*

Proof. We already saw necessity in example 1.4.5.

To prove sufficiency, suppose we know that $\sum_k M_{a_0k}^\dagger M_{a_0k} = \mathbf{1}$ for some outcome a_0. Since it is $\sum_{a,k} M_{ak}^\dagger M_{ak} = \mathbf{1}$, we have

$$\mathbf{1} = \sum_{a,k} M_{ak}^\dagger M_{ak} = \sum_k M_{a_0k}^\dagger M_{a_0k} + \sum_{a\neq a_0} \sum_k M_{ak}^\dagger M_{ak} =$$

$$= \mathbf{1} + \sum_{a\neq a_0} \sum_k M_{ak}^\dagger M_{ak}$$

This means that $\sum_{a\neq a_0} \sum_k M_{ak}^\dagger M_{ak} = \mathbf{0}$, and since all the operators are positive, $\sum_k M_{ak}^\dagger M_{ak} = \mathbf{0}$ for every $a \neq a_0$. This means that outcomes $a \neq a_0$ occur with zero probability, therefore the quantum instrument is actually a quantum channel. □

Now we can set the following axiom.

Axiom 1.4.7. *Every probabilistic transformation on a quantum system is given by a quantum instrument.*

In our treatment of quantum instruments, we started from unitary evolution and orthogonal measurements in a larger system and then we restricted our attention to a subsystem. Is it true that *every* quantum instrument can be obtained in this way? The answer is affirmative and comes from the following theorem.

Theorem 1.4.8 (Ozawa)**.** *Every quantum instrument on $\mathcal{H}_\mathrm{A}$, with Kraus operators $\{M_{ak}\}$, comes from a unitary evolution followed by a measurement in a larger system.*

Proof. Omitted. See [21]. □

1.5 Axiomatic approach to quantum operations

In the previous section we saw that quantum instruments are a collection of probabilistic transformations on the states of a system, called quantum operations. If the quantum instrument has only one quantum operation, it is a deterministic transformation on the states of a system: it is a quantum channel.

Since we are considering probabilistic transformations on the states, we free ourselves from the idea of a quantum instrument representing a measurement, so we allow an output system different from the input system.

In this section, we restrict ourselves to quantum operations rather than to quantum instruments. A quantum operation output encodes both the probability and the output state. To encode both these aspects in one expression, we allow quantum operations to be not trace-preserving. The trace of the output state is then the probability, as we noted above.

Therefore, using Kraus operators $\{M_{ak}\}$, a quantum operation is the (probabilistic) transformation

$$\rho \longmapsto \sum_k M_{ak} \rho M_{ak}^\dagger,$$

where the trace of the output state is the probability of occurrence of this quantum operation.

However, in this section, we want to follow a different approach to quantum operations, more axiomatic. We want to identify the minimal requirements for a map to be a quantum operation.

First of all, a quantum operation $\mathcal{C}$ must map states of a system A into (non-normalized) states of another system B. According to the probabilistic interpretation of trace, we must require $0 \leq \mathrm{tr}\mathcal{C}(\rho) \leq 1$. Thus, quantum operations must be trace-non-increasing. In particular, if the quantum operation is a quantum channel, it occurs with probability 1, so it has to be trace-preserving.

Then we require linearity for convex combinations. Suppose we have an ensemble of states $\rho = \sum_j p_j \rho_j$, we require

$$\mathcal{C}\left(\sum_j p_j \rho_j\right) = \sum_j p_j \mathcal{C}(\rho_j).$$

If we allow non-linear quantum operations, we can come to paradoxes, such as the so-called "Everett phone paradox" (see [16]).

Finally, the last requirement concerns the fact that the output of a quantum operation must be a positive operator. But we ask something stronger. Suppose a quantum operation $\mathcal{C}$ acts from A to B, but we consider an input state ρ which is a state of system AC, where C is another system. Suppose we apply $\mathcal{C}$ only to A, so subsystem C does not evolve. Then $\mathcal{C}(\rho)$ must be

a valid state of system BC, therefore the map $\mathcal{C} \otimes \mathcal{I}_{\mathrm{C}}$ must yield positive operators as output, for any system C, where $\mathcal{I}_{\mathrm{C}}$ is the identity channel on C. This requirement, called *complete positivity*, is reasonable also on physical grounds. Suppose, in addition to the physical system A on which we have control, there is a system C, of which we are unaware. Since we can control only system A, when we apply a quantum operation $\mathcal{C}$ on A, we are in fact applying $\mathcal{C} \otimes \mathcal{I}_{\mathrm{C}}$. Complete positivity is simply the requirement that a state of the combined system evolves to another state.

Remark 1.5.1. Complete positivity is stronger than positivity. Indeed, transposition T is positive, but not completely positive. To see it, consider $T_{\mathrm{A}} \otimes \mathbf{1}_{\mathrm{B}}$, where $\mathcal{H}_{\mathrm{AB}} \approx \mathbb{C}^d \otimes \mathbb{C}^d$, and consider the pure state $|\phi\rangle_{\mathrm{AB}} = \frac{1}{\sqrt{d}} \sum_{j=1}^{d} |j\rangle_{\mathrm{A}} |j'\rangle_{\mathrm{B}}$, where $\{|j\rangle_{\mathrm{A}}\}_{j=1}^{d}$ and $\{|j'\rangle_{\mathrm{B}}\}_{j=1}^{d}$ are orthonormal bases for $\mathcal{H}_{\mathrm{A}}$ and $\mathcal{H}_{\mathrm{B}}$ respectively. Then, if we act on $\rho = |\phi\rangle_{\mathrm{AB}} \langle\phi|_{\mathrm{AB}}$ with $T_{\mathrm{A}} \otimes \mathbf{1}_{\mathrm{B}}$, we have

$$\rho = \frac{1}{d} \sum_{j,k=1}^{d} |j\rangle_{\mathrm{A}} \langle k|_{\mathrm{A}} \otimes |j'\rangle_{\mathrm{B}} \langle k'|_{\mathrm{B}} \longmapsto$$

$$\longmapsto \rho' = \frac{1}{d} \sum_{j,k=1}^{d} |k\rangle_{\mathrm{A}} \langle j|_{\mathrm{A}} \otimes |j'\rangle_{\mathrm{B}} \langle k'|_{\mathrm{B}} .$$

After simple passages, we can see that the operator $d\rho'$ acts on pure states as

$$d\rho' \left(|\psi\rangle_{\mathrm{A}} |\varphi\rangle_{\mathrm{B}} \right) = |\varphi\rangle_{\mathrm{A}} |\psi\rangle_{\mathrm{B}}$$

Therefore $d\rho'$ is a swap operator, so one of its eigenvalues is -1, corresponding to swap-antisymmetric pure states.

Now we can give the following abstract definition of quantum operation.

Definition 1.5.2. A *quantum operation* $\mathcal{C}$ is a map from states of $\mathcal{H}_{\mathrm{A}}$ to states of $\mathcal{H}_{\mathrm{B}}$ such that

- it is trace-non-increasing: $0 \leq \mathrm{tr}\mathcal{C}(\rho) \leq 1$;
- it is linear for convex combinations: $\mathcal{C}\left(\sum_j p_j \rho_j\right) = \sum_j p_j \mathcal{C}(\rho_j)$;
- it is completely positive: $\mathcal{C} \otimes \mathcal{I}_{\mathrm{C}}(\rho)$ is positive for any system $\mathcal{H}_{\mathrm{C}}$.

Therefore we can give more abstract definitions of quantum channel and quantum instrument.

Definition 1.5.3. A *quantum channel* is a trace-preserving quantum operation.

Definition 1.5.4. A *quantum instrument* is a collection $\{\mathcal{C}_a\}$ of quantum operations, such that $\sum_a \mathcal{C}_a$ is a quantum channel (i.e. it is a trace-preserving quantum operation).

Quite surprisingly, these abstract definitions are equivalent to those given with Kraus operators. Indeed, we have the following theorem.

Theorem 1.5.5 (Kraus representation theorem)**.** *A map $\mathcal{C}$ from the states of $\mathcal{H}_{\mathrm{A}}$ to the states of $\mathcal{H}_{\mathrm{B}}$ is a quantum operation if and only if, for every state ρ*

$$\mathcal{C}(\rho) = \sum_k M_k \rho M_k^\dagger,$$

for a set of operators $\{M_k\}$ from $\mathcal{H}_{\mathrm{A}}$ to $\mathcal{H}_{\mathrm{B}}$, such that $\sum_k M_k \rho M_k^\dagger \leq 1$, where we have equality if and only if $\mathcal{C}$ is a quantum channel.

Proof. Omitted. The interested reader can find it in [17]. □

M_k's are Kraus operators. Now it is clear why Kraus operators for quantum instruments have two indices, unlike for the case of quantum channels. One index labels quantum operations in the quantum instrument, the other labels Kraus operators for a fixed quantum operation. In this way, the more Kraus operator are present, the more the quantum operation is "mixed", because it involves a coarse-graining (the sum over index k). Recall that our provisional definition of quantum measurement (definition 1.4.1) was atomic and every quantum operation had only one Kraus operator.

In this way, we can associate Kraus operators with every quantum operation. Is this association unique? The answer is negative, as the following counter-example shows [17]. This means that there can be different processes that give rise to the same dynamic for the system.

Example 1.5.6. Let $\mathcal{H} \approx \mathbb{C}^2$ (q-bit). Consider the Kraus operators

$$M_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad M_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

giving rise to the quantum operation

$$\rho \longmapsto M_1 \rho M_1^\dagger + M_2 \rho M_2^\dagger.$$

This corresponds to the physical situation in which we toss a fair coin, and, according to the outcome, we do nothing or we apply σ_3 to the system.

Now consider the Kraus operators

$$N_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad N_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

giving rise to the quantum operation

$$\rho \longmapsto N_1 \rho N_1^\dagger + N_2 \rho N_2^\dagger.$$

This corresponds to the physical situation of an orthogonal projection on the basis $\{|0\rangle, |1\rangle\}$. Although the physical situation is completely different from the previous one, the dynamic is the same. Indeed, note that $N_1 = \frac{1}{\sqrt{2}}(M_1 + M_2)$ and $N_2 = \frac{1}{\sqrt{2}}(M_1 - M_2)$. One has

$$\rho \longmapsto N_1 \rho N_1^\dagger + N_2 \rho N_2^\dagger =$$

$$= \frac{1}{2}\left[(M_1 + M_2)\,\rho\left(M_1^\dagger + M_2^\dagger\right) + (M_1 - M_2)\,\rho\left(M_1^\dagger - M_2^\dagger\right)\right] =$$

$$= M_1 \rho M_1^\dagger + M_2 \rho M_2^\dagger.$$

This shows that the quantum operation is actually the same.

In general, we can note that if we have n Kraus operators $\{M_k\}_{k=1}^n$, and we take a $n \times n$ unitary matrix U_{jl}, we have that $N_j = \sum_k U_{jk} M_k$ are other Kraus operators describing the same quantum operation. Indeed

$$\sum_j N_j \rho N_j^\dagger = \sum_j \sum_k U_{jk} M_k \rho \sum_l U_{jl}^* M_l^\dagger =$$

$$= \sum_{k,l} \delta_{kl} M_k \rho M_l^\dagger = \sum_k M_k \rho M_k^\dagger.$$

In fact, this is the only possibility that two sets of Kraus operators give rise to the same quantum operation.

Theorem 1.5.7. *Suppose $\{M_k\}_{k=1}^n$ and $\{N_l\}_{l=1}^m$ are Kraus operators giving rise to quantum operations $\mathcal{C}$ and $\mathcal{D}$ respectively. By appending some zero operators to the shortest set, we can always be in the case when $n = m$. Then we have $\mathcal{C} = \mathcal{D}$ if and only if there is a $n \times n$ unitary matrix U such that $M_k = \sum_{l=1}^n U_{kl} N_l$.*

Proof. We already saw sufficiency. The proof of necessity is omitted. See [17] for reference. □

Finally, we have a corollary that says how many Kraus operators we can have for a given quantum operation.

Corollary 1.5.8. *If a quantum operation $\mathcal{C}$ acts on the states of a d-dimensional Hilbert space, then it has at most d^2 Kraus operators.*

Proof. Immediate from the proof of theorem 1.5.7. See, for instance, [16]. □

Now we completed our exploration of operational formalism for quantum mechanics and we can move to the abstract version of operational formalism.

Chapter 2

Operational probabilistic theories

Now, we became familiar with operational formalism for quantum mechanics: POVMs, quantum channels, quantum instruments. Thus, it is time to begin our exploration of foundational issues. In the past, foundational questions were essentially related to the measurement problem, giving rise to different interpretations of quantum mechanics. However, recently the fields of quantum foundations and quantum information have started to merge, positively influencing each other [26, 27]. Correspondingly, interests in foundational topics moved towards general probabilistic theories, which are general theories with a probabilistic structure. In other words, general probabilistic theories describe what sets of experiments we can do with physical devices, and assign probabilities to the outcomes of such experiments.

Clearly, classical and quantum theory are included in this formalism. Why do we study operational probabilistic theories that are more general than quantum mechanics? We can single out at least three reasons.

1. We want to understand quantum mechanics better.
   Indeed, what are the features that single out quantum mechanics among all the other possible operational probabilistic theories?

2. We want to study extension of quantum mechanics.
   Suppose that some day quantum mechanics or some of its axioms will prove 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.

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 actually like?

In our treatment of general probabilistic theories, we will use a high-level language, derived from category theory. This formalism is particularly apt to capture the operational-informational background of a theory, namely, loosely speaking, the way in which "information is processed" [31]. In this vein, we will address our analysis in an abstract way, without resorting to the specific formalism of a given theory; instead, we will try to derive our consequences directly from the operational formalism.

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.

Suppose we have an experimenter in a laboratory. He can build up experiments 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 actually identifies a process that occurred between the input and the output system 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 translated into a formal language using graphical language, in which each device is represented as a box.

Many works have been done on this subject (see for instance [32, 33, 34, 35, 36, 37, 38]); in the present treatment, we will follow the line of reasoning of [39, 40] for the present and the next chapter, where informational axioms leading to quantum mechanics are identified.

The key idea is to impose some reasonable axioms to a general probabilistic theory to restrict ourselves to a class of theories. However, in the present treatment, we are not interested in deriving quantum mechanics, so we will not assume all the axioms presented in [40], but only the ones that are needed for the subsequent analysis of foundational aspects of thermodynamics.

2.1 Basic notions

In this section we introduce the basic elements of operational formalism and their graphical representation. As we will see, they have a direct experimental interpretation.

2.1.1 Systems and tests

In an operational theory, there are two primitive notions, which are the basis of every operational language: *systems* and *tests*.

We can have an intuition about their meaning thinking again to 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, systems somehow play the role of labels attached to input and output ports of a device.

We denote systems with capital letters in Roman character: A, B, etc. We will sometimes personify the systems saying "Alice" for system A, "Bob" for system B, and so on. This personification is especially meaningful for multi-party communication purposes: it stresses the fact that we can associate an actual experimenter that performs his own experiments with every system.

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.

Some physical devices have various outcomes, each outcomes corresponds to a particular event that occurred in the laboratory and that can be identified by the experimenter. Therefore, we can give the following characterization of a test.

Definition 2.1.1. 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.

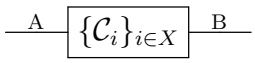
X is called *outcome set*.

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, then 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 he performs his experiment (a sequence of

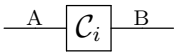
digits, a spot in a photographic plate, etc.). The outcome set X is the set that contains all the possible outcomes for a given test.

Graphically, we can represent a test as a box with in-coming and out-coming wires that represent input and output system respectively.



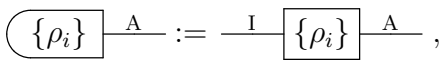
When there is no ambiguity, we will omit the outcome set X. From the graphical representation, it is apparent the role of systems as *labels*.

If we want to express that actually the specific event $\mathcal{C}_i$ occurred, then we will write



.

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

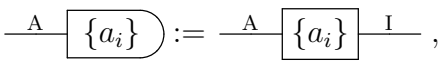


,

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 often use the Dirac-like notation $|\rho_i)_{\mathrm{A}}$ for the preparation-event ρ_i. The subscript A is intended to stress the fact that ρ_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*, that 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 to demolition measurements.

We will often use the Dirac-like notation $(a_i|_{\mathrm{A}}$ for the observation-event a_i. Again, the subscript A stresses the fact that a_i is related to system A, and

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

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

again we use round bracket because this definition is different and more general than the bra notion in quantum mechanics.

Finally, if we have a test from the trivial system to itself, we omit both the lines and the box.

$$p_i := \overset{\text{I}}{\text{---}}\boxed{p_i}\overset{\text{I}}{\text{---}}$$

Definition 2.1.2. We say that a test is *deterministic* if its outcome set has only one element.

If a test is deterministic, we omit braces and write simply $\mathcal{C}$ instead of $\{\mathcal{C}\}$.

In a non-deterministic test, we cannot to predict what particular outcome we will obtain. Instead, for a deterministic test, the outcome is completely determined. Since with non-deterministic tests we are not able to predict the outcome, we would like to set up a probabilistic structure that enables us at least to define probabilities for the various outcomes. We will address this issue soon, but first some other notions are needed.

2.1.2 Sequential and parallel composition

Since we are implementing a graphical language that 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 2.1.3. 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.

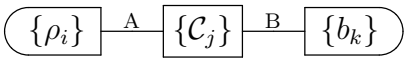
The graphical representation is rather intuitive: suppose we want to compose event $\mathcal{D}_j$ after event $\mathcal{C}_i$; we simply write

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

In this way, there is a natural ordering on tests, the one given by sequential composition. Indeed, some tests are performed first and other later. Using graphical language, this ordering goes from left to right: every box follows all the other on its left. However, we should not confuse this ordering with temporal ordering. We will come back later on this point in section 2.3.

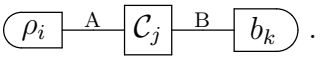
Now let us see an example of sequential composition of tests.

Example 2.1.4. Consider the diagram



It gives instructions to build up the experiment: first, we initialize system A with the preparation-test $\{\rho_i\}$ on A, 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 want to express which events actually occurred, we write

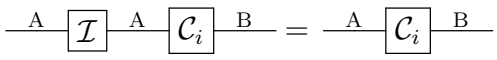


This means that the preparation-event ρ_i, the event $\mathcal{C}_j$ and the observation-event b_k occurred. We can represent the whole sequence in Dirac-like notation as $(b_k|\,\mathcal{C}_j\,|\rho_i)$.

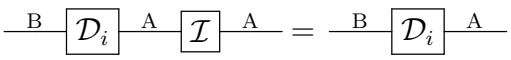
Let us now define the identity test.

Definition 2.1.5. 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

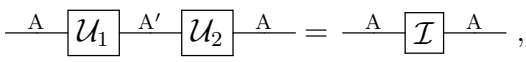


for every $\mathcal{D}_i$. According to this definition, it is clear that for every system A the identity test $\mathcal{I}_\mathrm{A}$ is unique.

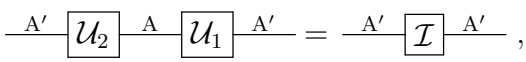
Applying the identity test is just like doing nothing. For this reason we will often omit the box for the identity test.

We sometimes want to "identify" similar system, namely systems that behave exactly in the same way from an operational point of view, but they are yet 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 (or by isomorphic Hilbert spaces to be precise).

Definition 2.1.6. 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

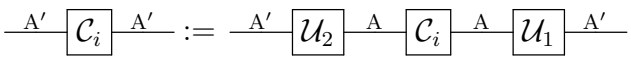


where $\mathcal{I}_{\mathrm{A}}$ is the identity test on A, and



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



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 AB.

Definition 2.1.7. If A and B are two systems, the corresponding *composite system* is AB. Moreover, system composition has the following properties.

1. $\mathrm{AI} = \mathrm{IA} = \mathrm{A}$ for every system A
2. $\mathrm{AB} \cong \mathrm{BA}$ for all systems A, B
3. $\mathrm{A}\,(\mathrm{BC}) = (\mathrm{AB})\,\mathrm{C}$ for all systems $\mathrm{A}, \mathrm{B}, \mathrm{C}$

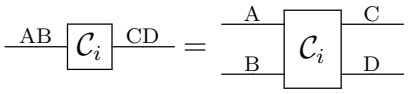
These properties have a rather intuitive meaning.

1. When we combine a system with "nothing", we still have the original system. In this case we will typically omit the line for the trivial system.

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

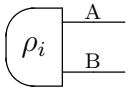
3. This particular form of "associativity" allows us to write simply ABC, without parentheses. Again, the order of composition is irrelevant.

Diagrammatically, we represent composite systems as a collection of lines one under another. We will typically omit the line for the trivial system. Therefore, we can represent an event $\mathcal{C}_i$ from system AB to system CD as a box with multiple lines, one for each system.



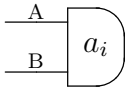
By property 2, it is completely irrelevant to write A rather than B on the upper input wire, and the same holds for every wire.

For composite systems we depict preparation-events as



, (2.1)

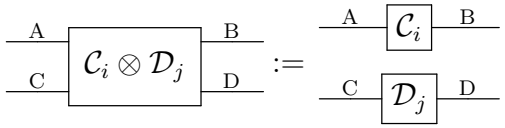
and observation-events as



. (2.2)

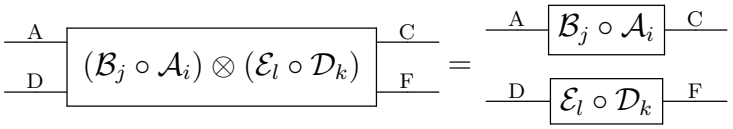
Now we can define parallel composition of tests.

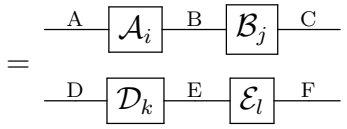
Definition 2.1.8. Let $\{\mathcal{C}_i\}_{i\in X}$ be a test from A to B, and let $\{\mathcal{D}_j\}_{j\in Y}$ be a test from C to D. The *parallel composition* $\{\mathcal{C}_i \otimes \mathcal{D}_j\}_{(i,j)\in X\times Y}$ is a test from AC to BD with outcome set $X \times Y$, and it is represented 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

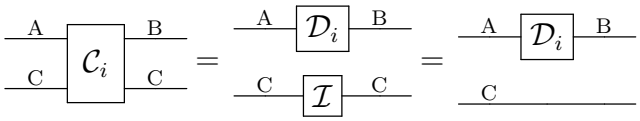




If we parallel-compose a test from A to B with the identity test $\mathcal{I}_{\mathrm{C}}$ on another system C, we have a test from AC to BC that actually acts only on A.

Definition 2.1.9. Consider a test $\{\mathcal{C}_i\}_{i\in X}$ from the composite system AC to BC. 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.

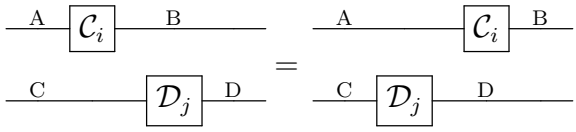
In other words a local test $\{\mathcal{C}_i\}_{i\in X}$ from AC to BC is such that $\mathcal{C}_i = \mathcal{D}_i \otimes \mathcal{I}_{\mathrm{C}}$, for some test $\{\mathcal{D}_i\}_{i\in X}$ from system A to system B.



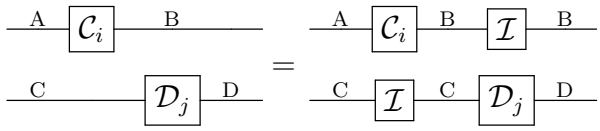
We will write simply $\mathcal{D}_i$ in formulas in place of $\mathcal{D}_i \otimes \mathcal{I}_{\mathrm{C}}$, for example we will write $\mathcal{D}_i \left|\rho\right)_{\mathrm{AC}}$ instead of $\mathcal{D}_i \otimes \mathcal{I}_{\mathrm{C}} \left|\rho\right)_{\mathrm{AC}}$.

As in quantum mechanics, we can prove that local tests on different systems commute.

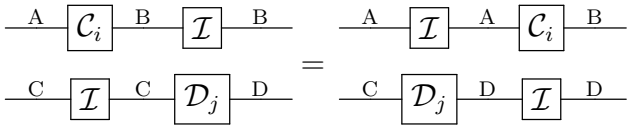
Proposition 2.1.10. *Let* $\{\mathcal{C}_i\}_{i\in X}$ *be a test from system* A *to system* B*, and let* $\{\mathcal{D}_j\}_{j\in Y}$ *be a test from system* C *to system* D*. Then we have*



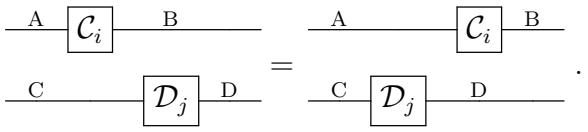
Proof. The proof is straightforward. Recall that we can insert the identity test when we have a line. In this way



Recall that every event commutes with the identity test.



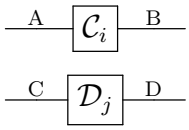
or, in other terms,



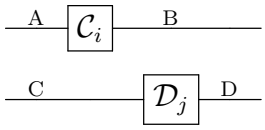
.

□

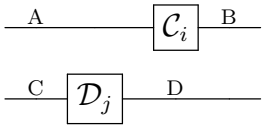
We are then entitled to write



in place of



or

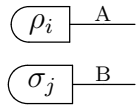


.

This shows that the parallel composition of two tests can be seen as a sequential composition of two local tests on different systems.

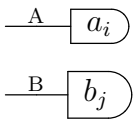
Note that we can compose preparation-tests only in parallel; the same holds for observation-tests, so we will write simply $|\rho_i)_{\mathrm{A}} |\sigma_j)_{\mathrm{B}}$ in place of $|\rho_i)_{\mathrm{A}} \otimes |\sigma_j)_{\mathrm{B}}$; and $(a_i|_{\mathrm{A}} (b_j|_{\mathrm{B}}$ in place of $(a_i|_{\mathrm{A}}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

$$|\rho_i)_{\mathrm{A}} |\sigma_j)_{\mathrm{B}} = \tag{2.3}$$



and

$$(a_i|_{\mathrm{A}} (b_j|_{\mathrm{B}} = \tag{2.4}$$



Remark 2.1.11. 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 2.1.12. An *operational theory* is given by a collections of systems, closed under composition, and a collection of tests, closed under sequential and parallel composition.

Although our graphical language can seem naive and not so sound, it has very strong foundations in category theory [31, 41, 42]. Therefore, we are entitled to use graphical language to prove theorems in abstract scenarios for operational theories.

For the reader familiar with category theory, an operational theory is a *strict symmetric monoidal category* (see [43]), where there is a parallel composition of systems, which is symmetric ($\mathrm{AB} \cong \mathrm{BA}$). Systems are *objects* and events are *arrows*. Every arrow has an input and an output object, and arrows can be sequentially composed. A test is a collection of arrows labelled by outcomes.

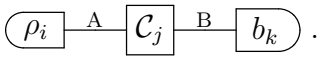
Now we can add the probabilistic ingredient to our theory: basically, we want to assign a number in the interval $[0, 1]$ to every test from the trivial system to itself.

Definition 2.1.13. An *operational-probabilistic theory* (probabilistic theory 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 composition of two events from the trivial system to itself is 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. In particular, we can associate a probability with every diagram with no external lines.

Example 2.1.14. Let us consider again



.

This is a diagrams 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 ρ_i, the event $\mathcal{C}_j$ and the observation-event b_k.

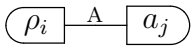
Henceforth we will assume that our operational theories are also probabilistic.

2.1.3 States, effects and transformations

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 totally ignoring the state of the system, or by taking the partial trace of one of Bell states [16].

The issue is now how to distinguish or identify different situations.

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 ρ_i and the observation-event a_j.

Remark 2.1.15. 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 ρ_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. In general, in a non-causal theory, the observation-event can influence the preparation-event, so, in principle, we are not allowed to say what event influenced the other.

If we have a preparation-event ρ_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 ρ_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, he can distinguish two unknown preparations of the system by examining the statistics he gets from performing,

in principle, all the possible measurements on the system. If he finds any difference in the statistics, then he concludes the preparations were different.

A very similar reasoning holds also for observation-events.

In this vein, we can introduce an equivalence relation between preparation-events (and similarly between observation-events). If ρ_i and σ_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 ρ_k on A we have $(a_i|\rho_k) = (b_j|\rho_k)$.

Definition 2.1.16. Equivalence classes of indistinguishable preparation-events are called *states*. The set of states of system A is denoted by $\mathsf{St}(\mathrm{A})$.

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

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. In particular, 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 2.1.17. 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 2.1.18. A state in a composite system AB is called *bipartite state*.

An effect in a composite system AB is called *bipartite effect*.

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

Bipartite states are depicted as in (2.1), bipartite effects are depicted as in (2.2). Product states are represented diagrammatically in (2.3), product effects are represented diagrammatically in (2.4).

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

Example 2.1.19. 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 ρ_i, with $\mathrm{tr}\,\rho_i \leq 1$. This is essentially a random preparation: a state ρ_i is prepared with a probability given by $\mathrm{tr}\,\rho_i$ (recall section 1.5). 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 2.3).

In this way, states and effects are identified with the corresponding functions. Therefore, two states ρ_0 and ρ_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 say that effects are *separating* for states and states are *separating* for effects.

Since states and effects are actually real-valued functions, 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 space, we have $\dim \mathsf{St}_{\mathbb{R}}\,(\mathrm{A}) = \dim \mathsf{Eff}_{\mathbb{R}}\,(\mathrm{A})$.

Example 2.1.20. 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 ($\mathcal{H} \approx \mathbb{C}^d$, for $d \geq 2$). $\mathsf{St}_{\mathbb{R}}\,(\mathrm{A})$ is the vector space of hermitian matrices of order d. It is a real vector space with dimension d^2. $\mathsf{Eff}_{\mathbb{R}}\,(\mathrm{A})$ is again the vector space of hermitian matrices of order d.

Remark 2.1.21. 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[3] of $[0, 1]$-valued functions, we get another $[0, 1]$-valued function.

[3]Recall that a convex combinations of points x_i's is defined as $\sum_i \lambda_i x_i$, where $\lambda_i \geq 0$

This is the first hint to the fact that $\mathsf{St}(\mathrm{A})$ and $\mathsf{Eff}(\mathrm{A})$ are 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 $\widehat{\mathcal{C}}_i$ as

$$\widehat{\mathcal{C}}_i : |\rho)_{\mathrm{A}} \longmapsto \mathcal{C}_i\,|\rho)_{\mathrm{A}}\,, \tag{2.5}$$

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 (2.5) 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 $v \in \mathsf{St}_{\mathbb{R}}(\mathrm{A})$, then we can express it as a linear combinations of states, $v = \sum_j \alpha_j \rho_j$, where $\alpha_j \in \mathbb{R}$ for every j. The most obvious linear extension of (2.5) 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 $v = \sum_j \alpha_j \rho_j$ and $v = \sum_j \beta_j \sigma_j$, where $\beta_j \in \mathbb{R}$ for every j. Our extension 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 (2.5) 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 AC to BC will induce a linear operator from $\mathsf{St}_{\mathbb{R}}(\mathrm{AC})$ to $\mathsf{St}_{\mathbb{R}}(\mathrm{BC})$. We then give the following definition.

Definition 2.1.22. 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.

Recall we already encountered the practice of appending an ancillary system and considering $\mathcal{C}_i \otimes \mathcal{I}_{\mathrm{C}}$ when we discussed complete positivity for quantum operations in section 1.5.

---

and $\sum_i \lambda_i = 1$.

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

Definition 2.1.23. 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 2.1.24. One may wonder why we gave such a definition of indistinguishable events, involving an ancillary system C. The most obvious way of defining indistinguishability would be 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 counter-example concerning quantum mechanics with real Hilbert space [44]. 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 discriminability* (see [39] for further details). Quantum mechanics with real Hilbert space does not fulfil 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.

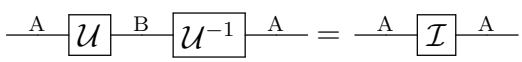
Definition 2.1.25. A deterministic transformation $\mathcal{C} \in \mathsf{Transf}(\mathrm{A},\mathrm{B})$ is called *channel*.

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

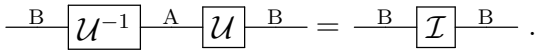
Among all possible channels, reversible channels are particularly important.

Definition 2.1.26. 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′.

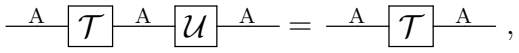
A strongly related topic is the one of invariant state, which we will use very often in the rest of this work.

Definition 2.1.27. A state $\chi \in \mathsf{St}(\mathrm{A})$ is called *invariant* if it is left invariant by the group $\mathbf{G}_{\mathrm{A}}$.

In other words, χ is invariant if and only if $\mathcal{U}\chi = \chi$ for every reversible channel on A.

Similarly to invariant states, we can consider channels with invariant *output*, they are called twirling channels.

Definition 2.1.28. A channel $\mathcal{T}$ on A is called a *twirling channel* if $\mathcal{U}\mathcal{T} = \mathcal{T}$ for every reversible channel $\mathcal{U}$ on A. Using diagrams,



for every $\mathcal{U} \in \mathbf{G}_{\mathrm{A}}$.

Note that since we require the output to be invariant, we apply the reversible channel after (in the order of sequential composition) channel $\mathcal{T}$.

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

Example 2.1.29. 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 $\{\mathcal{C}_k\}$ such that $\sum_k \mathcal{C}_k$ is a trace-preserving map. In this way, transformations are the $\mathcal{C}_k$'s, that are quantum operations (see section 1.5). 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.

A channel is a completely positive trace-preserving map from linear operators on n $\mathcal{H}_{\mathrm{A}}$ to linear operators on $\mathcal{H}_{\mathrm{B}}$ (see again section 1.5).

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.

We have only one invariant state. If $\mathcal{H} \approx \mathbb{C}^d$, then it is $\chi = \frac{1}{d}\mathbf{1}$. We will see in the next chapter that uniqueness is not a coincidence.

2.2 Pure conditioning

Even in an abstract probabilistic theory, it makes sense to talk about pure and mixed states, or, more generally, about pure and mixed transformations. The idea behind pure and mixed events is *coarse-graining*. Let us clarify this idea with the example of the roll of a die. In this random experiment, there are some atomic events, namely that cannot be further decomposed: 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 2.2.1. 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[4] $\{Y_i\}$ 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 one. It is clear that if $\{\mathcal{C}_i\}_{i\in X}$ is a coarse-graining of the test $\{\mathcal{D}_j\}_{j\in Y}$, then 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 summing all the states in a preparation-test; and we can get a deterministic effect summing all the effects in an observation-test.

[4]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.

Definition 2.2.2. A test $\{\mathcal{C}_i\}_{i\in X}$ such that the associated channel $\sum_{i\in X}\mathcal{C}_i$ is twirling is called a *twirling test*.

We can consider also refinements of single transformations.

Definition 2.2.3. 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 2.2.4. 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 $\sum_i p_i = 1$.

It is then reasonable to give the following definition.

Definition 2.2.5. A transformation $\mathcal{C}$ is *pure* (or *atomic*) if it has only trivial refinements.

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. 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.

Definition 2.2.6. A state ω is called *completely mixed* if any other state can refine it, namely for every state ρ there is a non-vanishing probability $p \in (0, 1]$ such that $\omega = p\rho + (1 - p)\,\sigma$, where σ is another state.

A completely mixed state expresses the fact that we have complete ignorance about the preparation of the system: the system could be in any of its allowed preparations.

Let us see some examples in quantum mechanics.

Example 2.2.7. If we diagonalize a density operator $\rho = \sum_j p_j |\psi_j\rangle \langle\psi_j|$, each density operator $p_j |\psi_j\rangle \langle\psi_j|$ is a refinement of ρ. More generally, a refinement of ρ is a state σ 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[5] of σ is contained in the support of ρ (see proposition A.1.1).

A pure state is a density operator ρ that is proportional (with non-vanishing proportional coefficient) to a rank-one projector. A state is completely mixed if and only if it is proportional (with non-vanishing proportional coefficient) to a full-rank density operator. Clearly, the invariant state $\chi = \frac{1}{d}\mathbf{1}$ is an example of completely mixed state. As we will see in the following chapter, this is not a coincidence.

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 ρ, as shown in section 1.5. A quantum operation is pure if and only it has only one Kraus operator (recall the provisional definition 1.4.1).

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

Lemma 2.2.8. *Let $\mathcal{U}$ be a reversible channel from* A *to* B. *Then $|\psi)_{\mathrm{A}}$ is pure if and only if $\mathcal{U}|\psi)_{\mathrm{A}}$ is pure.*

Proof. Necessity. Suppose, by contradiction, that $\mathcal{U}|\psi)_{\mathrm{A}}$ is mixed. Then it can be written as a coarse-graining of other states.

$$\mathcal{U}|\psi)_{\mathrm{A}} = \sum_i |\rho_i)_{\mathrm{B}} \tag{2.6}$$

Now, we apply $\mathcal{U}^{-1}$ to both sides of eq. (2.6). By linearity, we have

$$|\psi)_{\mathrm{A}} = \sum_i \mathcal{U}^{-1}|\rho_i)_{\mathrm{B}},$$

which is absurd because a pure state has been written as a coarse-graining of other states.

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

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

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 2.2.9. *Let $\mathcal{U}$ be a reversible channel from* A *to* B*. Then $(b|_{\mathrm{B}}$ is pure if and only if $(b|_{\mathrm{B}}\mathcal{U}$ is pure.*

Proof. The proof is analogous to the proof of lemma 2.2.8. □

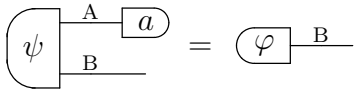
Now we move to the issue of composition of pure transformations. Is the composition still pure? Intuitively, such a composition should not destroy information. However, the issue is not so trivial, therefore it is worth setting an axiom, called *pure conditioning*.

Axiom 2.2.10 (Pure conditioning)**.** *Sequential and parallel compositions of pure transformations are pure transformations.*

From now on, we will consider only theories where this axiom holds. Quantum mechanics and classical mechanics are examples of theories that fulfil pure conditioning, because, for instance, the product of two pure states is a pure state. Actually, in classical mechanics, the only bipartite pure states are product of pure states.

Pure conditioning is a fairly reasonable property. If it did not hold, we would have a theory in which there is an information leakage when we compose transformations. This would constitute a serious limitation for experiments: in such a theory it would be virtually impossible to build up an experimental apparatus connecting various devices, because each connection would imply a loss of information.

In the following, we will often use a straightforward consequence of pure conditioning. Suppose we have a bipartite pure state $|\psi)_{\mathrm{AB}}$ of system AB. If we apply a pure effect $(a|_{\mathrm{A}}$ to A, then we get a pure state $|\varphi)_{\mathrm{B}}$ of system B.



Indeed, we are applying a pure effect on A and the identity on B, which are both pure. By pure conditioning, $|\varphi)_{\mathrm{B}}$ is pure.

2.3 Causality

In this section we will examine the issue of 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.

Requiring causality is equivalent to require that the experimenter's future choices do not influence his present experiments. Causality is implicit in most works on general probabilistic theories (see [32, 35, 36, 45, 46, 47, 48]), but there can be theories that do not fulfil causality requirement. Actually, theories with non-fixed causal structure might play an important role in physics. Indeed, according to the work by Hardy [49], a quantum theory with indefinite causal structure could be a route to the formulation of a quantum theory of gravity.

However, in our treatment we will assume causality.

Definition 2.3.1 (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 $\{a_j\}_{j\in Y}$.

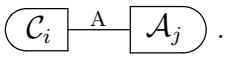
In other words, if $\{a_j\}_{j\in Y}$ and $\{b_k\}_{k\in Z}$ are two observation-tests, we have

$$\sum_{j\in Y} (a_j|\rho_i)_{\mathrm{A}} = \sum_{k\in Z} (b_k|\rho_i)_{\mathrm{A}} \,. \tag{2.7}$$

Loosely speaking, the preparation of the system does not depend on the choice of subsequent (or "future") measurements. In this way, the direction in which information flows, that 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 [50].

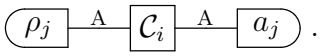
Example 2.3.2. Consider a theory in which the states of a system are quantum operations on that system. In particular, deterministic states are quantum channels. Then we can consider the channels of this theory to be quantum "supermaps", that map quantum channels into quantum channels.

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 lines. 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 theory. This is essentially the causality requirement (or axiom).

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

However, definition 2.3.1 is not so practical to work with, although it is operational. We will mostly use the following characterization.

Proposition 2.3.3. *A theory is causal if and only if for every system* A *there is a unique deterministic effect* $(e|_{\mathrm{A}}$.

Proof. Necessity. Suppose, by contradiction, that there are two deterministic effects e and e' for system A. Deterministic effects are particular examples of observation-tests. Eq. (2.7) then states that $(e|\rho_i)_{\mathrm{A}} = (e'|\rho_i)_{\mathrm{A}}$ for every $\rho_i \in \mathsf{St}(\mathrm{A})$. This means that $e = e'$.

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

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

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

We have 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 $\sum_{i\in X} a_i = e$, where e is the deterministic effect of A. This is a necessary condition for $\{a_i\}_{i\in X}$ to be an observation-test. In section 3.2 we will see that it is also sufficient if the theory satisfies another requirement: the purification postulate.

We saw in example (2.1.19) that in quantum mechanics there is only one deterministic effect, the identity operator. Hence quantum mechanics is a causal theory.

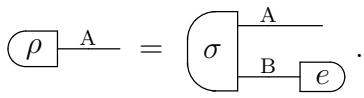
Let us see a straightforward corollary of uniqueness of the deterministic effect.

Corollary 2.3.4. *Let* A *and* B *be two systems. In a causal theory, if* $(e|_{\mathrm{A}}$ *and* $(e|_{\mathrm{B}}$ *are the deterministic effects of systems* A *and* B *respectively, then the deterministic effect for system* AB *is* $(e|_{\mathrm{A}}\,(e|_{\mathrm{B}}$.

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

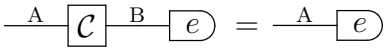
In a causal theory, we can define marginal states. Suppose we have a bipartite state of system AB and we are interested in the state of subsystem A. We want to throw away all the information concerning system B. This operation resembles marginalization in probability theory, whence the name. In quantum mechanics, this operation is simply taking the partial trace over B.

Definition 2.3.5. 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: $|\rho)_{\mathrm{A}} = (e|_{\mathrm{B}}\,|\sigma)_{\mathrm{AB}}$



In a causal theory, we have also useful properties for the characterization of channels and tests.

Proposition 2.3.6. *Let* $\mathcal{C} \in \mathsf{Transf}\,(\mathrm{A},\mathrm{B})$. $\mathcal{C}$ *is a channel if and only if* $(e|_{\mathrm{B}}\,\mathcal{C} = (e|_{\mathrm{A}}$.



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

Sufficiency. Suppose we have a test $\{\mathcal{C}_i\}_{i\in X}$ from system A to system B such that $\mathcal{C} := \mathcal{C}_{i_0}$ fulfils $(e|_{\rm B}\mathcal{C} = (e|_{\rm A}$. We want to prove that $\{\mathcal{C}_i\}_{i\in X}$ is actually 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 $(e|_{\rm B}\mathcal{C}' = (e|_{\rm A}$. Recalling the expression of $\mathcal{C}'$, we have

$$(e|_{\rm A} = (e|_{\rm B}\mathcal{C}' = (e|_{\rm B}\mathcal{C}_{i_0} + (e|_{\rm B}\sum_{i\neq i_0}\mathcal{C}_i = (e|_{\rm A} + (e|_{\rm B}\sum_{i\neq i_0}\mathcal{C}_i,$$

because $(e|_{\rm B}\mathcal{C}_{i_0} = (e|_{\rm A}$. This means $(e|_{\rm B}\sum_{i\neq i_0}\mathcal{C}_i = 0$, 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. □

In particular, if A is the trivial system, we have that a state $|\rho)_{\rm B}$ is deterministic if and only if $(e|\rho)_{\rm B} = 1$. 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

$$\sum_{i\in X}(e|_{\rm B}\mathcal{C}_i = (e|_{\rm A}\,.$$

This is a necessary condition. We will prove that in theories with purification it is also sufficient.

In a causal theory we have no signalling. This means that if we have a bipartite state, it is not possible for a party to communicate the outcome of a local measurement on its system to the other without exchanging physical systems.

Theorem 2.3.7. *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 $|\sigma)_{\rm AB}$. 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 $p_{ij} := (e|_{\rm AB}\,\mathcal{A}_i\otimes\mathcal{B}_j\,|\sigma)_{\rm AB}$ and the marginal probabilities as $p_i^{(\rm A)} := \sum_{j\in Y}(e|_{\rm AB}\,\mathcal{A}_i\otimes\mathcal{B}_j\,|\sigma)_{\rm AB}$ and $p_j^{(\rm B)} := \sum_{i\in X}(e|_{\rm AB}\,\mathcal{A}_i\otimes\mathcal{B}_j\,|\sigma)_{\rm AB}$. Each party cannot acquire any information about the outcomes of the other party 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.

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

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

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

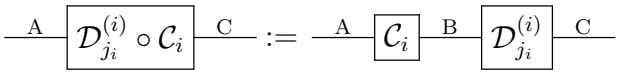
We see that Alice's marginal probability does not depend at all on the test performed by Bob, so she cannot get any information about the outcome of Bob's test based only on her system.

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 may choose a later test according to the result of a previous test. Suppose we perform first a test $\{\mathcal{C}_i\}_{i\in X}$ from A to B. According to the outcome i, we then perform different tests $\left\{\mathcal{D}_{j_i}^{(i)}\right\}_{j_i\in Y_i}$ 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 2.3.8. If $\{\mathcal{C}_i\}_{i\in X}$ is a test from A to B and, for every i, $\left\{\mathcal{D}_{j_i}^{(i)}\right\}_{j_i\in Y_i}$ 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 $\left\{\mathcal{D}_{j_i}^{(i)} \circ \mathcal{C}_i\right\}_{(i,j_i)\in Z}$.

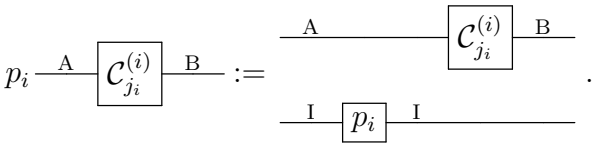
The graphical representation is as usual.



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, which we will use extensively hereafter.

Definition 2.3.9. If $\{p_i\}_{i\in X}$ is a set of probabilities[6] and, for every i, $\left\{\mathcal{C}_{j_i}^{(i)}\right\}_{j_i\in Y_i}$ is a test from A to B, we can construct the *randomized test* $\left\{p_i\mathcal{C}_{j_i}^{(i)}\right\}_{i\in X, j\in Y_i}$, which is a test from A to B whose events are defined as



2.3.1 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.

Definition 2.3.10. Let $\rho \in \mathsf{St}(\mathrm{A})$. We define the *norm* of ρ as

$$\|\rho\| := (e|\rho)\,.$$

It can be shown that this norm is related to the issue of distinguishing states, so it has an operational meaning [39].

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 2.3.11. *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 $\sum_{i\in X} a_i = e$, then

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

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. □

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

Definition 2.3.12. A state ρ such that $\|\rho\| = 1$ is called *normalized*.

We denote the set of normalized states of system A by $\mathsf{St}_1(\mathrm{A})$.

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

Lemma 2.3.13. *In a causal theory, a state is normalized if and only if it is deterministic.*

Proof. It is a trivial corollary of proposition 2.3.6, as we have already noted. □

Example 2.3.14. In quantum mechanics, we have

$$\|\rho\| = \operatorname{tr} \mathbf{1}\rho = \operatorname{tr} \rho.$$

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

For every state ρ, 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 ρ_{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 2.3.15. *A theory where every state is proportional to a deterministic state, that in general depends on the particular state we are considering, is causal.*

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

$$(e|\rho) = k\,(e|\overline{\rho}) = k = k\,(e'|\overline{\rho}) = (e'|\rho)\,.$$

Since ρ is arbitrary, $e = e'$ and the theory is therefore causal. □

In a causal theory, every non-normalized state ρ_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, $p_i = \|\rho_i\|$, but since $p_i \in [0,1]$, we can regard ρ_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 section 1.5). Recall the fact that $(e|\rho_i)$ gives the conditional probability of e given ρ_i. Since e is deterministic, the probability comes only from the preparation of ρ_i. Therefore states with vanishing norm cannot be prepared, so they are not true states.

The norm of states satisfies the following property.

Proposition 2.3.16. *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 equality if and only if $\mathcal{C}$ *is a channel.*

Proof. By definition, $\|\mathcal{C}\rho\|_{\mathrm{B}} = (e|_{\mathrm{B}}\,\mathcal{C}\,|\rho)_{\mathrm{A}}$. Since $(e|_{\mathrm{B}}\,\mathcal{C}$ is an effect of system A, we have $(e|_{\mathrm{B}}\,\mathcal{C}\,|\rho)_{\mathrm{A}} \leq (e|\rho)_{\mathrm{A}}$. Then we have $\|\mathcal{C}\rho\|_{\mathrm{B}} \leq \|\rho\|_{\mathrm{A}}$.

By proposition 2.3.6, $\mathcal{C}$ is a channel if and only if $(e|_{\mathrm{B}}\,\mathcal{C} = (e|_{\mathrm{A}}$, then

$$\|\mathcal{C}\rho\|_{\mathrm{B}} = (e|_{\mathrm{B}}\,\mathcal{C}\,|\rho)_{\mathrm{A}} = (e|\rho)_{\mathrm{A}} = \|\rho\|_{\mathrm{A}}.$$

□

Extending the norm to every element of $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$, we can use it to define a topology. In particular, 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, therefore $\mathsf{St}(\mathrm{A})$ is closed.

Assumption 2.3.17. *For all systems* A *the set* $\mathsf{St}(\mathrm{A})$ *is closed.*

Lemma 2.3.18. *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 [39]). 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 2.3.11.

Definition 2.3.19. Let $a \in \mathsf{Eff}(\mathrm{A})$. We define the *norm* of a as

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

Even in this case $0 \leq \|a\| \leq 1$. Clearly, for the deterministic effect, $\|e\| = 1$, because $(e|\rho) = 1$ if ρ is normalized.

We can also define a norm for general transformations [39].

Definition 2.3.20. Let $\mathcal{C} \in \mathsf{Transf}(\mathrm{A}, \mathrm{B})$. We define the *norm* of $\mathcal{C}$ as

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

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

We will prove later that also the sets of transformations is closed.

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

Proposition 2.3.21. *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 2.3.18), $p \in \mathsf{St}(\mathrm{I})$. 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 ρ and σ be two normalized states of system A. Then

$$\|p\rho + (1-p)\,\sigma\| = (e|p\rho + (1-p)\,\sigma) = p\,(e|\rho) + (1-p)\,(e|\sigma) =$$

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

□

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 realized from an operational point of view. 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 ρ_p from an operational point of view 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 ρ_i. In this way, we realize the preparation-test $\{p_0\rho_0, p_1\rho_1\}$. Note that each state is not normalized because it is not deterministic: the state ρ_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 ρ reflects a particular way of preparing ρ. We can give the following definition.

Definition 2.3.22. Let ρ be a normalized state. We say that a normalized state σ is *compatible* with ρ if we can write $\rho = p\sigma + (1-p)\,\tau$, where $p \in (0,1]$ and τ is another normalized state.

The set of all normalized states compatible with ρ is called the *face* identified by ρ, and we will denote it by F_ρ.

The states in the face identified by ρ are states that refine ρ.

Example 2.3.23. In quantum mechanics, given ρ, the states compatible with ρ are density operators whose support is contained in the support of ρ, because they refine ρ.

Clearly, a state σ is compatible with ρ if it is in some convex decomposition of ρ. This means in particular that all pure states are incompatible with each other, otherwise we would have a pure state that comes from a coarse-graining of other pure states, which is absurd. Therefore, if ψ is pure, then $F_\psi = \{\psi\}$.

A completely mixed state ω, instead, is a state such that every normalized state is in the face identified by ω. In symbols, $F_\omega = \mathsf{St}_1(\mathrm{A})$.

Lemma 2.3.24. *A state $\omega \in \mathsf{St}_1(\mathrm{A})$ is completely mixed if and only if $\mathsf{Span}(F_\omega) = \mathsf{St}_{\mathbb{R}}(\mathrm{A})$.*

Proof. Necessity is straightforward. If ω is completely mixed, we have $F_\omega = \mathsf{St}_1(\mathrm{A})$, therefore $\mathsf{St}(\mathrm{A}) \subset \mathsf{Span}(F_\omega)$. Since $\mathsf{Span}(\mathsf{St}(\mathrm{A})) = \mathsf{St}_{\mathbb{R}}(\mathrm{A})$, one has the thesis.

Sufficiency. Suppose we have $\mathsf{Span}(F_\omega) = \mathsf{St}_{\mathbb{R}}(\mathrm{A})$. Then, for a generic normalized state ρ, one has $\rho \in \mathsf{Span}(F_\omega)$. Hence it can be written as a linear combination of n states compatible with ω, $\rho = \sum_{i=1}^n a_i \rho_i$, where a_i's are real numbers and $\rho_i \in F_\omega$ for every i, but we do not assume that ρ_i's are normalized. Since $\rho_i \in F_\omega$, we can write

$$\omega = \rho_i + \sigma_i,$$

where σ_i is another state. Therefore

$$\omega = \frac{1}{n} \sum_{i=1}^n (\rho_i + \sigma_i).$$

Let us define $\sigma := \omega - p\rho$, where $p := \frac{1}{2na}$ and $a := \max_i \{a_i\}$. Clearly $a > 0$ because $\sum_{i=1}^n a_i \|\rho_i\| = 1$, so there must be at least a positive coefficient a_i. Moreover, since $\|\rho_i\| \leq 1$, we have

$$1 = \sum_{i=1}^n a_i \|\rho_i\| \leq \sum_{i=1}^n a_i,$$

therefore clearly it must be $a > \frac{1}{2n}$, otherwise it could not be $\sum_{i=1}^n a_i \geq 1$. Hence $0 < p < 1$. Recalling the expressions of ω and ρ in terms of the ρ_i's, one has

$$\sigma := \frac{1}{n} \sum_{i=1}^n (\rho_i + \sigma_i) - \frac{1}{2na} \sum_{i=1}^n a_i \rho_i = \frac{1}{n} \sum_{i=1}^n \left[\left(1 - \frac{a_i}{2a} \right) \rho_i + \sigma_i \right].$$

Clearly, $\frac{a_i}{2a} \leq \frac{1}{2}$, so $1-\frac{a_i}{2a}$ is positive. Taking the normalized state $\overline{\sigma}$ associated with σ, we have that $\overline{\sigma}$ is a convex combination of states, so it is a state. This shows that ω is a coarse-graining of ρ and $\overline{\sigma}$ ($\omega = p\rho + (1-p)\,\overline{\sigma}$), therefore $\rho \in F_\omega$. Since ρ is arbitrary, ω is a completely mixed state. □

Clearly, every mixed state can be ultimately decomposed into a convex combination of pure states, since the ultimate refinement for a mixed state comes from pure states.

Chapter 3

The purification postulate

In this chapter we analyse the purification postulate, the analogous of theorem 1.1.6. Here we will treat it as a postulate. In [40] it has been shown that the purification postulate is the key feature of every quantum theory. It expresses the idea that every physical process (even processes of stochastic nature!) can be ultimately described in terms of pure and reversible processes. In this way, the ignorance about a part is always compatible with maximal information about the whole [15]. This makes the purification postulate one of the main requirements to build foundations of thermodynamics and statistical mechanics [12].

The purification postulate is a principle of conservation of information: we can always recover maximal information by enlarging the system conveniently. This means that information was not destroyed, but was simply discarded because we considered a too small system. If we consider pieces of information contained in physical systems as fundamental blocks forming our world, then conservation of information is a reasonable requirement.

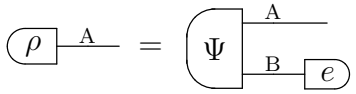
We will see that the purification postulate is the ultimate reason for the validity of extension theorems, such as those presented in chapter 1 (Naimark's theorem, Ozawa's theorem, etc.).

3.1 The purification postulate

In this section we will introduce the purification postulate and we will derive its first consequences.

Definition 3.1.1. Let $|\rho)_{\mathrm{A}}$ be a normalized state of system A. A *purification*

of $|\rho)_{\mathrm{A}}$ is a normalized bipartite *pure* state $|\Psi)_{\mathrm{AB}}$ such that $(e|_{\mathrm{B}}\,|\Psi)_{\mathrm{AB}} = |\rho)_{\mathrm{A}}$. Using diagrams,



System B is called *purifying system.*

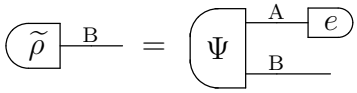
In other words, to purify a state, we must add an ancillary system B, that acts as environment, and we must consider a pure state of this bipartite system such that if we discard the ancillary system, we find the original state. Therefore, every normalized bipartite pure state is a purification of its marginals.

Clearly, it is not always possible to purify states. In classical theory, a bipartite state is pure if and only if it is the product of two pure states, hence it cannot have mixed marginals. Therefore, in classical theory, mixed states cannot be purified.

Remark 3.1.2. The notion of purification must be kept well distinct from the similar notion of *extension.* An extension of $\rho \in \mathsf{St}_1(\mathrm{A})$ is a normalized bipartite state $\sigma \in \mathsf{St}_1(\mathrm{AB})$ such that $(e|_{\mathrm{B}}\,|\sigma)_{\mathrm{AB}} = |\rho)_{\mathrm{A}}$. A purification is clearly an extension, but we require that the bipartite state is *pure*, whereas for a generic extension the state could in principle be mixed.

When we purify a state $|\rho)_{\mathrm{A}}$, we can associate a state of B with it: it is enough to take the marginal of the purification on the purifying system B.

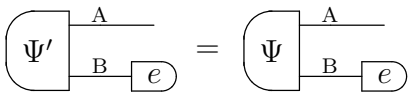
Definition 3.1.3. Let $|\Psi)_{\mathrm{AB}}$ be a purification of $|\rho)_{\mathrm{A}}$. The *complementary state* of $|\rho)_{\mathrm{A}}$ is the state $|\widetilde{\rho})_{\mathrm{B}} \in \mathsf{St}_1(\mathrm{B})$ defined as $|\widetilde{\rho})_{\mathrm{B}} = (e|_{\mathrm{A}}\,|\Psi)_{\mathrm{AB}}$, or



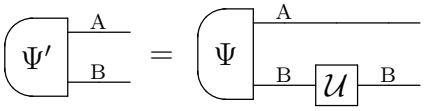
Clearly, if $|\Psi)_{\mathrm{AB}}$ is a purification of $|\rho)_{\mathrm{A}}$, if we apply a local reversible channel $\mathcal{U}_{\mathrm{B}}$ on the purifying system, we get another pure state $|\Psi')_{\mathrm{AB}}$ by lemma 2.2.8. Besides, $|\Psi')_{\mathrm{AB}}$ is such that $(e|_{\mathrm{B}}\,|\Psi')_{\mathrm{AB}} = |\rho)_{\mathrm{A}}$ because $\mathcal{U}_{\mathrm{B}}$ is a channel. Hence, $|\Psi')_{\mathrm{AB}}$ is another purification of $|\rho)_{\mathrm{A}}$. Now we have all the ingredients to state the purification postulate.

Axiom 3.1.4. *Every state has a purification, unique up to reversible channels on the purifying system.*

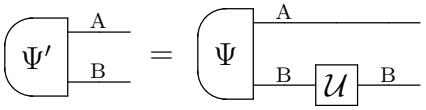
In other words, purification postulate states that if we have two purifications $|\Psi)_{\mathrm{AB}}$ and $|\Psi')_{\mathrm{AB}}$ of the same state $|\rho)_{\mathrm{A}}$, they are related by a local reversible channel on the purifying system B. Using diagrams, if



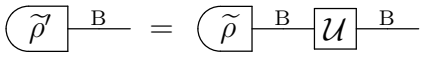
then



A straightforward consequence of uniqueness of purification up to reversible channels is that the complementary state of a state $|\rho)_{\mathrm{A}}$ is unique up to reversible channels. Indeed if we have two purifications $|\Psi)_{\mathrm{AB}}$ and $|\Psi')_{\mathrm{AB}}$,



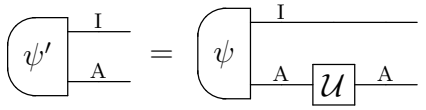
and taking the deterministic effect on A, we have



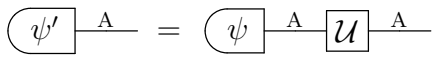
Now we prove an important result for theories with purification, which we will often use hereafter.

Proposition 3.1.5. *For any couple of pure states $\psi, \psi' \in \mathsf{St}_1(\mathrm{A})$ there is a reversible channel $\mathcal{U}$ on* A *such that $\psi' = \mathcal{U}\psi$.*

Proof. Every system is a purifying system for the trivial system I. Then ψ and ψ' are purifications of the same deterministic state of trivial system (which is number 1), therefore



and recalling that we can omit the lines for the trivial system, because IA = A, we finally obtain

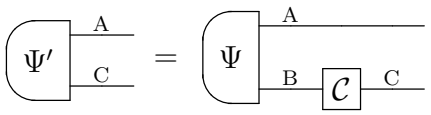


□

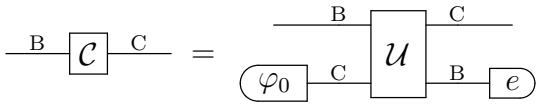
In other words, the action of the group $\mathbf{G}_{\mathrm{A}}$ is transitive on pure states of system A.

Now, we would like to see what we can say if we have two purifications of the same state, but with different purifying systems.

Lemma 3.1.6. *Let* $\Psi \in \mathsf{St}_1(\mathrm{AB})$ *and* $\Psi' \in \mathsf{St}_1(\mathrm{AC})$ *be two purifications of the state* $\rho \in \mathsf{St}_1(\mathrm{A})$*. Then there exist a channel* $\mathcal{C} \in \mathsf{Transf}(\mathrm{B}, \mathrm{C})$ *such that*

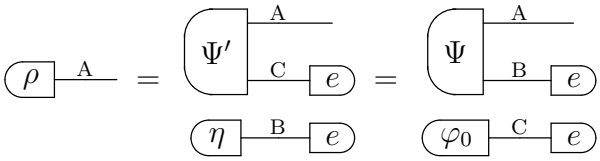


Moreover, $\mathcal{C}$ *has the form*

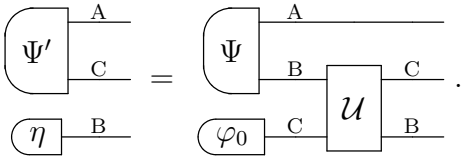


for some pure state $\varphi_0 \in \mathsf{St}_1(\mathrm{C})$ *and some reversible channel* $\mathcal{U}$ *on system* BC.

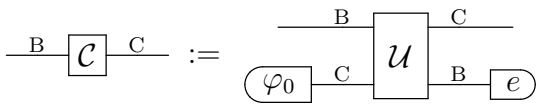
Proof. Let $|\eta)_{\mathrm{B}}$ and $|\varphi_0)_{\mathrm{C}}$ be arbitrary normalized pure states of system B and C respectively. Then $|\Psi')_{\mathrm{AC}}|\eta)_{\mathrm{B}}$ and $|\Psi)_{\mathrm{AB}}|\varphi_0)_{\mathrm{C}}$ are two purifications of ρ with the same purifying system BC. Indeed,



According to the purification postulate, we have



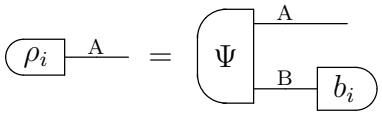
Taking the deterministic effect on system B we have the thesis, where



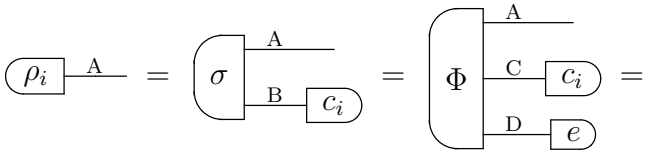
Note that $\mathcal{C}$ is a channel because $\mathcal{U}$ is a channel, although, in general, $\mathcal{C}$ is not reversible. □

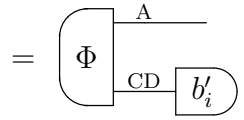
Now, let us move to analyse how we can induce a particular preparation of a mixed state by performing a measurement on the purifying system. This is an extremely important result, that will be used extensively throughout this work.

Theorem 3.1.7 (Steering property)**.** *Let* $\rho \in \mathsf{St}_1(\mathrm{A})$ *and let* $\Psi \in \mathsf{St}_1(\mathrm{AB})$ *be a purification of* ρ*. If* $\{\rho_i\}_{i\in X}$ *is a preparation-test for system* A *such that* $\sum_{i\in X}\rho_i = \rho$*, then there exists an observation-test* $\{b_i\}_{i\in X}$ *on the purifying system* B *such that*

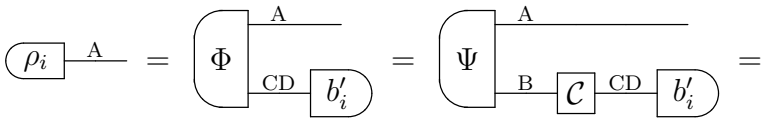


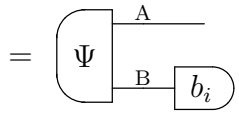
Proof. Let us consider $|X|$ normalized pure states $\{\varphi_i\}_{i\in X}$ of some system C such that there is an observation-test $\{c_i\}_{i\in X} \subseteq \mathsf{Eff}(\mathrm{C})$ such that $(c_j|\varphi_i)_{\mathrm{C}} = \delta_{ij}$. Now, let us consider the state $\sigma := \sum_{i\in X}\rho_i \otimes \varphi_i$, which is clearly an extension of ρ (to see it it is enough to take the deterministic effect on system C). Let us consider a purification $|\Phi)_{\mathrm{ACD}}$ of $|\sigma)_{\mathrm{AC}}$ with purifying system D. $|\Phi)_{\mathrm{ACD}}$ is also a purification of ρ, with purifying system CD. Then $|\rho_i)_{\mathrm{A}} = (c_i|_{\mathrm{C}}|\sigma)_{\mathrm{AC}}$, and





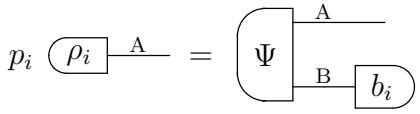
where $(b'_i|_{\mathrm{CD}} := (c_i|_{\mathrm{C}}(e|_{\mathrm{D}}$. So, we found a particular purification of ρ with a special purifying system such that we can obtain every ρ_i by applying a suitable effect on the purifying system. Now we want to prove that this holds for every purification of ρ. If $|\Psi)_{\mathrm{AB}}$ is another purification of $|\rho)_{\mathrm{A}}$, by lemma 3.1.6, we have



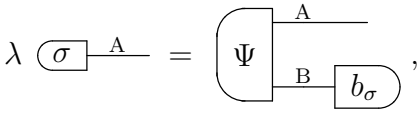


where b_i is an effect on B defined as $(b_i|_{\mathrm{B}} := (b_i'|_{\mathrm{CD}}\, \mathcal{C}$. □

In other words, working with normalized states, suppose we have the following convex decomposition of $\rho = \sum_{i \in X} p_i \rho_i$, where ρ_i is *normalized* for every $i \in X$. We can prepare each state ρ_i with the corresponding probability p_i by taking a purification of ρ and applying a suitable effect b_i on the purifying system. We express the fact that ρ_i is prepared with probability p_i by considering the non-normalized (or randomized) state $p_i \rho_i$.



Recall that a state lies in a convex decomposition of ρ if and only if it is in the face identified by ρ. Therefore, we can say that a state $|\sigma)_{\mathrm{A}}$ is compatible with $|\rho)_{\mathrm{A}}$ if and only if, given a purification $|\Psi)_{\mathrm{AB}}$ of $|\rho)_{\mathrm{A}}$, one has



where b_σ is an effect in the purifying system B and $\lambda \in (0, 1]$.

We might wonder: how large should the purifying system be? A first answer comes from the following corollary of the steering property.

Corollary 3.1.8. *Let $\rho \in \mathsf{St}_1(\mathrm{A})$ and let $\Psi \in \mathsf{St}_1(\mathrm{AB})$ be a purification of ρ. Then we have the following bound on the dimension of the vector space associated with the purifying system* B

$$\dim \mathsf{St}_{\mathbb{R}}(\mathrm{B}) \geq \dim \mathsf{Span}(F_\rho),$$

where F_ρ is the face identified by ρ.

In particular, if ρ is completely mixed, then

$$\dim \mathsf{St}_{\mathbb{R}}(\mathrm{B}) \geq \dim \mathsf{St}_{\mathbb{R}}(\mathrm{A}).$$

Proof. Let us consider the map $\widehat{f}$ from $\mathsf{Eff}_{\mathbb{R}}(\mathrm{B})$ to $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$ such that $b \mapsto (b|_{\mathrm{B}}\, |\Psi)_{\mathrm{AB}}$. According to the steering property, the range of $\widehat{f}$ contains F_ρ

because every effect on the purifying system induces a state of A compatible with ρ. In general, $\dim \mathsf{Ran}\, \widehat{f} \leq \dim \mathsf{Eff}_{\mathbb{R}}(\mathrm{B})$, a fortiori $\dim \mathsf{Span}(F_\rho) \leq \dim \mathsf{Eff}_{\mathbb{R}}(\mathrm{B})$. Since $\mathsf{Eff}_{\mathbb{R}}(\mathrm{B})$ is the dual space of $\mathsf{St}_{\mathbb{R}}(\mathrm{B})$, they have the same dimension, whence

$$\dim \mathsf{St}_{\mathbb{R}}(\mathrm{B}) \geq \dim \mathsf{Span}(F_\rho).$$

If we have a completely mixed state, by lemma 2.3.24, $\mathsf{Span}(F_\rho) = \mathsf{St}_{\mathbb{R}}(\mathrm{A})$, whence

$$\dim \mathsf{St}_{\mathbb{R}}(\mathrm{B}) \geq \dim \mathsf{St}_{\mathbb{R}}(\mathrm{A}).$$

□

This means that in the completely mixed case, the purifying system must be at least as large as system A. Note that this result is equivalent to corollary 1.1.7 for quantum mechanics.

Now we move to other aspects of the purification postulate.

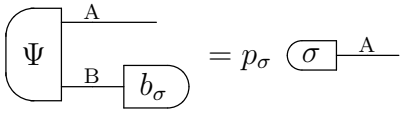
Definition 3.1.9. We say that two transformations $\mathcal{A}, \mathcal{A}' \in \mathsf{Transf}(\mathrm{A}, \mathrm{B})$ are *equal upon input* of the state $\rho \in \mathsf{St}_1(\mathrm{A})$ if $\mathcal{A}\sigma = \mathcal{A}'\sigma$ for every state σ compatible with ρ. In this case we will write $\mathcal{A} =_\rho \mathcal{A}'$.

This means that in quantum mechanics two quantum operations are equal upon input of the density operator ρ if they coincide on every density operator whose support is contained in the support of ρ.

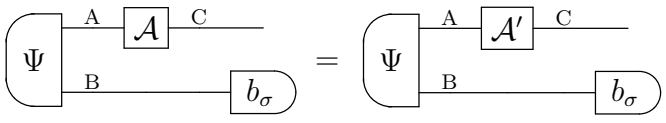
In general, we can prove the following proposition. Loosely speaking, one has that equality on purifications implies equality upon input.

Proposition 3.1.10. *Let $\rho \in \mathsf{St}_1(\mathrm{A})$ and let $\Psi \in \mathsf{St}_1(\mathrm{AB})$ be a purification of ρ. If $\mathcal{A}$ and $\mathcal{A}'$ are two transformations from* A *to* C *such that $\mathcal{A}|\Psi)_{\mathrm{AB}} = \mathcal{A}'|\Psi)_{\mathrm{AB}}$, then $\mathcal{A} =_\rho \mathcal{A}'$.*

Proof. By the steering property, a state $\sigma \in \mathsf{St}_1(\mathrm{A})$ is in the face identified by ρ if and only if there exist an effect b_σ on the purifying system B such that



where $p_\sigma \in (0, 1]$. Hence, we have $\mathcal{A} =_\rho \mathcal{A}'$ if and only if $\mathcal{A}\sigma = \mathcal{A}'\sigma$, namely if and only if



(3.1)

Clearly, if $\mathcal{A}\,|\Psi)_{\mathrm{AB}} = \mathcal{A}'\,|\Psi)_{\mathrm{AB}}$, then equality in (3.1) holds, therefore we conclude that $\mathcal{A} =_\rho \mathcal{A}'$. □

This proposition shows that if two local transformations are equal on a purification of a given state, then they are equal upon input of that state. Is the converse true? In other words, if two transformations are equal upon input of ρ, do they yield equal outputs if applied to a purification of ρ?

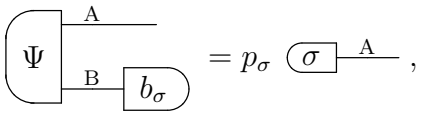
The answer, in general, is negative. The answer is affirmative for theories that fulfil local discriminability, but, for instance, in quantum mechanics with real Hilbert space, this property fails [44].

However, the converse of proposition 3.1.10 holds for effects. Essentially, the reason is that only one system is enough to distinguish effects, whereas, we need an ancillary system for general transformations (see subsection 2.1.3).

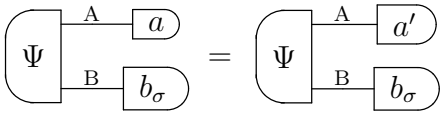
Proposition 3.1.11. *Let $\rho \in \mathsf{St}_1(\mathrm{A})$ and let $\Psi \in \mathsf{St}_1(\mathrm{AB})$ be a purification of ρ. If a and a' are two effects on* A*, then one has $(a|_{\mathrm{A}}\,|\Psi)_{\mathrm{AB}} = (a'|_{\mathrm{A}}\,|\Psi)_{\mathrm{AB}}$ if and only if $a =_\rho a'$.*

Proof. Necessity was already proven in proposition 3.1.10, it is enough to take C to be the trivial system I.

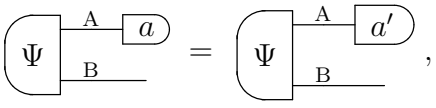
Let us move to sufficiency. Suppose $a =_\rho a'$. This means that $(a|\sigma) = (a'|\sigma)$ for every normalized state σ compatible with ρ. By the steering property, for every effect b_σ on B, we get a state σ compatible with ρ.



where $p_\sigma \in (0, 1]$. Hence, if $a =_\rho a'$, then



Since this equality holds for every effect b_σ, we conclude that



that is $(a|_{\mathrm{A}}\,|\Psi)_{\mathrm{AB}} = (a'|_{\mathrm{A}}\,|\Psi)_{\mathrm{AB}}$. □

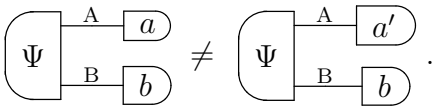
Now we move to a particular type of states.

Definition 3.1.12. We say that a state $\rho \in \mathsf{St}_1(\mathrm{AB})$ is *dynamically faithful* (*faithful* for short) for system A if, for any $\mathcal{A}, \mathcal{A}' \in \mathsf{Transf}(\mathrm{A}, \mathrm{C})$, we have $\mathcal{A} = \mathcal{A}'$ if $\mathcal{A}|\rho)_{\mathrm{AB}} = \mathcal{A}'|\rho)_{\mathrm{AB}}$.

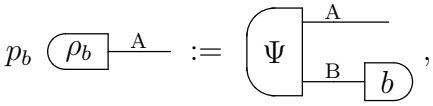
In this way, we can discriminate two transformations acting on A by checking their action on a dynamically faithful state for A. Note that the faithful state is not a state of A, but we must consider an ancillary system B.

Proposition 3.1.13. *If a pure state* $|\Psi)_{\mathrm{AB}}$ *is faithful for system* A*, then its marginal* ω *on* A *is completely mixed.*

Proof. Consider two distinct effects $a, a' \in \mathsf{Eff}(\mathrm{A})$. Then, since $|\Psi)_{\mathrm{AB}}$ is faithful, $(a|_{\mathrm{A}}|\Psi)_{\mathrm{AB}} \neq (a'|_{\mathrm{A}}|\Psi)_{\mathrm{AB}}$. $(a|_{\mathrm{A}}|\Psi)_{\mathrm{AB}}$ and $(a'|_{\mathrm{A}}|\Psi)_{\mathrm{AB}}$ are states of system B, therefore there is an effect b on B such that



Let us define $\rho_b \in \mathsf{St}_1(\mathrm{A})$ as



where $p_b \in (0, 1]$. By construction, $\rho_b \in F_\omega$, being ω the marginal of $|\Psi)_{\mathrm{AB}}$ on A, and $(a|\rho_b) \neq (a'|\rho_b)$. This means that F_ω is a separating set for $\mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$ (i.e. its elements are able to distinguish between elements of $\mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$). This means that F_ω is a spanning set for $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$, hence ω is completely mixed (see lemma 2.3.24). □

Once more, to state the converse, namely that if a state is completely mixed then a purification is faithful, local discriminability is necessary [40]. Nevertheless, we can state the following proposition.

Proposition 3.1.14. *Let* $|\Psi)_{\mathrm{AB}}$ *be a bipartite pure state with a completely mixed marginal* $|\omega)_{\mathrm{A}}$ *on* A*. If* $\mathcal{A}, \mathcal{A}' \in \mathsf{Transf}(\mathrm{A}, \mathrm{C})$ *and* $\mathcal{A}|\Psi)_{\mathrm{AB}} = \mathcal{A}'|\Psi)_{\mathrm{AB}}$*, then* $\mathcal{A}|\rho)_{\mathrm{A}} = \mathcal{A}'|\rho)_{\mathrm{A}}$ *for every* $\rho \in \mathsf{St}(\mathrm{A})$*.*

Proof. If $\mathcal{A}, \mathcal{A}' \in \mathsf{Transf}(\mathrm{A}, \mathrm{C})$ are such that $\mathcal{A}\,|\Psi)_{\mathrm{AB}} = \mathcal{A}'\,|\Psi)_{\mathrm{AB}}$, then, by proposition 3.1.10, this implies $\mathcal{A} =_{\omega} \mathcal{A}'$. Since ω is completely mixed, we have $\mathcal{A}\,|\rho)_{\mathrm{A}} = \mathcal{A}'\,|\rho)_{\mathrm{A}}$ for every $\rho \in \mathsf{St}(\mathrm{A})$. □

However, if we content ourselves with faithful states for effects, we can prove a complete equivalence between faithful states and completely mixed marginal.

Definition 3.1.15. We say that a state $\rho \in \mathsf{St}_1(\mathrm{AB})$ is *faithful for effects* of system A if, for any $a, a' \in \mathsf{Eff}(\mathrm{A})$, we have $a = a'$ if $a\,|\rho)_{\mathrm{AB}} = a'\,|\rho)_{\mathrm{AB}}$.

We can distinguish between two effects by checking their action on a faithful state for effects. The purification postulate guarantees the existence of pure faithful states for effects.

Proposition 3.1.16. *A pure state $|\Psi)_{\mathrm{AB}}$ is faithful for effects of system* A *if and only if its marginal $|\omega)_{\mathrm{A}}$ on* A *is completely mixed.*

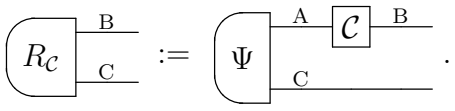
Proof. We already proved necessity in proposition 3.1.13: a faithful state is clearly a faithful state for effects.

Sufficiency. Suppose ω is completely mixed. We proved in proposition 3.1.14 that $(a|\rho) = (a'|\rho)$ for every $\rho \in \mathsf{St}(\mathrm{A})$. This means that $a = a'$. □

3.2 Choi correspondence

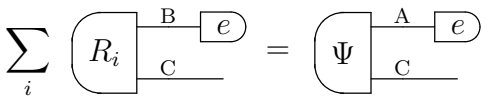
In this section we analyse an important correspondence between states and transformations. Although it may seem a technical result at first glance, it has a lot of important implications. The basic ingredients are purifications of completely mixed states.

Definition 3.2.1. If $\Psi \in \mathsf{St}_1(\mathrm{AC})$ is a purification of a completely mixed state of A, for every transformation $\mathcal{C} \in \mathsf{Transf}(\mathrm{A}, \mathrm{B})$, we define the *Choi state* $R_{\mathcal{C}} \in \mathsf{St}(\mathrm{BC})$ as



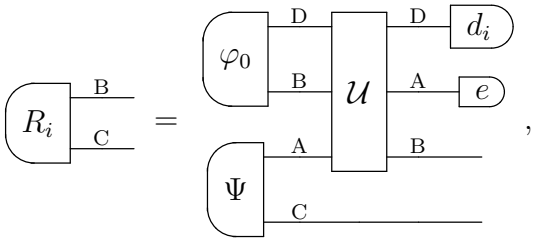
In other words, $|R_{\mathcal{C}})_{\mathrm{BC}} = \mathcal{C}\,|\Psi)_{\mathrm{AC}}$.

Lemma 3.2.2. *Let* $\{R_i\}_{i\in X}$ *be a preparation-test for system* BC *such that*



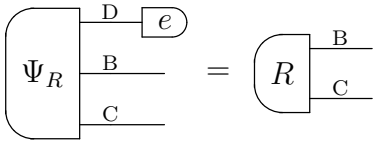
(3.2)

where $|\Psi)_{\mathrm{AC}}$ *is a pure state. Then, there exist a system* D*, a pure state* $\varphi_0 \in \mathsf{St}_1(\mathrm{BD})$*, a reversible channel* $\mathcal{U}$ *on system* ABD *and an observation-test* $\{d_i\}_{i\in X}$ *on system* D*, such that*

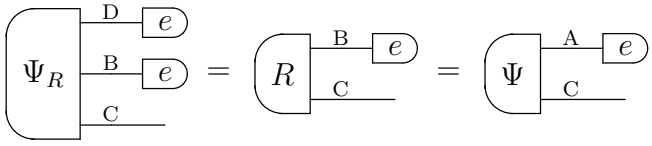


for any outcome $i \in X$*.*

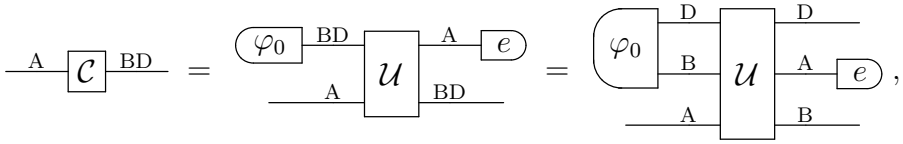
Proof. Let $|\Psi_R)_{\mathrm{DBC}}$ be a purification of the coarse-grained state $R = \sum_{i\in X} R_i$, with purifying system D.



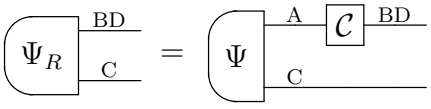
Let us take the deterministic effect on B. Recalling (3.2), we have



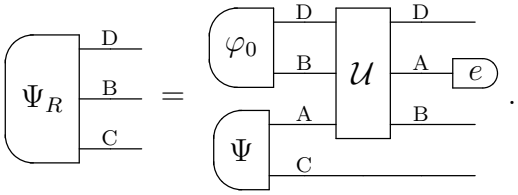
Hence $|\Psi_R)_{\mathrm{DBC}}$ and $|\Psi)_{\mathrm{AC}}$ have the same marginal on system C. By lemma 3.1.6, there exists a channel $\mathcal{C}$ from A to BD, defined as



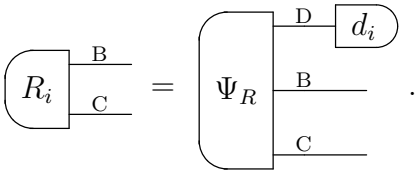
for some pure state $\varphi_0 \in \mathsf{St}_1(\mathrm{BD})$ and some reversible channel $\mathcal{U}$ on ABD; such that



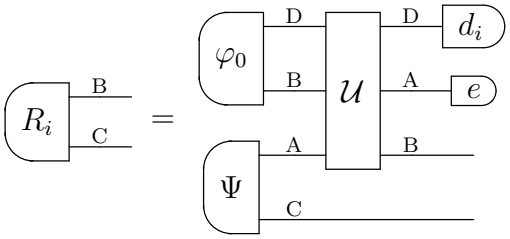
Then



Now, according to the steering property, there exists an observation-test $\{d_i\}_{i\in X}$ on D such that



We immediately get the thesis.



□

We have the following important theorem.

Theorem 3.2.3 (Choi correspondence)**.** *If $\mathcal{C} \in \mathsf{Transf}(\mathrm{A},\mathrm{B})$, the map $\mathcal{C} \mapsto R_{\mathcal{C}}$, for a given purification $|\Psi\rangle_{\mathrm{AC}}$ of a completely mixed state of* A*, has the following properties.*

1. *It defines a surjective map between tests $\{\mathcal{C}_i\}_{i\in X}$ from* A *to* B *and preparation-tests $\{R_i\}_{i\in X}$ of* BC*, such that*

$$\sum_{i\in X} (e|_{\mathrm{B}}\,|R_i\rangle_{\mathrm{BC}} = (e|_{\mathrm{A}}\,|\Psi\rangle_{\mathrm{AC}}\,. \tag{3.3}$$

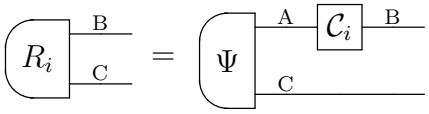
   The map is bijective if $|\Psi\rangle_{\mathrm{AC}}$ is faithful for A*.*

2. *The transformation $\mathcal{C}$ is pure if and only if $R_{\mathcal{C}}$ is pure.*

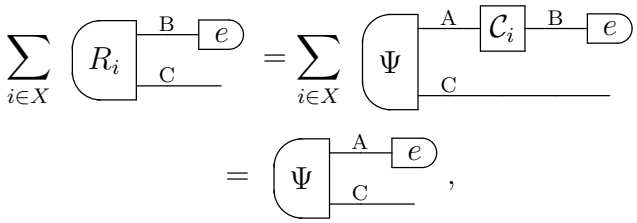
3. *The map $\mathcal{C} \mapsto R_{\mathcal{C}}$, from* $\mathsf{Transf}(\mathrm{A}, \mathrm{B})$ *to bipartite states $|R)_{\mathrm{BC}}$ of* BC *such that $(e|_{\mathrm{B}} |R)_{\mathrm{BC}} \in F_{\widetilde{\omega}}$, where $|\widetilde{\omega})_{\mathrm{C}} = (e|_{\mathrm{A}} |\Psi)_{\mathrm{AC}}$ is surjective. It is bijective if $|\Psi)_{\mathrm{AC}}$ is faithful for* A.

Proof. Let us prove the three items.

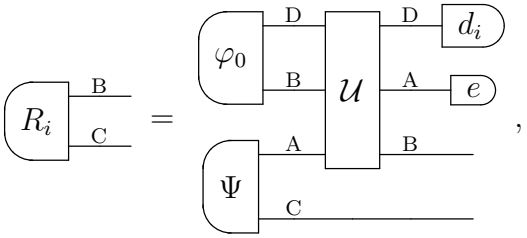
1. Let $\{\mathcal{C}_i\}_{i\in X}$ be a test from A to B; it must be $\sum_{i\in X} (e|_{\mathrm{B}} \mathcal{C}_i = (e|_{\mathrm{A}}$. We can consider the preparation-test $\{R_i\}_{i\in X}$, made up of Choi states associated with $\mathcal{C}_i$'s, with $R_i := R_{\mathcal{C}_i}$.



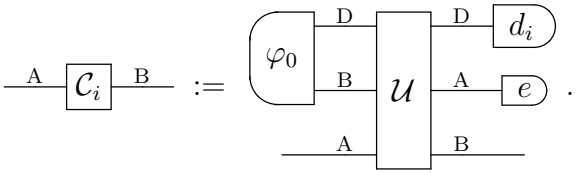
Clearly, it holds



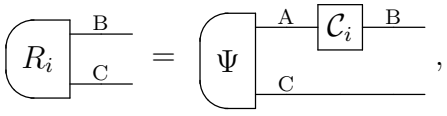
because $\sum_{i\in X} (e|_{\mathrm{B}} \mathcal{C}_i = (e|_{\mathrm{A}}$. Hence, eq. (3.3) is fulfilled. We have just shown that we can associate a preparation-test of system BC with every test $\{\mathcal{C}_i\}_{i\in X} \subset \mathsf{Transf}(\mathrm{A}, \mathrm{B})$. If two tests $\{\mathcal{C}_i\}_{i\in X}$ and $\{\mathcal{C}'_i\}_{i\in X}$ are such that $R_{\mathcal{C}_i} = R_{\mathcal{C}'_i}$ for every $i \in X$, then $\mathcal{C}_i |\rho)_{\mathrm{A}} = \mathcal{C}'_i |\rho)_{\mathrm{A}}$ for every $\rho \in \mathsf{St}(\mathrm{A})$, according to proposition 3.1.14. If $|\Psi)_{\mathrm{AC}}$ is faithful for A, then $\{\mathcal{C}_i\}_{i\in X} = \{\mathcal{C}'_i\}_{i\in X}$.
Let us now prove the converse, namely that we can associate a test from A to B with each preparation-test $\{R_i\}_{i\in X}$ on BC that satisfies eq. (3.3). In this case, by lemma 3.2.2, there exist a system D, a pure state $\varphi_0 \in \mathsf{St}_1(\mathrm{BD})$, a reversible channel on ABD and an observation-test $\{d_i\}_{i\in X}$ on D such that



for any outcome $i \in X$. Now we define a test $\{\mathcal{C}_i\}_{i\in X}$ from A to B, where



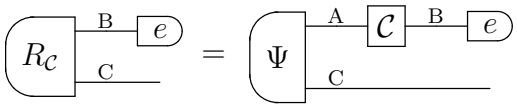
In this way,



and therefore item 1 is proven.

2. Item 2 is a straightforward consequence of item 1. Necessity is trivial, because if $\mathcal{C}$ is pure, then $R_{\mathcal{C}}$ is pure too, by pure conditioning, being the composition of a pure state with a pure transformation. Conversely, if $R_{\mathcal{C}}$ is pure, $\mathcal{C}$ must be pure, otherwise we would have a non-trivial refinement of $R_{\mathcal{C}}$.

3. Consider the Choi state $R_{\mathcal{C}}$ associated with $\mathcal{C} \in \mathsf{Transf}\,(\mathrm{A}, \mathrm{B})$. If we take the deterministic effect on system B, we have



Now, $(e|_{\mathrm{B}}\,\mathcal{C}$ is an effect on A, so, according to the steering property, it will induce a state of C which is in the refinement set of the marginal of $|\Psi)_{\mathrm{AC}}$ on C, that is $|\widetilde{\omega})_{\mathrm{C}}$. As in item 1, we do not have injectivity, because if $\mathcal{C}$ and $\mathcal{C}'$ are two transformations such that $R_{\mathcal{C}} = R_{\mathcal{C}'}$ we can conclude only that $\mathcal{C}\,|\rho)_{\mathrm{A}} = \mathcal{C}'\,|\rho)_{\mathrm{A}}$ for every $\rho \in \mathsf{St}\,(\mathrm{A})$. If $|\Psi)_{\mathrm{AC}}$ is faithful for A, then $\mathcal{C} = \mathcal{C}'$.
Let us prove surjectivity. Consider $|R)_{\mathrm{BC}}$ such that $(e|_{\mathrm{B}}\,|R)_{\mathrm{BC}}$ is compatible with $|\widetilde{\omega})_{\mathrm{C}}$. This means there exists a preparation-test $\{\widetilde{\omega}_i\}_{i\in X}$ for system C such that $\widetilde{\omega} = \sum_{i\in X}\widetilde{\omega}_i$ and $(e|_{\mathrm{B}}\,|R)_{\mathrm{BC}} = |\widetilde{\omega}_{i_0})_{\mathrm{C}}$ for some $i_0 \in X$. Now, consider an arbitrary set of normalized states $\{\rho_i\}_{i\in X}$ of system B and consider the collection of states $\{R_i\}_{i\in X}$ of system BC, defined as $R_{i_0} := R$ and $R_i := \rho_i \otimes \widetilde{\omega}_i$ if $i \neq i_0$. $\{R_i\}_{i\in X}$ is a preparation-test, and it can be obtained by randomization of the normalized states

$\overline{R}_i := \frac{R_i}{\|R_i\|}$ with probabilities $\|R_i\|$. Let us check if $\{R_i\}_{i\in X}$ satisfies eq. (3.3). We have

$$\sum_{i\in X} (e|_{\mathrm{B}} |R_i)_{\mathrm{BC}} = \sum_{i\in X} |\widetilde{\omega}_i)_{\mathrm{C}} = |\widetilde{\omega})_{\mathrm{C}}$$

because $(e|_{\mathrm{B}} |R)_{\mathrm{BC}} = |\widetilde{\omega}_{i_0})_{\mathrm{C}}$ and $(e|_{\mathrm{B}} |R_i)_{\mathrm{BC}} = |\widetilde{\omega}_i)_{\mathrm{C}}$ also for $i \neq i_0$. By definition, $|\widetilde{\omega})_{\mathrm{C}} = (e|_{\mathrm{A}} |\Psi)_{\mathrm{AC}}$, so eq. (3.3) is fulfilled. Therefore, using item 1, there is a test $\{\mathcal{C}_i\}_{i\in X} \subseteq \mathsf{Transf}(\mathrm{A},\mathrm{B})$ such that $R_i = R_{\mathcal{C}_i}$. In particular, $R = R_{i_0} = R_{\mathcal{C}_{i_0}}$. Thus, we managed to associate a transformation $\mathcal{C} := \mathcal{C}_{i_0}$ with the bipartite state $|R)_{\mathrm{BC}}$.

□

We see that the map is almost bijective, because it may happen we can associate two tests $\{\mathcal{C}_i\}_{i\in X}$ and $\{\mathcal{C}'_i\}_{i\in X}$ with the same preparation-test $\{R_i\}_{i\in X}$. However, these two tests are almost equal, because we have $\mathcal{C}_i |\rho)_{\mathrm{A}} = \mathcal{C}'_i |\rho)_{\mathrm{A}}$ for every $\rho \in \mathsf{St}(\mathrm{A})$ and for every $i \in X$. If the theory satisfies local discriminability, then this is enough to conclude that $\mathcal{C}_i = \mathcal{C}'_i$, but in general this is not true.

Clearly, the surjective map from states $\mathsf{St}(\mathrm{BC})$ to transformations $\mathsf{Transf}(\mathrm{A},\mathrm{B})$ can be extended to a linear map from the vector space $\mathsf{St}_{\mathbb{R}}(\mathrm{BC})$ to the vector space $\mathsf{Transf}_{\mathbb{R}}(\mathrm{A},\mathrm{B})$. This shows that $\dim \mathsf{Transf}_{\mathbb{R}}(\mathrm{A},\mathrm{B}) \leq \dim \mathsf{St}_{\mathbb{R}}(\mathrm{BC})$, so $\mathsf{Transf}_{\mathbb{R}}(\mathrm{A},\mathrm{B})$ is a finite-dimensional vector space too.

Now let us come to the implications of Choi correspondence. An immediate corollary states that the characterization of tests and observation-tests presented in section 2.3 is also sufficient.

Corollary 3.2.4. *Let* $\{\mathcal{C}_i\}_{i\in X} \subseteq \mathsf{Transf}(\mathrm{A},\mathrm{B})$ *be a collection of transformations from* A *to* B. $\{\mathcal{C}_i\}_{i\in X}$ *is a test if and only if*

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

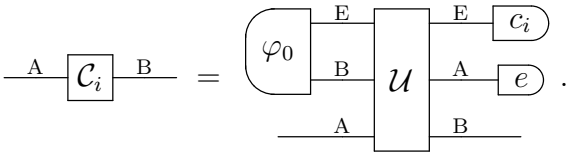
Proof. We already proved necessity in section 2.3.

Sufficiency. Consider the Choi state associated with the transformation $\mathcal{C}_i$, $|R_i)_{\mathrm{BC}} = \mathcal{C}_i |\Psi)_{\mathrm{AC}}$, where $|\Psi)_{\mathrm{AC}}$ is a purification of a completely mixed state of A. Clearly $\{R_i\}_{i\in X}$ satisfies eq. (3.3) by hypothesis, so we can associate a test with $\{R_i\}_{i\in X}$, such that R_i is the Choi state associated with $\mathcal{C}_i$ for every $i \in X$. But we already have that $\{\mathcal{C}_i\}_{i\in X}$ produces $\{R_i\}_{i\in X}$, therefore $\{\mathcal{C}_i\}_{i\in X}$ is a test. □

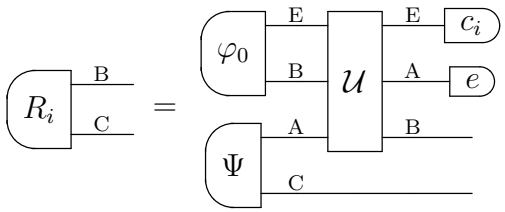
In particular, we have that $\{a_i\}_{i\in X} \subseteq \mathsf{Eff}(\mathrm{A})$ is an observation-test if and only if $\sum_{i\in X} a_i = e$ (here we have $\mathrm{B} = \mathrm{I}$).

Other corollaries show that every transformation comes from a reversible channel on a larger system. This is the abstract versions of Ozawa's theorem in quantum mechanics [21].

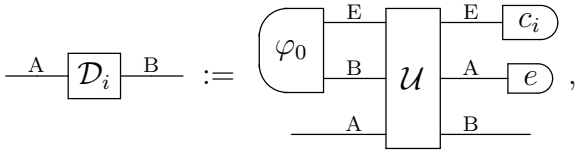
Corollary 3.2.5. *If the theory admits a faithful pure state for* A*, for every test* $\{\mathcal{C}_i\}_{i\in X}$ *from system* A *to system* B*, there exist a system* E*, a pure state* $\varphi_0 \in \mathsf{St}_1(\mathrm{BE})$*, a reversible channel* $\mathcal{U}$ *on* ABE*, and an observation-test* $\{c_i\}_{i\in X}$ *on system* E *such that for every* $i \in X$ *one has*



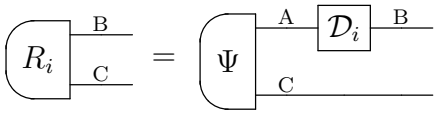
Proof. Let $|\Psi)_{\mathrm{AC}}$ be a faithful pure state for system A. Let us consider the Choi states $|R_i)_{\mathrm{BC}}$ associated with $\mathcal{C}_i$'s. We know that $\sum_{i\in X} (e|_{\mathrm{B}} |R_i)_{\mathrm{BC}} = (e|_{\mathrm{A}} |\Psi)_{\mathrm{AC}}$. According to lemma 3.2.2, there exist a system E, a pure state $\varphi_0 \in \mathsf{St}_1(\mathrm{BE})$, a reversible channel $\mathcal{U}$ on ABE, and an observation-test $\{c_i\}_{i\in X}$ on system E such that for every $i \in X$ one has



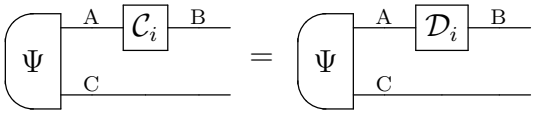
We define



whence

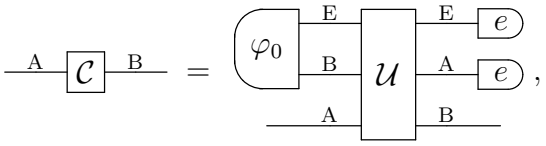


But $|R_i)_{\mathrm{BC}} = \mathcal{C}_i |\Psi)_{\mathrm{AC}}$, so we have



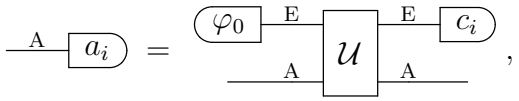
Since $|\Psi)_{\mathrm{AC}}$ is faithful for system A, we conclude that $\mathcal{C}_i = \mathcal{D}_i$ for every $i \in X$, whence the thesis follows. □

We can treat immediately two particular cases. The first case is when we have a deterministic test, i.e. a channel $\mathcal{C}$ from A to B. In this case, we replace the observation-test on E with the deterministic effect.



for some system E, some pure state $\varphi_0 \in \mathsf{St}_1(\mathrm{BE})$ and some reversible channel $\mathcal{U}$ on ABE. This is the abstract version of Stinespring's theorem [20].

The second case is when we have an observation-test $\{a_i\}_{i\in X}$ on A. In this case, system B is the trivial system, and

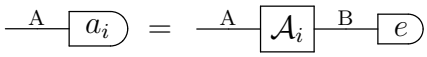


for some system E, some pure state $\varphi_0 \in \mathsf{St}_1(\mathrm{E})$ and some reversible channel $\mathcal{U}$ on AE. This is the abstract version of Naimark's theorem for quantum mechanics [19].

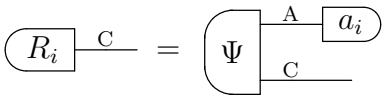
We see that the purification postulate is the actual reason that enables us to prove these important theorems even in an abstract scenario.

Actually, there is another characterization of observation-tests, which holds also when there are no faithful pure states for system A. We will use this characterization to diagonalize mixed states in chapter 5.

Corollary 3.2.6. *For every observation-test* $\{a_i\}_{i\in X}$ *on* A, *there is a system* B *and a test* $\{\mathcal{A}_i\}_{i\in X} \subset \mathsf{Transf}(\mathrm{A}, \mathrm{B})$ *such that every* $\mathcal{A}_i$ *is pure and* $(a_i|_{\mathrm{A}} = (e|_{\mathrm{B}}\, \mathcal{A}_i$.

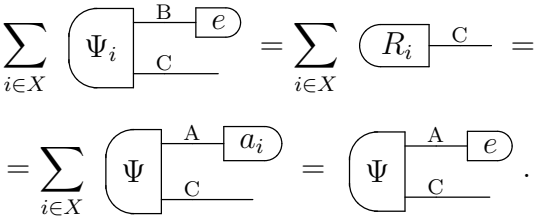


Proof. Let $|\Psi)_{\mathrm{AC}}$ be a purification of a completely mixed state of A. Let $|R_i)_{\mathrm{C}}$ be the Choi state associated with $(a_i|_{\mathrm{A}}$.

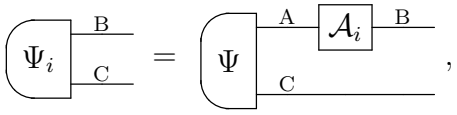


(3.4)

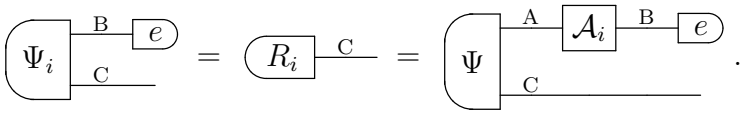
Let us consider a purification $|\Psi_i)_{\mathrm{BC}}$ of $|R_i)_{\mathrm{C}}$. If we take the deterministic effect on B, and we sum over i, we have



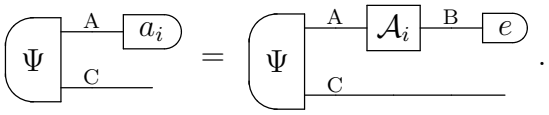
Therefore, since $|\Psi_i)_{\mathrm{BC}}$'s fulfil eq. (3.3), by Choi correspondence there exists a test $\{\mathcal{A}_i\}_{i\in X} \subset \mathsf{Transf}\,(\mathrm{A},\mathrm{B})$ such that



where every $\mathcal{A}_i$ is pure because $|\Psi_i)_{\mathrm{BC}}$ is pure (see theorem 3.2.3). If we take the deterministic effect on system B, we get



Recalling the definition of R_i (eq. (3.4)), we have



Since $|\Psi)_{\mathrm{AC}}$ is faithful for effects of system A, it immediately follows that $(a_i|_{\mathrm{A}} = (e|_{\mathrm{B}}\,\mathcal{A}_i$. □

Now we have some results about the topological properties of sets of transformations, channels and states.

Proposition 3.2.7. *Given two systems* A *and* B, $\mathsf{Transf}\,(\mathrm{A},\mathrm{B})$ *is compact in the operational norm.*

Proof. The proof is slightly too technical and lengthy, so we invite the interested reader to refer to [39]. □

Proposition 3.2.8. *Given two systems* A *and* B, *the set of channels from* A *to* B *is compact in the operational norm.*

Proof. Since $\mathsf{Transf}_{\mathbb{R}}(\mathrm{A},\mathrm{B})$ is finite-dimensional and the set of channels is limited (indeed $\|\mathcal{C}\| \leq 1$ for every channel $\mathcal{C}$ from A to B), it is enough to show that it is closed. Consider a Cauchy sequence of channels $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$. Since $\mathsf{Transf}(\mathrm{A},\mathrm{B})$ is closed, $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$ converges to a transformation $\mathcal{C}$. We must show that $\mathcal{C}$ is a channel. Indeed, since each $\mathcal{C}_n$ is a channel, we have $(e|_{\mathrm{B}}\,\mathcal{C}_n = (e|_{\mathrm{A}}$. Hence, for every state $\rho \in \mathsf{St}(\mathrm{A})$, one has

$$(e|_{\mathrm{B}}\,\mathcal{C}\,|\rho)_{\mathrm{A}} = \lim_{n\to+\infty} (e|_{\mathrm{B}}\,\mathcal{C}_n\,|\rho)_{\mathrm{A}} = (e|\rho)_{\mathrm{A}}$$

This shows that $(e|_{\mathrm{B}}\,\mathcal{C} = (e|_{\mathrm{A}}$, so $\mathcal{C}$ is a channel, according to proposition 2.3.6. □

Then we can prove a result about reversible channels acting on a system A.

Proposition 3.2.9. *The group* $\mathbf{G}_{\mathrm{A}}$ *of reversible channels acting on* A *is a compact Lie group.*

Proof. Let $\{\mathcal{U}_n\}_{n\in\mathbb{N}}$ be a sequence of reversible channels converging to a channel $\mathcal{U}$. Let us show that $\mathcal{U}$ is reversible. Consider the sequence $\{\mathcal{U}_n^{-1}\}_{n\in\mathbb{N}}$. Since the set of channels is compact, we can choose a subsequence $\{\mathcal{U}_{n_k}^{-1}\}$ converging to a channel $\mathcal{C}$. Every subsequence of $\{\mathcal{U}_n\}_{n\in\mathbb{N}}$ converges to $\mathcal{U}$, therefore

$$\mathcal{U}\mathcal{C} = \lim_{k\to+\infty} \mathcal{U}_{n_k}\mathcal{U}_{n_k}^{-1} = \mathcal{I}.$$

Similarly one proves that also $\mathcal{C}\mathcal{U} = \mathcal{I}$, so $\mathcal{U}$ is reversible and $\mathcal{C}$ is its inverse, $\mathcal{U}^{-1} = \mathcal{C}$. This proves that $\mathbf{G}_{\mathrm{A}}$ is closed and therefore compact.

Since $\mathbf{G}_{\mathrm{A}}$ has a faithful finite-dimensional representation (see [53]) it is a Lie group. □

Finally, we have a result about the topology of pure states.

Proposition 3.2.10. *The set of pure states of a system* A *is compact.*

Proof. Again, to prove compactness it is enough to prove closure. Let us consider a sequence of pure states $\{\psi_n\}_{n\in\mathbb{N}}$ converging to some state ρ. We want to show that ρ is pure. In a theory with purification, the set of pure states is transitive under the action of the group of reversible channels. Therefore, we can write $\psi_n = \mathcal{U}_n\varphi_0$ for some pure state φ_0, for every $n \in \mathbb{N}$. Since $\mathbf{G}_{\mathrm{A}}$

is compact, we can consider the subsequence $\{\mathcal{U}_{n_k}\}$, which converges to the reversible channel $\mathcal{U}$. Therefore,

$$\rho = \lim_{k\to+\infty} \psi_{n_k} = \lim_{k\to+\infty} \mathcal{U}_{n_k}\varphi_0 = \mathcal{U}\varphi_0.$$

Therefore ρ is pure. □

Remark 3.2.11. For a generic probabilistic theory, without purification, in general, it is not true that the set of pure states is closed. Let us consider the convex hull of the subset of the 3-dimensional real affine $\mathbb{A}(\mathbb{R}^3)$ space

$$S = \left\{(x, y, 0) \in \mathbb{A}\left(\mathbb{R}^3\right) : x^2 + y^2 = 1\right\} \cup \{(1, 0, \pm 1)\}.$$

S is the union of a circumference and two points, therefore its convex hull is the union of two cones having vertexes in $(1, 0, 1)$ and $(1, 0, -1)$ and the same circle $\{(x, y, 0) \in \mathbb{A}(\mathbb{R}^3) : x^2 + y^2 \leq 1\}$ as basis. The set of pure states is made of the two vertexes $(1, 0, 1)$ and $(1, 0, -1)$ and of the points of the circumference $\{(x, y, 0) \in \mathbb{A}(\mathbb{R}^3) : x^2 + y^2 = 1\}$, except the point $(1, 0, 0)$, which is mixed. But we can find a sequence of pure states on the circumference that converges to $(1, 0, 0)$; thus in this theory there is a sequence of pure states converging to a mixed state. The set of pure states is not closed.

The last consequence of Choi correspondence concerns invariant states. We are able to prove that the invariant state exists and it is unique. In addition it is a completely mixed state.

However, we need a lemma first.

Lemma 3.2.12. *For every system* A*, there exists a twirling test* $\{p_i\mathcal{U}_i\}_{i\in X}$*, where* p_i*'s are probabilities and* $\mathcal{U}_i$*'s are reversible channels. One of the reversible channels* $\mathcal{U}_i$ *can always be taken to be the identity.*

Proof. Let $d\mathcal{W}$ be the normalized Haar measure on the Lie group $\mathbf{G}_{\mathrm{A}}$, namely such that $\int_{\mathbf{G}_{\mathrm{A}}} d\mathcal{W} = 1$. Let us define a channel $\mathcal{T}$ as

$$\mathcal{T} := \int_{\mathbf{G}_{\mathrm{A}}} \mathcal{W}\, d\mathcal{W}. \tag{3.5}$$

Let us show that $\mathcal{T}$ is a twirling channel (see definition 2.1.28). Let $\mathcal{U} \in \mathbf{G}_{\mathrm{A}}$, then

$$\mathcal{U}\mathcal{T} = \int_{\mathbf{G}_{\mathrm{A}}} \mathcal{U}\mathcal{W}\, d\mathcal{W} = \int_{\mathbf{G}_{\mathrm{A}}} \mathcal{W}'\, d\mathcal{W}' = \mathcal{T},$$

where we used invariance of the Haar measure. Since reversible channels span a finite-dimensional vector space, their convex hull is a finite-dimensional convex set. Then, by Carathéodory's theorem (theorem A.1.2), the integral in eq. (3.5) can be replaced by a finite convex combination of reversible channels

$$\mathcal{T} = \sum_{i \in X} p_i \mathcal{U}_i.$$

This shows that $\{p_i \mathcal{U}_i\}_{i \in X}$ is a twirling test. Since $\mathcal{U}\mathcal{T} = \mathcal{T}$, for every $\mathcal{T}$, to have identity among $\mathcal{U}_i$'s, it is sufficient to choose one of the $\mathcal{U}_i$'s, say $\mathcal{U}_{i_0}$, and apply $\mathcal{U}_{i_0}^{-1}$ after $\mathcal{T}$. In this way, we have

$$\mathcal{T} = \sum_{i \in X} p_i \mathcal{U}_{i_0}^{-1} \mathcal{U}_i = p_i \mathcal{I} + \sum_{i \neq i_0} p_i \widetilde{\mathcal{U}}_i,$$

where $\widetilde{\mathcal{U}}_i := \mathcal{U}_{i_0}^{-1} \mathcal{U}_i$. The lemma is proven. □

Now we are ready to state the following proposition.

Proposition 3.2.13. *For every system* A*, there is a unique invariant state* $|\chi)_{\mathrm{A}}$*. Moreover* χ *is completely mixed.*

Proof. Let channel $\mathcal{T}$ be defined as in lemma 3.2.12. For every couple of pure states ψ and ψ', there is a reversible channel such that $\psi' = \mathcal{U}\psi$. Let us evaluate $\mathcal{T}\psi'$.

$$\mathcal{T}\psi' = \mathcal{T}\mathcal{U}\psi = \int_{\mathbf{G}_{\mathrm{A}}} \mathcal{W}\mathcal{U}\psi \, d\mathcal{W} = \int_{\mathbf{G}_{\mathrm{A}}} \mathcal{W}'\psi \, d\mathcal{W}' = \mathcal{T}\psi,$$

where we used invariance of the Haar measure. This shows that $\mathcal{T}$ is constant on pure states. Define $\chi := \mathcal{T}\psi$. Since every (normalized) mixed state can be written as a convex combination of (normalized) pure states, we have

$$\mathcal{T}\rho = \mathcal{T}\left(\sum_i p_i \psi_i\right) = \sum_i p_i \mathcal{T}\psi_i = \left(\sum_i p_i\right) \chi = \chi.$$

Therefore $\mathcal{T}$ is constant on all states. In particular, if ρ is an invariant state, we have

$$\mathcal{T}\rho = \int_{\mathbf{G}_{\mathrm{A}}} \mathcal{W}\rho \, d\mathcal{W} = \left(\int_{\mathbf{G}_{\mathrm{A}}} d\mathcal{W}\right) \rho = \rho.$$

But $\mathcal{T}\rho = \chi$. This shows that the invariant state exists (because $\chi = \mathcal{T}\rho$ for any state ρ) and it is unique (because every invariant state is equal to χ).

Let us show that χ is completely mixed. By lemma 3.2.12, we have that $\mathcal{T}$ can be expressed as a sum of the transformations of a twirling test containing identity. Therefore, for every ρ

$$\chi = \mathcal{T}\rho = p_{i_0}\rho + \sum_{i \neq i_0} p_i \mathcal{U}_i \rho.$$

This means that every ρ is compatible with χ, so χ is completely mixed. □

Now we have concluded our introduction to probabilistic theories; in the next chapter we will start examining the core of this work: entanglement and mixedness in the context of a general probabilistic theory.

Chapter 4

Entanglement in theories with purification

In this chapter, we enter into the core of the present work and into its original part. We follow the route delineated by Thirring [11], according to which the foundations of (quantum) statistical mechanics should be based on an ordering of states according to their mixedness. We have already mentioned that pure-state entanglement also plays a central role in the foundations of statistical mechanics [12]. It is indeed possible to define also an ordering of entangled pure states in terms of their entanglement.

At first glance, entanglement and mixedness may seem the farthest concepts. Indeed, pure-state entanglement has been proven to be a valuable resource to implement information protocols, such as quantum teleportation or dense coding (see [16, 17]). Mixedness, instead, means incompleteness of information, so it does not appear as the most valuable of resources. Yet their equivalence is a well-established fact in quantum mechanics (see [17, 59, 60, 61, 62]). Therefore, it is the same to build foundations of thermodynamics on mixedness or on entanglement, at least in quantum mechanics.

But what about general probabilistic theories? Can we prove the same equivalence between entanglement and mixedness? In this chapter we will try to answer this question, and our analysis will be carried out in the framework of general probabilistic theories that satisfy causality, pure conditioning, and, above all, purification. The important and original results we present set forth the equivalence between entanglement and mixedness even in an abstract framework.

As a side result, we will prove for the first time an important theorem con-

cerning a particular type of communication protocols in an abstract framework.

4.1 Entanglement and mixedness

We begin presenting the important relationship between entangled states and mixed marginals in general probabilistic theories. However, we must first give the definition of entangled states for bipartite states.

Definition 4.1.1. A normalized bipartite state $\sigma \in \mathsf{St}_1(\mathrm{AB})$ is said *separable* if it can be written as a convex combination of product states, namely $|\sigma)_{\mathrm{AB}} = \sum_i p_i |\rho_i)_{\mathrm{A}} |\tau_i)_{\mathrm{B}}$, where $p_i \geq 0$ for every i and $\sum_i p_i = 1$, and $\rho_i \in \mathsf{St}_1(\mathrm{A})$ and $\tau_i \in \mathsf{St}_1(\mathrm{B})$.

A bipartite state is *entangled* if it is not separable.

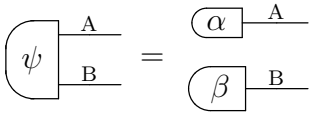
From now on, we will concentrate exclusively on the pure state case.

Note that a bipartite pure state is separable if and only if it is a product state, because it cannot be written as a non-trivial convex combination of product states.

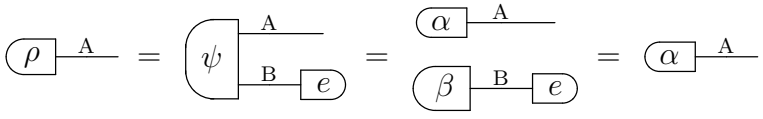
We have the following important proposition concerning pure product states.

Proposition 4.1.2. *A bipartite pure state $|\psi)_{\mathrm{AB}}$ is a product state if and only if one of its marginals, say $|\rho)_{\mathrm{A}}$, is pure.*

Proof. Necessity. Suppose $|\psi)_{\mathrm{AB}} = |\alpha)_{\mathrm{A}} |\beta)_{\mathrm{B}}$, where both $|\alpha)_{\mathrm{A}}$ and $|\beta)_{\mathrm{B}}$ must be pure, otherwise $|\psi)_{\mathrm{AB}}$ would be mixed.

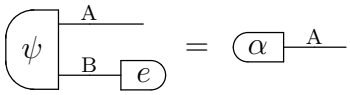


If we apply the deterministic effect to system B, we get the marginal state $|\rho)_{\mathrm{A}}$.

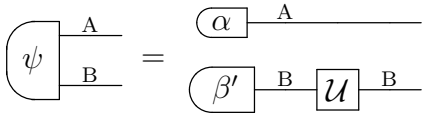


We immediately see that $|\rho)_{\mathrm{A}} = |\alpha)_{\mathrm{A}}$, hence $|\rho)_{\mathrm{A}}$ is pure. Similarly, one proves that also $|\rho)_{\mathrm{B}}$ is pure, therefore actually both marginals of $|\psi)_{\mathrm{AB}}$ are pure, not only one.

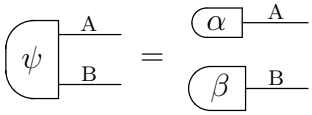
Sufficiency. Now suppose that the bipartite pure state $|\psi)_{\mathrm{AB}}$ has a pure marginal $|\rho)_{\mathrm{A}} = |\alpha)_{\mathrm{A}}$.



This means that $|\psi)_{\mathrm{AB}}$ is a purification of $|\alpha)_{\mathrm{A}}$ with purifying system B. By pure conditioning, another purification of $|\alpha)_{\mathrm{A}}$ with the same purifying system, is $|\alpha)_{\mathrm{A}}|\beta')_{\mathrm{B}}$, where $|\beta')_{\mathrm{B}}$ is a pure state of system B. Therefore, $|\psi)_{\mathrm{AB}}$ differs from $|\alpha)_{\mathrm{A}}|\beta')_{\mathrm{B}}$ by a local reversible channel $\mathcal{U}$ on system B.



Now, $\mathcal{U}|\beta')_{\mathrm{B}}$ is another pure state of system B, say $|\beta)_{\mathrm{B}}$. Therefore, $|\psi)_{\mathrm{AB}}$ is the parallel composition of $|\alpha)_{\mathrm{A}}$ and $|\beta)_{\mathrm{B}}$, hence it is the product of two pure states.



□

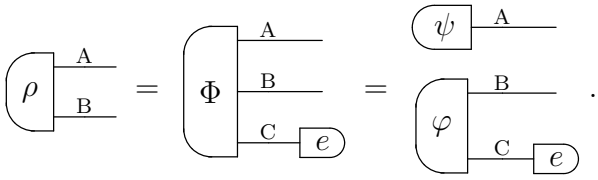
Note that we did not have to assume the purification postulate to prove necessity. Therefore, a bipartite pure product state has pure marginals in every causal theory.

Proposition 4.1.2 says that a bipartite pure state is entangled if and only if it has a mixed marginal. Actually, both marginals are mixed.

Remark 4.1.3. Following the proof of proposition 4.1.2, we can prove that if $\psi \in \mathsf{St}_1(\mathrm{A})$ is pure and $\rho \in \mathsf{St}_1(\mathrm{AB})$ is an extension of ψ, then $\rho = \psi \otimes \sigma$, for some $\sigma \in \mathsf{St}_1(\mathrm{B})$. Indeed, if ρ is pure, by proposition 4.1.2, we have $\rho = \psi \otimes \psi'$, where ψ' is a pure normalized state of system B.

If ρ is mixed, we can consider one of its purification, say $|\Phi)_{\mathrm{ABC}}$. $|\Phi)_{\mathrm{ABC}}$ is also a purification of ψ with purifying system BC. By pure conditioning, $|\Phi)_{\mathrm{ABC}} = |\psi)_{\mathrm{A}}|\varphi)_{\mathrm{BC}}$, where $|\varphi)_{\mathrm{BC}}$ is a pure normalized state of BC. Indeed, we showed in the proof of proposition 4.1.2 that every purification of a pure state is a product of pure states. Since $|\Phi)_{\mathrm{ABC}}$ is a purification of $|\rho)_{\mathrm{AB}}$, we

have



We have then $\rho = \psi \otimes \sigma$, where $\sigma = (e|_{\mathrm{C}} |\varphi)_{\mathrm{BC}}$. Note that σ is mixed, otherwise, by pure conditioning, ρ would be pure. This implies also that φ is entangled, because it has a mixed marginal.

The result expressed in proposition 4.1.2 clearly holds also for quantum mechanics. However, in quantum mechanics, we can take advantage of its specific formalism, in particular of Schmidt decomposition, to prove it. We report this proof to give a comparison with the proof of proposition 4.1.2.

Proposition 4.1.4. *A bipartite pure state $|\psi\rangle_{\mathrm{AB}}$ is a product state if and only if one of its marginals, say $\rho_{\mathrm{A}} = \mathrm{tr}_{\mathrm{B}} |\psi\rangle_{\mathrm{AB}} \langle\psi|_{\mathrm{AB}}$, is pure.*

Proof. Necessity. Let us suppose $|\psi\rangle_{\mathrm{AB}}$ is a product state, $|\psi\rangle_{\mathrm{AB}} = |\alpha\rangle_{\mathrm{A}} |\beta\rangle_{\mathrm{B}}$; then it is immediate that $\rho_{\mathrm{A}} = |\alpha\rangle_{\mathrm{A}} \langle\alpha|_{\mathrm{A}}$, hence ρ_{A} is pure. Similarly one proves that ρ_{B} is pure too.

Sufficiency. Let us suppose ρ_{A} is pure, then it has only one non-vanishing eigenvalue, which is 1. The other marginal ρ_{B} has only one non-vanishing eigenvalue too. Therefore there is only one term in the Schmidt decomposition of $|\psi\rangle_{\mathrm{AB}}$, whence $|\psi\rangle_{\mathrm{AB}}$ is a product state. In particular, suppose $\rho_{\mathrm{A}} = |\alpha\rangle_{\mathrm{A}} \langle\alpha|_{\mathrm{A}}$ and $\rho_{\mathrm{B}} = |\beta\rangle_{\mathrm{B}} \langle\beta|_{\mathrm{B}}$, then the Schmidt decomposition of $|\psi\rangle_{\mathrm{AB}}$ is $|\psi\rangle_{\mathrm{AB}} = |\alpha\rangle_{\mathrm{A}} |\beta\rangle_{\mathrm{B}}$, so $|\psi\rangle_{\mathrm{AB}}$ is the tensor product of its marginals. □

We can note that in this proof the assumptions of causality, pure conditioning and purification are not apparent, but they are hidden behind the formalism.

According to proposition 4.1.2, we can see that the purification postulate is a sufficient condition for the existence of entangled states: every theory which satisfies the purification postulate admits entangled pure states, as shown in the following corollary.

Corollary 4.1.5. *If ρ is mixed and Ψ is one of its purifications, then Ψ is entangled.*

Proof. Ψ has a mixed marginal, which is ρ, therefore it is entangled. □

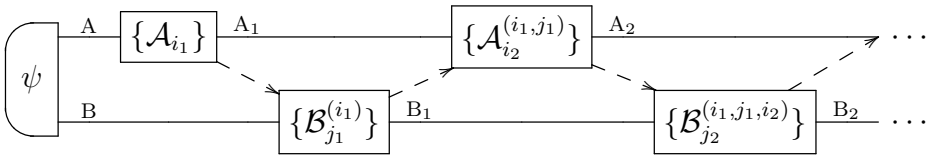
Recall that the purification postulate is just a sufficient condition for entanglement. A general probabilistic theory may admit entangled states even without the purification postulate, even though this is not true for classical theory.

4.2 The relation "to be more entangled than"

We have just seen that pure-state entanglement and mixed marginals are two equivalent notions. In this section, we would like to define an order on the set of bipartite pure states, according to their entanglement. This is important not only for the foundations of thermodynamics, but also because pure-state entanglement can be used as a resource, therefore we would like to compare two pure states and say which is more entangled (and hence more useful as a resource).

In quantum mechanics, LOCC[1] protocols have been proven not to increase entanglement of a bipartite pure state [55, 56]. Therefore, it is natural to say that a pure state is more entangled than another one if the former can be transformed into the latter by means of an LOCC protocol.

An LOCC protocol is a communication protocol in which there are two parties that are allowed to perform only local tests. They are allowed to use classical communication to communicate the outcome of each (local) test to the other party. The situation becomes more intriguing if we think the two parties to be very far away from each other. An LOCC protocol has the following form.



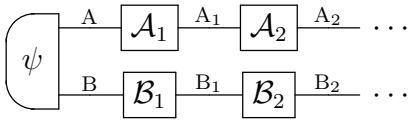
Here, round brackets explicitly show the dependence of the test by previous outcomes via classical communication, more or less like the notation for conditioning (see section 2.3). Classical communication is represented as a dashed arrow, where the tip specifies the direction in which classical communication goes. Hence, the test $\left\{\mathcal{B}_{j_1}^{(i_1)}\right\}$ follows the test $\{\mathcal{A}_{i_1}\}$ temporally, therefore it is represented after (i.e. to the right of) $\{\mathcal{A}_{i_1}\}$ in the diagram.

[1]Local Operations and Classical Communication

Clearly, there is no apparent reason why LOCC protocols should begin with Alice applying a test on her system. The protocol could start with Bob as well, this is only a matter of convention. We will adopt this convention for the rest of this work, unless explicitly stated.

Let us see some examples of particular types of LOCC protocols.

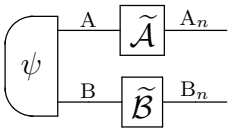
Example 4.2.1. If we consider an LOCC protocol made up only of deterministic tests (i.e. channels), classical communication is not necessary, because each party knows with certainty the output of the various transformations of the other party. In this case, there is no reason for Bob to perform his channels after Alice, because he must not wait for classical communication from Alice. For this reason, we will write Bob's channels under Alice's ones in diagrams.



(4.1)

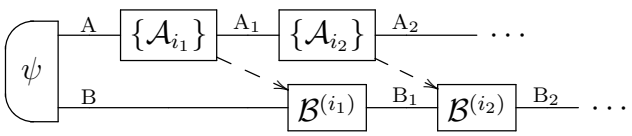
In this diagram, $\mathcal{A}_i$'s (for $i = 1, \ldots, n$) are Alice's channels, whereas $\mathcal{B}_j$'s (for $j = 1, \ldots, m$) are Bob's channels.[2]

Such a protocol with only channels is equivalent to a protocol in which Alice and Bob apply one channel each. It is enough to take the sequential composition of the channels of the two parties.



In this diagram, $\widetilde{\mathcal{A}} = \mathcal{A}_n\mathcal{A}_{n-1}\ldots\mathcal{A}_1$ and $\widetilde{\mathcal{B}} = \mathcal{B}_n\mathcal{B}_{n-1}\ldots\mathcal{B}_1$, where $\mathcal{A}_i$'s and $\mathcal{B}_j$'s are the channels of (4.1). Clearly, $\widetilde{\mathcal{A}} \in \mathsf{Transf}(\mathrm{A}, \mathrm{A}_n)$ and $\widetilde{\mathcal{B}} \in \mathsf{Transf}(\mathrm{B}, \mathrm{B}_n)$, where A_n and B_n are the output systems of the last channels.

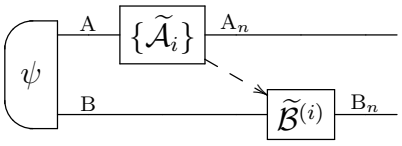
Example 4.2.2. We can consider LOCC protocols in which Alice can use classical communication with Bob, but Bob is not allowed to communicate with Alice. In this way, Bob can use only channels on his system.



[2]Actually, without loss of generality we can assume $n = m$. Indeed, consider the case when $n < m$. We can always add $m - n$ identity channels to Alice's channels, such that the sets of Alice's and Bob's channels have the same cardinality.

In this case, since there is no classical communication from Bob to Alice, Alice can perform her tests at the same time as Bob's channels.

Once more, we can consider the sequential composition of Alice's tests and Bob's channels. In this way, we end up with a protocol in which Alice performs only one test, communicates her outcome to Bob, and Bob applies a channel according to the outcome.



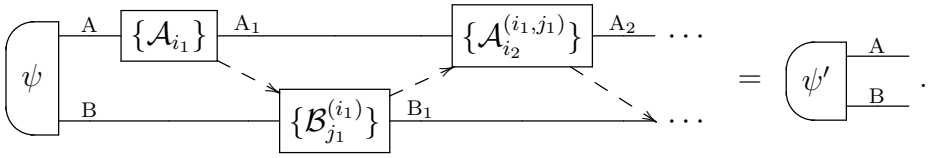
Here $\left\{\widetilde{\mathcal{A}}_i\right\}$ is the sequential composition of all Alice's tests, whereas $\widetilde{\mathcal{B}}^{(i)}$ is the sequential composition of all Bob's channels and it depends on Alice's outcome i.

This type of protocols, in which only one round of classical communication is allowed, are called *1-way LOCC protocols*.

Thanks to LOCC protocols, we can give the following definition.

Definition 4.2.3. Let $|\psi\rangle_{\mathrm{AB}}$ and $|\psi'\rangle_{\mathrm{AB}}$ be two pure bipartite states. We say that $|\psi\rangle_{\mathrm{AB}}$ is *more entangled* than $|\psi'\rangle_{\mathrm{AB}}$ if $|\psi\rangle_{\mathrm{AB}}$ can be transformed into $|\psi'\rangle_{\mathrm{AB}}$ by an LOCC protocol.

Note that this is a fully operational definition, without any references to the mathematical structure of the theory. In this way, this definition is as valid in quantum mechanics as in other conceivable theory that admits entangled states. Using diagrams, we can express the fact that $|\psi\rangle_{\mathrm{AB}}$ is more entangled than $|\psi'\rangle_{\mathrm{AB}}$ as



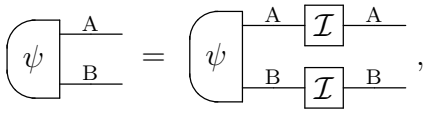
Note that when transforming one entangled state into another, the last output systems must be A and B, in order to have an equality in the above diagram.

4.2.1 Mathematical properties

Intuitively, the relation "to be more entangled than" (entanglement relation for short) resembles an order relation. Let us check if it satisfies the properties of an order.

Reflexive property $|\psi)_{\mathrm{AB}}$ is more entangled than $|\psi)_{\mathrm{AB}}$.

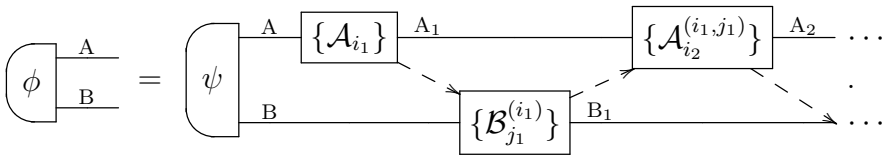
Indeed,



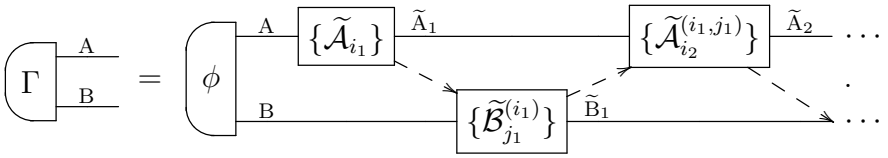
and this is an LOCC protocol in which the two parties apply the identity channel.

Transitive property If $|\psi)_{\mathrm{AB}}$ is more entangled than $|\phi)_{\mathrm{AB}}$ and $|\phi)_{\mathrm{AB}}$ is more entangled than $|\Gamma)_{\mathrm{AB}}$, then $|\psi)_{\mathrm{AB}}$ is more entangled than $|\Gamma)_{\mathrm{AB}}$.

Indeed, if $|\psi)_{\mathrm{AB}}$ is more entangled than $|\phi)_{\mathrm{AB}}$, then



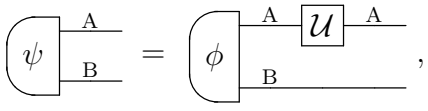
Besides, if $|\phi)_{\mathrm{AB}}$ is more entangled than $|\Gamma)_{\mathrm{AB}}$, then



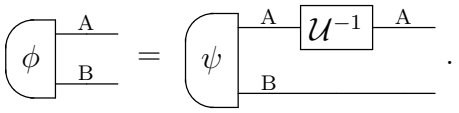
We can compose the two LOCC protocols, and the resulting protocol is still an LOCC protocol from $|\psi)_{\mathrm{AB}}$ to $|\Gamma)_{\mathrm{AB}}$.

Antisymmetric property This property fails, namely, if $|\psi)_{\mathrm{AB}}$ is more entangled than $|\phi)_{\mathrm{AB}}$ and $|\phi)_{\mathrm{AB}}$ is more entangled than $|\psi)_{\mathrm{AB}}$, we cannot conclude that $|\psi)_{\mathrm{AB}} = |\phi)_{\mathrm{AB}}$. Now we show a counter-example.

Consider $|\psi)_{\mathrm{AB}} \neq |\phi)_{\mathrm{AB}}$, such that



where $\mathcal{U}$ is a reversible channel. We can regard the right-hand side as an LOCC protocol with only one reversible channel $\mathcal{U}$ on A. This shows that $|\phi)_{\mathrm{AB}}$ is more entangled than $|\psi)_{\mathrm{AB}}$. However, on the other hand,



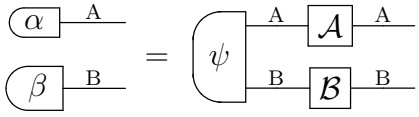
At the right-hand side we have another LOCC protocol, with reversible channel $\mathcal{U}^{-1}$. This shows that $|\psi)_{\mathrm{AB}}$ is more entangled than $|\phi)_{\mathrm{AB}}$. In this counterexample, we have that $|\psi)_{\mathrm{AB}}$ is more entangled than $|\phi)_{\mathrm{AB}}$ and vice versa, but, by hypothesis, $|\psi)_{\mathrm{AB}} \neq |\phi)_{\mathrm{AB}}$. This fact shows that antisymmetric property generally fails.

We have just shown that the entanglement relation is not an order, since it does not fulfil antisymmetric property. Such a binary relation is called a *preorder* [57].

Definition 4.2.4. A binary relation $\lesssim$ is called a *preorder* if it is reflexive and transitive.

Clearly, orders and equivalence relations are both preorders. Let us show now that our relation between bipartite pure states is *not* an equivalence relation, by showing that $|\psi)_{\mathrm{AB}}$ is more entangled than the product state $|\alpha)_{\mathrm{A}}|\beta)_{\mathrm{B}}$, but clearly the converse does not hold.

Example 4.2.5. We may consider



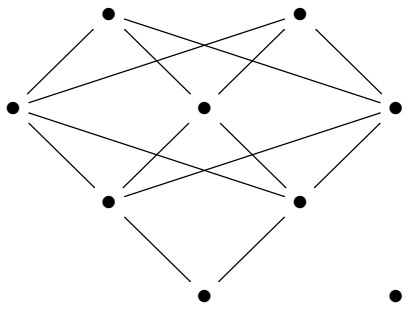
where $\mathcal{A}$ and $\mathcal{B}$ are channels that prepare $|\alpha)_{\mathrm{A}}$ and $|\beta)_{\mathrm{B}}$ respectively, namely $\mathcal{A} = |\alpha)_{\mathrm{A}}(e|_{\mathrm{A}}$ and $\mathcal{B} = |\beta)_{\mathrm{B}}(e|_{\mathrm{B}}$. The right-hand side is an LOCC protocol, then $|\psi)_{\mathrm{AB}}$ is more entangled than $|\alpha)_{\mathrm{A}}|\beta)_{\mathrm{B}}$, but the converse is not true, because we cannot eliminate parallel composition with local operations and classical communication.

This fact tells us that our relation is not an equivalence relation. According to proposition A.1.3, each genuine preorder (i.e. such that is not an equivalence relation) can be turned into an order on a quotient set. In our case, we consider the equivalence relation "to be as entangled as", which is the equivalence relation associated with our preorder.

Definition 4.2.6. We say that $|\psi)_{AB}$ is *as entangled as* $|\psi')_{AB}$ if $|\psi)_{AB}$ is more entangled than $|\psi')_{AB}$ and $|\psi')_{AB}$ is more entangled than $|\psi)_{AB}$.

It would be interesting to characterize this equivalence relation better. We have seen that if $|\psi)_{AB}$ differs from $|\psi')_{AB}$ by a local reversible channel, then $|\psi)_{AB}$ is as entangled as $|\psi')_{AB}$. In quantum mechanics it is true also the converse, namely that if $|\psi)_{AB}$ is as entangled as $|\psi')_{AB}$, it must differ from $|\psi')_{AB}$ by a local unitary channel. But what can we say about a general probabilistic theory? We will give an answer in chapter 7.

Summing up, all we did in this section was to prove that it is possible to quantify entanglement of a bipartite pure state in a very primitive sense, that is we can order bipartite pure states according to their entanglement and to establish a sort of hierarchy among them [58], as it is shown in the following graphical example.



In this graph, each node represents a bipartite pure state. From the top to the bottom, we go from the most entangled states to the least entangled ones. Each row represents an equivalence class. We draw a solid line between nodes that are connected by the order relation[3] $\leq$ and there is no other node between them.
In the bottom row, there is an isolated node. It is not in the same equivalence class as the other node in the same row, otherwise it would be connected to the nodes of the row above. This means that we cannot compare its entanglement to any other node: the order $\leq$ is in general not total.

[3]Here the attentive reader should have noted that we did a little abuse of notation, because the order is defined between rows (equivalence classes), not between points. The notation used here was aimed only at stressing that if two points are connected by a line, then we can say that a point is "strictly" more entangled than the other.

4.3 The relation "to be more mixed than"

In this section we introduce a relation on the set of states, that orders them according to their mixedness, more or less in the same way we ordered bipartite pure states according to their entanglement. The rest of this chapter will be devoted to proving the equivalence of these two relations.

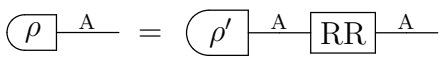
We can give the following definition.

Definition 4.3.1. Let ρ and ρ' be two states of system A. We say that ρ is *more mixed* than ρ' if ρ can be written as

$$\rho = \sum_i p_i \mathcal{U}_i \rho',$$

where p_i's are probabilities ($p_i \geq 0$ for every i, and $\sum_i p_i = 1$), and $\mathcal{U}_i$ are reversible channels on system A for every i.

We will call $\sum_i p_i \mathcal{U}_i$ *random reversible channel* (RR). Using diagrams, if ρ is more mixed than ρ', we write

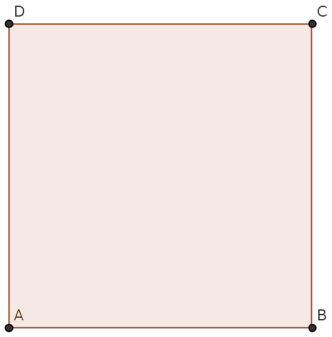


where the RR-box indicates a random reversible channel.

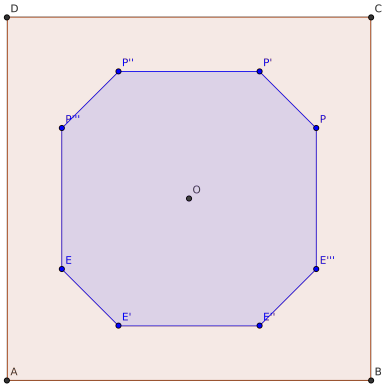
We see that ρ is more mixed than ρ' if the former can be obtained from the latter by a random deterministic (and reversible) evolution. Suppose we start with ρ', but we ignore its evolution. We only know that it evolves with a reversible channel randomly chosen from the ensemble $\{p_i \mathcal{U}_i\}$, where ρ' evolves with $\mathcal{U}_i$ with probability p_i. Then the state after this unknown evolution is precisely $\rho = \sum_i p_i \mathcal{U}_i \rho'$.

We can have a pictorial insight about the relation "to be more mixed than" (mixedness relation for short), as shown in the following example.

Example 4.3.2. Let us consider a theory with only 4 pure states, which are the vertexes of a square. Then the set of states of this theory is the convex hull of these 4 points, that is a square. All the points in the square, except the 4 vertexes, are mixed states.



The group of reversible channels maps pure states into pure states, this means that it is the dihedral group D_4, which has 8 elements. Let us pick up a generic (mixed) state P of this theory: it is a point of the square. We want to identify the states that are more mixed than P. According to the definition, we apply all the reversible channels $\{\mathcal{U}_i\}_{i=1}^{8}$ to P and then we take the convex hull of these 8 points (in blue).



Therefore, the blue octagon is the set of states that are more mixed than P. Note that pure states are not in this set; this seems quite obvious, but we will prove it in the next subsection (see example 4.3.3). Note that the centre O of the square is invariant under D_4, therefore the convex hull generated by the action on O of the group of reversible channels is made only of O itself.

This means that there are no states more mixed than it. It is a maximal element.

In general, the invariant state χ is always a maximal element, therefore we say that it is a *maximally mixed state*. Indeed, suppose that a state ρ is more mixed than χ.

$$\rho = \sum_i p_i \mathcal{U}_i(\chi) = \sum_i p_i \chi = \chi$$

This proves that χ is a maximal element and that it is the only element in its equivalence class, because there are no other states more mixed than χ.

4.3.1 Mathematical properties

Similarly to the case of entanglement, the mixedness relation resembles an order relation. Let us check its properties.

Reflexive property ρ is more mixed than ρ. Indeed, we have $\rho = \mathcal{I}(\rho)$, and $\mathcal{I}$ is a particular random reversible channel, made up only of the identity channel.

Transitive property If ρ is more mixed than ρ' and ρ' is more mixed than ρ'', then ρ is more mixed than ρ''.

Indeed, if ρ is more mixed than ρ', then $\rho = \sum_i p_i \mathcal{U}_i \rho'$; if ρ' is more mixed than ρ'', then $\rho' = \sum_j \widetilde{p}_j \widetilde{\mathcal{U}}_j \rho''$. Combining this two statements, one has

$$\rho = \sum_i p_i \mathcal{U}_i \left(\sum_j \widetilde{p}_j \widetilde{\mathcal{U}}_j \rho'' \right) = \sum_{i,j} p_i \widetilde{p}_j \mathcal{U}_i \widetilde{\mathcal{U}}_j \rho'' = \sum_{i,j} q_{ij} \mathcal{V}_{ij} \rho''.$$

Here we set $\mathcal{V}_{ij} := \mathcal{U}_i \widetilde{\mathcal{U}}_j$, which is clearly a reversible channel for every i and j; and $q_{ij} := p_i \widetilde{p}_j$. From the properties of p_i and $\widetilde{p}_j$, it follows that q_{ij}'s are probabilities too. Hence, we showed that ρ is more mixed than ρ''.

Antisymmetric property Even in this case, antisymmetry fails. The counter-example is constructed in a very close way to the one in subsection 4.2.1.

Consider $\rho \neq \rho'$, such that $\rho = \mathcal{U}(\rho')$. This is a particular random reversible channel, so ρ is more mixed than ρ'. On the other hand, $\rho' = \mathcal{U}^{-1}(\rho)$, whence ρ' is more mixed than ρ, but, by hypothesis, $\rho \neq \rho'$.

Note that if we take the deterministic effect on system B in the counter-example in subsection 4.2.1, we get exactly this counter-example for the marginal on system A.

Even in this case, the mixedness relation is a preorder. However, it is not an equivalence relation as the following example shows.

Example 4.3.3. Let us show that every state ρ is more mixed than a pure state ψ. By definition, ρ can be expressed as a convex combination of pure states.

$$\rho = \sum_i p_i \varphi_i$$

As a consequence of purification postulate, for every φ_i there exists a reversible channel $\mathcal{U}_i$ such that $\mathcal{U}_i(\psi) = \varphi_i$. Then,

$$\rho = \sum_i p_i \mathcal{U}_i(\psi),$$

whence ρ is more mixed than ψ. The converse is not true, because a random reversible channel is a coarse-graining and a pure state cannot be the coarse-graining of a mixed state.

Thus, we can define an equivalence relation and then an order on the quotient set, as we have done with entangled states.

Definition 4.3.4. We say that ρ is *as mixed as* ρ' if ρ is more mixed than ρ' and ρ' is more mixed than ρ.

Again, it would be interesting to characterize this equivalence relation better. We have seen that if ρ differs from ρ by a reversible channel, then ρ is as mixed as ρ'. Once more, in quantum mechanics it is true also the converse. But what can we say about a general theory? We will give an answer to this question in chapter 7.

Note that a graphical representation of the order between mixed states is exactly the same as the one on page 98.

4.4 Equivalence between entanglement and mixedness

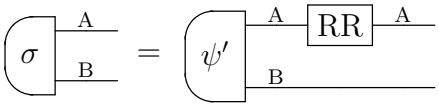
In this section we prove that the entanglement and the mixedness relations are equivalent in a general probabilistic causal theory with the purification postulate, with a further assumption. This result is already known in quantum mechanics (see [17]). In this way, entanglement and mixedness show up as two sides of the same physical phenomenon even in a general probabilistic theory, and the order between entangled states is equivalent to the order between mixed states. This will enable us to choose the more apt relation according to our specific needs.

4.4.1 More mixed implies more entangled

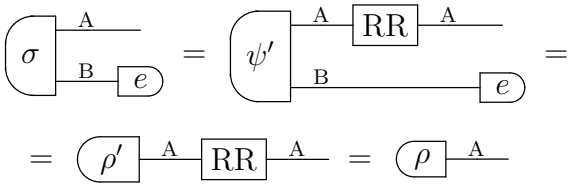
One implication in the equivalence requires less effort to be proven.

Lemma 4.4.1. *Let* $|\psi)_{\mathrm{AB}}$ *and* $|\psi')_{\mathrm{AB}}$ *be two bipartite pure states, and let* $|\rho)_{\mathrm{A}}$ *and* $|\rho')_{\mathrm{A}}$ *be their marginals on system* A*. If* $|\rho)_{\mathrm{A}}$ *is more mixed than* $|\rho')_{\mathrm{A}}$*, then* $|\psi)_{\mathrm{AB}}$ *is more entangled than* $|\psi')_{\mathrm{AB}}$*.*

Proof. By hypothesis, $|\rho)_{\mathrm{A}} = \sum_i p_i \mathcal{U}_{i,\mathrm{A}} |\rho')_{\mathrm{A}}$, for some reversible channels $\mathcal{U}_i$'s acting on A. Let us define $|\sigma)_{\mathrm{AB}} = \sum_i p_i \mathcal{U}_i |\psi')_{\mathrm{AB}}$.

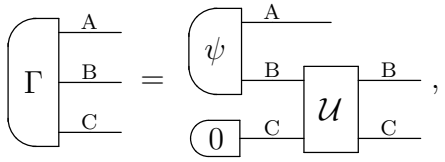


Clearly, $|\sigma)_{\mathrm{AB}}$ is mixed, because it is given by a coarse-graining. By hypothesis, the marginal of $|\psi')_{\mathrm{AB}}$ on A is $|\rho')_{\mathrm{A}}$, hence $|\sigma)_{\mathrm{AB}}$ is an extension of $|\rho)_{\mathrm{A}}$, as the following diagrams show.

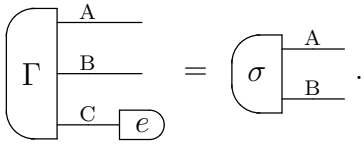


Let us consider a purification $|\Gamma)_{\mathrm{ABC}}$ of $|\sigma)_{\mathrm{AB}}$; $|\Gamma)_{\mathrm{ABC}}$ is clearly a purification of $|\rho)_{\mathrm{A}}$ too, with purifying system BC, because $|\sigma)_{\mathrm{AB}}$ is an extension of $|\rho)_{\mathrm{A}}$. Now, consider $|\psi)_{\mathrm{AB}} |0)_{\mathrm{C}}$, where $|0)_{\mathrm{C}}$ is some pure state of system C.

$|\psi)_{\mathrm{AB}}\,|0)_{\mathrm{C}}$ is a purification of $|\rho)_{\mathrm{A}}$, with purifying system BC. By uniqueness of purification (up to reversible channels on the purifying system), there exists a reversible channel $\mathcal{U}$, acting on BC, such that $|\Gamma)_{\mathrm{ABC}} = \mathcal{U}_{\mathrm{BC}}\,|\psi)_{\mathrm{AB}}\,|0)_{\mathrm{C}}$.

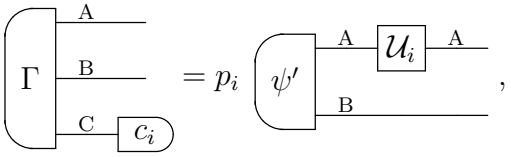


where

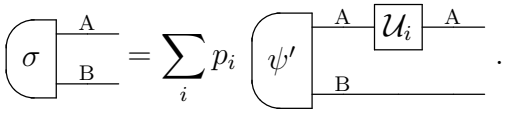


In this way, Alice and Bob are able to transform $|\psi)_{\mathrm{AB}}$ into $|\Gamma)_{\mathrm{ABC}}$ only via local operations. Indeed, it is enough that Bob adds a system C and applies the suitable reversible channel $\mathcal{U}$ on BC.

By the steering property, there exists an observation-test $\{c_i\}$ on C, such that for every i one has



because



Hence, if Bob performs an observation-test on C and gets outcome i, he prepares $p_i \mathcal{U}_{i,\mathrm{A}}\,|\psi')_{\mathrm{AB}}$. Now, he calls Alice to communicate her his outcome. Alice applies $\mathcal{U}_{i,\mathrm{A}}^{-1}$ on A and she obtains $p_i\,|\psi')_{\mathrm{AB}}$. Summing over i, that is doing a coarse-graining, Alice is able to get $|\psi')_{\mathrm{AB}}$.

In this way, we were able to transform $|\psi)_{\mathrm{AB}}$ into $|\psi')_{\mathrm{AB}}$ by means of an LOCC protocol. Therefore, $|\psi)_{\mathrm{AB}}$ is more entangled than $|\psi')_{\mathrm{AB}}$. □

The expression $p_i\,|\psi')_{\mathrm{AB}}$ in the proof means that $|\psi')_{\mathrm{AB}}$ has been generated with probability p_i. There are several possible events, but in all of them $|\psi')_{\mathrm{AB}}$ is always prepared. If we sum over i, one gets the "average" state produced by the protocol, but $\sum_i p_i\,|\psi')_{\mathrm{AB}} = |\psi')_{\mathrm{AB}}$. Notice that the

coarse-graining is not on repeated tests, but it is the average over possible outcomes.

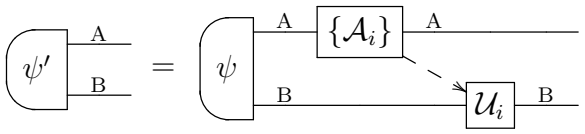
Remark 4.4.2. In the statement of the lemma, we stressed the fact ρ is more mixed than ρ' on system A. So far, there is no guarantee that even $|\rho)_{\mathrm{B}}$ is more mixed than $|\rho')_{\mathrm{B}}$.

4.4.2 More entangled implies more mixed

It is much more challenging to prove the converse implication, namely that if $|\psi)_{\mathrm{AB}}$ is more entangled than $|\psi')_{\mathrm{AB}}$, then $|\rho)_{\mathrm{A}}$ is more mixed than $|\rho)_{\mathrm{A}}$ in the framework of a general probabilistic theory.

It is not so hard to prove it if we consider a special case: when $|\psi)_{\mathrm{AB}}$ can be transformed into $|\psi')_{\mathrm{AB}}$ by means of a 1-way LOCC protocol (see example 4.2.2). We note that in the proof of lemma 4.4.1 the resulting LOCC protocol was in fact a 1-way LOCC protocol. In this way, 1-way LOCC protocols seem to play an important role as far as entanglement is concerned.

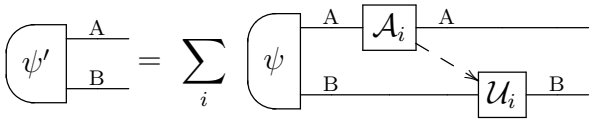
Lemma 4.4.3. *Let $|\psi)_{\mathrm{AB}}$ and $|\psi')_{\mathrm{AB}}$ be bipartite pure states. Suppose that $|\psi)_{\mathrm{AB}}$ is more entangled than $|\psi')_{\mathrm{AB}}$ and that there exists a 1-way LOCC protocol with classical communication from* A *to* B *with a reversible channel $\mathcal{U}$ on* B*, such that it transforms $|\psi)_{\mathrm{AB}}$ into $|\psi')_{\mathrm{AB}}$, namely*[4]



(4.2)

Then $|\rho)_{\mathrm{B}}$ is more mixed than $|\rho')_{\mathrm{B}}$, where $|\rho)_{\mathrm{B}}$ and $|\rho')_{\mathrm{B}}$ are the marginals on system B *of $|\psi)_{\mathrm{AB}}$ and $|\psi')_{\mathrm{AB}}$ respectively.*

Proof. Note that we can rewrite (4.2) as

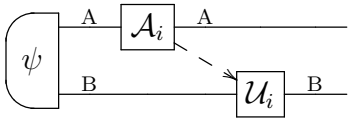


(4.3)

because we are summing over all possible events in the test $\{\mathcal{A}_i\}$.

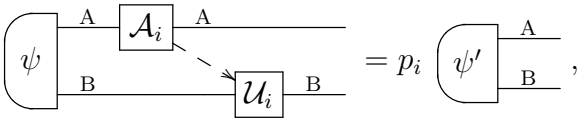
[4]Here, for the sake of simplicity, instead of writing $\mathcal{U}^{(i)}$, as in section 4.2, we write simply $\mathcal{U}_i$.

Now, (4.3) is a coarse-graining of

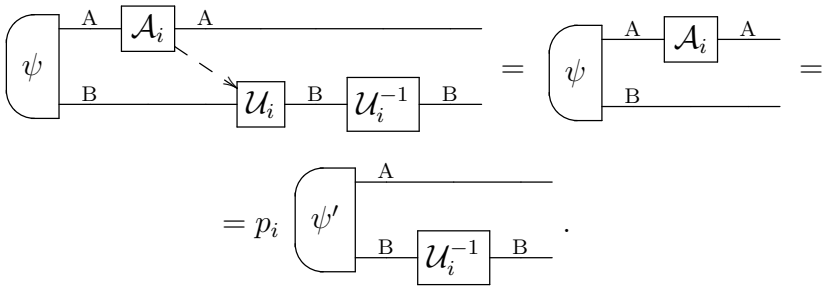


. (4.4)

Then, (4.4) is a refinement of $|\psi'\rangle_{\mathrm{AB}}$, and since $|\psi'\rangle_{\mathrm{AB}}$ is pure, such refinement is trivial.

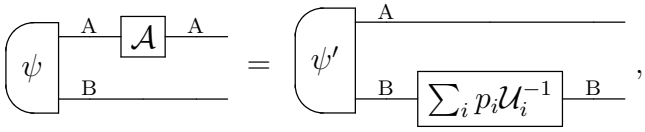


where $p_i \in (0, 1]$, and $\sum_i p_i = 1$. If we now apply $\mathcal{U}_i^{-1}$ to B, we have



.

Now we sum over i



,

where $\mathcal{A}$ is the channel associated with the coarse-graining of the test $\{\mathcal{A}_i\}$. Clearly $\sum_i p_i \mathcal{U}_i^{-1}$ is a random reversible channel. Finally, let us apply the deterministic effect on A, getting $|\rho\rangle_{\mathrm{B}} = \sum_i p_i \mathcal{U}_i^{-1} |\rho'\rangle_{\mathrm{B}}$. Hence $|\rho\rangle_{\mathrm{B}}$ is more mixed than $|\rho'\rangle_{\mathrm{B}}$. □

Note that if the 1-way LOCC protocol goes from A to B, we can show that $|\rho\rangle_{\mathrm{B}}$ is more mixed than $|\rho'\rangle_{\mathrm{B}}$ if $|\psi\rangle_{\mathrm{AB}}$ is more entangled than $|\psi'\rangle_{\mathrm{AB}}$. Thus, the direction in which classical communication goes is essential in establishing which marginal is more mixed than the other. If the target of the communication is Bob, than we can say something about mixedness of his state; if the target is Alice, then we can evaluate mixedness of her marginal.

Even if there exists a 1-way protocol from A to B, there is no guarantee that there exists a 1-way protocol in the other direction, namely from B to

A. In other words, even if $|\rho)_{\mathrm{B}}$ is more mixed than $|\rho')_{\mathrm{B}}$, we cannot say that $|\rho)_{\mathrm{A}}$ is more mixed than $|\rho')_{\mathrm{A}}$. A priori we might have a rather paradoxical situation in which $|\psi)_{\mathrm{AB}}$ is more entangled than $|\psi')_{\mathrm{AB}}$, but, for example, in A we have that $|\rho')_{\mathrm{A}}$ is more mixed than $|\rho)_{\mathrm{A}}$, whereas in B $|\rho)_{\mathrm{B}}$ is more mixed than $|\rho')_{\mathrm{B}}$.

4.5 Lo-Popescu theorem

We have just seen that we can prove the equivalence between entanglement and mixedness if we restrict ourselves to 1-way LOCC protocols with a reversible channel. It would be fine if we could prove this equivalence with full generality, namely also for generic LOCC protocols.

Fortunately, in quantum theory there is a theorem that guarantees that every LOCC protocol can be reduced to a 1-way LOCC protocol with a unitary channel [17, 63].

Theorem 4.5.1 (Lo-Popescu). *If $|\psi\rangle_{\mathrm{AB}}$ can be transformed into $|\phi\rangle_{\mathrm{AB}}$ by an LOCC protocol, then it can be transformed into $|\phi\rangle_{\mathrm{AB}}$ 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. For the sake of simplicity, we will analyse only the case when all quantum operations are pure, as the ones presented in the provisional definition 1.4.1, where there is only one Kraus operator. We will defer the general case until the treatment of Lo-Popescu theorem for general probabilistic theories.

Let us start from the bipartite pure state $|\psi\rangle_{\mathrm{AB}}$ and let us consider its Schmidt decomposition[5] $|\psi\rangle_{\mathrm{AB}} = \sum_j \sqrt{\lambda_j}\, |j\rangle_{\mathrm{A}} |j\rangle_{\mathrm{B}}$. Suppose Bob applies a quantum instrument with Kraus operators $\{M_j\}$, which can be expressed in his Schmidt basis[6] as

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

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

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

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

$$|\psi_j\rangle_{\mathrm{AB}} = \frac{1}{p_j}\left(\mathbf{1}_{\mathrm{A}} \otimes M_j\right)|\psi\rangle_{\mathrm{AB}} = \frac{1}{p_j}\sum_{k,l}\sqrt{\lambda_l}M_{j,kl}\,|l\rangle_{\mathrm{A}}\,|k\rangle_{\mathrm{B}}\,, \tag{4.5}$$

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

$$p_j = \langle\psi|_{\mathrm{AB}}\,M_j^\dagger M_j\,|\psi\rangle_{\mathrm{AB}} = \sum_{k,l}\lambda_l\,|M_{j,kl}|^2\,. \tag{4.6}$$

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

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

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

$$|\varphi_j\rangle_{\mathrm{AB}} = \frac{1}{p_j}\left(N_j \otimes \mathbf{1}_{\mathrm{B}}\right)|\psi\rangle_{\mathrm{AB}} = \frac{1}{p_j}\sum_{k,l}\sqrt{\lambda_l}M_{j,kl}\,|k\rangle_{\mathrm{A}}\,|l\rangle_{\mathrm{B}}\,, \tag{4.7}$$

where p_j is still given by eq. (4.6). Comparing eq. (4.5) with eq. (4.7), we see that $|\psi_j\rangle_{\mathrm{AB}}$ and $|\varphi_j\rangle_{\mathrm{AB}}$ are the same state, up to exchanging the role of system A and system B. Therefore they have the same Schmidt coefficients. By lemma 1.1.4, they differ by a tensor product of unitary operators, namely $|\psi_j\rangle_{\mathrm{AB}} = U_{j,\mathrm{A}} \otimes V_{j,\mathrm{B}}\,|\varphi_j\rangle_{\mathrm{AB}}$.
Therefore, when Bob applies a quantum instrument with Kraus operators $\{M_j\}$, this is equivalent to the situation in which 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.

If the original LOCC protocol is a multi-round one, whenever Bob performs a measurement and communicates the 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. □

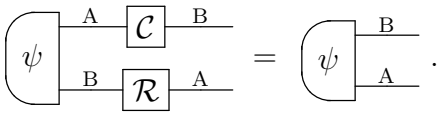
Clearly, this proof cannot be fitted easily into an abstract operational framework, because, at first glance, it seems to heavily rely on the mathematical structure of quantum theory. However, we can see that the core of the proof is the fact that we can exchange the role of system A and system B. This is a purely operational fact, that can be implemented even without any references to Hilbert spaces. However, we must note that this exchange of systems is allowed by Schmidt decomposition, therefore it holds only when system AB is in a pure state. Indeed, when we write a Schmidt decomposition of a bipartite pure state, say $|\psi\rangle_{\mathrm{AB}} = \sum_j \sqrt{p_j}\,|j\rangle_{\mathrm{A}}\,|j'\rangle_{\mathrm{B}}$, p_j's are the eigenvalues of the two marginals ρ_{A} and ρ_{B} and $|j\rangle_{\mathrm{A}}$ and $|j'\rangle_{\mathrm{B}}$ are the eigenvectors of ρ_{A} and ρ_{B} respectively (cf. subsection 1.1.1). Therefore, we see that we can associate every eigenvector of ρ_{A} with a corresponding eigenvector of ρ_{B}.

This is not true when system AB is in a mixed state, because Schmidt decomposition does not exist for mixed states.

Now, it seems reasonable to assume that such an exchange of subsystems for bipartite pure states is possible even in an abstract scenario. We might be tempted to state that in system AB, system A and system B are operationally equivalent (see definition 2.1.6), which is the most naive way to implement a sort of equivalence between A and B. However, this is not true even in quantum mechanics: when $\mathcal{H}_{\mathrm{A}}$ and $\mathcal{H}_{\mathrm{B}}$ have different dimensions, it is not possible to find a unitary operator from $\mathcal{H}_{\mathrm{A}}$ to $\mathcal{H}_{\mathrm{B}}$.

Therefore, we make a weaker assumption, as stated in the following axiom.

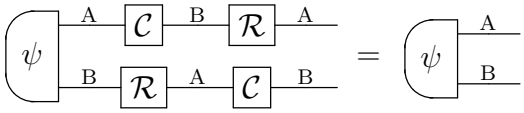
Axiom 4.5.2. *For any bipartite pure state* $|\psi)_{\mathrm{AB}}$*, there exist two channels* $\mathcal{C} \in \mathsf{Transf}\,(\mathrm{A},\mathrm{B})$ *and* $\mathcal{R} \in \mathsf{Transf}\,(\mathrm{B},\mathrm{A})$ *such that*



First of all, we must note that these two channels, in general, depend on the (pure) state of the system, namely if the state changes, the channels change too. The only important point is that such two channels exist, it is not even necessary that $\mathcal{C}$ and $\mathcal{R}$ are reversible.

These channels simply implement an exchange of the roles of A and B. This exchange is achieved by means of a local protocol, so it does not involve any "quantum" communication between the two parties.

A first trivial consequence of the axiom is that if we apply the channels twice, we undo the exchange of the two subsystems.

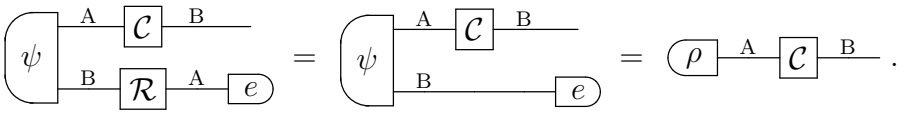


As one could expect, these two channels transform one marginal state of a bipartite pure state into the other one.

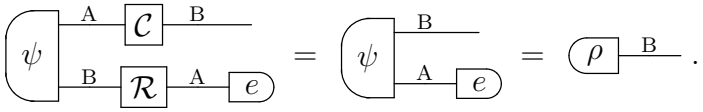
Proposition 4.5.3. *Let* $|\psi)_{\mathrm{AB}}$ *be a bipartite pure state and* $\mathcal{C}$ *and* $\mathcal{R}$ *its associated channels. If* $|\rho)_{\mathrm{A}}$ *and* $|\rho)_{\mathrm{B}}$ *are the marginal states on systems* A *and* B *respectively, then* $\mathcal{C}\,|\rho)_{\mathrm{A}} = |\rho)_{\mathrm{B}}$ *and* $\mathcal{R}\,|\rho)_{\mathrm{B}} = |\rho)_{\mathrm{A}}$.

Proof. Let us prove the statement for marginal $|\rho)_{\mathrm{A}}$, the other one can be proven in an analogous way.

Let us apply $\mathcal{C}$ and $\mathcal{R}$ to the pure state $|\psi)_{\mathrm{AB}}$, and let us take the deterministic effect on system A. Since $\mathcal{R}$ is a channel, we have



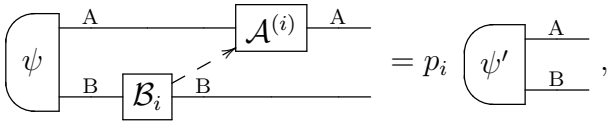
On the other hand, if we recall the action of $\mathcal{C}$ and $\mathcal{R}$ on $|\psi)_{\mathrm{AB}}$, we can write



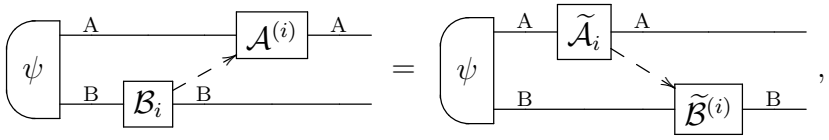
This proves that $\mathcal{C}\,|\rho)_{\mathrm{A}} = |\rho)_{\mathrm{B}}$. □

Now, let us move to the core of this section: an abstract version of Lo-Popescu theorem. Essentially, the key idea in this respect is presented in the following lemma.

Lemma 4.5.4. *Consider a bipartite pure state* $|\psi)_{\mathrm{AB}}$ *as input of a 1-way LOCC protocol with classical communication from* B *to* A*, such as*

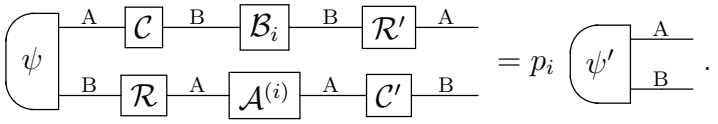


where $\mathcal{B}_i$ is a transformation in a test on B, *$\mathcal{A}^{(i)}$ is a channel on* A *which depends on the outcome of the test on* B, *and $p_i \in (0,1]$. Then, this protocol is equivalent to a 1-way LOCC protocol with classical communication from* A *to* B. *Using diagrams,*

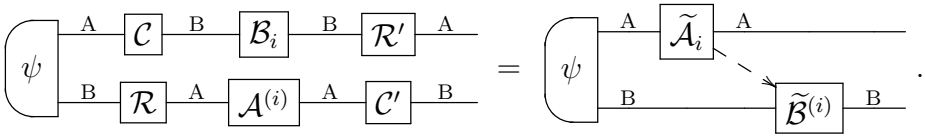


where $\widetilde{\mathcal{A}}_i$ is now a transformation in a test on A *and $\widetilde{\mathcal{B}}^{(i)}$ is a channel on* B *which depends on the outcome of the test on* A.

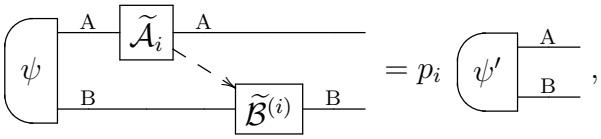
Proof. We must distinguish two cases. The first and easiest case is when $\mathcal{A}^{(i)}$ and $\mathcal{B}_i$ are pure. In this case, by pure conditioning, we know that after $\mathcal{A}^{(i)}$ and $\mathcal{B}_i$ the state will still be pure. Therefore, according to axiom 4.5.2, there exist two channels $\mathcal{R}' \in \mathsf{Transf}\,(\mathrm{B},\mathrm{A})$ and $\mathcal{C}' \in \mathsf{Transf}\,(\mathrm{A},\mathrm{B})$ such that they interchange the two systems. Therefore, we are entitled to build up a 1-way LOCC protocol from A to B in a similar fashion to the proof of theorem 4.5.1. In this vein, we exchange the two systems before applying $\mathcal{A}^{(i)}$and $\mathcal{B}_i$ and then we exchange the two systems again after $\mathcal{A}^{(i)}$and $\mathcal{B}_i$. This will not clearly affect the final output of the protocol, because $\mathcal{A}^{(i)}$ is still applied on A and $\mathcal{B}_i$ is still applied on B. Thus we have



On the other hand, on the left-hand side, we can regard $\{\mathcal{R}'\mathcal{B}_i\mathcal{C}\}$ as a test $\left\{\widetilde{\mathcal{A}}_i\right\}$ on system A, and $\mathcal{C}'\mathcal{A}^{(i)}\mathcal{R}$ as a channel $\widetilde{\mathcal{B}}^{(i)}$ on system B, which will depend on the outcome of $\left\{\widetilde{\mathcal{A}}_i\right\}$. Now classical communication goes from A to B. Using diagrams, we can write

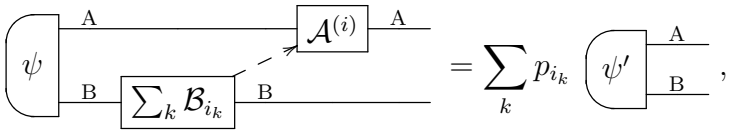


In this way we prove that

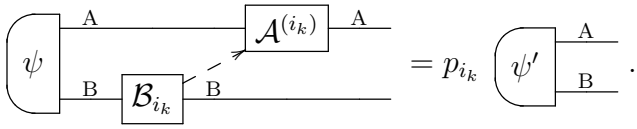


whence this new 1-way LOCC protocol, with classical communication in the opposite direction, has the same output as the original one, thus proving the equivalence between the two protocols.

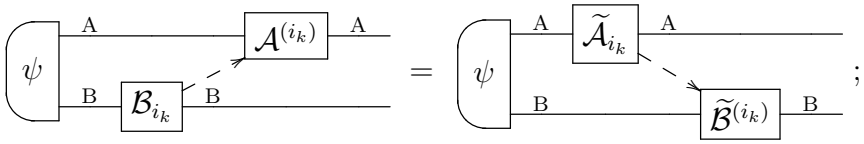
If $\mathcal{A}_i$ and $\mathcal{B}_i$ are not pure, some more passages are necessary. Suppose $\mathcal{B}_i$ is not pure (the cases when $\mathcal{A}^{(i)}$ is not pure or $\mathcal{B}_i$ and $\mathcal{A}^{(i)}$ are both not pure are analogous). This means that $\mathcal{B}_i = \sum_k \mathcal{B}_{i_k}$, where $\mathcal{B}_{i_k}$'s are pure.



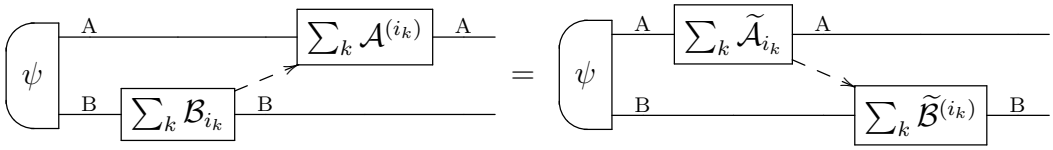
where p_{i_k} is the probability associated with each pure transformation $\mathcal{B}_{i_k}$. We can consider a refinement of the LOCC protocol, namely



We have already proved that in this case the statement of the lemma holds



then we can redo the coarse-graining over $\mathcal{B}_{i_k}$ and in this way we manage to prove the statement of the lemma even in the non-pure case.

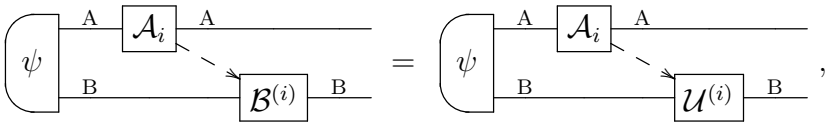


□

So far, we have proved that in every 1-way LOCC protocol, classical communication can be inverted. Note that lemma 4.5.4 admits a really straightforward generalization when we have a generic transformation $\mathcal{A}_j^{(i)}$ on A instead of a channel $\mathcal{A}^{(i)}$. Indeed, the fact that $\mathcal{A}^{(i)}$ was a channel was actually unessential in the proof. This slight generalization will be used in proving the abstract version of Lo-Popescu theorem.

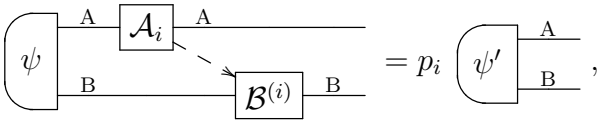
Nevertheless, we still miss a part of Lo-Popescu theorem, namely that we can always use reversible channels. This is stated in the following lemma.

Lemma 4.5.5. *Every 1-way LOCC protocol from* A *to* B *is equivalent to a 1-way LOCC protocol from* A *to* B *with a reversible channel on* B.

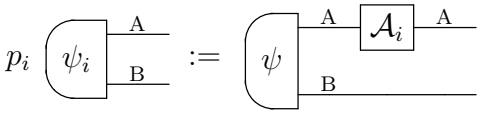


where $\mathcal{A}_i$ *is a transformation on* A, $\mathcal{B}^{(i)}$ *is a channel on* B *and* $\mathcal{U}^{(i)}$ *is a reversible channel on* B.

Proof. Let $|\psi')_{\mathrm{AB}}$ be the output of such a protocol, where $\mathcal{B}^{(i)}$ need not to be reversible.

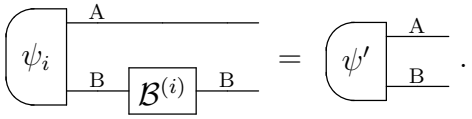


where $p_i \in (0, 1]$. Let us define $|\psi_i)_{\mathrm{AB}}$ as

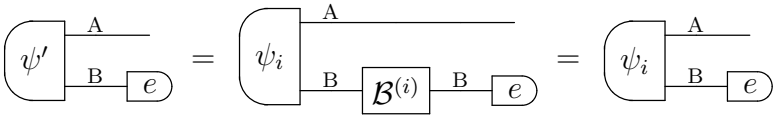


(4.8)

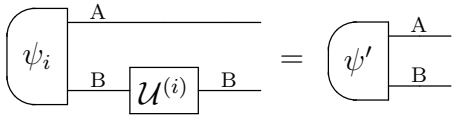
then



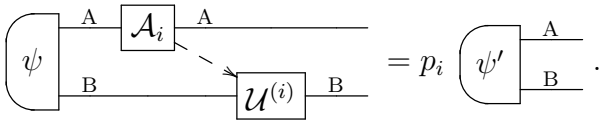
Now let us take the deterministic effect on B. Recalling that $\mathcal{B}^{(i)}$ is a channel, we have



This shows that $|\psi_i)_{\mathrm{AB}}$ and $|\psi')_{\mathrm{AB}}$ have the same marginal on A. Hence, they must differ by a reversible channel $\mathcal{U}^{(i)}$ on the purifying system B.



Multiplying both sides by p_i, and recalling (4.8), we finally get

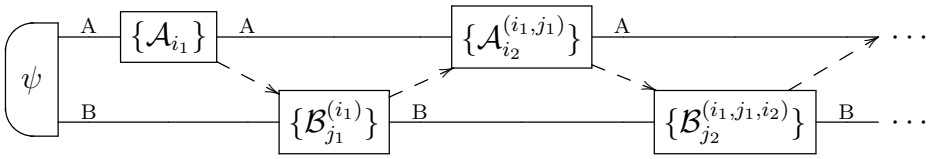


This proves the statement of the lemma. □

Now we are ready to state and prove the abstract version of Lo-Popescu theorem.

Theorem 4.5.6 (Lo-Popescu, abstract version)**.** *If $|\psi)_{\mathrm{AB}}$ can be transformed into $|\phi)_{\mathrm{AB}}$ by an LOCC protocol, then it can be transformed into $|\phi)_{\mathrm{AB}}$ by a 1-way LOCC protocol, where Alice performs a test, she communicates her outcome to Bob, and Bob applies a reversible channel on his system.*

Proof. Let us consider a generic LOCC protocol.[7]

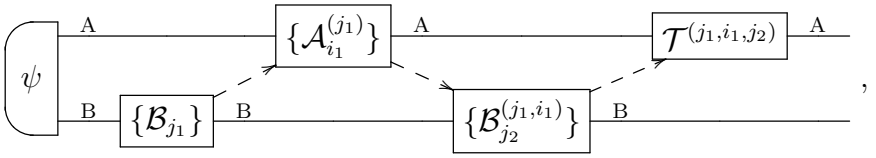


In general, the protocol is not made of all pure transformations. However, we can decompose each non-pure transformation into a sum of pure ones. Taking all the sums out of the protocol, we end up with a pure LOCC protocol, with some sums in front of it. In this way, we can examine the resulting pure LOCC protocol. We can regard this protocol as a sequential composition of several 1-way LOCC protocols, corresponding to the various rounds of classical communication. Let us focus on each of these 1-way LOCC protocols. We proceed in this way:

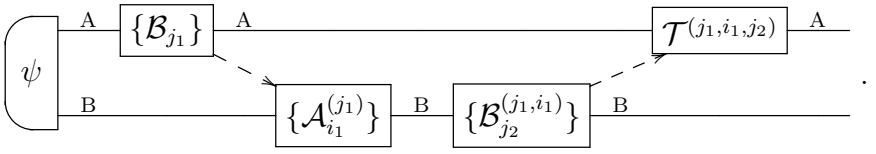
- if classical communication goes from Alice to Bob and Bob applies a channel, we do nothing;
- whenever we encounter a test on Bob's system, followed by a channel on Alice's system, we swap them, according to lemma 4.5.4, and now classical communication goes from Alice to Bob;

[7]Here, for the sake of simplicity, we suppose all the tests are on A or on B.

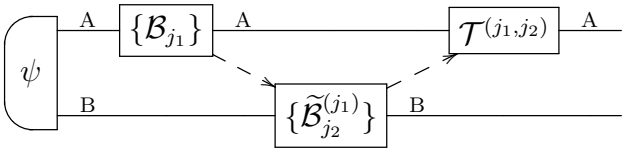
- if a test on Bob's system is followed by a non-deterministic test on Alice's system, this is a bit tricky situation, so it is better to see a simple example to understand what we do in this case. Suppose we have the LOCC protocol



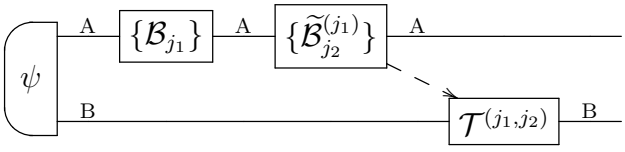
where $\mathcal{T}^{(j_1,i_1,j_2)}$ is some operation on Alice's system, that may be a channel or a non-deterministic test. We can swap $\{\mathcal{B}_{j_1}\}$ and $\left\{\mathcal{A}_{i_1}^{(j_1)}\right\}$, according to the generalization of lemma 4.5.4. Now the protocol becomes[8]



Let us take the sequential composition of tests $\left\{\mathcal{A}_{i_1}^{(j_1)}\right\}$ and $\left\{\mathcal{B}_{j_2}^{(j_1,i_1)}\right\}$ on Bob's system, this is a test $\left\{\widetilde{\mathcal{B}}_{j_2}^{(j_1)}\right\}$.

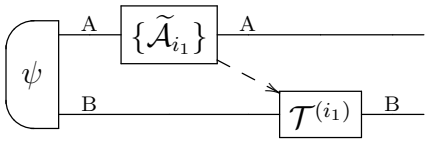


Then, we swap $\left\{\widetilde{\mathcal{B}}_{j_2}^{(j_1)}\right\}$ and $\mathcal{T}^{(j_1,j_2)}$.

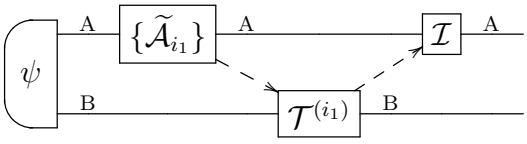


[8]Here we do slight abuses of notation, and for instance we call $\{\mathcal{B}_{j_1}\}$ the swapped test, even if it is on A, just to make it clearer that we swapped it.

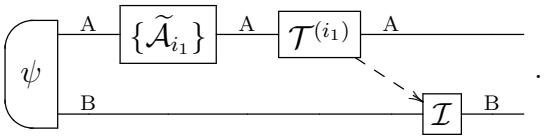
Again, we take the sequential composition of tests on Alice's system, let us call it $\left\{\widetilde{\mathcal{A}}_{i_1}\right\}$.



If $\mathcal{T}^{(i_1)}$ is a channel, we are done, because we are in the standard form for a 1-way LOCC protocol (see example 4.2.2). If $\mathcal{T}^{(i_1)}$ is a non-deterministic test, we still need a further passage. We can regard this case as if $\mathcal{T}^{(i_1)}$ is followed by the identity on Alice's system, irrespective of any Bob's outcome.

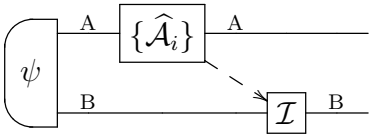


Now we swap $\mathcal{T}^{(i_1)}$ and $\mathcal{I}$, getting[9]



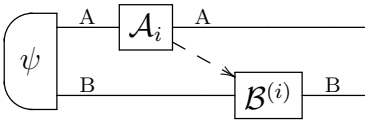
.

Finally, we take the sequential composition $\left\{\widehat{\mathcal{A}}_i\right\}$ of the two tests on Alice's system and we have



.

Since the protocol is pure, every operation we did, implying the swap of the two systems, is licit. Eventually, we drop to the standard form of a 1-way LOCC protocol



,

[9]Beware that here $\mathcal{I}$ is not, in general, the identity channel on B; it is only the "swapped version" of the identity channel on B, obtained by composing the swapping channels $\mathcal{C}$ and $\mathcal{R}$.

and we have classical communication from A to B. By lemma 4.5.5, we are entitled to replace the channel on B with a reversible channel. Now we have constructed our 1-way LOCC protocol with reversible channel, that is completely equivalent to the initial multi-way LOCC protocol. □

Clearly, by lemma 4.5.4, if we construct a 1-way LOCC protocol from Alice to Bob, there exists another one from Bob to Alice, so the role of Alice has nothing special.

Coming back to our discussion about mixedness, recalling lemma 4.4.3, this result states that if $|\psi)_{\mathrm{AB}}$ is more entangled than $|\psi')_{\mathrm{AB}}$, then $|\rho)_{\mathrm{A}}$ is always more mixed than $|\rho')_{\mathrm{A}}$ *and* $|\rho)_{\mathrm{B}}$ is always more mixed than $|\rho')_{\mathrm{B}}$. We can sum up all the results of this chapter in the following theorem.

Theorem 4.5.7 (Equivalence between entanglement and mixedness)**.** *Let $|\psi)_{\mathrm{AB}}$ and $|\psi')_{\mathrm{AB}}$ be bipartite pure states, let $|\rho)_{\mathrm{A}}$, $|\rho)_{\mathrm{B}}$ and $|\rho')_{\mathrm{A}}$, $|\rho')_{\mathrm{B}}$ be their marginals. The following statements are equivalent.*

1. *$|\psi)_{\mathrm{AB}}$ is more entangled than $|\psi')_{\mathrm{AB}}$.*
2. *$|\rho)_{\mathrm{A}}$ is more mixed than $|\rho')_{\mathrm{A}}$.*
3. *$|\rho)_{\mathrm{B}}$ is more mixed than $|\rho')_{\mathrm{B}}$.*

Now we have established the complete equivalence between entanglement and mixedness. This enables us to choose the relation we prefer in order to study its properties. We will choose the mixedness relation, which is simpler to treat and has some powerful tools related to it.

Chapter 5

Diagonalizing mixed states

In this chapter we want to build up some further tools to characterize and study the mixedness relation. In particular, our aim is to develop an abstract theory of majorization for mixed states, which we will do in the following two chapters. In quantum theory, it is a well established fact that mixedness relation is equivalent to majorization [11, 17, 60, 61, 62]. Majorization, as we will see in section 6.2, is a preorder that can be defined between sets of eigenvalues of density operators. Therefore, the issue is then how to give an abstract definition of eigenvalues of a mixed state.

When we diagonalize a density operator ρ, we can write $\rho = \sum_j p_j |j\rangle \langle j|$, where $\rho |j\rangle = p_j |j\rangle$ for every $|j\rangle$. Clearly, from an operational point of view, we cannot exploit the idea of eigenvalues and eigenvectors, because in an abstract scenario mixed states are not operators. Nevertheless, we see that when diagonalizing ρ, we are actually writing ρ as a convex combination of pure states, which makes sense in a general theory. However, ρ can be expressed as a convex combination of pure states in many other ways. What special feature distinguishes diagonalization from all the other convex combinations?

In quantum mechanics, this feature is the fact that eigenvectors are *orthogonal* pure states. We can translate this mathematical property into operational language as the fact that an experimenter can *perfectly distinguish* an eigenvector from all the other eigenvectors of a given density operator.

Therefore, this chapter is mainly devoted to the introduction of the notion of perfect distinguishability among states, and to the exploration of its consequences as far as the abstract version of the diagonalization of mixed states is concerned. Eventually, we come to the second central original result

of the present work: we develop a new protocol to diagonalize mixed states even in an abstract framework. Diagonalization will be defined as a convex decomposition of a mixed state in terms of perfectly distinguishable pure states.

5.1 Perfect distinguishability

We begin this section with the definition of perfectly distinguishable states.

Definition 5.1.1. We say that the normalized states $\{\rho_i\}$ are *perfectly distinguishable* if there exists an observation-test $\{a_j\}$ such that $(a_j|\rho_i) = \delta_{ij}$.

We say that the observation-test $\{a_j\}$ is *perfectly distinguishing*.

This definition is aimed at imitating the definition of orthogonal states in quantum mechanics.

In quantum mechanics, the role of reversible channels is played by unitary operators. Unitary operators have two key features: they are invertible and they preserve scalar products. It is interesting to see if reversible channels preserve the abstract version of orthogonality. The answer is affirmative.

Lemma 5.1.2. *Let $\{\rho_i\}$ be a set of perfectly distinguishable states. If $\mathcal{U}$ is reversible channel, then $\{\mathcal{U}\rho_i\}$ is another set of perfectly distinguishable states.*

Proof. Let $\{a_i\}$ be the perfectly distinguishing test for $\{\rho_i\}$. Then $\{a_i\mathcal{U}^{-1}\}$ is a perfectly distinguishing test for $\{\mathcal{U}\rho_i\}$. Indeed, for every i and j we have

$$\left(a_i\mathcal{U}^{-1}|\mathcal{U}\rho_j\right) = (a_i|\,\mathcal{U}^{-1}\mathcal{U}\,|\rho_j) = (a_i|\rho_j) = \delta_{ij}.$$

□

For the rest of this work, we will assume the following axiom related to perfect distinguishability.

Axiom 5.1.3. *For every normalized pure state $|\psi)$ there exists a pure effect $(a|$ such that $(a|\psi) = 1$.*

This means that there exists a pure effect that yields 1 on $|\psi)$, as if it were the deterministic effect. We will show in a while that this pure effect is unique, but some other results are necessary.

Example 5.1.4. In quantum mechanics, axiom 5.1.3 holds. Indeed, for every pure state $|\psi\rangle\langle\psi|$ we can associate a (unique) pure effect, given by the projector on the subspace spanned by $|\psi\rangle$. It is exactly $|\psi\rangle\langle\psi|$. In this way,

$$\text{tr } |\psi\rangle\langle\psi|\,\psi\rangle\langle\psi| = \text{tr } |\psi\rangle\langle\psi| = 1.$$

This shows that in quantum theories the duality between pure states and pure effects is stronger than a simple duality between the associated vector spaces.

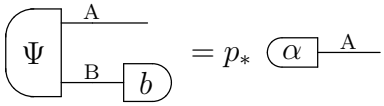
Now we prove an important proposition, which will play a central role in all our treatment of diagonalization.

Proposition 5.1.5. *Let ρ be a normalized mixed state of system* A. *Let p_* be the maximum weight with which a pure state appears in a convex decomposition of ρ into pure states,*

$$p_* := \max\{p \in (0,1] : \rho = p\alpha + (1-p)\,\sigma, \alpha \text{ pure}\}\,.$$

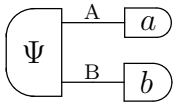
Then, if we purify ρ with purifying system B, *and $\widetilde{\rho}$ is the complementary state (see definition 3.1.3), there is a pure state of* B *that appears in a convex decomposition of $\widetilde{\rho}$ into pure states with the same weight p_*.*

Proof. By hypothesis, we can write $\rho = p_*\alpha + (1-p_*)\,\sigma$, where α is a normalized pure state of A and σ is another normalized state of A, which can be possibly mixed. Let us purify $|\rho)_{\text{A}}$, and let $|\Psi)_{\text{AB}}$ be one of its purifications. According to the steering property, there exists an effect b, which in principle may be not pure, on B such that it prepares α with probability p_*.



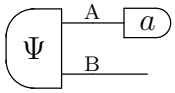
(5.1)

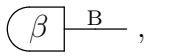
Let a be a pure effect such that $(a|\alpha) = 1$, it exists by axiom 5.1.3. If we apply a on A in (5.1), we get



Now, let us change perspective slightly. Suppose we apply a on A to the state $|\Psi)_{\text{AB}}$ (without b on system B!). By the steering property and pure

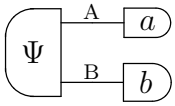
conditioning, since a is pure, it induces a pure state β of system B in a convex decomposition of $\widetilde{\rho}$ such that

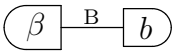


$= q$


, (5.2)

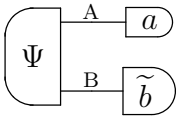
where $q \in (0, 1]$. If we now apply b (the effect on B that prepares α on A), we have



$= p_* = q$


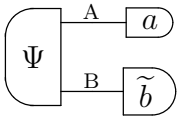
.

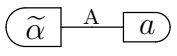
Since $(b|\beta) \in [0, 1]$, then it must be $q \geq p_*$. Now, let us apply $\widetilde{b}$, a pure effect such that $\left(\widetilde{b}|\beta\right) = 1$, in (5.2).



$= q$

On the other hand, $\widetilde{b}$ will induce some pure state[1] $\widetilde{\alpha}$ compatible with ρ on A, with probability $\widetilde{p} \in (0, 1]$. Using diagrams,

$q =$


$= \widetilde{p}$


,

whence $q \leq \widetilde{p}$. But, by hypothesis, $\widetilde{p} \leq p_*$, so we have the chain of inequalities $p_* \leq q \leq \widetilde{p} \leq p_*$. It follows that we actually have $p_* = q = \widetilde{p}$.

This proves that there exists a pure state β in B such that it has the same weight p_* in a convex decomposition of $\widetilde{\rho}$. □

The pure state β is such that $(b|\beta) = 1$, because $p_* = q$, where b is the effect that prepares α, and β is prepared by a.

Remark 5.1.6. In our previous line of reasoning, the starting point was the maximum probability p_* with which a pure state appears in a convex decomposition of ρ. This quantity is not as simple to define as it may appear at first glance. Actually, we must proceed in the following way.

[1]Recall that $\widetilde{b}$ is pure.

- We consider all the convex decompositions of ρ into pure states. We take the supremum of the weights in each decomposition.
- Then we take the supremum over the set of suprema.

In other words, p_* is defined as

$$p_* := \sup_{\{p_i\}} \sup_i \{p_i\} .$$

If the vector space associated with the states of our theory is finite-dimensional, then each convex decomposition can be taken to have a finite number of terms, according to Carathéodory's theorem (see theorem A.1.2). Therefore, in this case, which is the case we are interested in, the inner supremum is in fact a maximum.

$$p_* = \sup_{\{p_i\}} \max_i \{p_i\} .$$

But what about the outer supremum? It is important that it is in fact a maximum, because in that case we can associate a pure state α with it and proposition 5.1.5 makes sense.

This is true if and only if the set of pure states is closed. Indeed, let us consider a converging sequence of probabilities $\{p_{i,n}\}_{n\in\mathbb{N}}$ and a converging sequence of pure states $\{\alpha_{i,n}\}_{n\in\mathbb{N}}$, such that $\rho = \sum_i p_{i,n}\alpha_{i,n}$ for every $n \in \mathbb{N}$. We would like to replace this sequence with its limit (this is equivalent to impose a closure condition), so that we can write $\rho = \sum_i p_i\alpha_i$, where α_i is a pure state for every i, and p_i and α_i are the limits of the two sequences. So, we require that every converging sequence of pure states converges to a pure state, which means that the set of pure states is closed. But we already know that this is true in a theory with purification (see proposition 3.2.10).

Now let us turn to prove the uniqueness of the effect defined in axiom 5.1.3. As we will show soon, this problem is strongly related to its dual problem: suppose we know that $(a|\psi) = 1$, where a is a pure effect. Do there exist other (possibly mixed) states such that $(a|\rho) = 1$? The answer is negative: it must be $\rho = \psi$. First of all, let us prove that such a ρ must be pure. We need a preparatory lemma.

Lemma 5.1.7. *If* $(a|\rho) = 1$, *then*[2] $a =_\rho e$.

[2]Recall that $a =_\rho e$ means that a is equal to the deterministic effect upon input of ρ (see definition 3.1.9).

Proof. Let σ be a state compatible with ρ. This means that $\rho = p\sigma + (1-p)\,\tau$, for some $p \in (0,1]$. If $(a|\rho) = 1$, then

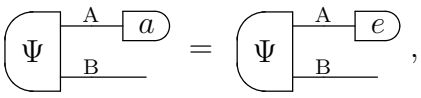
$$1 = p\,(a|\sigma) + (1-p)\,(a|\tau)$$

Since $(a|\sigma) \leq 1$ and $(a|\tau) \leq 1$, we have that the right-hand side is less than $p+1-p = 1$ and it is equal to 1 if and only if $(a|\sigma) = (a|\tau) = 1$. This means that a behaves as the deterministic effect when applied to all the states in the face identified by ρ. □

Now we are ready to prove that if $(a|\rho) = 1$, then ρ must be pure.

Lemma 5.1.8. *Let a be a pure effect. If we have a normalized state ρ such that $(a|\rho) = 1$, then ρ is pure.*

Proof. By lemma 5.1.7, if $(a|\rho) = 1$, then $a =_\rho e$. According to proposition 3.1.11, we have



where Ψ is a purification of ρ. Since a is pure, by pure conditioning it will induce a pure state on system B, so the complementary state $\widetilde{\rho}$ is pure. By proposition 4.1.2, ψ is the product of its marginals, $\psi = \rho \otimes \widetilde{\rho}$, whence ρ is pure. □

The next step is the proof of the dual uniqueness statement, which is easier.

Proposition 5.1.9. *Let a be a pure effect such that $(a|\alpha) = 1$ for some normalized pure state α. If $(a|\rho) = 1$, then $\rho = \psi$.*

Proof. Clearly, by lemma 5.1.8, ρ must be pure. Let $\alpha' \neq \alpha$ be another pure state such that $(a|\alpha') = 1$. If we consider the mixed state $\sigma = \frac{1}{2}\,(\alpha + \alpha')$, we have $(a|\sigma) = \frac{1}{2}\,(a|\alpha) + \frac{1}{2}\,(a'|\alpha) = 1$; but this is impossible, because, by lemma 5.1.8, σ must be pure. Therefore $\alpha = \alpha'$. □

This result enables us to say that the effect b on B that prepares α on A with probability p_* can be taken to be pure. Indeed, recall the proof of proposition 5.1.5, where $\widetilde{b}$ is a *pure* effect that prepares the pure state $\widetilde{\alpha}$ on A with probability $\widetilde{p}$. But since $\widetilde{p} = p_*$, then $(a|\widetilde{\alpha}) = 1$, therefore, by

proposition 5.1.9 $\widetilde{\widetilde{\alpha}} = \alpha$. Hence, the pure effect $\widetilde{b}$ actually prepares α with probability p_*. This shows α can be always prepared with probability p_* using a pure effect on B.

To see the uniqueness of the pure effect associated with a pure state, we apply the same argument of the proof of proposition 5.1.5 to the case when ρ is the invariant state χ. We can follow the same line of reasoning thanks to the following lemma.

Lemma 5.1.10. *Let χ be the invariant state of system* A *and let φ be a normalized pure state. Then*

$$p_{\max} := p_\varphi = \max\{p : \exists\, \sigma, \chi = p\varphi + (1-p)\,\sigma\}$$

does not depend on φ.

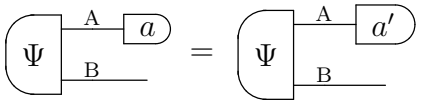
Proof. Since for any couple of pure states φ and ψ there is a reversible channel $\mathcal{U}$ such that $\psi = \mathcal{U}\varphi$, applying $\mathcal{U}$ to χ yields $\mathcal{U}\chi = p\psi + (1-p)\,\psi$. Because χ is invariant, one has $\chi = p\psi + (1-p)\,\psi$. This shows that p_φ actually does not depend on the particular pure state φ. □

Since χ is completely mixed, every (pure) state is compatible with it. p_φ is nothing but p_*, relative to the pure state φ. The surprising result is that p_* is the same for all pure states. Note that $p_{\max}$ is non-vanishing.

Now we can prove uniqueness.

Proposition 5.1.11. *For every normalized pure state α there is a unique pure effect a such that $(a|\alpha) = 1$.*

Proof. Suppose, by contradiction, that there are two pure effects a and a' such that $(a|\alpha) = (a'|\alpha) = 1$. Let $|\Psi)_{\mathrm{AB}}$ be a purification of the invariant state, and let b the pure effect that prepares α with probability $p_{\max}$. Then, recalling the proof of proposition 5.1.5, a and a' prepare two pure states on B, say β and β', such that $(b|\beta) = (b|\beta') = 1$. Hence, by proposition 5.1.9, $\beta = \beta'$. Therefore



Since χ is completely mixed, $|\Psi)_{\mathrm{AB}}$ is faithful for effects of system A, therefore $a = a'$. □

Now, let us see some consequences of proposition 5.1.11. The most obvious consequence is that we can associate a pure effect with every normalized pure state and vice versa. In this way, we expect that the action of the group of reversible channels is transitive also on the set of pure effects $\{a\}$ such that $\|a\| = 1$.

Corollary 5.1.12. *If a and a' are two pure effects with $\|a\| = \|a'\| = 1$, then there is a reversible channel such that $a' = a\mathcal{U}$.*

Proof. Let α and α' be the normalized pure states associated with a and a' respectively. We know that there is a reversible channel $\mathcal{U}$ such that $\alpha = \mathcal{U}\alpha'$. Now, we have $1 = (a|\alpha) = (a|\mathcal{U}\alpha') = (a'|\alpha)$. Since $a\mathcal{U}$ is a pure effect (see lemma 2.2.9), by the uniqueness of the pure effect associated with α, we conclude that $a' = a\mathcal{U}$. □

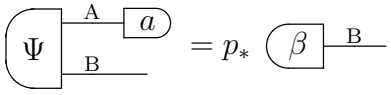
Let α and α' be two normalized pure states and a and a' their associated effects. If $\alpha' = \mathcal{U}\alpha$, notice that the corresponding effects are related by $a' = a\mathcal{U}^{-1}$. In other words, if we go from a pure state to another by $\mathcal{U}$, we go from one corresponding effect to the other by $\mathcal{U}^{-1}$.

5.2 Diagonalizing mixed states

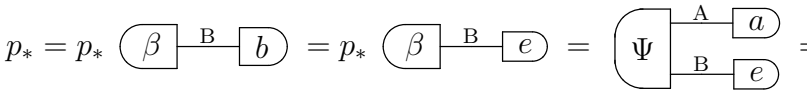
In this section we deal with the issue of diagonalizing mixed states, namely of writing a mixed state as a convex combination of perfectly distinguishable pure states. The starting point for diagonalization is a straightforward corollary of proposition 5.1.5.

Corollary 5.2.1. *Consider $\rho = p_*\alpha + (1 - p_*)\,\sigma$, where p_* is defined in proposition 5.1.5, and let a be a pure effect such that $(a|\alpha) = 1$. Then $(a|\rho) = p_*$.*

Proof. According to the proof of proposition 5.1.5, if we apply a to a purification of ρ, it prepares a pure state β on B with probability p_*.



If we now apply b (the effect that prepares α) to B, we know that $(b|\beta) = 1$, so we can put the deterministic effect in place of b. Hence



$$= \quad \rho \;\overset{\text{A}}{—}\; a \;.$$

This shows that $(a|\rho) = p_*$. □

This corollary enables us to define p_* in an alternative way.

Proposition 5.2.2. *Let $\rho \in \mathsf{St}_1(\mathrm{A})$. Define $p^* := \max_a (a|\rho)$, where the maximum is taken on the set of pure effects such that $\|a\| = 1$. Then $p^* = p_*$.*

Proof. By corollary 5.2.1, clearly one has $p^* \geq p_*$. Suppose, by contradiction, that $p^* > p_*$. In this way, since p^* is the maximum, there exists a pure effect a' such that $(a'|\rho) > p_*$. Let $(a'|\rho) = \lambda$.

$$\lambda = \quad \rho \;\overset{\text{A}}{—}\; a' \quad = \quad \Psi \;\begin{matrix}\overset{\text{A}}{—}\; a' \\ \underset{\text{B}}{—}\; e\end{matrix} \;,$$

where $|\Psi)_{\mathrm{AB}}$ is a purification of ρ. Now, a' prepares a pure state β' of B with probability λ. Indeed, if we take the pure effect b' that yields 1 when applied to β', we have

$$\lambda = \lambda \quad \beta' \;\overset{\text{B}}{—}\; e \quad = \lambda \quad \beta' \;\overset{\text{B}}{—}\; b' \quad = \quad \Psi \;\begin{matrix}\overset{\text{A}}{—}\; a' \\ \underset{\text{B}}{—}\; b'\end{matrix} \quad =$$

$$= \quad \Psi \;\begin{matrix}\overset{\text{A}}{—}\; a' \\ \underset{\text{B}}{—}\; e\end{matrix} \quad = \quad \rho \;\overset{\text{A}}{—}\; a' \;,$$

which means

$$\Psi \;\begin{matrix}\overset{\text{A}}{—}\; a' \\ \underset{\text{B}}{—}\; b'\end{matrix} \quad = \lambda.$$

Now, b' prepares on A a pure state α' with probability q such that

$$\lambda = \quad \Psi \;\begin{matrix}\overset{\text{A}}{—}\; a' \\ \underset{\text{B}}{—}\; b'\end{matrix} \quad = q \quad \alpha' \;\overset{\text{A}}{—}\; a' \;.$$

This implies $q \geq \lambda > p_*$. Therefore $q > p_*$, and the pure state α' appears in a convex decomposition of ρ with a weight strictly greater than p_*, and this contradicts the fact that p_* is the maximum weight for a pure state. □

This alternative way of defining p_*, using effects instead of states, strengthens the idea of duality between states and effects.

The result expressed in corollary 5.2.1 has deep and important implications. Since $(a|\rho) = p_*$, this means that $(a|\sigma) = 0$, provided[3] $p_* \neq 1$, where $\rho = p_*\alpha + (1 - p_*)\,\sigma$. Indeed,

$$p_* = (a|\rho) = p_*\,(a|\alpha) + (1 - p_*)\,(a|\sigma) = p_* + (1 - p_*)\,(a|\sigma)\,,$$

because $(a|\alpha) = 1$. This implies that α and σ are perfectly distinguishable from each other. Indeed, we can build up an observation-test $\{a, e - a\}$ that distinguishes between α and σ. This is a hint of the fact that we are on the right track to write ρ as a convex combination of perfectly distinguishable pure states. Besides, if $(a|\sigma) = 0$, then $(a|\tau) = 0$ for any state τ compatible with σ. To show it, it is enough to recall that $\tau \in F_\sigma$ if and only if there exists $p \in (0, 1]$ such that $\sigma = p\tau + (1 - p)\,\tau'$. If $(a|\sigma) = 0$, then

$$0 = p\,(a|\tau) + (1 - p)\,(a|\tau')\,.$$

It must be $(a|\tau) = 0$, because the right-hand side is a sum of two non-negative numbers. In particular, $(a|\alpha) = 0$ for any pure state α in the face identified by σ.

Corollary 5.2.3. *Every pure state is perfectly distinguishable from some other pure state.*

Proof. Let us consider the invariant state χ. For every normalized pure state ψ, we have $\chi = p_{\max}\psi + (1 - p_{\max})\,\sigma$, where σ is another normalized state. By corollary 5.2.1, if a is the pure effect such that $(a|\psi) = 1$, then $(a|\sigma) = 0$. If σ is pure, then ψ is perfectly distinguishable from σ by means of the observation-test $\{a, e - a\}$.

If σ is mixed, than $(a|\varphi) = 0$, for every pure state φ in the face identified by σ. Therefore ψ is perfectly distinguishable from φ by the observation-test $\{a, e - a\}$. □

We can proceed in this manner on the way of diagonalizing ρ.

- Once we have determined $p_* =: p_1$ and we have found $\alpha =: \alpha_1$ such that $\rho = p_1\alpha_1 + (1 - p_1)\,\sigma_1$, we repeat the previous procedure for the state σ_1.

[3] If $p_* = 1$, then ρ is pure, and we are done. Therefore, it is licit to assume $p_* \neq 1$.

- We find p_2, defined as the maximum weight such that we can write $\sigma_1 = p_2\alpha_2 + (1-p_2)\,\sigma_2$, where α_2 is a pure state and σ_2 is another state; and so on.

This process must end sooner or later when we find that the remaining state σ_i is pure. At each step i, α_i is perfectly distinguishable from σ_i and therefore it is perfectly distinguishable from each of the pure states in the face identified by σ_i.

At the end of this procedure, we can write ρ as $\rho = \sum_i p_i\alpha_i$, where $p_i \geq p_{i+1}$ for every i and $(a_i|\alpha_j) = 0$ for every $j > i$, where a_i is defined as the pure effect such that $(a_i|\alpha_i) = 1$.

Our next goal is to show that all the α_i's are perfectly distinguishable. Therefore, we move on to consider the following issue.

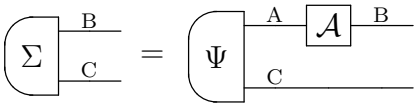
Suppose somebody prepares a normalized pure state taken from the set $\{\alpha_i\}$. We know that for every pure state α_i there is a (unique) pure effect a_i such that $(a_i|\alpha_i) = 1$, and that $(a_i|\alpha_j) = 0$ for every $j > i$. Are we able to distinguish the pure states perfectly, namely to identify with certainty which state has been prepared?

The answer is affirmative, but we have to construct a distinguishing protocol carefully. The key idea will be to switch from pure effects to pure transformations, that occur with the same probability as the effects, according to corollary 3.2.6. We will focus on transformations rather than on effects because we want to do a procedure that involves several iterations of a basic procedure. If we use effects, we will not be able to iterate the procedure, because effects destroy the system we are examining.

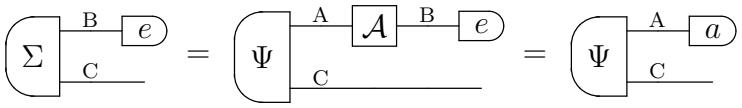
Before explaining the procedure, we need a result about the relationship between effects and their associated transformations.

Proposition 5.2.4. *Let a be an effect such that $(a|\rho) = 1$, for some normalized state $\rho \in \mathsf{St}_1(\mathrm{A})$. Then there exists a transformation $\mathcal{T}$ on* A *such that $(a| = (e|\,\mathcal{T}$ and $\mathcal{T} =_\rho \mathcal{I}$.*

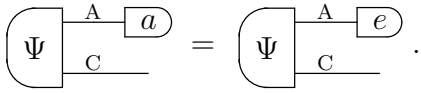
Proof. Let us consider a purification $\Psi \in \mathsf{St}_1(\mathrm{AC})$ of ρ. By corollary 3.2.6, there exists a system B and a pure transformation $\mathcal{A} \in \mathsf{Transf}(\mathrm{A},\mathrm{B})$ such that $(a|_{\mathrm{A}} = (e|_{\mathrm{B}}\,\mathcal{A}$. Let us apply $\mathcal{A}$ on system A; the resulting state $|\Sigma)_{\mathrm{BC}}$ will be pure, because $\mathcal{A}$ is pure.



Let us now take the deterministic effect on system B.

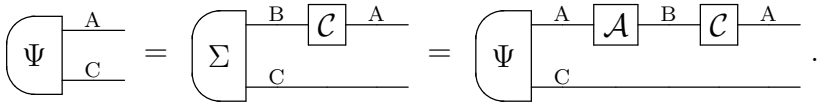


By lemma 5.1.7, $(a|\rho) = 1$ implies $a =_\rho e$. According to proposition 3.1.11, we have then



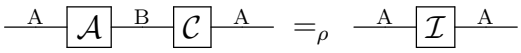
.

This means that $|\Sigma)_{\mathrm{BC}}$ and $|\Psi)_{\mathrm{AC}}$ are two purifications of the same state (their marginal on C). Then, there exists a channel $\mathcal{C}$ from B to A to such that

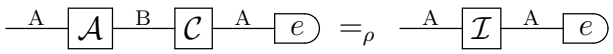


.

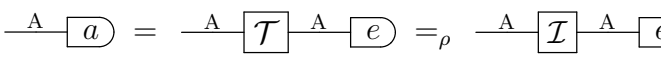
Since equality on purifications implies equality on input (see proposition 3.1.10), we have that $\mathcal{C}\mathcal{A} =_\rho \mathcal{I}_{\mathrm{A}}$.



Let us take the deterministic effect on A.



Since $\mathcal{C}$ is a channel, the left-hand side is nothing but $(a|_{\mathrm{A}}$. If we define a transformation $\mathcal{T}$ on A as $\mathcal{T} := \mathcal{C}\mathcal{A}$, we then have



.

The proposition is proven. □

Note that, unlike in corollary 3.2.6, here the transformation is *on* A, whereas in corollary 3.2.6, the transformation is from A to B, where system B need not to be equal to A. Therefore, for a generic effect $(a|_{\mathrm{A}}$ we have $(a|_{\mathrm{A}} = (e|_{\mathrm{B}}\,\mathcal{A}$, where $\mathcal{A}$ is a pure transformation from A to a system B. However, if $(a|_{\mathrm{A}}$ is such that $\|a\| = 1$, namely there is a normalized state ρ such that $(a|\rho) = 1$, the transformation can be taken to act on A, without an additional system B. However, in general, in this case the transformation is not pure.

Now we can explain our procedure of discriminating the pure states $\{\alpha_i\}_{i=1}^n$, such that $(a_i|\alpha_i) = 1$, and that $(a_i|\alpha_j) = 0$ for every $j > i$.

- Let us consider the observation-test $\{a_1, e - a_1\}$, which perfectly distinguishes α_1 from the other states, and we apply the associated test $\{\mathcal{A}_1, \mathcal{A}_1^\perp\}$, according to corollary 3.2.6. If the outcome is $\mathcal{A}_1$, then we conclude that the state is α_1. If not, the state is one of the other ones. Now we consider $\rho_1 = \frac{1}{n-1}\sum_{i=2}^n \alpha_i$. Since $(e - a_1|\rho_1) = 1$, because $(a_1|\alpha_1) = 1$, then $\mathcal{A}_1^\perp$ will leave all the pure states $\{\alpha_i\}_{i=2}^n$ invariant, according to proposition 5.2.4. We are ready to repeat the test.

- This time we consider the observation-test $\{a_2, e - a_2\}$, which perfectly distinguishes α_2 from the remaining states, and we apply the associated test $\{\mathcal{A}_2, \mathcal{A}_2^\perp\}$. If the outcome is $\mathcal{A}_2$, then the state is α_2. If not, we consider $\rho_2 = \frac{1}{n-2}\sum_{i=3}^n \alpha_i$. Since $(e - a_2|\rho_2) = 1$, then $\mathcal{A}_2^\perp$ leaves all the pure states $\{\alpha_i\}_{i=3}^n$ invariant. Now we repeat the procedure again.

- In this way, repeating the procedure several times, we are able to identify the state with certainty.

We therefore managed to "diagonalize" every normalized mixed state ρ, namely we managed to write it as a convex combination of perfectly distinguishable pure states.

Remark 5.2.5. Note that, in general, the perfectly distinguishing test for the pure states $\{\alpha_i\}$ is *not* made of the pure effects $\{a_i\}$ such that $(a_i|\alpha_i) = 1$. In other words, in general, the perfectly distinguishing test is *not* pure. Indeed, suppose that $\{\alpha_i\}_{i=1}^n$ are perfectly distinguishable pure states, and that if we add some pure state α_{n+1}, the states $\{\alpha_i\}_{i=1}^{n+1}$ are perfectly distinguishable too. Suppose we want to build up the perfectly distinguishing tests for these two sets by putting together all the pure effects a_i associated with the pure states. In this way, the perfectly distinguishing test for $\{\alpha_i\}_{i=1}^n$ would be $\{a_i\}_{i=1}^n$, and we have $\sum_{i=1}^n a_i = e$. The perfectly distinguishing test for the pure states $\{\alpha_i\}_{i=1}^{n+1}$ is $\{a_i\}_{i=1}^{n+1}$, and again $\sum_{i=1}^{n+1} a_i = e$. Comparing the two sums we have

$$e = \sum_{i=1}^{n+1} a_i = \sum_{i=1}^{n} a_i + a_{n+1} = e + a_{n+1}.$$

This means that $a_{n+1} = 0$, but this is absurd because by hypothesis $(a_{n+1}|\alpha_{n+1}) = 1$.

Therefore not all the perfectly distinguishing tests are pure.

We can now give the following definition.

Definition 5.2.6. Let ρ be a normalized mixed state. A *diagonalization* of ρ is a convex decomposition of ρ into perfectly distinguishable normalized pure states.

We call the weights in a diagonalization of ρ the *eigenvalues* of ρ relative to that diagonalization.

We have proven that every mixed state can be diagonalized.

In general, the diagonalization of ρ will not be unique, since it is not unique even in quantum mechanics. However, in quantum mechanics two different diagonalizations of a mixed state can differ only by the choice of the eigenvectors (by a change of basis), whereas the eigenvalues are the same. In a general theory, we do not know if this is true; and this is the reason why we were forced to specify that the eigenvalues are relative to a specific diagonalization.

However, if recall the procedure of diagonalization, we see that each eigenvalue is defined as the maximum weight with which a pure state appears in a convex decomposition of the given mixed state. Since we know that the maximum of a set is unique, we are tempted to state that the eigenvalues of a mixed state are uniquely defined. If the mixed state is non-degenerate, namely if at any step of the procedure the maximum is achieved by a unique pure state, this is true. And of course this is always true for the largest eigenvalue p_1. Problems arise if somewhere the maximum is achieved by more than one state. In this way, when we diagonalize, we have to choose one of these states and the eigenvalues found at later steps may in principle depend on the choice of the pure state. So, if for non-degenerate mixed states the eigenvalues are well-defined, for the other states the question remains open, at least in our procedure of diagonalization.

Nevertheless, the most serious issue is that there might be other diagonalization procedures which are different from the one we presented in this section. Indeed, we showed a procedure to diagonalize mixed states, but we cannot tell whether this is the only procedure of diagonalization. Clearly, different procedures of diagonalization can yield different eigenvalues, even if the state is non-degenerate.

Our procedure of diagonalizing mixed states gives us also some results about the relationship between our diagonalization of ρ and the corresponding diagonalization of its complementary state $\widetilde{\rho}$. In proposition 5.1.5, we showed that the pure effect a, where $(a|\alpha) = 1$ and α appears in a convex decomposition of ρ with weight p_*, prepares a pure state β of B such that

β appears in a convex decomposition of $\widetilde{\rho}$ with the same weight p_*. Actually, since every step in our procedure of diagonalization involves a p_* of a particular mixed state, we see that at the end we manage to write $\widetilde{\rho}$ as a convex combination of pure states with the same weights as ρ. But do we actually get a diagonalization of $\widetilde{\rho}$? In other words, are the pure states in this convex decomposition of $\widetilde{\rho}$ perfectly distinguishable? The answer is clearly affirmative. Indeed, applying a similar argument to the one in corollary 5.2.1 to system B, one gets that $(b|\widetilde{\rho}) = p_*$, where b is the pure effect associated with β. Then one can redo the same procedure to diagonalize $\widetilde{\rho}$. This means that the pure states appearing in the convex decomposition of $\widetilde{\rho}$ are perfectly distinguishable. Then we can say that the complementary state $\widetilde{\rho}$ has a diagonalization with the same eigenvalues as the corresponding diagonalization of ρ.

Chapter 6

A two-level system

This chapter is devoted to analyse the first consequences of our procedure of diagonalization of mixed states, as far as mixedness relation is concerned. In this vein, we will focus on the simplest with perfectly distinguishable pure states, a 2-level system. The main tool to study mixedness relation using the eigenvalues we have just defined is given by majorization, a widely used tool in statistics. We will develop this topic first in full generality, and then applying it to the case of a 2-level system. Related to this subject, we have Schur-concave functions, which directly measure mixedness by assigning a real number to each mixed state according to its eigenvalues. Among them we have entropies, that emerge in our treatment as measures of mixedness.

There have been several approaches towards a definition of entropy in general probabilistic theories [64, 65, 66]. In particular, Barnum *et al.* [64] base their definition on states, whereas Short and Wehner [65] rely on effects. Our innovative approach, based on diagonalization of mixed states, has the merit of unifying these two approaches, because, as a consequence of proposition 5.2.2, we can achieve diagonalization of mixed states starting from pure states or pure effects indifferently.

6.1 A two-level system

In this section we see some consequences of the results of the the previous chapter for a 2-level system, which is the simplest system where we can take advantage of the procedure of diagonalization. We will assume axiom 5.1.3, as well as all the results of the previous chapter.

Before introducing the definition of a 2-level system, we define what a maximal set of perfectly distinguishable states is.

Definition 6.1.1. Let $\{\rho_i\}_{i=1}^n$ be a set of perfectly distinguishable states. We say that $\{\rho_i\}_{i=1}^n$ is *maximal* if there is *no* state ρ_{n+1} such that the states $\{\rho_i\}_{i=1}^{n+1}$ are perfectly distinguishable.

Intuitively, a 2-level system, is a system where every maximal set of perfectly distinguishable states is made of two elements.

Definition 6.1.2. A *2-level system* (or *g-bit*[1]) is a system whose maximal sets of perfectly distinguishable states have 2 elements.

Actually, we will focus on maximal sets of perfectly distinguishable *pure* states. We know that such sets exist because every pure state is perfectly distinguishable from another pure state (see corollary 5.2.3). In general, the maximal sets of perfectly distinguishable pure states are not pairwise disjoint. Suppose ψ_1 is perfectly distinguishable from ψ_2 by means of the observation-test $\{a_1, a_2\}$. Nothing forbids that ψ_1 is perfectly distinguishable also from another pure state ψ_2' by an observation-test $\{a_1', a_2'\}$. Clearly, at least one of the two observation-tests is *not* pure. Indeed, if they were both pure, then $a_1 = a_1'$, because there is only one pure effect that yields 1 on ψ_1. Since $a_1 + a_2 = e$ and $a_1' + a_2' = e$, this would imply also $a_2 = a_2'$. Therefore $(a_2|\psi_2) = (a_2|\psi_2')$. By proposition 5.1.9, then $\psi_2 = \psi_2'$.

Thus, a sufficient condition to have pairwise disjoint maximal sets of perfectly distinguishable pure states is that all the perfectly distinguishing tests are *pure*.

The most important consequence of the definition of a 2-level system is about diagonalization. Since in a 2-level system we have at most two perfectly distinguishable pure states, every mixed state diagonalization has two terms.

In this way, we can write $\rho = p\psi + (1-p)\,\psi'$, where ψ and ψ' are perfectly distinguishable pure states. We will use the convention $p \geq 1 - p$, namely $p \geq \frac{1}{2}$ (and clearly $p \leq 1$). We see that diagonalizations in a 2-level system are somehow "rigid": it is necessary to specify only one of the weight, because the other one is completely determined. This "rigidity" will enable us to easily prove some specific results for a 2-level system, such as the following proposition.

[1]Generalized bit

Proposition 6.1.3. *Every diagonalization of the invariant state χ has $p = \frac{1}{2}$, namely there exist two perfectly distinguishable pure states ψ and ψ' such that $\chi = \frac{1}{2}\psi + \frac{1}{2}\psi'$.*

Proof. Let us consider a diagonalization of the invariant state $\chi = p\psi + (1-p)\,\psi'$, where $\frac{1}{2} \leq p \leq 1$. Let us apply a reversible channel $\mathcal{U}$, such that $\mathcal{U}\psi = \psi'$, to χ. According to lemma 5.1.2, $\psi'' = \mathcal{U}\psi'$ will be perfectly distinguishable from $\mathcal{U}\psi = \psi'$. Then we another diagonalization of χ.

$$\chi = p\psi' + (1-p)\,\psi'' = p\psi + (1-p)\,\psi'$$

We want to prove that, actually, $\psi'' = \psi$. Now,

$$\psi = \frac{1}{p}\chi - \frac{1-p}{p}\psi' = \chi + \frac{1-p}{p}\Delta$$

$$\psi'' = \frac{1}{1-p}\chi - \frac{p}{1-p}\psi' = \chi + \frac{p}{1-p}\Delta,$$

where $\Delta := \chi - \psi'$. Since $\frac{1-p}{p} > 0$ and $\frac{p}{1-p} > 0$, then ψ and ψ'' are in the same half-line with initial point χ and direction Δ. If $\frac{1-p}{p} > \frac{p}{1-p}$, which means $0 < p < \frac{1}{2}$, then ψ'' lies in the segment with extremal points χ and ψ. But this is impossible, because we are assuming $\frac{1}{2} \leq p \leq 1$. If, instead, $\frac{1-p}{p} < \frac{p}{1-p}$, which means $\frac{1}{2} < p < 1$, ψ can be can be written as a convex combination of χ and ψ''. This is impossible, because ψ is pure. The only left possibility is that $\frac{1-p}{p} = \frac{p}{1-p}$, which means $p = \frac{1}{2}$: in this case $\psi'' = \psi$. □

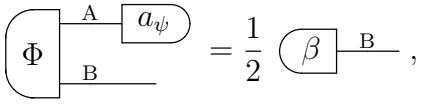
We have just proven that every diagonalization of χ has only the eigenvalue $\frac{1}{2}$ with multiplicity 2. In particular, we have that maximal sets of perfectly distinguishable pure states that diagonalize the invariant state are related to each other by reversible channels.

Now, we would like to prove something stronger, namely that whenever we take a maximal set of perfectly distinguishable pure states, their convex combination with equal weights yields the invariant state.

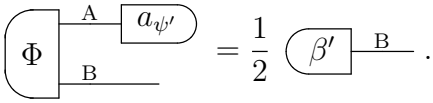
We need an additional assumption. To identify what the best choice is, let us prove some additional results.

Lemma 6.1.4. *In a 2-level system there is a pure state which is perfectly distinguishable from another pure state by a pure observation-test.*

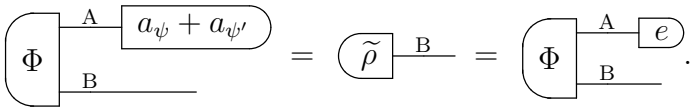
Proof. Let us consider a diagonalization of the invariant state $\chi = \frac{1}{2}(\psi + \psi')$, obtained according to the procedure of the previous chapter. Let us prove that ψ can be perfectly distinguished from ψ' by a pure observation-test, namely made of the pure effects a_ψ and $a_{\psi'}$ associated with ψ and ψ' respectively. Since the two pure states have the same weight, we can apply all the results of the previous chapter to both of them. Then we know that $(a_\psi|\psi) = (a_{\psi'}|\psi') = 1$ and $(a_\psi|\psi') = (a_{\psi'}|\psi) = 0$. We want to prove that $a_\psi + a_{\psi'} = e$; thanks to the purification postulate this is a sufficient condition for $\{a_\psi, a_{\psi'}\}$ to be a test. Let us consider a purification $|\Phi)_{\mathrm{AB}}$ of $|\chi)_{\mathrm{A}}$. According to the steering property, a_ψ and $a_{\psi'}$ induce pure states on B such that



and



From the results of the previous chapter, we know that the complementary state is $\widetilde{\rho} = \frac{1}{2}(\beta + \beta')$. So



Since $|\chi)_{\mathrm{A}}$ is completely mixed, $|\Phi)_{\mathrm{AB}}$ is faithful for effects of system A. Therefore we conclude that $a_\psi + a_{\psi'} = e$, so a_ψ and $a_{\psi'}$ make up an observation-test. □

A straightforward corollary is the following.

Corollary 6.1.5. *In a 2-level system* every *pure state is perfectly distinguishable from another pure state by a pure observation-test.*

Proof. Consider the state ψ defined in the proof of lemma 6.1.4. ψ is perfectly distinguishable from ψ' by the pure observation-test $\{a_\psi, a_{\psi'}\}$. Now, every pure state can be obtained by applying a suitable reversible channel $\mathcal{U}$ to ψ. By lemma 5.1.2, the set $\{\mathcal{U}\psi, \mathcal{U}\psi'\}$ is a set of perfectly distinguishable pure states. The perfectly distinguishing test is $\{a_\psi \mathcal{U}^{-1}, a_{\psi'}\mathcal{U}^{-1}\}$ and it is also pure. This proves that every pure state can be perfectly distinguished from another pure state by means of a pure observation-test. □

Therefore, if a pure state is perfectly distinguishable from more than one pure state, then in one case perfect distinguishability is achieved by means of a pure observation-test; in the other cases the perfectly distinguishing test is not pure.

Proposition 6.1.6. *The following statements are equivalent in a 2-level system.*

1. *Given two maximal sets of perfectly distinguishable pure states* $\{\psi, \psi'\}$ *and* $\{\varphi, \varphi'\}$*, there exists a reversible channel* $\mathcal{U}$ *such that* $\{\varphi, \varphi'\} = \{\mathcal{U}\psi, \mathcal{U}\psi'\}$.

2. *Every perfectly distinguishing test is pure.*

Proof. Let us prove the two implications.

Suppose that every perfectly distinguishing test is pure. Let $\{\psi, \psi'\}$ and $\{\varphi, \varphi'\}$ be two maximal sets of perfectly distinguishable pure states. We know that there is a reversible channel $\mathcal{U}$ such that $\varphi = \mathcal{U}\psi$. $\mathcal{U}$ maps perfectly distinguishable sets into perfectly distinguishable sets. Therefore, $\{\mathcal{U}\psi, \mathcal{U}\psi'\} = \{\varphi, \mathcal{U}\psi'\}$ is a set of perfectly distinguishable pure states. As we noted above, if every perfectly distinguishing test is pure, then every pure state is perfectly distinguishable from only one pure state. Now, since φ is perfectly distinguishable only from φ', this means that $\varphi' = \mathcal{U}\psi'$.

Let us prove the converse implication. Suppose ψ can be perfectly distinguished from ψ' by a pure observation-test. Then, if $\{\varphi, \varphi'\}$ is another maximal set of perfectly distinguishable pure states, there is a reversible channel such that $\{\varphi, \varphi'\} = \{\mathcal{U}\psi, \mathcal{U}\psi'\}$. Since ψ and ψ' are perfectly distinguishable by the pure observation-test $\{a_\psi, a_{\psi'}\}$, then φ and φ' are perfectly distinguishable by the pure observation-test $\{a_\psi \mathcal{U}^{-1}, a_{\psi'} \mathcal{U}^{-1}\}$. In this way, the (pure) states in every maximal set of perfectly distinguishable pure states are distinguishable by a pure observation-test. □

We want to assume one of the two statements, to derive some further results. Since they are equivalent, we can choose either one. It will turn out to be more convenient to make the following assumption.

Assumption 6.1.7. *Given two maximal sets of perfectly distinguishable pure states* $\{\psi, \psi'\}$ *and* $\{\varphi, \varphi'\}$*, there exists a reversible channel* $\mathcal{U}$ *such that* $\{\varphi, \varphi'\} = \{\mathcal{U}\psi, \mathcal{U}\psi'\}$.

Note that with this assumption, the definition of a 2-level system is somewhat redundant. Indeed, if every maximal set of perfectly distinguishable pure states can be obtained from one maximal set of perfectly distinguishable pure states made of two elements, all maximal sets of perfectly distinguishable pure states have two elements.

Remark 6.1.8. With this assumption, the proof that every diagonalization of the invariant state χ has $\frac{1}{2}$ eigenvalues is simpler. Indeed, let us consider a diagonalization of the invariant state $\chi = p\psi + (1-p)\psi'$. According to assumption 6.1.7, there is a reversible channel such that $\mathcal{U}\psi = \psi'$ and $\mathcal{U}\psi' = \psi$. Basically, $\mathcal{U}$ only permutes ψ and ψ'. If χ is invariant, we have

$$\chi = \mathcal{U}\chi = p\psi' + (1-p)\psi.$$

Now we have two diagonalizations of χ. Taking $(a_\psi|\chi)$ on both of them, where $(a_\psi|\psi) = 1$ and $(a_\psi|\psi') = 0$, one gets $p = 1-p$, whence $p = \frac{1}{2}$.

Now we prove an important result for the invariant state.

Proposition 6.1.9. *Let ψ and ψ' two perfectly distinguishable pure states. Then $\frac{1}{2}(\psi + \psi') = \chi$.*

Proof. Let us consider a diagonalization of χ. According to proposition 6.1.3, we have $\chi = \frac{1}{2}(\varphi + \varphi')$, where φ and φ' are perfectly distinguishable pure states. By assumption 6.1.7, there is a reversible channel such that $\mathcal{U}\varphi = \psi$ and $\mathcal{U}\varphi' = \psi'$. Then we have

$$\chi = \mathcal{U}\chi = \frac{1}{2}(\mathcal{U}\varphi + \mathcal{U}\varphi') = \frac{1}{2}(\psi + \psi').$$

□

In this way, we have proven that the only degenerate state is the invariant state χ and that all its diagonalizations have the same eigenvalues. For all the other mixed states ρ, p is uniquely defined as the maximum weight with which a pure state appears in a convex decomposition of ρ. So, their eigenvalues are uniquely-defined, at least in the procedure of diagonalization presented in section 5.2, as we noticed in the previous chapter. But what about different procedures of diagonalization? Are the eigenvalues always the same for a given state? The answer will be given in section 6.3.

Remark 6.1.10. The fact that χ can be diagonalized using every couple of perfectly distinguishable pure states, clearly shows that, in general, the diagonalization of a mixed state is not unique, even for a 2-level system.

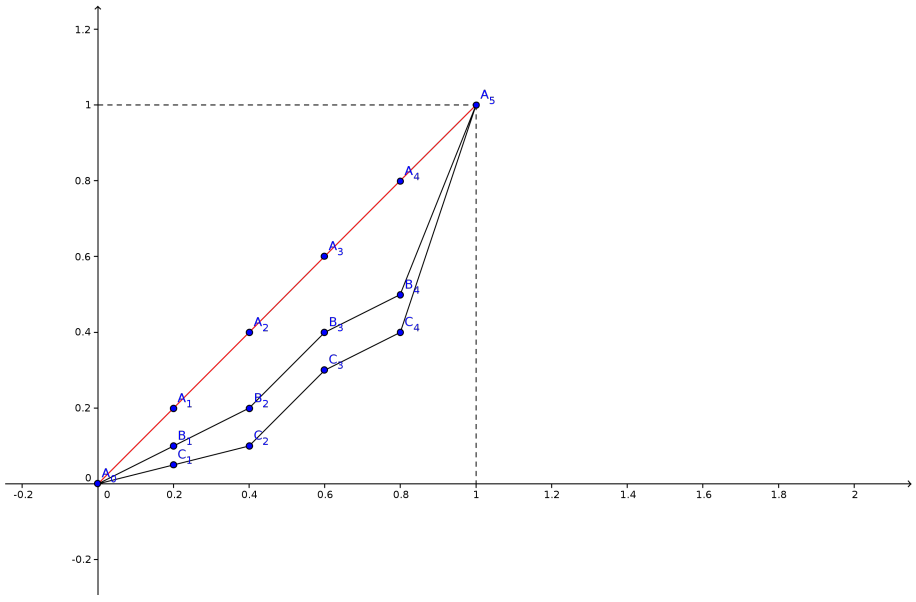
6.2 Majorization and its properties

Finally we introduce the main practical tool to have a better insight about the mixedness relation, taking advantage of the diagonalization procedure. This was essentially the main reason why we introduced and discussed diagonalization of mixed states. And this is also why it is so important that eigenvalues are well-defined, so that they completely characterize a given mixed state.

6.2.1 Heuristic introduction

We want to quantify how mixed a probability distribution is. Intuitively, a probability distribution is "more mixed" than another one if it is flatter, or in other terms, more equal. Some early studies about mixedness in statistics were in fact inspired by economic studies.

For instance, Lorenz [67] studied the concept of income inequality. Let us consider a population of n individual, and let x_i be the wealth of i-th individual. Let us order individuals from poorest to richest: in this way, 1 is the poorest individual, whereas n is the richest. Now we plot the points $\left(\frac{k}{n}, \frac{S_k}{S_n}\right)$, where $k = 0, \ldots, n$ and $S_k = \sum_{i=1}^{k} x_i$, with $S_0 = 0$. Hence S_k is the total wealth of the poorest k individuals. Finally, we join these points with a polygonal chain starting from $(0, 0)$ and ending in $(1, 1)$.



In this figure, we can see three different wealth distributions. The red line is a straight line and corresponds to the situation in which wealth is evenly distributed. The more a polygonal chain is bent in the middle, the more uneven is the distribution. The extreme case is when we have a polygonal chain made of the lower side of the $[0,1]\times[0,1]$ square, and of the left vertical side of the same square. This situation means that all the wealth is owned by a single individual.

Let us consider the B-line and the C-line. Let x_i be the wealth of each point in the B-line and let y_i be the wealth of each point in the C-line. Since the C-line is more bent than B-line, this means that[2]

$$\sum_{i=1}^{k} x_i \geq \sum_{i=1}^{k} y_i$$

and obviously $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$. Therefore we can see that an income distribution is more even (or more mixed) than another if the following two conditions are fulfilled.

- $\sum_{i=1}^{k} x_i \geq \sum_{i=1}^{k} y_i$ for every $k = 1, \ldots, n-1$
- $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$

6.2.2 Majorization

The concept of majorization captures the idea that a vector is more random than another.

We introduce a notation [68].

Notation. 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 $x_{[i]}$.

We define $x_\uparrow$ as the increasing rearrangement of x. We denote the i-th component of $x_\uparrow$ as $x_{(i)}$.

In this way, $x_{[i]} \geq x_{[i+1]}$ for every i, whereas $x_{(i)} \leq x_{(i+1)}$ for every i. With the heuristic motivation in mind, let us give the following definition.

Definition 6.2.1 (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

[2]Recall that x_i are in increasing order.

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

Our intention is to apply the definition of majorization to the eigenvalues of mixed states, therefore in this specific case the latter requirement is always fulfilled because we know that $\sum_{i=1}^{n} x_{[i]} = \sum_{i=1}^{n} y_{[i]} = 1$. Note that our procedure of diagonalization naturally yielded eigenvalues in decreasing order.

However, there is nothing special about decreasing order as opposed to the increasing order. Indeed, the majorization conditions in definition 6.2.1 are equivalent to

- $\sum_{i=1}^{k} x_{(i)} \geq \sum_{i=1}^{k} y_{(i)}$
- $\sum_{i=1}^{n} x_{(i)} = \sum_{i=1}^{n} y_{(i)}$

These two conditions are exactly the same we met at the end of our heuristic introduction in subsection 6.2.1.

To see the equivalence of these two conditions for majorization, first of all notice that $\sum_{i=1}^{n} x_{(i)} = \sum_{i=1}^{n} y_{(i)}$ and $\sum_{i=1}^{n} x_{[i]} = \sum_{i=1}^{n} y_{[i]}$ are exactly the same condition; in particular

$$\sum_{i=1}^{n} x_{[i]} = \sum_{i=1}^{k} x_{[i]} + \sum_{i=1}^{n-k} x_{(i)}, \tag{6.1}$$

and similarly for the sum with $y_{[i]}$. Suppose we know that $\sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]}$. Then, if we subtract $\sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]}$ from $\sum_{i=1}^{n} x_{[i]} = \sum_{i=1}^{n} y_{[i]}$, recalling eq. (6.1), we get $\sum_{i=1}^{n-k} x_{(i)} \geq \sum_{i=1}^{n-k} y_{(i)}$. Similarly one proves that the increasing-order conditions imply the decreasing-order conditions.

Thus, decreasing (or increasing) order has nothing special, but we prefer resorting to decreasing order when studying majorization.

Example 6.2.2. Let x be a vector of n non-negative real numbers such that

$$x = \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix},$$

and $\sum_{i=1}^{n} p_i = 1$. We will call this type of vectors *vectors of probabilities*. Clearly, when comparing two vectors of probabilities, we need not to check the condition $\sum_{i=1}^{n} x_{[i]} = \sum_{i=1}^{n} y_{[i]}$, because it is automatically satisfied, because $\sum_{i=1}^{n} x_{[i]} = \sum_{i=1}^{n} y_{[i]} = 1$. Without loss of generality, we can assume $p_i \geq p_{i+1}$ for every $i = 1, \ldots, n-1$. Then x is majorized by

$$y = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Indeed, $p_1 \leq 1$, otherwise it could not be $\sum_{i=1}^{n} p_i = 1$, for the p_i's are all non-negative. In addition,

$$\sum_{i=1}^{k} p_i \leq 1 + \underbrace{0 + \ldots + 0}_{k-1} = 1$$

for $k = 2, \ldots, n-1$. This shows that $x \preceq y$.

Example 6.2.3. Let x be a vector of n probabilities

$$x = \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix}.$$

Then x majorizes

$$y = \begin{pmatrix} \frac{1}{n} \\ \vdots \\ \frac{1}{n} \end{pmatrix}.$$

Indeed, it is well known (from the properties of arithmetic mean) that $\frac{k}{n} \leq \sum_{i=1}^{k} p_i$, for every $k < n$. This shows that $y \preceq x$, for every such x. This means that y is the minimum of vectors of probability according to majorization relation.

Since we are working with rearrangements of a vector, it will be useful to analyse this topic better. We will make use of permutations. Indeed, suppose we want to translate into mathematical language the fact that we are

exchanging the entries of a vector in $\mathbb{R}^n$ to order it. We need a representation of the symmetric group S_n.

Suppose we want to permute the entries of a vector x by the permutation $\pi \in S_n$. 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 Π 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 6.2.4. A square matrix Π 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.

Example 6.2.5. Let us consider the permutation matrix

$$\Pi = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

This matrix maps e_1 into e_3, e_2 into e_1 and e_3 into e_2. The associated permutation is then

$$\pi = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix},$$

namely the cycle $\begin{pmatrix} 1 & 3 & 2 \end{pmatrix}$.

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

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

6.2.3 Mathematical properties

Clearly, majorization is a binary relation on the set of vectors in $\mathbb{R}^n$. Let us analyse its properties.

Reflexive property It is obvious from the definition of majorization that $x \preceq x$ for every $x \in \mathbb{R}^n$.

We can say even something more. If we consider any rearrangement of x, namely Πx, where Π is a permutation matrix, then $x \preceq \Pi x$. This holds because, as far as majorization is concerned, only the values of the entries of x are important, not the order in which they appear.

Transitive property Suppose $x \preceq y$ and $y \preceq z$. Then we have $x \preceq z$, for every $x, y, z \in \mathbb{R}^n$. Indeed, if $x \preceq y$, then $\sum_{i=1}^k x_{[i]} \leq \sum_{i=1}^k y_{[i]}$ for every $k = 1, \ldots, n-1$ and $\sum_{i=1}^n x_{[i]} = \sum_{i=1}^n y_{[i]}$. If $y \preceq z$, then $\sum_{i=1}^k y_{[i]} \leq \sum_{i=1}^k z_{[i]}$ for every $k = 1, \ldots, n-1$ and $\sum_{i=1}^n y_{[i]} = \sum_{i=1}^n z_{[i]}$. Combining everything, one has $\sum_{i=1}^k x_{[i]} \leq \sum_{i=1}^k z_{[i]}$ for every $k = 1, \ldots, n-1$ and $\sum_{i=1}^n x_{[i]} = \sum_{i=1}^n z_{[i]}$.

Antisymmetric property Once more antisymmetric property fails. Suppose we know that $x \preceq y$ and $y \preceq x$. What can we conclude?

If $x \preceq y$, then $\sum_{i=1}^k x_{[i]} \leq \sum_{i=1}^k y_{[i]}$ for every $k = 1, \ldots, n-1$ and $\sum_{i=1}^n x_{[i]} = \sum_{i=1}^n y_{[i]}$. If the converse also holds, then $\sum_{i=1}^k y_{[i]} \leq \sum_{i=1}^k x_{[i]}$ for every $k = 1, \ldots, n-1$. This means that we actually have $\sum_{i=1}^k x_{[i]} = \sum_{i=1}^k y_{[i]}$ for every $k = 1, \ldots, n$ and this means that $x_{[i]} = y_{[i]}$ for every i. In other words, $x_\downarrow = y_\downarrow$, whence x and y differ only by a rearrangement of their entries. We conclude that $y = \Pi x$, for some permutation matrix Π.

Again, majorization is a genuine preorder,[3] therefore we can turn it into an order by taking the quotient with respect to vectors that differ by a permutation matrix, i.e. that differ only by a rearrangement of their entries. Indeed, the associated equivalence relation is $x \sim y$ if $x = \Pi y$. From an operative point of view, this can be achieved by rearranging each vector in decreasing order before comparing it to any other vector. In this way $\preceq$ becomes an order between sets of numbers and no more between vectors.

However, the order is not total, as the following example shows.

Example 6.2.6. Let us consider two vectors x and y in $\mathbb{R}^3$.

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

[3] It is not an equivalence relation, as shown for instance in example 6.2.2.

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 order is not total.

Majorization has a close relationship with doubly stochastic matrices, which we will use in the following. Fist of all, let us define what a doubly stochastic matrix is.

Definition 6.2.7. 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).

Clearly, the identity matrix and permutation matrices are trivial examples of doubly stochastic matrices. A less trivial example is the following.

Example 6.2.8. The $n \times n$ matrix

$$M = \begin{pmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & \ddots & \vdots \\ \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}$$

is a doubly stochastic matrix.

In order to prove some results about doubly stochastic matrices, we need an alternative and more compact characterization of them.

Lemma 6.2.9. *Let P be a square matrix of order n with non-negative entries. If $u \in \mathbb{R}^n$ is such that*

$$u = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix},$$

then P is doubly stochastic if and only if

$$Pu = u$$

and

$$u^t P = u^t,$$

where u^t is the transpose of u.

Proof. Let us see what the two conditions mean. $Pu = u$ means $\sum_{j=1}^{n} P_{ij}u_j = u_i$, but $u_i = u_j = 1$, so actually this relation reads $\sum_{j=1}^{n} P_{ij} = 1$. Similarly, $u^t P = u^t$ means $\sum_{i=1}^{n} u_i P_{ij} = u_j$, that is $\sum_{i=1}^{n} P_{ij} = 1$. □

Vector u plays also a special role in majorization. Indeed, if $x \preceq u$, then $x = u$. To see it, it is enough to recall example 6.2.3, where u plays the role of

$$\begin{pmatrix} \frac{1}{n} \\ \vdots \\ \frac{1}{n} \end{pmatrix},$$

and we showed that such vector is the minimum in majorization relation.

Thanks to lemma 6.2.9, we can easily show that the product of two doubly stochastic matrices is still a doubly stochastic matrix.

Lemma 6.2.10. *Let P_1 and P_2 be two doubly stochastic matrices. Then their product $P = P_1P_2$ is a doubly stochastic matrix.*

Proof. Clearly, $P = P_1P_2$ is a square matrix with non-negative elements. Then it is enough to check the two conditions of lemma 6.2.9. We have

$$Pu = P_1P_2u = P_1u = u,$$

where we used the fact that both P_1 and P_2 are doubly stochastic. Very similarly,

$$u^tP = u^tP_1P_2 = u^tP_2 = u^t.$$

Hence P is a doubly stochastic matrix. □

We will make use also of the following result by Birkhoff [70], which we will not prove. For a proof see [68].

Theorem 6.2.11 (Birkhoff)**.** *Doubly stochastic matrices are the convex hull of permutation matrices, where permutation matrices are the extreme points of it.*

This means that doubly stochastic matrices can be thought as performing a random permutation. Indeed, every doubly stochastic matrix P can be expressed as a convex combination of permutation matrices Π_i.

$$P = \sum_i \lambda_i \Pi_i$$

We are ready to state the equivalence between majorization and doubly stochastic matrices. The first thing we will do is to have a characterization of doubly stochastic matrices in terms of majorization.

Proposition 6.2.12. *A square matrix P of order n is doubly stochastic if and only if $Px \preceq x$ for every $x \in \mathbb{R}^n$.*

Proof. Sufficiency. Suppose $Px \preceq x$ for every $x \in \mathbb{R}^n$. In particular, $Pu \preceq u$, where u is defined as in lemma 6.2.9 As we noted above, it must be $Pu = u$ because u is the minimum in majorization relation.
Now let us take $x = e_j$, the j-th vector of the canonical basis for $\mathbb{R}^n$. We know that $Pe_j \preceq e_j$, which means

$$\begin{pmatrix} P_{1j} \\ \vdots \\ P_{nj} \end{pmatrix} \preceq e_j$$

In particular, this implies $\sum_{i=1}^n P_{ij} = 1$, namely $u^t P = u^t$. It also implies that $\min_i P_{ij} \geq 0$ for every j, so P has non-negative entries. We conclude that P is a doubly stochastic matrix.

Necessity. Suppose P is doubly stochastic. Let us consider $y = Px$. We can suppose, without loss of generality that x and y are arranged in decreasing order. If this is not the case, let us consider the permutation matrices Π and Σ such that Πx and Σy are arranged in decreasing order. Then, it is enough to replace P in the following argument by $\Sigma P \Pi^{-1}$ and x and y by Πx and Σy. $\Sigma P \Pi^{-1}$ is still a doubly stochastic matrix, because it is the product of doubly stochastic matrices. So, let us assume a decreasing order for x and y. We have $y_i = \sum_{j=1}^n P_{ij} x_j$. Summing over i ranging from 1 to $k < n$, we have

$$\sum_{i=1}^k y_i = \sum_{i=1}^k \sum_{j=1}^n P_{ij} x_j = \sum_{j=1}^n \sum_{i=1}^k P_{ij} x_j = \sum_{j=1}^n t_j x_j,$$

where we set

$$t_j := \sum_{i=1}^k P_{ij}.$$

Clearly, $0 \leq t_j \leq 1$ because P is doubly stochastic, and

$$\sum_{j=1}^n t_j = \sum_{j=1}^n \sum_{i=1}^k P_{ij} = \sum_{i=1}^k \sum_{j=1}^n P_{ij} = k.$$

Let us calculate $\sum_{i=1}^{k} y_i - \sum_{i=1}^{k} x_i$.

$$\sum_{i=1}^{k} y_i - \sum_{i=1}^{k} x_i = \sum_{i=1}^{n} t_i x_i - \sum_{i=1}^{k} x_i = \sum_{i=1}^{n} t_i x_i - \sum_{i=1}^{k} x_i + x_k \left(k - \sum_{i=1}^{n} t_i \right) =$$

$$= \sum_{i=1}^{k} (x_i - x_k)(t_i - 1) + \sum_{i=k+1}^{n} t_i (x_i - x_k).$$

In the first sum, $x_i - x_k \geq 0$ because x is arranged in decreasing order and $i \leq k$, whereas $t_i - 1 \leq 0$. Hence the first sum is non-positive. In the second sum, $t_i \geq 0$, whereas $x_i - x_k \leq 0$ because x is arranged in decreasing order, but now $i \geq k$. This shows that $\sum_{i=1}^{k} y_i - \sum_{i=1}^{k} x_i \leq 0$, that is $\sum_{i=1}^{k} y_i \leq \sum_{i=1}^{k} x_i$ for every $k < n$. Moreover, we have

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \sum_{j=1}^{n} P_{ij} x_j = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} P_{ij} \right) x_j = \sum_{j=1}^{n} x_j.$$

This proves that $y \preceq x$. □

Now, it is time to give a characterization of majorization using doubly stochastic matrices. This will be a key result which we will use several times henceforth. First of all, we need the following definition.

Definition 6.2.13. A *T-transform* is a doubly stochastic matrix defined as

$$T = \lambda \mathbf{1} + (1 - \lambda) Q,$$

where $\lambda \in [0, 1]$ and Q is a permutation matrix associated with a transposition.[4]

Let us see what the effect of a T-transform is. Suppose $x \in \mathbb{R}^n$; if Q exchanges i-th entry with j-th entry, then

$$T \begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_j \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ \lambda x_i + (1 - \lambda) x_j \\ \vdots \\ \lambda x_j + (1 - \lambda) x_i \\ \vdots \\ x_n \end{pmatrix}.$$

[4]Recall that a transposition is a permutation that exchanges only two elements of a set.

Essentially, the T-transform mixed the i-th with the j-th entry.

Remark 6.2.14. Note that a T-transform is a doubly stochastic matrix because it is a convex combination of two permutation matrices (see theorem 6.2.11).

Let us now see an example of a T-transform.

Example 6.2.15. Consider

$$P = \begin{pmatrix} \frac{3}{4} & 0 & \frac{1}{4} \\ 0 & 1 & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} \end{pmatrix}.$$

We can write it as

$$P = \frac{3}{4}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

P is clearly a T-transform, where $\lambda = \frac{3}{4}$ and

$$Q = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

is the permutation matrix associated with the transposition $\begin{pmatrix} 1 & 3 \end{pmatrix}$.

Now we need the following technical lemma.

Lemma 6.2.16. *If $x \preceq y$, then x can be derived from y by successive applications of a finite number of T-transforms.*

Proof. First of all, suppose that x can be obtained from y by permuting its entries by a permutation matrix Π. We know that the associated permutation can be decomposed as a product of a finite number transposition, namely Π is a product of transposition matrices. But transposition matrices are T-transforms (with $\lambda = 0$), so in this case we proved that x can be obtained from y by applying a finite number of T-transforms.

Now, let us suppose that x cannot be obtained from y by simply permuting its entries. Without loss of generality, we can suppose x and y are arranged in decreasing order. If this is not the case, it is sufficient to consider

the permutation that achieves decreasing order: every permutation matrix is a product of a finite number of T-transforms.
So, if x and y are arranged in decreasing order and $x \neq y$, let j be the largest index such that $x_j < y_j$ and let k be the smallest index greater than j such that $x_k > y_k$. Since $x \preceq y$, j and k must exist because the smallest index i for which $x_i \neq y_i$ must satisfy $x_i < y_i$, whereas the largest index such that $x_i \neq y_i$ must satisfy $x_i > y_i$. Clearly we have $y_j > x_j \geq x_k > y_k$ because x and y are arranged in decreasing order. Let us define

$$\delta := \min\{y_j - x_j, x_k - y_k\}$$

(note that $\delta > 0$),

$$1 - \lambda := \frac{\delta}{y_j - y_k}$$

(note that $0 < \lambda < 1$ because $\delta < y_j - y_k$) and

$$\left(y^{(1)}\right)^t := \begin{pmatrix} y_1 & \dots & y_{j-1} & y_j - \delta & y_{j+1} & \dots & y_{k-1} & y_k + \delta & y_{k+1} & \dots & y_n \end{pmatrix}.$$

After some passages we can rewrite this expression as

$$\left(y^{(1)}\right)^t = \lambda y^t + (1-\lambda) \begin{pmatrix} y_1 & \dots & y_{j-1} & y_k & y_{j+1} & \dots & y_{k-1} & y_j & y_{k+1} & \dots & y_n \end{pmatrix}.$$

Therefore $y^{(1)} = Ty$, where T is the T-transform $T = \lambda \mathbf{1} + (1-\lambda) Q$, where Q interchanges the j-th and the k-th entries. By proposition 6.2.12, $y^{(1)} \preceq y$. We have also $x \preceq y^{(1)}$, indeed

$$\sum_{i=1}^{l} x_i \leq \sum_{i=1}^{l} y_i = \sum_{i=1}^{l} y_i^{(1)}$$

for $l = 1, \dots, j-1$; and $x_j \leq y_j^{(1)}$ and $y_i^{(1)} = y_i$ for $i = j+1, \dots, k-1$; and

$$\sum_{i=1}^{l} x_i \leq \sum_{i=1}^{l} y_i = \sum_{i=1}^{l} y_i^{(1)}$$

for $l = k, \dots, n-1$ and finally

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} y_i^{(1)}.$$

Let $v, w \in \mathbb{R}^n$. Let us denote by $d(v, w)$ the number of non-vanishing entries of the vector $v - w$. If $\delta = y_j - x_j$, then $y_j^{(1)} = x_j$, whereas $y_k^{(1)} = x_k$ if $\delta = x_k - y_k$. Therefore $d\left(x, y^{(1)}\right) \leq d(x, y) - 1$. Therefore $y^{(1)}$ is "closer" than y to x, because it has more entries equal to the entries of x. Therefore, we went from y to $y^{(1)}$ by a T-transform , and still we have $x \preceq y^{(1)}$. Now we can iterate the procedure again, going from $y^{(1)}$ to $y^{(2)}$ (such that $x \preceq y^{(2)}$) by another T-transform, and $y^{(2)}$ will be even "closer" to x, and so on. At any step k, $d\left(x, y^{(k)}\right)$ is strictly decreasing and $x \preceq y^{(k)}$, so sooner or later we will obtain x, only by a composition of a finite number of T-transforms. □

From the proof of this lemma, it is apparent that at most $n - 1$ T-transforms will be necessary. Indeed, in the worst case, $d(x, y) = n$; this means that all the entries of x and y are different. Therefore $d\left(x, y^{(1)}\right) \leq n - 1$ and, in the worst case, $d\left(x, y^{(1)}\right) = n - 1$. So, after k iterations, in the worst case it will be $d\left(x, y^{(k)}\right) = n - k$. In particular, after $n - 1$ iterations, we have $d\left(x, y^{(n-1)}\right) \leq 1$. But, since $x \preceq y^{(n-1)}$, we cannot have $d\left(x, y^{(n-1)}\right) = 1$, otherwise it cannot be $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$. Hence, it must actually be $d\left(x, y^{(n-1)}\right) = 0$. Thus, we showed that we need at most $n - 1$ T-transforms.

Now we can prove the main result.

Theorem 6.2.17. *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. Sufficiency. Suppose we have $x = Py$. Then, by proposition 6.2.12, we have $Py \preceq y$, so $x \preceq y$.

Necessity. Suppose $x \preceq y$. By lemma 6.2.16, we know that we can obtain x from y by a product of a finite number of T-transforms. Since T-transforms are doubly stochastic, their product is still a doubly stochastic matrix. Therefore we can obtain x from y by applying a suitable doubly stochastic matrix P, $x = Py$. □

This means that if a vector x is majorized by y, then the former can be obtained from the latter by means of a random permutation. In this way, thinking again of income inequalities in subsection 6.2.1, we can regard Robin Hood's actions as doubly stochastic matrices, that increase "mixedness" by redistributing wealth among the individuals of the population.

Remark 6.2.18. Even if a product of T-transforms is a doubly stochastic matrix, this does not mean that every doubly stochastic matrix is a product

of T-transforms. However, this is surely true if the doubly stochastic matrix is in fact a permutation matrix, because every permutation is a product of transpositions, which are T-transforms with $\lambda = 0$.

Suppose $x \preceq y$, then we now that there exists a doubly stochastic matrix P such that $x = Py$. One might wonder if such doubly stochastic matrix P is unique. The following counter-example shows that it is not true in general.

Example 6.2.19. Consider

$$x = \begin{pmatrix} 4 \\ 3 \\ 2 \end{pmatrix} \qquad y = \begin{pmatrix} 5 \\ 3 \\ 1 \end{pmatrix}.$$

We can easily check that $x \preceq y$. We can choose as intertwining doubly stochastic matrix

$$P_1 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Indeed,

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 5 \\ 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \\ 2 \end{pmatrix}.$$

But we can also choose

$$P_2 = \begin{pmatrix} \frac{3}{4} & 0 & \frac{1}{4} \\ 0 & 1 & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} \end{pmatrix}$$

(which is a T-transform, see example 6.2.15), indeed

$$\begin{pmatrix} \frac{3}{4} & 0 & \frac{1}{4} \\ 0 & 1 & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} \end{pmatrix} \begin{pmatrix} 5 \\ 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \\ 2 \end{pmatrix}.$$

Actually, we can even take any convex combination of P_1 and P_2. Thus, the intertwining doubly stochastic matrix is not unique in general.

After all these results, we can move to a more physical situation in which we can apply our theory of majorization to analyse mixedness in a 2-level system.

6.3 Mixedness relation in a two-level system

In this section we will show that in a 2-level system, mixedness relation between mixed states is equivalent to majorization relation between the eigenvalues. In this chapter we deliberately choose to prove theorems in a more general way than it would be required for a 2-level system, because then proofs can be adapted also for a d-level system (see chapter 7).

Theorem 6.3.1. *Let ρ and σ be two mixed states of a 2-level system, such that ρ is more mixed than σ. Then, if x and y are respectively the vectors of the eigenvalues of diagonalizations of ρ and σ, we have $x \preceq y$.*

Proof. If ρ is more mixed than σ, we have $\rho = \sum_k \lambda_k \mathcal{U}_k \sigma$. Suppose $\rho = \sum_{j=1}^2 p_j \psi_j$ and $\sigma = \sum_{j=1}^2 q_j \varphi_j$. Then, $\rho = \sum_k \lambda_k \mathcal{U}_k \sigma$ becomes

$$\sum_{j=1}^{2} p_j \psi_j = \sum_k \lambda_k \sum_{j=1}^{2} q_j \mathcal{U}_k \varphi_j.$$

Let us consider the (pure) effect a_i such that $(a_i|\psi_j) = \delta_{ij}$. We get

$$p_i = \sum_{j=1}^{2} q_j \sum_k \lambda_k \left(a_i\right| \mathcal{U}_k \left|\varphi_j\right).$$

This expression can be rewritten as $p_i = \sum_{j=1}^2 P_{ij} q_j$, where

$$P_{ij} := \sum_k \lambda_k \left(a_i\right| \mathcal{U}_k \left|\varphi_j\right) =: \sum_k \lambda_k M_{k,ij},$$

and $M_{k,ij} = \left(a_i\right| \mathcal{U}_k \left|\varphi_j\right)$. We want to prove that $M_{k,ij}$'s are entries of a doubly stochastic matrix M_k for every k.

We have $M_{k,ij} \geq 0$ because $\left(a_i\right| \mathcal{U}_k \left|\varphi_j\right) \in [0,1]$. If we compute $\sum_{i=1}^2 M_{k,ij}$, we get

$$\sum_{i=1}^{2} \left(a_i\right| \mathcal{U}_k \left|\varphi_j\right) = \left(\sum_{i=1}^{2} a_i \right| \mathcal{U}_k \left|\varphi_j\right) = \left(e\right| \mathcal{U}_k \left|\varphi_j\right) = \\ = (e|\varphi_j) = 1,$$

where we used the fact that $\{a_i\}_{i=1}^2$ is an observation-test and that $\mathcal{U}_k$ is a (reversible) channel. We have to prove also that $\sum_{j=1}^2 M_{k,ij} = 1$.

$$\sum_{j=1}^{2} \left(a_i\right| \mathcal{U}_k \left|\varphi_j\right) = \left(a_i\right| \mathcal{U}_k \left| \sum_{j=1}^{2} \varphi_j \right)$$

Now, $\sum_{j=1}^{2} \varphi_j = 2\chi$ (see proposition 6.1.9), whence

$$\sum_{j=1}^{2} (a_i| \mathcal{U}_k |\varphi_j) = 2\,(a_i| \mathcal{U}_k |\chi) = 2\,(a_i|\chi)\,,$$

because χ is invariant. Now, $\chi = \frac{1}{2}(\psi_1 + \psi_2)$, therefore $(a_i|\chi) = \frac{1}{2}$. Finally we get $\sum_{j=1}^{2} M_{k,ij} = 2 \cdot \frac{1}{2} = 1$.

This proves that M_k is doubly stochastic. By theorem 6.2.11, the set of doubly stochastic matrices is convex, therefore $P = \sum_k \lambda_k M_k$ is a doubly stochastic matrix. In this way, we have that $x = Py$. By theorem 6.2.17, this means that $x \preceq y$. □

Remark 6.3.2. Note that to prove this theorem we actually exploited some results about 2-level systems, namely that the invariant state can be diagonalized with respect to every maximal set of perfectly distinguishable pure states.

A straightforward consequence of this theorem is that eigenvalues of a mixed state do not depend on the procedure of diagonalization.

Corollary 6.3.3. *Let ρ be a mixed state of a 2-level system. Then all diagonalizations of ρ have the same eigenvalues.*

Proof. Let $\rho = p\psi + (1-p)\,\psi'$ and $\rho = q\varphi + (1-q)\,\varphi'$ be two diagonalizations of ρ, and let x and y be the vectors of the eigenvalues associated with the two diagonalizations. We know that ρ is more mixed than ρ. This implies $x \preceq y$, but also $y \preceq x$, therefore $x = \Pi y$, for some permutation matrix Π. This means that x and y differ only by a rearrangement of their entries, whence the eigenvalues of ρ are uniquely defined. □

Thus, we are entitled to talk about the eigenvalues of a state and not of a specific diagonalization.

Loosely speaking, theorem 6.3.1 states that if ρ is more mixed than σ, then its eigenvalues are "flatter" than the ones of σ.

In a 2-level system, majorization is particularly simple: since the sum of eigenvalues of mixed states is always 1 and we have only two eigenvalues, we actually need to check only one condition about the largest eigenvalue. Indeed,

$$\begin{pmatrix} p \\ 1-p \end{pmatrix} \preceq \begin{pmatrix} q \\ 1-q \end{pmatrix}$$

if and only if $p \leq q$. It is then clear that in this case majorization is a total order, because either $p \leq q$, or $q \leq p$.

We would like to prove also the converse of theorem 6.3.1, in this way we would set forth the equivalence between mixedness relation and majorization.

Theorem 6.3.4. *Let $\rho = \sum_{i=1}^{2} p_i \psi_i$ and $\sigma = \sum_{i=1}^{2} q_i \varphi_i$ be diagonalizations of the mixed states ρ and σ. Let x and y respectively be the vectors of their eigenvalues. If $x \preceq y$, then ρ is more mixed than σ.*

Proof. If $x \preceq y$, then, by theorem 6.2.17, one has $x = Py$ for some doubly stochastic matrix P. Now, by Birkhoff's theorem, $P = \sum_k \lambda_k \Pi_k$, where Π_k's are permutation matrices, therefore $x = \sum_k \lambda_k \Pi_k y$. In particular, this means that $p_i = \sum_k \lambda_k \sum_{j=1}^{2} \Pi_{k,ij} q_j$. Therefore we have

$$\rho = \sum_{i=1}^{2} p_i \psi_i = \sum_{i=1}^{2} \sum_k \lambda_k \sum_{j=1}^{2} \Pi_{k,ij} q_j \psi_i = \sum_k \lambda_k \sum_{j=1}^{2} q_j \sum_{i=1}^{2} \Pi_{k,ij} \psi_i.$$

Our goal is to prove that $\rho = \sum_k \lambda_k \mathcal{U}_k \sigma$, namely

$$\rho = \sum_k \lambda_k \sum_j q_j \mathcal{U}_k \varphi_j.$$

Therefore it is enough to prove that $\sum_{i=1}^{2} \Pi_{k,ij} \psi_i$ can be written as $\mathcal{U}_k \varphi_j$. If we recall the definition of a permutation matrix, then there exists a permutation $\pi_k \in S_2$ such that $\Pi_{k,ij} = \delta_{i,\pi_k(j)}$. In this way,

$$\sum_{i=1}^{2} \Pi_{k,ij} \psi_i = \sum_{i=1}^{2} \delta_{i,\pi_k(j)} \psi_i = \psi_{\pi_k(j)}.$$

Therefore, $\psi_{\pi_k(j)}$ can be obtained by a permutation of the pure states ψ_i's. According to assumption 6.1.7, we can obtain the ψ_i's from the φ_j's by applying a suitable reversible channel. Hence, for every k there is a reversible channel $\mathcal{U}_k$ such that $\psi_{\pi_k(j)} = \mathcal{U}_k \varphi_j$. In this way we manage to write

$$\rho = \sum_k \lambda_k \sum_j q_j \mathcal{U}_k \varphi_j.$$

This proves that ρ is more mixed than σ. □

Remark 6.3.5. Again, to prove this theorem, we actually exploited some results about 2-level systems. In this case we used assumption 6.1.7.

In this way, we proved the equivalence between mixedness relation and majorization.

Therefore, the fact that a mixed state is more mixed than another one depends only on the eigenvalues of the two states and not on the pure states that appear in the diagonalization. In other words, mixedness depends only only the way pure states are mixed together to get a mixed state, and not on the specific pure states involved.

Since mixedness relation is completely equivalent to majorization, it inherits the properties of majorization. In particular, since majorization is a total order for a 2-level system, mixedness relation is a total order on the set of states of a 2-level system. Again, this is a consequence of the "rigidity" of 2-level systems.

Example 6.3.6. We can see, in another way, that every state is more mixed than a pure state. Indeed, we can associate a vector

$$x = \begin{pmatrix} p \\ 1-p \end{pmatrix}$$

with $\frac{1}{2} \leq p \leq 1$ with every state, whereas we can associate the vector

$$y = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

with every pure state. Clearly $x \preceq y$ because $p \leq 1$.

Example 6.3.7. Now we can prove that the invariant state is not only a maximal element, but actually the maximum according to the mixedness relation. Indeed, the vector of eigenvalues of χ is

$$y = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix},$$

whereas, as usual, the vector of eigenvalues of a generic state is

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

with $\frac{1}{2} \leq p \leq 1$. Then $y \preceq x$ because $p \geq \frac{1}{2}$. This shows that the invariant state is the most mixed state of all states.

As we can see, majorization is a practical tool to see if a state is more mixed than another one, since it involves numbers rather than abstract entities like states.

In subsection 4.3.1 we had an open issue, namely to characterize the equivalence relation "to be as mixed as". Now we will answer this open question.

Proposition 6.3.8. *In a 2-level system, ρ is as mixed as σ if and only if $\rho = \mathcal{U}\sigma$, for some reversible channel $\mathcal{U}$.*

Proof. We already proved sufficiency in subsection 4.3.1.

Let us move to necessity. If ρ is as mixed as σ, then $x \preceq y$ and $y \preceq x$, where x and y are the vectors of the eigenvalues associated with ρ and σ. According to the results of subsection 6.2.3, this happens if and only if x and y differ by a permutation matrix, namely if ρ and σ have the same eigenvalues. Thus, $\rho = p\psi + (1-p)\,\psi'$ and $\sigma = p\varphi + (1-p)\,\varphi'$. By assumption 6.1.7, there exists a reversible channel $\mathcal{U}$ such that $\mathcal{U}\varphi = \psi$ and $\mathcal{U}\varphi' = \psi'$. Therefore,

$$\mathcal{U}\sigma = p\mathcal{U}\varphi + (1-p)\,\mathcal{U}\varphi' = p\psi + (1-p)\,\psi' = \rho.$$

□

In particular we see that two equally mixed states have the same eigenvalues.

6.4 Schur-concave functions

In the previous sections we saw how majorization can simplify our analysis of mixedness of states. However, majorization requires performing some checks on the set of the eigenvalues. We would like to have a more immediate tool to say when a state is more mixed than another one by simply assigning a number to states and comparing those numbers. Clearly, this new tool must be completely equivalent to majorization.

The first notion we need is that of order-preserving functions.

Definition 6.4.1. Let $(A, \leq)$ be a (pre)ordered subset of $\mathbb{R}^n$. A function $f : A \longrightarrow \mathbb{R}$ is said to be *order-preserving* or *isotonic* if $x \leq y$ implies $f(x) \leq f(y)$.

In this definition f associates a real number to each vector. Our goal is to find some functions that preserve the order given by majorization. This is the basic requirement. In this way, the order between vectors is translated into the natural order between real numbers. However, we must note that, the natural order on real numbers is total, whereas majorization gives rise only to a partial order (unless we are in $\mathbb{R}^2$). This discrepancy in the characters of these two order relations will cause some troubles in our search of such functions.

Definition 6.4.2. A real-valued function $f : A \longrightarrow \mathbb{R}^n$ that preserves majorization is called *Schur-convex function*. In other words, a function is Schur-convex if $x \preceq y$ implies $f(x) \leq f(y)$, for every $x, y \in A$.

A function f is called *Schur-concave* if $-f$ is Schur-convex.

Schur-concave functions reverse the order given by majorization, namely if $x \preceq y$, then $f(x) \geq f(y)$.

We can give an equivalent characterization of Schur-convex functions using doubly stochastic matrices, according to theorem 6.2.17.

Proposition 6.4.3. *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 $x \in \mathbb{R}^n$.

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 $x \in \mathbb{R}^n$.

Proof. By theorem 6.2.17, $Px \preceq x$, for every doubly stochastic matrix P. Therefore f is Schur-convex if and only if $f(Px) \leq f(x)$. Similarly one proves the statement for Schur-concave functions. □

Since we have in mind to apply these new tools to vectors of eigenvalues, we will assume that the set A, which will be the set of vectors of eigenvalues, is *symmetric*, namely if $x \in A$, then also $\Pi x \in A$, for every permutation matrix Π. Indeed, every vector of eigenvalues can be rearranged and the resulting vector is still a vector of eigenvalues.

Remark 6.4.4. If a function is Schur-convex on A, then $f(\Pi x) = f(x)$. Indeed, $x \preceq \Pi x \preceq x$,[5] then $f(x) \leq f(\Pi x) \leq f(x)$ and this chain of inequalities implies $f(x) = f(\Pi x)$; we say that such a function is *symmetric*. This means that every Schur-convex function is actually defined on equivalence classes given by the equivalence relation associated with majorization, namely $x \sim y$ if $x = \Pi y$. Thinking of mixed states, this means that Schur-convex functions take the same value on equally mixed states, which have the same eigenvalues.

We have seen that in a 2-level system ρ is *more* mixed than σ if and only if $x \preceq y$, so, loosely speaking, in some sense if x is *less* than y. In this way, we see that majorization relation is in the "opposite" direction of mixedness relation. We are instead looking for a practical tool that tells us immediately when a state is more mixed than another. In this vein, we want some function such that $f(x) \geq f(y)$ if ρ is more mixed than σ, namely if $x \preceq y$. Hence we are looking for a Schur-concave function.

When we want to prove that some function is Schur-concave, it is sometimes useful to note that we can consider $\mathbb{R}^2$ instead of $\mathbb{R}^n$. This is a consequence of lemma 6.2.16. Indeed, suppose $x \preceq y$, where $x, y \in \mathbb{R}^n$. If x and y differ by more than two entries, we can apply some T-transforms in order to make them differ only by two entries. Recall that x and y cannot differ by less than two entries, unless $x = \Pi y$ for some permutation matrix Π. Therefore, it is enough to prove that $f(x) \geq f(y)$ when x and y differ by only two entries, namely when only two arguments of f are free. Therefore, since f is symmetric (see remark 6.4.4), and so there is no privileged argument, it is sufficient to prove that $f(x_1, x_2, x_3, \ldots, x_n)$ is Schur-concave in x_1 and x_2, or, in other words, when its domain is a subset of $\mathbb{R}^2$.

Let us introduce a sufficient condition for Schur-concavity which we will use later.

Proposition 6.4.5. *If f is symmetric and concave, then it is Schur-concave.*

Proof. According to what we have said above, it is enough to prove the statement in $\mathbb{R}^2$. Suppose $x \preceq y$, then, by theorem 6.2.17, $x = Py$, where P is a doubly stochastic matrix. By Birkhoff theorem, P is a convex combination of permutation matrices. In $\mathbb{R}^2$, there are only two permutation matrices

$$\mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad Q = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

[5]Note that this does not imply that $\Pi x = x$, since $\preceq$ is a preorder and not an order.

So,

$$P = \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + (1-\lambda) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \lambda & 1-\lambda \\ 1-\lambda & \lambda \end{pmatrix},$$

where $\lambda \in [0,1]$. Then, we have

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \lambda & 1-\lambda \\ 1-\lambda & \lambda \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix},$$

namely

$$\begin{cases} x_1 = \lambda y_1 + (1-\lambda) y_2 \\ x_2 = (1-\lambda) y_1 + \lambda y_2 \end{cases}.$$

Since f is concave,

$$f(x_1, x_2) = f(\lambda y_1 + (1-\lambda) y_2, (1-\lambda) y_1 + \lambda y_2) \geq \lambda f(y_1, y_2) + (1-\lambda) f(y_2, y_1).$$

Since f is symmetric, $f(y_1, y_2) = f(y_2, y_1)$, so $\lambda f(y_1, y_2) + (1-\lambda) f(y_2, y_1) = f(y_1, y_2)$. We proved that $f(x_1, x_2) \geq f(y_1, y_2)$, which means that f is Schur-concave. □

Actually, we want a complete equivalence between the order induced by f and majorization. Requiring that f is Schur-concave means that if $x \preceq y$, then $f(x) \geq f(y)$. Now we require also that $f(x) \geq f(y)$ implies $x \preceq y$, and here the troubles come. Clearly, if $f(x) \geq f(y)$ for every Schur-concave function, then $x \preceq y$. Indeed, the majorization conditions in the increasing order form are all Schur-concave functions by definition. We would like to restrict ourselves only to a proper subset of Schur-concave functions. We will see that it is enough to check the condition $f(x) \geq f(y)$ only for a particular class of Schur-concave functions.

Corollary 6.4.6. *Let I be an interval and let $g : I \longrightarrow \mathbb{R}$ be a concave function. Then $f : I^n \longrightarrow \mathbb{R}$ defined as*

$$f(x) = \sum_{i=1}^{n} g(x_i)$$

is Schur-concave.

Proof. Such an $f(x)$ is clearly symmetric and concave, so proposition 6.4.5 applies. □

Definition 6.4.7. A Schur-concave function $f : I^n \longrightarrow \mathbb{R}$ is said to be *concave* and *separate-variable* if $f(x) = \sum_{i=1}^{n} g(x_i)$, where $g : I \longrightarrow \mathbb{R}$ is a concave function.

Note that the summation ranges up to n.

Example 6.4.8. Let us consider a subset A of $\mathbb{R}^n$, where every vector is a vector of probabilities.

$$x = \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix}$$

Let us consider *Shannon entropy* H, defined as

$$H(x) := -\sum_{i=1}^{n} p_i \log_a p_i,$$

where $a > 1$ and we set $0 \log_a 0 = 0$. Here $g = -x \log_a x$ is concave, because its second derivative is always negative. Therefore H is a concave and separate-variable Schur-concave function.

In particular, we have that

$$H(1, \ldots, 0) \leq H(p_1, \ldots, p_n) \leq H\left(\frac{1}{n}, \ldots, \frac{1}{n}\right), \tag{6.2}$$

because in examples 6.2.2 and 6.2.3 we proved that

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

In particular, eq. (6.2) implies $H(x) \leq -n \cdot \frac{1}{n} \log_a \frac{1}{n}$, namely

$$0 \leq H \leq \log_a n.$$

The class of concave and separate-variable Schur-concave functions is enough to give an equivalent characterization of majorization.

Theorem 6.4.9. *Let $x, y \in \mathbb{R}^n$. Then $x \preceq y$ if and only if $f(x) \geq f(y)$, for every continuous concave and separate-variable Schur-concave function f.*

Proof. Necessity is trivial, because it directly follows from corollary 6.4.6.

Sufficiency. Suppose $f(x) = \sum_{i=1}^{n} g(x_i)$, and that $\sum_{i=1}^{n} g(x_i) \geq \sum_{i=1}^{n} g(y_i)$. Let us take, for fixed $k = 1, \ldots, n$, $g(z) = \min\{z - y_{(k)}, 0\}$. Then we have

$$\sum_{i=1}^{k} g\left(y_{(i)}\right) = \sum_{i=1}^{k} \left(y_{(i)} - y_{(k)}\right) = \sum_{i=1}^{k} y_{(i)} - k y_{(k)}$$

because $y_{(i)} \leq y_{(k)}$ if $i \leq k$, and we have

$$\sum_{i=k+1}^{n} g\left(y_{(i)}\right) = 0.$$

Combining these two sums, we have

$$\sum_{i=1}^{n} g\left(y_{(i)}\right) = \sum_{i=1}^{k} y_{(i)} - k y_{(k)}.$$

Since $g(z)$ is continuous and concave, by hypothesis we know that

$$\sum_{i=1}^{n} g\left(x_{(i)}\right) \geq \sum_{i=1}^{n} g\left(y_{(i)}\right).$$

We also know that $g(z) \leq 0$ and $g(z) \leq z - y_{(k)}$, whence

$$\sum_{i=1}^{k} y_{(i)} - k y_{(k)} = \sum_{i=1}^{n} g\left(x_{(i)}\right) \leq \sum_{i=1}^{k} g\left(x_{(i)}\right) \leq \sum_{i=1}^{k} \left(x_{(i)} - y_{(k)}\right) =$$

$$= \sum_{i=1}^{k} x_{(i)} - k y_{(k)}.$$

Therefore we conclude that

$$\sum_{i=1}^{k} x_{(i)} \geq \sum_{i=1}^{k} y_{(i)} \tag{6.3}$$

for every $k = 1, \ldots, n$. We must show that if $k = n$ equality holds. To do that, let us take $g(z) = -z$, which is continuous and concave. In this way, by hypothesis, we have $-\sum_{i=1}^{n} x_i \geq -\sum_{i=1}^{n} y_i$, which means $\sum_{i=1}^{n} x_i \leq \sum_{i=1}^{n} y_i$. Therefore, recalling that (6.3) implies $\sum_{i=1}^{n} x_{(i)} \geq \sum_{i=1}^{n} y_{(i)}$, we have $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$. This proves that $x \preceq y$. □

6.4.1 Rényi entropies

In this subsection we present a widely used one-parameter family of Schur-concave functions, *Rényi entropies.* They were introduced by Rényi [75] as a generalization of Shannon entropy. Rényi entropies $H_\alpha(x)$ are defined as

$$H_\alpha(x) := \frac{1}{1-\alpha}\log_a \sum_{i=1}^n x_i^\alpha,$$

where x is a vector of probabilities, $\alpha \geq 0$ and $\alpha \neq 1$ and $a > 1$, with the convention that $0^\alpha = 0$ for every such α.

It is interesting to study some particular cases and limits.

- $\alpha = 0$ In this case, $H_0(x)$ is the logarithm of the number of non-vanishing entries of x.

- $\alpha \to 1$ We want to study the limit $\lim_{\alpha\to 1} H_\alpha(x)$. In this case we have a $\frac{0}{0}$ indeterminate form. Using de l'Hôpital rule, we have

$$\lim_{\alpha\to 1} H_\alpha(x) = \lim_{\alpha\to 1} \frac{1}{1-\alpha}\log_a \sum_{i=1}^n x_i^\alpha \overset{\text{H}}{=} -\lim_{\alpha\to 1} \frac{\sum_{i=1}^n x_i^\alpha \log_a x_i}{\sum_{j=1}^n x_j^\alpha} =$$

$$= -\sum_{i=1}^n x_i \log_a x_i,$$

  which is Shannon entropy.

- $\alpha \to +\infty$ We want to study the limit $\lim_{\alpha\to+\infty} H_\alpha$. It is useful to recall the definition of α-norm ($\alpha \geq 1$) of a vector.

$$\|x\|_\alpha = \left(\sum_{i=1}^n |x_i|^\alpha\right)^{\frac{1}{\alpha}} \tag{6.4}$$

  In particular, $\|x\|_\infty = \lim_{\alpha\to+\infty}\|x\|_\alpha = \max_i |x_i|$. Therefore

$$H_\alpha(x) = \frac{\alpha}{1-\alpha}\log_a \|x\|_\alpha,$$

  and

$$H_\infty(x) := \lim_{\alpha\to+\infty} H_\alpha(x) = -\log_a \|x\|_\infty = -\log_a \max_i x_i.$$

  This is called *min-entropy.*

Rényi entropies satisfy some interesting properties. First of all, they are Schur-concave functions.

Proposition 6.4.10. *$H_\alpha(x)$ is Schur-concave for every $\alpha \geq 0$.*

Proof. We already know that the thesis holds for $\alpha = 1$, because we have Shannon entropy, which is Schur-concave. First of all it is useful to rewrite Rényi entropies (with $\alpha \neq 1$) as

$$H_\alpha(x) = \frac{\alpha}{1-\alpha} \log_a \|x\|_\alpha ,$$

even[6] for $0 \leq \alpha < 1$. Note that $\|x\|_\alpha$ is a symmetric function of x, namely its value does not change if we permute the entries of x, therefore $H_\alpha(x)$ is a symmetric function of x. If we manage to prove that $H_\alpha(x)$ is also concave, we are done, according to proposition 6.4.5. We must distinguish two cases.

- Suppose $0 \leq \alpha < 1$ In this case $\frac{\alpha}{1-\alpha} \log_a$ is an increasing function, and $\|x\|_\alpha$ is concave; therefore $\frac{\alpha}{1-\alpha} \log_a \|x\|_\alpha$ is concave.
- Suppose $\alpha > 1$. In this case $\frac{\alpha}{1-\alpha} \log_a$ is a decreasing function, and $\|x\|_\alpha$ is convex because it is a norm; therefore $\frac{\alpha}{1-\alpha} \log_a \|x\|_\alpha$ is concave.

We managed to prove that $H_\alpha(x)$ is always concave, therefore it is also Schur-concave. □

Now we want to study the relationship between Rényi entropies with different α. We need the following lemma first.

Lemma 6.4.11. *Let $\{p_i\}_{i=1}^n$ and $\{q_i\}_{i=1}^n$ be two sets of probabilities. Then the quantity*

$$H(p_i \parallel q_i) := -\sum_{i=1}^{n} p_i \log_a \frac{q_i}{p_i}, \tag{6.5}$$

where $a > 1$ and $0 \log_a 0 = 0$, is always non-negative and vanishes if and only if $p_i = q_i$ for every i.

[6] We are in fact extending the notation of eq. (6.4) also for $0 \leq \alpha < 1$. However, in this case, $\|\bullet\|_\alpha$ is no more a norm because triangle inequality fails.

Proof. The fundamental inequality for the natural logarithm says that $\ln x \leq x - 1$ and equality holds if and only if $x = 1$. Therefore

$$\log_a x \leq \frac{1}{\ln a}(x-1).$$

Applying this inequality to eq. (6.5), we obtain

$$H(p_i \parallel q_i) = -\sum_{i=1}^{n} p_i \log_a \frac{q_i}{p_i} \geq \frac{1}{\ln a} \sum_{i=1}^{n} p_i \left(1 - \frac{q_i}{p_i}\right) =$$

$$= \frac{1}{\ln a} \sum_{i=1}^{n} (p_i - q_i) = 0.$$

Therefore $H(p_i \parallel q_i) \geq 0$.

We have equality instead of inequality if and only if $\frac{q_i}{p_i} = 1$ for every i, namely if and only if $p_i = q_i$ for every i. □

Actually, $H(p_i \parallel q_i)$ is not just a formal means to prove some statements. It is interesting on its own, because in classical probability theory it is useful to measure the "distance" between two probability distributions. $H(p_i \parallel q_i)$ is called *relative entropy* of $\{p_i\}$ to $\{q_i\}$.

Proposition 6.4.12. *The family of Rényi entropies is decreasing in α.*

Proof. The family of Rényi entropies is differentiable in α. Therefore, let us compute $\frac{\partial H_\alpha}{\partial \alpha}$.

$$\frac{\partial H_\alpha}{\partial \alpha} = \frac{1}{(1-\alpha)^2} \log_a \sum_{i=1}^{n} x_i^\alpha + \frac{1}{1-\alpha} \frac{\sum_{i=1}^{n} x_i^\alpha \log_a x_i}{\sum_{j=1}^{n} x_j^\alpha}.$$

Let us set

$$z_i := \frac{x_i^\alpha}{\sum_{j=1}^{n} x_j^\alpha}.$$

Note that $z_i \geq 0$ and $\sum_{i=1}^{n} z_i = 1$, so the z_i's can be thought as probabilities. Then

$$\frac{\partial H_\alpha}{\partial \alpha} = \frac{1}{(1-\alpha)^2} \left[(1-\alpha) \sum_{i=1}^{n} z_i \log_a x_i + \log_a \sum_{i=1}^{n} x_i^\alpha\right] =$$

$$= \frac{1}{(1-\alpha)^2} \left[\sum_{i=1}^{n} z_i \log_a x_i - \alpha \sum_{i=1}^{n} z_i \log_a x_i + \log_a \sum_{i=1}^{n} x_i^\alpha\right].$$

Let us evaluate $-\alpha \sum_{i=1}^{n} z_i \log_a x_i + \log_a \sum_{i=1}^{n} x_i^\alpha$. Using the properties of logarithms,

$$-\alpha \sum_{i=1}^{n} z_i \log_a x_i + \log_a \sum_{i=1}^{n} x_i^\alpha = -\sum_{i=1}^{n} z_i \log_a x_i^\alpha + \log_a \sum_{i=1}^{n} x_i^\alpha$$

We replace x_i^α in the first logarithm with $z_i \sum_{j=1}^{n} x_j^\alpha$. Then

$$-\sum_{i=1}^{n} z_i \log_a x_i^\alpha = -\sum_{i=1}^{n} z_i \log_a \left(z_i \sum_{j=1}^{n} x_j^\alpha \right) = -\sum_{i=1}^{n} \left(z_i \log_a z_i + z_i \log_a \sum_{j=1}^{n} x_j^\alpha \right) =$$

$$= -\log_a \sum_{i=1}^{n} x_i^\alpha - \sum_{i=1}^{n} z_i \log_a z_i,$$

because the z_i's sum to 1. Then

$$-\alpha \sum_{i=1}^{n} z_i \log_a x_i + \log_a \sum_{i=1}^{n} x_i^\alpha = -\sum_{i=1}^{n} z_i \log_a z_i.$$

Plugging this result into the expression of $\frac{\partial H_\alpha}{\partial \alpha}$, we have

$$\frac{\partial H_\alpha}{\partial \alpha} = -\frac{1}{(1-\alpha)^2} \left(-\sum_{i=1}^{n} z_i \log_a \frac{x_i}{z_i} \right).$$

By lemma 6.4.11, the round bracket is always non-negative, therefore $\frac{\partial H_\alpha}{\partial \alpha}$ is non-positive, whence the entropies are decreasing in α. □

This means that if $\alpha_1 \leq \alpha_2$ then $H_{\alpha_1}(x) \geq H_{\alpha_2}(x)$. In particular,

$$H_0(x) \geq H_1(x) \geq \ldots \geq H_\infty(x).$$

Now, it is clear why $H_\infty(x)$ is called min-entropy: it takes the smallest value among all Rényi entropies.

The proof of proposition 6.4.12 gives us also some information about when all Rényi entropies are equal.

Corollary 6.4.13. *All Rényi entropies of x are equal if and only if all the non-vanishing entries of x are equal.*

Proof. The entropies are equal, i.e. constant in α, if and only if $\frac{\partial H_\alpha}{\partial \alpha} = 0$. Recalling the proof of proposition 6.4.12 and lemma 6.4.11, this happens if and only if $x_i = z_i$ for every i. This means that

$$x_i = \frac{x_i^\alpha}{\sum_{j=1}^n x_j^\alpha} \tag{6.6}$$

for every i. This equation is trivially solved if $x_i = 0$, for some i. Note that not all entries of x vanish, otherwise x would not be a vector of probabilities. Therefore, at least one entry is non-vanishing.

If only one entry is non-vanishing, then it must be 1 and 1 solves eq. (6.6).

Suppose there are more than one non-vanishing entries. Let us take two of them. Without loss of generality, we can take them to be for $i = 1$ and for $i = 2$. Eq. (6.6) reads

$$\begin{cases} x_1 = \frac{x_1^\alpha}{\sum_{j=1}^n x_j^\alpha} \\ x_2 = \frac{x_2^\alpha}{\sum_{j=1}^n x_j^\alpha} \end{cases} .$$

Taking the ratio between these two equations, one gets

$$\frac{x_1}{x_2} = \left(\frac{x_1}{x_2}\right)^\alpha ,$$

whose non vanishing solution is $\frac{x_1}{x_2} = 1$, namely $x_1 = x_2$.

Repeating this procedure, with x_{i_0} in place of x_2, where x_{i_0} is any other non-vanishing entry of x, one gets $x_{i_0} = x_1$. This means that all the non-vanishing entries of x are equal. □

This corollary highlights some important facts about Rényi entropies.

- If $x = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}$, this is a trivial case when all the non-vanishing entries are equal. Therefore all the Rényi entropies are equal and to know what $H_\alpha(x)$ is, it is enough to compute only one of them, say[7] H_0.
$$H_\alpha(x) = H_0(x) = \log_a 1 = 0$$
- $x = \begin{pmatrix} \frac{1}{k} & \dots & \frac{1}{k} & 0 & \dots & 0 \end{pmatrix}$, where we have k non-vanishing entries, is another case when all Rényi entropies are equal. Therefore
$$H_\alpha(x) = H_0(x) = \log_a k$$

[7] Alternatively, we already did the calculation for Shannon entropy in example 6.4.8.

- If $x = \begin{pmatrix} \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix}$, again all Rényi entropies are equal.

$$H_\alpha(x) = H_0(x) = \log_a n$$

In particular, recalling that $x = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}$ is the maximum in majorization relation, whereas $x = \begin{pmatrix} \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix}$ is the minimum, one has

$$0 \leq H_\alpha(x) \leq \log_a n$$

for every $\alpha \geq 0$ and for every x.

One might wonder if this family of functions gives a complete characterization of majorization. In other terms, if $H_\alpha(x) \geq H_\alpha(y)$ for every $\alpha \geq 0$, can we conclude that $x \preceq y$? The answer is negative, as the following counterexample shows [78].

Example 6.4.14. In this example we will show that there exist two vectors of probabilities in $\mathbb{R}^8$ such that neither $x \preceq y$, nor $y \preceq x$, but $H_\alpha(x) > H_\alpha(y)$ for every $\alpha \geq 0$. As we can see, the pathology arises because majorization is not a total order.

Consider

$$x = \begin{pmatrix} \frac{2}{9} & \frac{2}{9} & \frac{2}{9} & \frac{2}{9} & \frac{1}{36} & \frac{1}{36} & \frac{1}{36} & \frac{1}{36} \end{pmatrix}$$
$$y = \begin{pmatrix} \frac{4}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & 0 & 0 \end{pmatrix}$$

Notice that $\frac{2}{9} \leq \frac{4}{9}$, so $y \not\preceq x$ and $\frac{2}{9}+\frac{2}{9}+\frac{2}{9}+\frac{2}{9} \geq \frac{4}{9}+\frac{1}{9}+\frac{1}{9}+\frac{1}{9}$ because $\frac{8}{9} \geq \frac{7}{9}$, so $x \not\preceq y$. Therefore, neither $x \preceq y$, nor $y \preceq x$.

Let us see what we can say about Rényi entropies. We have, for instance,

$$H_0(x) = \log_a 8 \qquad H_0(y) = \log_a 6$$

$$H_1(x) = \log_a 9 - \frac{1}{3}\log_a 4 \qquad H_1(y) = \log_a 9 - \frac{4}{9}\log_a 4$$

$$H_\infty(x) = \log_a 9 - \log_a 2 \qquad H_\infty(y) = \log_a 9 - \log_a 4$$

Therefore we have $H_\alpha(x) > H_\alpha(y)$ at least for $\alpha = 0, 1, +\infty$. Let us show that $H_\alpha(x) > H_\alpha(y)$ for every $\alpha \geq 0$, $\alpha \neq 1$. If $H_\alpha(x) < H_\alpha(y)$ for some α, then, by continuity in α, there exists α_0 such that $H_{\alpha_0}(x) = H_{\alpha_0}(y)$. Therefore, to prove that $H_\alpha(x) > H_\alpha(y)$ for every α, it is enough to prove that $H_\alpha(x) = H_\alpha(y)$ is impossible. Now,

$$H_\alpha(x) = \frac{1}{1-\alpha}\log_a \left[4\left(\frac{2}{9}\right)^\alpha + 4\left(\frac{1}{36}\right)^\alpha\right]$$

and

$$H_\alpha(y) = \frac{1}{1-\alpha}\log_a\left[\left(\frac{4}{9}\right)^\alpha + 5\left(\frac{1}{9}\right)^\alpha\right].$$

Let us try to solve the equation $H_\alpha(x) = H_\alpha(y)$ for $\alpha \geq 0$ and $\alpha \neq 1$.

$$4\left(\frac{2}{9}\right)^\alpha + 4\left(\frac{1}{36}\right)^\alpha = \left(\frac{4}{9}\right)^\alpha + 5\left(\frac{1}{9}\right)^\alpha$$

$$4 \cdot 8^\alpha + 4 = 16^\alpha + 5 \cdot 4^\alpha$$

We set $x = 2^\alpha$, then the equation becomes

$$x^4 - 4x^3 + 5x^2 - 4 = 0.$$

Factorizing the left-hand side, one gets

$$(x-2)\left[x(x-1)^2 + 2\right] = 0$$

We have a solution for $x = 2$, which means $\alpha = 1$, but we are assuming $\alpha \neq 1$. The other factor is, instead, always positive for $x > 0$, as it is our case. Therefore the equation $H_\alpha(x) = H_\alpha(y)$ has no solution. We then conclude that $H_\alpha(x) > H_\alpha(y)$ for every $\alpha \geq 0$.

Therefore we have that all Rényi entropies evaluated on x are greater than the corresponding entropies evaluated on y; yet x and y are "incomparable" in the majorization relation.

Rényi entropies enjoy also another property: *additivity*.

Definition 6.4.15. Given two vectors $x, y \in \mathbb{R}^n$, we define their *dyadic product* $x \otimes y$ as $(x \otimes y)_{ij} := x_i y_j$.

We want to see how Rényi entropies behave on dyadic products.

Proposition 6.4.16 (Additivity)**.** *We have* $H_\alpha(x \otimes y) = H_\alpha(x) + H_\alpha(y)$ *for every* $x, y \in \mathbb{R}^n$ *and for every* $\alpha \geq 0$.

Proof. Let us compute $H_\alpha(x \otimes y)$ explicitly.

$$H_\alpha(x \otimes y) = \frac{1}{1-\alpha}\log_a \sum_{i,j=1}^n (x_i y_j)^\alpha = \frac{1}{1-\alpha}\log_a \left(\sum_{i=1}^n x_i^\alpha\right)\left(\sum_{j=1}^n y_j^\alpha\right) =$$

$$= \frac{1}{1-\alpha}\log_a \sum_{i=1}^n x_i^\alpha + \frac{1}{1-\alpha}\log_a \sum_{j=1}^n y_j^\alpha = H_\alpha(x) + H_\alpha(y)$$

□

6.5 Schur-concave functions for a two-level system

In this section we want to apply the formalism of Schur-concave functions to a 2-level system, to relate mixedness relation to inequalities between real numbers. In a 2-level system, all is simpler because majorization is a total order on vectors of probabilities in $\mathbb{R}^2$. Indeed, we have that $x \preceq y$ if and only if $p \leq q$, where $p = \max x_i$ and $q = \max y_i$.

Therefore, in a 2-level system, we have $x \preceq y$ if and only if $f(x) \geq f(y)$, where f is some decreasing function of the maximum of a vector.[8]

Such a function is clearly Schur-concave (even in $\mathbb{R}^n$!). Indeed, if $x \preceq y$, then $\max_i x_i \leq \max_i y_i$; therefore $f(\max_i x_i) \geq f(\max_i y_i)$ if f is a decreasing function.

Therefore in $\mathbb{R}^2$ one function is enough to infer majorization between vectors of probabilities. For instance, we can use min-entropy, which is indeed a decreasing function of the maximum of a vector. Min-entropy $H_\infty(x)$ is enough to give a complete characterization of mixedness for a 2-level system.

Instead of referring to the min-entropy as the min-entropy of the vector x of the eigenvalues of ρ, we prefer calling it the min-entropy of ρ and writing it as

$$S_\infty(\rho) := H_\infty(x),$$

in order to comply with literature. For a 2-level system, we can say that ρ is more mixed than σ if and only if $S_\infty(\rho) \geq S_\infty(\sigma)$.

What about all the other Schur-concave functions? The answer comes from the following proposition.

Proposition 6.5.1. *Let ρ and σ be two mixed states of a 2-level system, and let f be a Schur-concave function. Then ρ is more mixed than σ if and only if $f(x) \geq f(y)$, where x and y are the vectors of the eigenvalues of ρ and σ respectively.*

Proof. Necessity is immediate, and follows from the definition of Schur-concave function. Indeed, if ρ is more mixed than σ, then $x \preceq y$. Since f is Schur-concave, we conclude that $f(x) \geq f(y)$.

Let us move to sufficiency. Suppose we know that $f(x) \geq f(y)$. Then, by definition of Schur-concave function, we know that $y \not\preceq x$. Since the

[8] Alternatively, we can choose an increasing function of the minimum of a vector.

majorization order is total for a 2-level system, we conclude that $x \preceq y$, and therefore ρ is more mixed than σ. □

This proves that the choice of min-entropy, has nothing special, although it was the simplest one. We can choose whatever Schur-concave function we want, for example any of the Rényi entropies.

As done with min-entropy, we slightly change the notation to explicitly show that Schur-concave functions must be thought as functions of the mixed state itself. Therefore we will write

$$f(\rho) := f(x),$$

where x is the vector of the eigenvalues of ρ. Actually, when f is a Rényi entropy, we prefer writing $S_\alpha(\rho)$, instead of $H_\alpha(\rho)$, to better comply with common use in literature.

We conclude this section with a theorem that sums up all the results we have got so far for a 2-level system.

Theorem 6.5.2. *The following statements are equivalent in a 2-level system.*

1. *ρ is more mixed than σ.*
2. *$x \preceq y$, where x and y are the vectors of the eigenvalues of ρ and σ respectively.*
3. *$f(\rho) \geq f(\sigma)$ for some Schur-concave function f.*

We then see that for a 2-level system, it is not necessary to check the conditions of theorem 6.4.9, namely to consider all concave and separate-variable Schur-concave functions. It is indeed enough to consider only one Schur-concave function, regardless of the fact that it may be separate-variable or not.

Chapter 7

General measures of mixedness

In this chapter we extend and generalize the results obtained for a 2-level system to a d-level system, with d arbitrary. Here, the overall picture is richer and more complicated, because majorization is no more a total order.

We will choose a particular Schur-concave function, Shannon entropy, which is concave and separate-variable, and we will study its properties. These property will be virtually identical to the well-known properties of quantum von Neumann entropy, such as subadditivity or triangle inequality, but here we will derive them starting from our procedure of diagonalization. Therefore their validity is more general than the mere field of ordinary quantum mechanics.

Thanks to these properties, we will prove an inequality for Shannon entropy that strongly resembles a second law of thermodynamics. This inequality was already known in quantum mechanics, but here we prove it in a more general framework.

7.1 A d-level system

Now we want to generalize the results concerning 2-level systems for a generic d-level system, following the conceptual scheme of the previous chapter.

Definition 7.1.1. A *d-level system* is a system where all maximal sets of perfectly distinguishable states have d elements.

d is called the *dimension* of the system.

Again, we will focus on maximal sets of perfectly distinguishable *pure* states.

Clearly, in a d-level system, every diagonalization of a mixed state has at most d terms.

As we did in section 6.1, we will make an assumption very close to assumption 6.1.7.

Assumption 7.1.2. *Given two maximal sets of perfectly distinguishable pure states $\{\psi_i\}_{i=1}^d$ and $\{\varphi_i\}_{i=1}^d$, there exists a reversible channel $\mathcal{U}$ such that $\varphi_i = \mathcal{U}\psi_i$ for every i.*

Thanks to this assumption, we can prove several results. Again, it is immediately clear that all maximal sets of perfectly distinguishable pure states have the same cardinality, because they all come from one set of perfectly distinguishable pure states by applying reversible channels.

Proposition 7.1.3. *Every diagonalization of the invariant state $\chi = \sum_{i=1}^d p_i\psi_i$ has $p_i = \frac{1}{d}$, for every i.*

Proof. Suppose $\{a_i\}$ is the perfectly distinguishing test, then $p_i = (a_i|\chi)$. Let us consider all the possible permutations of the pure states $\{\psi_i\}$. For instance, if $\pi \in S_d$, we can consider the permuted states $\{\psi_{\pi(i)}\}$, which are obviously still perfectly distinguishable. By assumption 7.1.2, there is a reversible channel $\mathcal{U}_\pi$ that achieves this permutation, namely $\psi_{\pi(i)} = \mathcal{U}_\pi\psi_i$. Let us apply $\mathcal{U}_\pi$ to χ.

$$\chi = \sum_{j=1}^d p_j \mathcal{U}_\pi\psi_j = \sum_{j=1}^d p_j\psi_{\pi(j)}$$

Now let us apply a_i to χ. We have $(a_i|\chi) = \sum_{j=1}^d p_j\delta_{i,\pi(j)}$, whence $p_i = p_{\pi^{-1}(i)}$. Since this holds for every $\pi \in S_d$, one has $p_i = p_j$ for every j. This implies that the weights are equal, therefore $p_i = \frac{1}{d}$. □

Note that, in this case, we had to make assumption 7.1.2 to prove that the eigenvalues of the invariant state are all equal. Instead, for a 2-level system, we managed to prove it without any additional assumption. Again, this is essentially due to the fact that dimension" 2 is somehow "rigid".

In particular, proposition 7.1.3 implies that the pure states that appear in every diagonalization of the invariant state χ form a maximal set of perfectly distinguishable pure states because there are exactly d perfectly distinguishable terms. One can wonder about the converse: is it true that if we take any set with d perfectly distinguishable pure states, and we consider their convex combination with equal weights $\frac{1}{d}$, we obtain the invariant state?

Proposition 7.1.4. *Let $\{\psi_i\}_{i=1}^d$ be a maximal set of perfectly distinguishable pure states. Then $\frac{1}{d}\sum_{i=1}^d \psi_i$ is the invariant state χ.*

Proof. Let us consider a diagonalization of χ, say $\chi = \frac{1}{d}\sum_{i=1}^d \varphi_i$. Here, $\{\varphi_i\}_{i=1}^d$ is a maximal set of perfectly distinguishable pure states. By assumption 7.1.2, there is a reversible channel $\mathcal{U}$ such that $\mathcal{U}\varphi_i = \psi_i$ for every i. Then we have

$$\chi = \mathcal{U}\chi = \frac{1}{d}\sum_{i=1}^d \mathcal{U}\varphi_i = \frac{1}{d}\sum_{i=1}^d \psi_i.$$

□

Following a line of reasoning very close to the proof of lemma 6.1.4, one can prove that every maximal set of perfectly distinguishable pure state can be distinguished by a perfectly distinguishing pure test.

7.2 Majorization in a d-level system

Now we have all the ingredients to state the complete equivalence between majorization and mixedness even for a d-level system, with $d > 2$.

Recalling the analogous result for a 2-level system, this is a direct consequence of the assumption that we can obtain all maximal sets of perfectly distinguishable pure states starting from a single one by applying suitable reversible channels.

Theorem 7.2.1. *Let ρ and σ be two mixed states of a d-level system, and let x and y be the vectors of the eigenvalues of two diagonalizations of ρ and σ respectively. The following statements are equivalent.*

1. *ρ is more mixed than σ*
2. *$x \preceq y$*

Proof. The proof is completely identical to the proof of theorems 6.3.1 and 6.3.4, provided we replace 2 with d. □

Once more, mixedness depends only on the way pure states are mixed together and not on what the particular pure states are.

We also have the analogous of corollary 6.3.3 and the proof is completely identical.

Corollary 7.2.2. *Let ρ be a mixed state of a d-level system. Then all diagonalizations of ρ have the same eigenvalues.*

Now we are completely entitled to talk about eigenvalues of a mixed state, regardless of a specific diagonalization. Note that we can regard the vector of the eigenvalues of ρ as a probability distribution.

If $d > 2$, the mixedness relation is not a total order, because even majorization is not a total order, as we showed in example 6.2.6.

Recalling examples 6.2.2 and 6.2.3, we have another way to see that every mixed state is more mixed than a pure state, because

$$\begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_d \end{pmatrix} \preceq \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Moreover, we can say that the invariant state is not only maximal, but actually the maximum in mixedness relation. Indeed,

$$\begin{pmatrix} \frac{1}{d} \\ \vdots \\ \frac{1}{d} \end{pmatrix} \preceq \begin{pmatrix} p_1 \\ \vdots \\ p_d \end{pmatrix}.$$

In particular, this tells us that every purification of the invariant state is more entangled than any other bipartite pure state, according to the equivalence between mixedness and entanglement we established in section 4.4.

Now we can give a final answer to the open question about when a mixed state is as mixed as another one.

Proposition 7.2.3. *In a d-level system, ρ is as mixed as σ if and only if $\rho = \mathcal{U}\sigma$, for some reversible channel $\mathcal{U}$.*

Proof. We already proved sufficiency in subsection 4.3.1.

Let us move to necessity. If ρ is as mixed as σ, then $x \preceq y$ and $y \preceq x$, where x and y are the vectors of eigenvalues associated with ρ and σ. According to results of subsection 6.2.3, this happens if and only if x and y differ by a permutation matrix, namely if ρ and σ have the same eigenvalues. Thus, $\rho = \sum_{i=1}^{d} p_i \psi_i$ and $\sigma = \sum_{i=1}^{d} p_i \varphi_i$. By assumption 7.1.2, there exists a

reversible channel $\mathcal{U}$ such that $\mathcal{U}\varphi_i = \psi_i$ for every i. Therefore,

$$\mathcal{U}\sigma = \sum_{i=1}^{d} p_i \mathcal{U}\varphi_i = \sum_{i=1}^{d} p_i \psi_i = \rho.$$

□

We can see that equally mixed states have the same eigenvalues.

This result is also an answer to the equivalent question about entangled pure states. We then have that a bipartite pure state $|\psi)_{\mathrm{AB}}$ is as entangled as $|\phi)_{\mathrm{AB}}$ if and only if $|\psi)_{\mathrm{AB}} = \mathcal{U}\,|\phi)_{\mathrm{AB}}$, for some local reversible channel (on A or on B) $\mathcal{U}$.

7.3 Shannon entropy for d-level systems

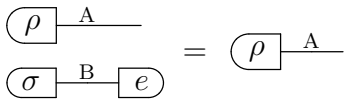
Once we have established the complete equivalence between mixedness and majorization, we can try to apply the formalism of Schur-concave functions to a generic d-level system. However, if $d > 2$ then majorization order is not total. Therefore we cannot use only one Schur-concave function to infer mixedness. If f is a Schur-concave function and we have $f(\rho) \geq f(\sigma)$, we can say that $y \not\preceq x$, so σ is not more mixed than ρ. But we cannot conclude that ρ is more mixed than σ because the order is not total, maybe ρ and σ are not "comparable". Therefore we have to use theorem 6.4.9 to state a complete equivalence with mixedness relation.

Therefore, in this section we decide to abandon the proposal of studying general Schur-concave functions, but we focus on Shannon entropy. In particular, we want to analyse and characterize the relationship between Shannon entropy and bipartite states. Namely, if we know the Shannon entropy of a bipartite state, what can we say about the Shannon entropy of its two marginals?

Intuitively, according to the well known results in quantum mechanics [16, 17, 80], one would expect that Shannon entropy will be subadditive even in general probabilistic theories, but it will be additive if and only if the bipartite state is a product state.

Let us begin our treatment from the simplest case, that is from product states. Suppose we have two systems, A and B that are d-level systems, but possibly with different d, i.e. $d_{\mathrm{A}} \neq d_{\mathrm{B}}$. Suppose we know that system AB

is in the product state $\rho_{\mathrm{A}} \otimes \sigma_{\mathrm{B}}$. Clearly the two marginals are ρ_{A} and σ_{B} respectively, as the following diagram shows for the marginal on system A.



The problem is now how to characterize the eigenvalues of $\rho \otimes \sigma$ in terms of the eigenvalues of ρ and σ. Let us first deal with the trivial case when everything is pure. If ρ and σ are both pure, then $\rho \otimes \sigma$ is pure by pure conditioning. There is nothing to prove, for $S(\rho \otimes \sigma) = 0$ and $S(\rho) + S(\sigma) = 0 + 0 = 0$, because everything is pure.

Let us then focus on the non-trivial situation. Suppose we have the two diagonalizations $\rho = \sum_{i=1}^{d_{\mathrm{A}}} p_i \alpha_i$ and $\sigma = \sum_{j=1}^{d_{\mathrm{B}}} q_j \beta_j$. Suppose α_i's are perfectly distinguishable by the observation-test[1] $\{a_k\}$, whereas β_j's are perfectly distinguishable by the observation-test $\{b_l\}$. Then

$$\rho \otimes \sigma = \sum_{i=1}^{d_{\mathrm{A}}} \sum_{j=1}^{d_{\mathrm{B}}} p_i q_j \alpha_i \otimes \beta_j.$$

Clearly, the pure states $\alpha_i \otimes \beta_j$ are perfectly distinguishable by the observation-test $\{a_k \otimes b_l\}$. Let us first show that $\{a_k \otimes b_l\}$ is an observation-test.

$$\sum_{k,l} a_k \otimes b_l = \sum_k a_k \otimes \sum_l b_l = e_{\mathrm{A}} \otimes e_{\mathrm{B}} = e_{\mathrm{AB}}$$

Now, let us show that it is perfectly distinguishing.

$$(a_k|_{\mathrm{A}} (b_l|_{\mathrm{B}} |\alpha_i)_{\mathrm{A}} |\beta_j)_{\mathrm{B}} = \delta_{ik}\delta_{jl} = \delta_{(i,j),(k,l)}.$$

Therefore, $\rho \otimes \sigma = \sum_{i=1}^{d_{\mathrm{A}}} \sum_{j=1}^{d_{\mathrm{B}}} p_i q_j \alpha_i \otimes \beta_j$ is really a diagonalization of $\rho \otimes \sigma$ and the vector of the eigenvalues of this diagonalization is the dyadic product of the two vectors of the eigenvalues of ρ and σ.

Now we make the following rather weak and reasonable assumption.

Assumption 7.3.1. *The composition of two d-level systems with dimension d_{A} and d_{B} is still a system where all maximal sets of perfectly distinguishable states have the same cardinality.*

[1]Note that if some p_i vanishes, then a_k's are fewer than d_{A}, and this is the reason why we do not specify the range of k. The same holds also for b_l's.

This assumptions says nothing about the dimension of the composite system. With this assumption we know that in AB the eigenvalues do not depend on the specific diagonalization. We can then say that the eigenvalues of $\rho\otimes\sigma$ are p_iq_j so the vector of the eigenvalues of $\rho\otimes\sigma$ is exactly the dyadic product of the vectors of the eigenvalues of ρ and σ. We know that Rényi entropies are additive, so, in particular,

$$H(x\otimes y)=H(x)+H(y),$$

which means that

$$S(\rho\otimes\sigma)=S(\rho)+S(\sigma).$$

Actually, this holds for every Rényi entropy, as we proved in proposition 6.4.16.

Proposition 7.3.2. *If we have a product state $\rho_{\mathrm{A}}\otimes\sigma_{\mathrm{B}}$, then $S_\alpha(\rho_{\mathrm{A}}\otimes\sigma_{\mathrm{B}})=S_\alpha(\rho_{\mathrm{A}})+S_\alpha(\sigma_{\mathrm{B}})$ for every $\alpha\geq 0$.*

As we noted in the previous section, we can regard the eigenvalues of a mixed state as a probability distribution. We consider the probability distribution associated with a state of AB as the joint probability distribution of the ones associated with its marginals. If we have a product state, then the probability distributions of its two marginals are independent: indeed, $\widetilde{p}_{ij}=p_iq_j$, where $\widetilde{p}_{ij}$'s are the eigenvalues of $\rho\otimes\sigma$. In other words, we do not have correlations between the two marginals.

What happens if $|\rho)_{\mathrm{AB}}$ is not a product state?

First of all, we want to define some observables that are functions of a state. Let ρ be diagonalized as $\rho=\sum_{i=1}^{d}p_i\psi_i$. If $f:[0,1]\longrightarrow\mathbb{R}$ is a function, in general, we can define, relative to the given diagonalization,

$$(f(\rho)|:=\sum_{i=1}^{d}f(p_i)(a_i|,$$

where $(a_i|$'s are the effects that make up the perfectly distinguishing test for the pure states ψ_i's. Clearly, p_i is the weight naturally associated with a_i, namely the weight with which ψ_i appears in the diagonalization.

We can exploit this new formalism to express Shannon entropy in a new form.

Example 7.3.3. Let us consider Shannon entropy. Recall that $S(\rho) = -\sum_{i=1}^{d} p_i \log_a p_i$. Now, let us compute

$$(-\log_a \rho|\rho) = -\sum_{i=1}^{d} \log_a p_i \, (a_i| \sum_{j=1}^{d} p_j \, |\psi_j) =$$

$$= -\sum_{i,j=1}^{d} p_j \log_a p_i \delta_{ij} = -\sum_{i=1}^{d} p_i \log_a p_i.$$

This is precisely Shannon entropy $S(\rho)$. Therefore $S(\rho) = -(\log_a \rho|\rho)$.

Now we want to define relative entropy.

Definition 7.3.4. Let ρ and σ be two normalized states. The *relative entropy* of ρ to σ is

$$S(\rho \parallel \sigma) := (\log_a \rho - \log_a \sigma|\rho),$$

where $a > 1$.

Thanks to this definition and to the following lemma, we can prove some further important results about Shannon entropy. The proofs will be virtually identical to the ones in [17].

Lemma 7.3.5 (Klein's inequality)**.** *Let ρ and σ be two normalized states. One has $S(\rho \parallel \sigma) \geq 0$ and $S(\rho \parallel \sigma)$ vanishes if and only if $\rho = \sigma$.*

Proof. Let $\rho = \sum_{i=1}^{d} p_i \psi_i$ and $\sigma = \sum_{i=1}^{d} q_i \varphi_i$ be[2] diagonalizations of ρ and σ respectively. Let a_i and b_i be pure effects such that $(a_i|\psi_j) = \delta_{ij}$ and $(b_i|\varphi_j) = \delta_{ij}$. Recall that a_i's and b_i's are pure because we are considering a maximal set of perfectly distinguishable pure states. Now, let us compute $S(\rho \parallel \sigma)$ explicitly. In particular, as we showed above,

$$(\log_a \rho|\rho) = \sum_{i=1}^{d} p_i \log_a p_i,$$

and

$$(\log_a \sigma|\rho) = \sum_{j=1}^{d} \log_a q_j \, (b_j| \sum_{i=1}^{d} p_i \, |\psi_i) = \sum_{i,j=1}^{d} (b_j|\psi_i) \, p_i \log_a q_j.$$

[2]Here, for the sake of simplicity, we are tacitly assuming that all p_i's and q_i's are non-vanishing, but the proof can be easily fitted to the general case.

Then

$$S(\rho \parallel \sigma) = \sum_{i=1}^{d} p_i \left(\log_a p_i - \sum_{j=1}^{d} P_{ij} \log_a q_j \right),$$

where we set $P_{ij} := (b_j|\psi_i)$. P_{ij}'s are the entries of a doubly stochastic matrix. Indeed $P_{ij} \geq 0$ because $(b_j|\psi_i) \in [0,1]$. Then

$$\sum_{j=1}^{d} P_{ij} = \sum_{j=1}^{d} (b_j|\psi_i) = (e|\psi_i) = 1;$$

and, finally,

$$\sum_{i=1}^{d} P_{ij} = \sum_{i=1}^{d} (b_j|\psi_i) = d\,(b_j|\chi) = d \cdot \frac{1}{d} = 1,$$

where we used the fact that the invariant state χ can be diagonalized with respect to every maximal set of perfectly distinguishable pure states. Since the logarithm is concave, $\sum_{j=1}^{d} P_{ij} \log_a q_j \leq \log_a \sum_{j=1}^{d} P_{ij} q_j$ and one has equality if and only if there is a j such that $P_{ij} = 1$. In this way,

$$S(\rho \parallel \sigma) \geq \sum_{i=1}^{d} p_i \left(\log_a p_i - \log_a \sum_{j=1}^{d} P_{ij} q_j \right) = -\sum_{i=1}^{d} p_i \log_a \frac{r_i}{p_i},$$

where we set $r_i := \sum_{j=1}^{d} P_{ij} q_j$. Since $r_i \geq 0$ and $\sum_{i=1}^{d} r_i = 1$ because P is doubly stochastic, by lemma 6.4.11, the right-hand side is always non negative, therefore $S(\rho \parallel \sigma) \geq 0$. Moreover, $S(\rho \parallel \sigma) = -\sum_{i=1}^{d} p_i \log_a \frac{r_i}{p_i}$ if and only if there is a j such that $P_{ij} = 1$, namely P_{ij}'s are the entries of a permutation matrix. In this case, $-\sum_{i=1}^{d} p_i \log_a \frac{r_i}{p_i}$ vanishes if and only if $r_i = p_i$ for every i. Relabelling the pure states that appear the diagonalization of σ, one can take P_{ij} to be δ_{ij} (recall that P is now a permutation matrix). This means that $q_i = p_i$ for every i, and $P_{ij} = (b_j|\psi_i) = \delta_{ij}$. This implies that $\psi_i = \varphi_i$ for every i by proposition 5.1.9. Then $\sigma = \sum_{i=1}^{d} p_i \psi_i$. Therefore we conclude that $S(\rho \parallel \sigma) = 0$ if and only if $\rho = \sigma$. □

Now we can apply this result to prove subadditivity of Shannon entropy, namely that $S(\rho_{\mathrm{AB}}) \leq S(\rho_{\mathrm{A}}) + S(\rho_{\mathrm{B}})$, where ρ_{AB} is a generic bipartite state and ρ_{A} and ρ_{B} are its two marginals on system A and B respectively.

Proposition 7.3.6 (Subadditivity)**.** *Shannon entropy is* subadditive*, namely*

$$S(\rho_{\mathrm{AB}}) \leq S(\rho_{\mathrm{A}}) + S(\rho_{\mathrm{B}}),$$

and it is additive if and only if ρ_{AB} is a product state.

Proof. Let us apply Klein's inequality when $\rho = \rho_{\mathrm{AB}}$ and $\sigma = \rho_{\mathrm{A}} \otimes \rho_{\mathrm{B}}$. We have

$$-(\log_a \rho|\rho)_{\mathrm{AB}} \leq -(\log_a \sigma|\rho)_{\mathrm{AB}}.$$

The left-hand side is Shannon entropy of ρ, so we have $S(\rho) \leq -(\log_a \sigma|\rho)_{\mathrm{AB}}$. Let us characterize $-(\log_a \sigma|\rho)_{\mathrm{AB}}$. Recalling that $\sigma = \rho_{\mathrm{A}} \otimes \rho_{\mathrm{B}}$, we have

$$(\log_a \sigma|_{\mathrm{AB}} = \sum_{i,j=1}^{d} \log_a p_i q_j \, (a_i|_{\mathrm{A}} \, (b_j|_{\mathrm{B}}, \tag{7.1}$$

where $\rho_{\mathrm{A}} = \sum_{i=1}^{d} p_i \alpha_i$ and $\rho_{\mathrm{B}} = \sum_{j=1}^{d} q_j \beta_j$ are diagonalizations of ρ_{A} and ρ_{B} respectively, with perfectly distinguishing tests $\{a_i\}$ and $\{b_j\}$. Now, let us rewrite eq. (7.1) as

$$\sum_{i,j=1}^{d} \log_a p_i q_j \, (a_i|_{\mathrm{A}} \, (b_j|_{\mathrm{B}} = \sum_{i,j=1}^{d} \left(\log_a p_i \, (a_i|_{\mathrm{A}} \, (b_j|_{\mathrm{B}} + \log_a q_j \, (a_i|_{\mathrm{A}} \, (b_j|_{\mathrm{B}}\right) =$$

$$= \sum_{i=1}^{d} \log_a p_i \, (a_i|_{\mathrm{A}} \left(\sum_{j=1}^{d} b_j \right|_{\mathrm{B}} + \sum_{j=1}^{d} \log_a q_j \left(\sum_{i=1}^{d} a_i \right|_{\mathrm{A}} (b_j|_{\mathrm{B}} =$$

$$= \sum_{i=1}^{d} \log_a p_i \, (a_i|_{\mathrm{A}} \, (e|_{\mathrm{B}} + \sum_{j=1}^{d} \log_a q_j \, (e|_{\mathrm{A}} \, (b_j|_{\mathrm{B}}.$$

Now we are ready to compute $-(\log_a \sigma|\rho)_{\mathrm{AB}}$.

$$-(\log_a \sigma|\rho)_{\mathrm{AB}} = -\sum_{i=1}^{d} \log_a p_i \, (a_i|_{\mathrm{A}} \, (e|_{\mathrm{B}} \, |\rho)_{\mathrm{AB}} - \sum_{j=1}^{d} \log_a q_j \, (e|_{\mathrm{A}} \, (b_j|_{\mathrm{B}} \, |\rho)_{\mathrm{AB}} =$$

$$= -\sum_{i=1}^{d} \log_a p_i \, (a_i|\rho)_{\mathrm{A}} - \sum_{j=1}^{d} \log_a q_j \, (b_j|\rho)_{\mathrm{B}} = -\sum_{i=1}^{d} p_i \log_a p_i - \sum_{j=1}^{d} q_j \log_a q_j =$$

$$= S(\rho_{\mathrm{A}}) + S(\rho_{\mathrm{B}}).$$

We have then $S(\rho_{\mathrm{AB}}) \leq S(\rho_{\mathrm{A}}) + S(\rho_{\mathrm{B}})$. Throughout the proof, equality holds if and only if $\rho = \sigma$, namely $\rho_{\mathrm{AB}} = \rho_{\mathrm{A}} \otimes \rho_{\mathrm{B}}$. This proves that Shannon entropy is additive if and only if a bipartite state is a product state. □

One might wonder if subadditivity is merely a coincidence of the functional expression of Shannon entropy or has a more fundamental nature. If its nature were more fundamental, this would mean that a generic bipartite state is not more mixed than the product of its marginals. The following counter-example [11] shows that subadditivity does not hold for Rényi entropies with $\alpha \neq 1$, so it is a mere coincidence due to the functional form of Shannon entropy.

Example 7.3.7. Let us consider the composition of two g-bits and the diagonalization of the bipartite state

$$\rho_{\mathrm{AB}} = (pq + \varepsilon)\, \alpha_1 \otimes \beta_1 + [p\,(1-q) - \varepsilon]\, \alpha_1 \otimes \beta_2 +$$

$$+ [(1-p)\, q - \varepsilon]\, \alpha_2 \otimes \beta_1 + [(1-p)\,(1-q) + \varepsilon]\, \alpha_2 \otimes \beta_2,$$

where $p, q \in (0,1)$, but $p, q \neq \frac{1}{2}$. Actually, we have a family of states, each one labelled by an $\varepsilon \geq 0$. If $\varepsilon = 0$, then ρ_{AB} is a product state and its marginals are

$$\rho_{\mathrm{A}} = p\alpha_1 + (1-p)\, \alpha_2 \qquad \rho_{\mathrm{B}} = q\beta_1 + (1-q)\, \beta_2. \tag{7.2}$$

According to subadditivity, the maximum Shannon entropy is achieved when the state is a product state, that is when $\varepsilon = 0$. If this property holds for every Rényi entropy, then one must have $S_\alpha(\rho_{\mathrm{AB}}) \leq S_\alpha(\rho_{\mathrm{A}}) + S_\alpha(\rho_{\mathrm{B}})$ for every $\alpha \geq 0$, where ρ_{A} and ρ_{B} are defined in eq. (7.2). This means that $S_\alpha(\rho_{\mathrm{AB}})$ has a maximum for $\varepsilon = 0$. Let us see if the function

$$g(\varepsilon) = (pq + \varepsilon)^\alpha + [p\,(1-q) - \varepsilon]^\alpha + [(1-p)\, q - \varepsilon]^\alpha + [(1-p)\,(1-q) + \varepsilon]^\alpha,$$

which is differentiable in ε, has a maximum for $\varepsilon = 0$. If $\alpha = 0$, $g(\varepsilon) = 4$ and therefore $S_0(\rho_{\mathrm{AB}}) = S_0(\rho_{\mathrm{A}}) + S_0(\rho_{\mathrm{B}})$ always. This clearly contradicts subadditivity. We can now suppose $\alpha > 0$. Computing the first derivative of $g(\varepsilon)$, and evaluating it for $\varepsilon = 0$, we obtain

$$g'(0) = \alpha \left[(pq)^{\alpha-1} - [p\,(1-q)]^{\alpha-1} - [(1-p)\, q]^{\alpha-1} + [(1-p)\,(1-q)]^{\alpha-1}\right]$$

Note that the expression is well-defined even for $0 < \alpha < 1$, because $p, q \in (0,1)$. It is convenient to set $t := p^{\alpha-1}$, $u := q^{\alpha-1}$, $v := (1-p)^{\alpha-1}$ and $w := (1-q)^{\alpha-1}$. In this way, one has $g'(0) = 0$ if and only if

$$(t - v)\,(u - w) = 0.$$

This happens if and only if $t = v$ or $u = w$. Let us solve then

$$p^{\alpha-1} = (1-p)^{\alpha-1},$$

that is

$$\left(\frac{p}{1-p}\right)^{\alpha-1} = 1.$$

A trivial solution is when $p = 1 - p$, namely $p = \frac{1}{2}$, but this is excluded by hypothesis. Then it must be $\alpha = 1$. A similar line of reasoning applies also for q. This proves that we really have an extreme of $g(\varepsilon)$ for $\varepsilon = 0$ if and only if $\alpha = 1$.

We conclude that the *only* subadditive Rényi entropy is Shannon entropy ($\alpha = 1$).

Hence, subadditivity does not come from a fundamental reason concerning mixedness.

Let us now explore some other properties of Shannon entropy. These will be essentially corollaries of subadditivity.

Recall that if a bipartite state ρ_{AB} is pure, then its two marginals have the same eigenvalues (see section 5.2). This clearly means that they have the same Shannon entropy, so subadditivity reads $S(\rho_{\mathrm{A}}) \geq 0$, with equality if and only if the bipartite pure state is in fact a product state, for we have $S(\rho_{\mathrm{A}}) + S(\rho_{\mathrm{B}}) \geq S(\rho_{\mathrm{AB}})$, but $S(\rho_{\mathrm{AB}}) = 0$ and $S(\rho_{\mathrm{A}}) = S(\rho_{\mathrm{B}})$, because the two marginals have the same eigenvalues. This is nothing but the statement that the marginal of a bipartite pure state can be mixed and in particular it is pure ($S(\rho_{\mathrm{A}}) = 0$) if and only if the bipartite pure state is a product state (see proposition 4.1.2).

But we can say something more even in the case when the bipartite state is not a pure state, thanks to *triangle inequality*.

Proposition 7.3.8 (Triangle inequality)**.** *Shannon entropy fulfils* triangle inequality*, namely*

$$S(\rho_{\mathrm{AB}}) \geq |S(\rho_{\mathrm{A}}) - S(\rho_{\mathrm{B}})|.$$

Proof. Let us consider a purification of $|\Psi\rangle_{\mathrm{ABE}}$ of $|\rho\rangle_{\mathrm{AB}}$ with purifying system E. Let us apply subadditivity to system AE.

$$S(\rho_{\mathrm{AE}}) \leq S(\rho_{\mathrm{A}}) + S(\rho_{\mathrm{E}})$$

Since ABE is in a pure state, as we noticed above, the marginals on each couple of systems (e.g. AE and B) have the same Shannon entropy. Therefore $S(\rho_{AE}) = S(\rho_B)$ and $S(\rho_E) = S(\rho_{AB})$. Hence,

$$S(\rho_B) \leq S(\rho_A) + S(\rho_{AB}),$$

that is

$$S(\rho_{AB}) \geq S(\rho_B) - S(\rho_A).$$

Since the situation is symmetrical for A and B, one concludes that

$$S(\rho_{AB}) \geq |S(\rho_A) - S(\rho_B)|.$$

□

If we sum up all the result, we have

$$|S(\rho_A) - S(\rho_B)| \leq S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B).$$

7.4 A derivation of the second law of thermodynamics

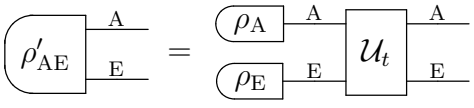
Subadditivity is the key ingredient in deriving an inequality that can be interpreted as a form of second law of thermodynamics, even though, in our framework, the physical interpretation is still open. However, the important point is that this inequality is derived from a function measuring mixedness, which shows that measures of mixedness play a central role in thermodynamics.

Our proof follows the one given in [16] very closely. Suppose we are in a physical theory in which time evolution for an isolated system is given by a one-parameter subgroup of reversible channels $\{\mathcal{U}_t\}_{t\in\mathbb{R}}$.

Consider a system A and the environment E and suppose they are completely uncorrelated at the initial time $t = 0$, i.e. system AE is in a product state $\rho_A \otimes \rho_E$. According to additivity of Shannon entropy, we have

$$S(\rho_A \otimes \rho_E) = S(\rho_A) + S(\rho_E).$$

Now consider system AE at a time $t > 0$. In general, the state $\rho'_{AE} = \mathcal{U}_t(\rho_A \otimes \rho_E)$ will be no more a product state.



However, since reversible channels preserve mixedness, we are sure that ρ'_{AE} is as mixed as $\rho_{\mathrm{A}} \otimes \rho_{\mathrm{E}}$ (see proposition 7.2.3), and this means that $S(\rho'_{\mathrm{AE}}) = S(\rho_{\mathrm{A}} \otimes \rho_{\mathrm{E}})$. By subadditivity,

$$S(\rho'_{\mathrm{AE}}) \leq S(\rho'_{\mathrm{A}}) + S(\rho'_{\mathrm{E}}),$$

where ρ'_{A} and ρ'_{E} are the marginals of ρ'_{AE} on the system and on the environment respectively. Putting all together, we have

$$S(\rho'_{\mathrm{A}}) + S(\rho'_{\mathrm{E}}) \geq S(\rho_{\mathrm{A}}) + S(\rho_{\mathrm{E}}).$$

This inequality can be read as "the sum of the entropies of the system and of the environment is non-decreasing". This sounds like a second law of thermodynamics. Nevertheless, this is essentially a mathematical result; so far there is no physical meaning behind Shannon entropy and there is even less meaning in thinking at the sum of Shannon entropies of the system and of the environment as the entropy of the universe. Indeed, one may argue that the actual final entropy of the universe is $S(\rho'_{\mathrm{AE}})$ and not $S(\rho_{\mathrm{A}}) + S(\rho_{\mathrm{E}})$.

Conclusions

Starting from the purification postulate, we developed a theory for quantifying and measuring entanglement and mixedness. As a central original result, we managed to prove that these two seemingly distinct aspects are in fact two sides of the same phenomenon. This sets forth a close relationship between entanglement and statistical mixtures, and assigns a primary role in the foundations of statistical physics to pure-state entanglement. In this vein, even stochastic processes, namely processes naturally involving mixedness, can be realized in a pure and reversible fashion exploiting entanglement.

As a side, but not minor, result, we managed to prove an abstract version of Lo-Popescu theorem, an important result about a class of communication protocols, LOCC protocols.

In a general probabilistic theory, mixed states are far from being density operators. However, we managed to devise a diagonalization procedure that enables us to define the analogy of eigenvalues. This was achieved by introducing a further axiom regarding distinguishability of states.

In this way, we managed to associate a set of eigenvalues with every mixed state even in general probabilistic theories. We then exploited eigenvalues to turn the issue of quantifying mixedness from an operational issue to an issue concerning real numbers.

Thus, the theory of majorization came to the aid of us, and we proved it to be an equivalent means to compare mixedness of states. Since we were looking for a more immediate tool to measure mixedness of a state, we turned our attention to Schur-concave functions as measures of mixedness. In particular, we proved some interesting properties for a particular Schur-concave function, Shannon entropy, which are completely analogous to the properties of quantum von Neumann entropy. Eventually, these properties naturally led to a proof of an inequality that can be interpreted as a second law of thermodynamics. In this way, we showed that measures of mixedness, although they

may appear as a completely abstract and mathematical tool, play a central role in defining thermodynamic inequalities. Therefore, entanglement and mixedness seem to be the key ingredients to build the foundations of thermodynamics.

Acknowledgements

This thesis was mainly done at IIIS (Institute for Interdisciplinary Information Sciences), Tsinghua University, Beijing, to which I am very grateful for the warm hospitality during a short visit in December 2013 and during my stay from the beginning of April 2014 to the middle of June 2014.

This work and my stay in Beijing were funded by Scuola Galileiana di Studi Superiori (Galilean School of Higher Education); this work was also supported by the Foundational Questions Institute through the large grant "The fundamental principles of information dynamics".

I am particularly grateful to prof. Giulio Chiribella, who supervised this work with passion and dedication during my stay in Beijing. My gratitude goes also to prof. Pieralberto Marchetti, who supervised all this thesis with genuine dedication and interest. Special thanks also to prof. Luca Salasnich, who acted as a referee for this work. I would like to thank also prof. Antonio Saggion for stimulating discussions about the interplay between thermodynamics and information theory.

My thanks also to all professors I encountered during my five years at the University of Padova.

Appendix A

Some useful mathematical results

In this appendix we collect some interesting (and not too difficult) mathematical results that are used in this work, but that are not completely related to the subjects presented.

A.1 Some theorems

First of all, we prove a proposition concerning the support of a density operator in a finite-dimensional Hilbert space.

Proposition A.1.1. *Suppose ρ and σ are positive operators such that $\rho \leq \sigma$. Then the support of ρ is contained in the support of σ.*

Proof. Recall that the support of an operator is the orthogonal complement of its kernel. If V and W are subspaces of a Hilbert space $\mathcal{H}$, then we have $V^\perp \leq W^\perp$ if and only if $W \leq V$. Therefore it is enough to prove that $\ker \sigma \leq \ker \rho$.

Suppose, by contradiction, that there is a vector $|\psi\rangle$ such that $\langle\psi|\,\sigma\,|\psi\rangle = 0$, but $\langle\psi|\,\rho\,|\psi\rangle \neq 0$. Since ρ is a positive operator, $\langle\psi|\,\rho\,|\psi\rangle > 0$. If $\rho \leq \sigma$, then $\tau := \sigma - \rho$ is another positive operator. We can compute $\langle\psi|\,\tau\,|\psi\rangle$.

$$\langle\psi|\,\tau\,|\psi\rangle = \langle\psi|\,\sigma - \rho\,|\psi\rangle = \langle\psi|\,\sigma\,|\psi\rangle - \langle\psi|\,\rho\,|\psi\rangle =$$

$$= -\langle\psi|\,\rho\,|\psi\rangle < 0$$

This contradicts the fact that τ is a positive operator. We then conclude that $\ker \sigma \leq \ker \rho$, namely the support of ρ is contained in the support of σ. □

The next result we present is a theorem by Carathéodory about convex geometry [81, 82].

Theorem A.1.2 (Carathéodory)**.** *Let $\mathbb{A}\left(\mathbb{R}^d\right)$ be a d-dimensional real affine space. If a point x lies in the convex hull of a set $S \subseteq \mathbb{A}\left(\mathbb{R}^d\right)$, then it lies in the convex hull of at most $d+1$ points of S.*

Proof. If x is in the convex hull of S, then it can be written as a convex combination of a finite number of elements in S.

$$x = \sum_{i=1}^{n} \lambda_i x_i,$$

where $x_i \in S$ and $\lambda_i \geq 0$ for every i, and $\sum_i \lambda_i = 1$. If $n \leq d+1$, there is nothing to prove. Suppose then that $n > d+1$. This means that the $n-1 > d$ vectors $x_i - x_1$ (for $i \geq 2$) must be linearly dependent. Thus, there exist $n-1$ not all vanishing real numbers α_i such that $\sum_{i=2}^{n} \alpha_i (x_i - x_1) = 0$. If we now define $\alpha_1 = -\sum_{i=2}^{n} \alpha_i$, we have $\sum_{i=1}^{n} \alpha_i = 0$ and $\sum_{i=1}^{n} \alpha_i (x_i - x_1) = 0$. Therefore, we can write, for any $\mu \in \mathbb{R}$,

$$x = \sum_{i=1}^{n} \lambda_i x_i - \mu \sum_{i=1}^{n} \alpha_i (x_i - x_1) = \sum_{i=1}^{n} (\lambda_i - \mu\alpha_i) x_i,$$

where we used the fact that $\sum_{i=1}^{n} \alpha_i = 0$. Let us choose μ as

$$\mu := \min_{1 \leq i \leq n} \left\{ \frac{\lambda_i}{\alpha_i} : \alpha_i > 0 \right\}.$$

Notice that there exists at least one α_i which is positive because α_i's are not all 0 and $\sum_i \alpha_i = 0$. In this way, clearly $\mu > 0$, and $\lambda_i - \mu\alpha_i \geq 0$ and one has equality for the index i that achieves the minimum. In this way $x = \sum_{i=1}^{n} (\lambda_i - \mu\alpha_i) x_i$, and the right-hand side is actually a sum with $n-1$ term (one of them vanishes). Besides, $\sum_{i=1}^{n} (\lambda_i - \mu\alpha_i) = 1$ because $\sum_{i=1}^{n} \alpha_i = 0$. We managed to write x as a convex combination of $n' = n-1$ points. If $n' > d+1$, then we repeat this procedure several times, until we achieve $n' = d+1$. □

We conclude with a mathematical result about preorders.

Proposition A.1.3. *Let $(X, \lesssim)$ be a preordered set, where $\lesssim$ is not an equivalence relation. If $x, y \in X$, define*[1] *$x \sim y$ 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, form the fact that $x \sim y$, it follows that $x \lesssim y$ and $y \lesssim x$; and from the fact that $y \sim z$, it follows that $y \lesssim z$ 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.

Now, 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 $[x] \leq [x]$ means $x \lesssim x$, and this is true because $\lesssim$ is reflexive. It is transitive, since if $[x] \leq [y]$ and $[y] \leq [z]$, it means that $x \lesssim y$ 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 $[x] = [y]$.

□

[1] If $\lesssim$ were an equivalence relation, then $x \sim y$, for every $y \in X$, and there would be only one equivalence class. In this case, the result of proposition A.1.3 would be degenerate.

Bibliography

[1] J. C. Maxwell, *The theory of heat*, Dover Publications: Mineola, Reprint 2001.

[2] H. Leff, A. F. Rex (ed.), *Maxwell demon 2: entropy, classical and quantum information, computing*, IOP Publishing: London, 2003.

[3] L. Szilard, *On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings*, Syst. Res. **9**, 301 (1929, Reprint 1964).

[4] R. Landauer, *Irreversibility and heat generation in the computing process*, IBM J. Res. Dev. **5**, 183 (1961).

[5] R. Faraldo, A. Saggion, *L'osservatore in termodinamica* (The observer in thermodynamics), unpublished research note.

[6] E. A. Guggenheim, *Thermodynamics*, Elsevier: Amsterdam, 1967.

[7] C. H. Bennet, *Logical reversibility of computation*, IBM J. Res. Dev. **17**, 525 (1973).

[8] E. Fredkin, T. Toffoli, *Conservative logic*, Int. Journ. Theor. Phys. **21**, 219 (1982).

[9] J. A. Wheeler, *Information, physics, quantum: the search for links*, in *Complexity, entropy and the physics of information*, W. H. Zurek (ed.), Addison-Wesley: Boston, 1990, p. 5.

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

[11] W. Thirring, *Quantum mathematical physics*, Springer: New York, 2003.

[12] S. Popescu, A. J. Short, A. Winter, *Entanglement and the foundations of statistical mechanics*, Nature Physics **2**, 754 (2006).

[13] J. Gemmer, A. Otte, G. Mahler, *Quantum approach to a derivation of the second law of thermodynamics*, Phys. Rev. Lett. **86**, 1927 (2001).

[14] K. Huang, *Statistical mechanics*, Wiley: Hoboken, 1988.

[15] E. Schrödinger, *Discussion of probability relations between separated systems*, Proc. Camb. Phil. Soc. **31**, 555 (1935).

[16] J. Preskill, *Lecture notes on quantum computation*, `http://www.theory.caltech.edu/people/preskill/ph229/`.

[17] M. A. Nielsen, I. L. Chuang, *Quantum computation and quantum information*, Cambridge University Press: Cambridge, 2010.

[18] L. Maccone, L. Salasnich, *Fisica moderna. Meccanica quantistica, caos e sistemi complessi* (Modern physics. Quantum mechanics, chaos and complex systems), Carocci: Roma, 2008.

[19] M. A. Naimark, Iza. Akad. Nauk USSR, Ser. Mat. **4**, 277 (1940).

[20] W. F. Stinespring, *Positive functions on C^*-algebras*, Proc. Amer. Math. Soc. **6**, 211 (1955).

[21] M. Ozawa, *Quantum measuring processes of continuous observables*, J. Math. Phys. **25**, 79 (1984).

[22] K. Kraus, *States, effects and operations: fundamental notions of quantum theory*, Lecture Notes in Physics, Springer: Berlin, 1983.

[23] K.-E. Hellwig, K. Kraus, *Pure operations and measurements*, Commun. Math. Phys. **11**, 214 (1969).

[24] K.-E. Hellwig, K. Kraus, *Operations and measurements II*, Commun. Math. Phys. **16**, 142 (1970).

[25] P. Janotta, H. Hinrichsen, *Generalized probability theories: what determines the structure of quantum physics?*, arXiv:1402.6562 [quant-ph].

[26] L. Hardy, R. Spekkens, *Why physics needs quantum foundations*, Phys. Can. **66** (2), 73 (2010).

[27] P. Ball, *Physics: Quantum quest*, Nature **501**, 154 (2013).

[28] G. W. Mackey, *Mathematical foundations of quantum mechanics*, Dover Publications: Mineola, Reprint 2004.

[29] G. Ludwig, *Foundations of quantum mechanics*, Springer: Berlin, 1983.

[30] G. Ludwig, *An axiomatic basis for quantum Mechanics, Vol. 1: Derivation of Hilbert Space Structure*, Springer: Berlin, 1985.

[31] B. Coecke, *Quantum picturalism*, Contemporary Physics **51**, 59 (2010).

[32] L. Hardy, *Quantum theory from five reasonable axioms*, arXiv:quant-ph/0101012.

[33] G. M. D'Ariano, *Probabilistic theories: what is special about quantum mechanics?*, in *Philosophy of Quantum Information and Entanglement*, A. Bokulich, G. Jaeger (eds.), Cambridge University Press: Cambridge, 2010, pp. 85–126.

[34] P. Goyal, K. H. Knuth, J. Skilling, *Origin of complex quantum amplitudes and Feynman's rules*, Phys. Rev. A **81**, 022109 (2010).

[35] B. Dakić, C. Bruckner, *Quantum theory and beyond: is entanglement special?*, in *Deep beauty: Understanding the quantum world through mathematical innovation*, H. Halvorson (ed.), Cambridge University Press: Cambridge, 2011, pp. 365–392.

[36] L. Masanes, M. Müller, *A derivation of quantum theory from physical requirements*, New J. Phys. **13**, 063001 (2011).

[37] L. Hardy, *Reformulating and reconstructing quantum theory*, arXiv:1104.2066 [quant-ph].

[38] L. Masanes, M. P. Müller, R. Augusiak, D. Perez-Garcia, *A digital approach to quantum theory*, PNAS **110**, 16373 (2013).

[39] G. Chiribella, G. M. D'Ariano, P. Perinotti, *Probabilistic theories with purification*, Phys. Rev. A **81**, 062348 (2010).

[40] G. Chiribella, G. M. D'Ariano, P. Perinotti, *Informational derivation of quantum theory*, Phys. Rev. A **84**, 012311 (2011).

[41] B. Coecke, *Kindergarten quantum mechanics*, arXiv:quant-ph/0510032.

[42] P. Selinger, *A survey of graphical languages for monoidal categories*, Springer Lecture Notes in Physics **813**, 289 (2011).

[43] S. Mac Lane, *Categories for the working mathematician*, Springer: New York, 1971.

[44] W. K. Wootters, *Local accessibility of quantum states*, in *Complexity, entropy and the physics of information*, W. H. Zurek (ed.), Addison-Wesley: Boston, 1990, p. 39.

[45] S. Popescu, D. Rohrlich, *Quantum nonlocality as an axiom*, Found. Phys. **3**, 379 (1994).

[46] J. Barrett, *Information processing in generalized probabilistic theories*, Phys. Rev. A **75**, 032304 (2007).

[47] H. Barnum, J. Barrett, M. Leifer, A. Wilce, *A generalized no-broadcasting theorem*, Phys. Rev. Lett. **99**, 240501 (2007).

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

[49] L. Hardy, *Towards quantum gravity: a framework for probabilistic theories with non-fixed causal structure*, J. Phys. A **40**, 3081 (2007).

[50] G. Chiribella, G. M. D'Ariano, P. Perinotti, *Transforming quantum operations: quantum supermaps*, Europhysics Letters **83**, 30004 (2008).

[51] M.-D. Choi, *Completely positive linear maps on complex matrices*, Lin. Alg. Appl. **10**, 285 (1975).

[52] A. Jamiołkowski, *Linear transformations which preserve trace and positive semidefiniteness of operators*, Rep. Math. Phys. **3**, 275 (1972).

[53] G. B. Folland, *A course in abstract harmonic analysis*, CRC Press: Boca Raton, 1995.

[54] M. Horodecki, P. Horodecki, J. Oppenheim, *Reversible transformations from pure to mixed states, and the unique measure of information*, Phys. Rev. A **67,** 062104 (2003).

[55] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, W. K. Wootters, *Purification of noisy entanglement and faithful teleportation via noisy channels*, Phys. Rev. Lett. **76**, 722 (1996).

[56] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, W. K. Wootters, *Mixed-state entanglement and quantum error correction*, Phys. Rev. A **54**, 3824 (1996).

[57] M. Cailotto, *Algebra e geometria lineari e quadratiche, prima parte* (Lecture notes on linear and quadratic algebra and geometry, 1st part), `http://www.math.unipd.it/~maurizio/`.

[58] A. Facchini, *Algebra e matematica discreta* (Algebra and discrete mathematics), Zanichelli: Bologna, 2000.

[59] M. A. Nielsen, *Conditions for a class of entanglement transformations*, Phys. Rev. Lett. **83** (2), 436 (1999).

[60] A. Uhlmann, *Sätze über dichtematrizen*, Wiss. Z. Karl-Marx-Univ. Leipzig **20**, 633 (1971).

[61] A. Uhlmann, *Endlich-dimensionale dichtematrizen I*, Wiss. Z. Karl-Marx-Univ. Leipzig **21**, 421 (1972).

[62] A. Uhlmann, *Endlich-dimensionale dichtematrizen II*, Wiss. Z. Karl-Marx-Univ. Leipzig **22**, 139 (1973).

[63] H.-K. Lo, S. Popescu, *Concentrating local entanglement by local actions: beyond mean values*, Phys. Rev. A **63**, 022301 (2001).

[64] H. Barnum et al., *Entropy and information causality in general probabilistic theories*, New J. Phys. **12**, 033024 (2010).

[65] A. J. Short, S. Wehner, *Entropy in general physical theories*, New J. Phys. **12**, 033023 (2010).

[66] G. Kimura, K. Nuida, H. Imai, *Distinguishability measures and entropies for general probabilistic theories*, Rep. Math. Phys. **66**, 175 (2010).

[67] M. O. Lorenz, *Methods of measuring concentration of wealth*, J. Amer. Statist. Assoc. **9**, 209 (1905).

[68] A. W. Marshall, I. Olkin, B. C. Arnold, *Inequalities: theory of majorization and its applications*, Springer: New York, 2011.

[69] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge University Press: Cambridge, 1952.

[70] G. Birkhoff, *Tres observaciones sobre el algebra lineal*, Univ. Nac. Tucumán Rev. Ser. A **5**, 147 (1946).

[71] A. M. Ostrowski, *Sur quelques applications des fonctions convexes et concaves au sens de I. Schur*, J. Math. Pures Appl. **31**, 253 (1952).

[72] R. F. Muirhead, *Some methods applicable to identities and inequalities of symmetric algebraic functions of n letters*, Proc. Edinburgh Math. Soc. **21**, 144 (1903).

[73] T. Ando, *Majorization, doubly stochastic matrices, and comparison of eigenvalues*, Hokkaido University: Sapporo, 1982.

[74] C. E. Shannon, *A mathematical theory of communication*, Bell System Tech. J. **27**, 379 (1948).

[75] A. Rényi, *On measures of information and entropy*, Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability 1960, 547 (1961).

[76] T. M. Cover, J. A. Thomas, *Elements of information theory*, Wiley: Hoboken, 2006.

[77] S. Kullback, R. A. Leibler, *On information and sufficiency*, Ann. Math. Stat. **22**, 79 (1951).

[78] W. Hässelbarth, *The incompleteness of Rényi entropies*, Theor. Chim. Acta **70**, 119 (1986).

[79] H. Umegaki, *Conditional expectation in an operator algebra. IV. Entropy and information,* Kodai Math. Sem. Rep. **14**, 59 (1962).

[80] H. Araki, E. H. Lieb, *Entropy inequalities*, Comm. Math. Phys. **18**, 160 (1970).

[81] C. Carathéodory, *Über den variabilitätsbereich der fourier'schen konstanten von positiven harmonischen funktionen*, Rendiconti del Circolo Matematico di Palermo **32**, 193 (1911).

[82] E. Steinitz, *Bedingt konvergente Reihen und konvexe Systeme*, J. Reine Angew. Math. **143**, 128 (1913).