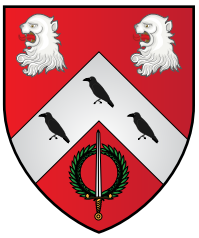
UNIVERSITY OF OXFORD

ST. ANNE'S COLLEGE



CARLO MARIA SCANDOLO

INFORMATION-THEORETIC FOUNDATIONS OF THERMODYNAMICS IN GENERAL PROBABILISTIC THEORIES

A THESIS SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

TRINITY 2018

To my parents, Antonio and Silvana

Contents

Abstract

In this thesis we study the informational underpinnings of thermodynamics and statistical mechanics. To this purpose, we use an abstract framework—general probabilistic theories—, capable of describing arbitrary physical theories, which allows one to abstract the informational content of a theory from the concrete details of its formalism. In this framework, we extend the treatment of microcanonical thermodynamics, namely the thermodynamics of systems with a well-defined energy, beyond the known cases of classical and quantum theory. We formulate two requirements a theory should satisfy to have a well-defined microcanonical thermodynamics. We adopt the recent approach of resource theories, where one studies the transitions between states that can be accomplished with a restricted set of physical operations. We formulate three different resource theories, differing in the choice of the restricted set of physical operations.

To bridge the gap between the objective dynamics of particles and the subjective world of probabilities, one of the core issues in the foundations of statistical mechanics, we propose four information-theoretic axioms. They are satisfied by quantum theory and more exotic alternatives, including a suitable extension of classical theory where classical systems interact with each other creating entangled states. The axioms identify a class of theories where every mixed state can be modelled as the reduced state of a pure entangled state. In these theories it is possible to introduce well-behaved notions of majorisation, entropy, and Gibbs states, allowing for an information-theoretic derivation of Landauer's principle. The three resource theories define the same notion of resource if and only if, on top of the four axioms, the dynamics of the underlying theory satisfy a condition called "unrestricted reversibility". Under this condition we derive a duality between microcanonical thermodynamics and pure bipartite entanglement.

Chapter 1

Introduction

Thermodynamics is a powerful phenomenological paradigm encompassing several scientific disciplines, from physics to chemistry, up to biology and engineering. Its principles form a framework every experimental observation must adhere to. Its two most important laws express energy conservation (the first law), and the existence of irreversible processes, or, loosely speaking, of an arrow of time (the second law). Whilst it is fairly easy and reasonable to accept a principle such as energy conservation, the second law and its consequent arrow of time have caused a great bewilderment among scientists and philosophers, since it was difficult to find an explanation for the origin of irreversibility.

It was also necessary to find a place in the structure of physics for the new concepts introduced by thermodynamics: work, heat, temperature, etc. Were they fundamental or could they be derived from other concepts? Several theories of thermodynamics were proposed. The first leaned towards the view that heat and temperature are primitive notions, but later, the classic works by Maxwell [1, 2], Boltzmann [3], and Gibbs [4] undertook a reduction of the laws of thermodynamics to the laws of the underlying dynamics of particles and fields. This reduction led to the establishment of statistical mechanics as the standard paradigm for the foundations for thermodynamics [5–8]. This has worked even for quantum systems, where quantum statistical mechanics was able to predict new, genuinely quantum, phenomena, such as Bose-Einstein condensation, later observed in a laboratory [9], and related to other important phenomena: superfluidity and superconductivity [10]. For this reason, in the following we will use the terms “statistical mechanics” and “thermodynamics” nearly

as synonyms.

However, the statistical paradigm in turn led to novel questions, the central one now being how to reconcile the use of statistical notions (mixed states), associated with the incomplete knowledge of an agent, with the picture of Nature provided by classical and quantum mechanics, where the fundamental dynamics are deterministic (and reversible). Different proposals have been made for classical statistical mechanics, the best known of which are ergodic theory [11–15], and Jaynes' maximum entropy approach [16], then extended to quantum statistical mechanics [17]. Deffner and Zurek refer to these attempts as to a

> "'half-way' house, populated by fictitious but useful concepts such as *ensembles*" [18].

Quantum theory, instead, offers a *radically new* opportunity. As originally noted by Schrödinger [19], a system and its environment can be jointly in a pure state, whilst the system is individually in a mixed state. Here the mixed state does *not* represent an ensemble of identical systems, but rather the state of a *single* quantum system. Based on this idea, Popescu, Short, and Winter [20], and Goldstein, Lebowitz, Tumulka, and Zanghì [21] proposed that entanglement could be the starting point for a new, genuinely quantum foundation of statistical mechanics. The idea was that, when the environment is large enough, the system is approximately in the equilibrium state for the typical joint pure states of the system and the environment. This idea has been explored in a variety of settings [22–29], whose common inspiration is the idea that quantum entanglement can provide a new foundation for statistical mechanics, and ultimately, thermodynamics. Furthermore, even irreversibility can be explained because some degrees of freedom are traced out.

The success of the statistical mechanical paradigm is deeply tied to the fact that the physical systems under investigation are composed of an enormously large number of particles, which guarantees the applicability of statistical methods. However, more recently, the scope of thermodynamics in the quantum regime has been extended from quantum gases to microscopic systems far from the thermodynamic limit, so crucial for the development of nanotechnology [30–32]. In this new regime one studies the thermodynamic transformations and the fluctuations of quantum systems with very few particles, a scenario often called the *single-shot* regime. Clearly one cannot use the standard tools of statistical mechan-

ics, and a new way to address this new regime is to adopt a resource-theoretic approach [33–35], where one starts from a subset of quantum operations that are "free" or "easy to implement", and characterises the transitions that can be accomplished by these free operations. This establishes a preorder on quantum states based on their value as thermodynamic resources, from the most valuable to the least valuable (the so-called "free states"), which correspond to equilibrium states. This emergence of thermalisation through the repeated applications of some transformations is the reason why resource theories have been so successful in the study of quantum thermodynamics [30–32]. In this approach thermodynamic potentials often emerge as functions assigning a value to resources compatibly with the resource preorder. Many recent results in quantum thermodynamics have been obtained in this way [36–55].

The approach to quantum thermodynamics based on resource theories uses a lot of concepts and techniques from (quantum) information theory. This should not surprise, as from the early development of statistical mechanics it became clear that thermodynamic concepts are intimately tied to information-theoretic ones, as shown by the paradigmatic examples of Maxwell's demon [56], and the closely related Szilard engine [57]. In these examples, the knowledge possessed by a (microscopic) observer was used to set up a physical process violating the second law of thermodynamics. This was clearly paradoxical, and called for an explanation in order to reaffirm the validity of the famed second law. Since the paradoxes were all based on the information possessed by an observer, if a solution was to be found, it would involve some information-theoretic concepts. The correct solution came many years later, when Landauer found out something very surprising: the act of erasing and overwriting the memory of a computing device has a physical effect [58]. More precisely, if we erase an unknown bit at temperature T, there is an associated heat dissipation of $kT \ln 2$, where k is Boltzmann constant. This opened the way to a solution of the paradoxes, which was found by Bennett [59,60]. He understood that if an observer is to act on a physical system based on the result of their previous observation, they have to *store* their observation somewhere. Since infinite storage does not exist, at some point they will have to erase their memory, spending energy to do it, according to Landauer's result. In conclusion, microscopic observers such as Maxwell's demon *cannot* be used for cyclic work extraction at no additional cost, because their usage would involve some energy dissipation at some stage of the protocol.

In this thesis, in accordance with the recent trend in theoretical physics of grounding physics on information theory, we study the information-theoretic foundations of thermodynamics and statistical mechanics. However, we do it in a different way from before: instead of analysing them in quantum theory, we do it in arbitrary physical theories. This is a novel area of research, and this DPhil thesis is the first doctoral thesis on this topic.

If quantum theory is the ultimate theory of Nature, at least in the microscopic domain, why do we bother to study the foundations of thermodynamics in general physical theories? There are manifold answers. One is that one of the major trends in theoretical physics has always been to generalise known results and broaden their scope, therefore it is natural to study thermodynamics in its full generality. This may seem a truly ambitious and hard feat. However this is precisely in the spirit of thermodynamics: being a very general phenomenological paradigm, it should be theory-independent in its essence, therefore it should be possible to define and study it in abstract terms. Another answer is that some features of quantum theory, and by extension, of quantum thermodynamics, are best understood when looked at them "from the outside". Contrasting quantum behaviour with the behaviour one observes in general theories can provide a new insights into why the world is quantum, this time from a thermodynamic angle. Another reason is that, working in an abstract way, we can capture the information-theoretic essence of thermodynamics, without "being distracted" by the concrete details of the formalism of a specific physical theory.

Clearly, the first thing we need is a theory-independent way to address physical theories, an abstract framework that allows us to describe all their common traits, without plunging deep into the details of their specific formalisms. Fortunately, such a framework exists, and it is that of *general probabilistic theories* [61–66], which identifies the two main ingredients of any physical theory to be its *compositional structure* (how to build experiments) and its *probabilistic structure* (how to assign probabilities to experimental observations). This is even more obvious in the variant of the formalism known as *operational probabilistic theories* [67–74], arisen from the marriage of the graphical language of category theory [75–80] with probability theory. As opposed to the original framework for general probabilistic theories, based on the convex geometry of states [62, 64, 66, 81, 82], the focus of operational probabilistic theories is on physical processes and

their composition. This is the approach adopted in this thesis: since thermodynamics is a theory concerned with processes and transitions, it is natural to resort to a formalism where processes play centre stage.

Historically, general probabilistic theories were introduced as a framework from which to derive quantum theory by imposing suitable information-theoretic principles, leading to various quantum reconstructions [61,68,70, 72,73,83–89]. However, their scope is broader than just this: besides helping us gain an operational understanding of quantum theory and why Nature is quantum, general probabilistic theories are important also for studying extensions and restrictions of quantum structures [90]. Indeed, several proposals for a theory of quantum gravity have called for a modification of the quantum laws to a more general form (see e.g. [91]). On the other side, sometimes one considers sub-theories of quantum theory, arising for instance from an experimental limitation on the states or operations one can implement in a laboratory. In this case, what is the resulting theory like? To give an example, Bartlett, Rudolph, and Spekkens studied Gaussian quantum theory, and found out that it admits a semiclassical explanation as an epistemically restricted theory [92,93].

In our quest for the foundations of thermodynamics, we will be inspired by the recent results obtained in quantum thermodynamics for microscopic systems. Therefore we adopt the resource-theoretic approach, which was shown to be not at all specific to quantum theory, but rather applicable to a broad range of theories and situations [94–99].

We will find out that not all physical theories, which in principle may be extremely counter-intuitive, are suitable to support a sensible thermodynamics. Therefore we have to introduce some axioms in order to restrict ourselves to thermodynamically relevant theories. Specifically, inspired by the results about typicality [20,21], we start from entanglement [67,100], turning it into an axiomatic foundation for statistical mechanics. We explore the hypothesis that the physical systems admitting a well-behaved statistical mechanics are exactly those where, at least in principle, mixed states can be modelled as the local states of larger systems, globally in a pure state [101]. This modelling is possible in quantum theory, where it provides the stepping stone for the derivation of the microcanonical and canonical states in [18, 20, 21]. But the foundational role of entanglement is not limited to quantum theory. We show that even classical statistical mechanics, where entanglement is absent, can find a new foundation if classical theory is regarded as part of a larger physical theory where

classical mixed states can be obtained as marginals of pure states of non-classical composite systems [101]. Remarkably, the mere fact that classical systems *could* be entangled with some other physical systems determines some of their properties, and opens the way to the use of typicality arguments like in the quantum case. The same approach is applicable to several extensions of quantum theory, including quantum theory with superselection rules [101–105], and variants of quantum theory with real amplitudes [106–109]. In this framework, we demand the validity of four information-theoretic axioms, informally stated as follows:

Causality [67] No signal can be sent from the future to the past.

Purity Preservation [90] The composition of two pure transformations is a pure transformation.

Pure Sharpness [110] Every system has at least one pure sharp observable.

Purification [67] Every state can be modelled as the marginal of a pure state. Such a modelling is unique up to local reversible transformations.

We call the theories satisfying these axioms *sharp theories with purification*, a notable example being quantum theory itself. We show that the validity of above axioms implies that these theories have some nearly quantum behaviour (e.g. the existence of entanglement), yet they need not be quantum [101, 105, 111]. Their key feature is that they admit a level of description in which all processes are pure and reversible, and all measurements are sharp. As such, we believe them to play a really fundamental role in physics, and quantum theory is an example of this. These axioms enforcing purity at the fundamental level are also interesting from a thermodynamic point of view. For example, Causality can be related to the ability to discard systems, and therefore to restrict ourselves to a smaller subsystem of a larger system. On the other side, Purification, being the foundation for all extension and dilation theorems [67, 71], can be thought of as the ability for a thermodynamic observer to enlarge their system in order to always have an isolated system.

We study the simplest situation: microcanonical thermodynamics, describing a system with fixed energy, first in arbitrary physical theories,

and then in sharp theories with purification. We formulate two requirements a theory should satisfy to have a well-defined microcanonical thermodynamics, and we show that the axioms of sharp theories with purification guarantee that they are satisfied. The following step is to introduce a resource-theoretic treatment of thermodynamics in this regime. Clearly, it is natural to choose the microcanonical state as free, but what about the choice of free operations? We have essentially three possibilities [105]:

random reversible channels arising from reversible dynamics with randomly fluctuating parameters;

noisy operations generated by preparing ancillas in the microcanonical state, turning on a reversible dynamic, and discarding the ancillas;

unital channels defined as the processes that preserve the microcanonical state.

In sharp theories with purification the three sets of operations satisfy some remarkable inclusion relations like in quantum theory, with random reversible channels included in the set of noisy operations, and noisy operations included in the set of unital channels.

We show that the preorder induced by unital channel is completely characterised by a suitable majorisation criterion [105]. As a consequence, the functions that assign a value to states compatibly with the preorder, which are measures of mixedness, bear a close resemblance to entropies; in more mathematical terms they are Schur-concave functions [112]. In this setting we show that it is possible to put forward a definition of Shannon-von Neumann entropy with similar properties to its quantum counterpart, which allows us to prove an operational version of Landauer's principle [101].

If we want majorisation to completely characterise the preorder of all the three resource theories, the physical theory must satisfy an additional axiom, called "unrestricted reversibility", which comes in three equivalent flavours in sharp theories with purification [105]:

Permutability [70] Every permutation of every maximal set of perfectly distinguishable pure states can be implemented by a reversible transformation.

Strong Symmetry [87, 113] For every two maximal sets of perfectly distinguishable pure states, there exists a reversible transformation converting the states in one set into the states in the other.

Reversible controllability [114] For every pair of systems A and B, it is possible to reversibly implement any control-reversible transformation.

When unrestricted reversibility holds, the three resource theories identify the same notion of resource—purity—and in this case we can prove a duality between purity and pure-state entanglement [100].

Published work The core of this work is taken from [101, 105], with some minor parts from [111, 115]. Specifically, most of the material presented in chapter 4 comes from [101], whereas chapter 5 contains material from both [101] (the part about entropies, mixedness monotones, and Landauer's principle) and [105] (the rest).

Structure The thesis is structured as follows: in chapter 2 we introduce the basic framework of general probabilistic theories, presented mainly in the operational-probabilistic variant. In the same chapter we introduce the first axiom, Causality, stating that information propagates from the past to the future, and we analyse its consequences. Causality will remain a standing assumption throughout the rest of the thesis. In chapter 3 we present the main tool we use to study thermodynamics in general physical theories, namely resource theories. Sharp theories with purification are introduced in chapter 4, where their general properties are studied in detail. The key thermodynamic results are exposed in chapter 5: we start from microcanonical thermodynamics, examined in great detail both in arbitrary physical theories and in sharp theories with purification. Then we introduce thermal states, by which we obtain an information-theoretic derivation of Landauer's principle. Finally conclusions are drawn in chapter 6, with an outlook on further directions of research.

In this thesis we assume that the reader is already familiar with the basic framework and terminology of quantum mechanics and elementary quantum information theory, in particular mixed states and quantum channels. Good references in this respect are [116–119]. The other, more

advanced, concepts will be thoroughly explained when they are introduced.

1.1 List of publications and preprints

The work presented in this thesis contains material from the following publications and preprints:

1. G. Chiribella, C. M. Scandolo, *Microcanonical thermodynamics in general physical theories*, New J. Phys. **19** (12), 123043 (2017) [105].

2. G. Chiribella, C. M. Scandolo, *Entanglement as an axiomatic foundation for statistical mechanics*, arXiv:1608.04459 [quant-ph] (2016) [101].

3. H. Barnum, C. M. Lee, C. M. Scandolo, J. H. Selby, *Ruling out Higher-Order Interference from Purity Principles*, Entropy **19** (6), 253 (2017) [111].

4. C. M. Scandolo, R. Salazar, J. K. Korbicz, P. Horodecki, *Is it possible to be objective in every physical theory?* arXiv:1805.12126 [quant-ph] (2018) [115].

Chapter 2

General probabilistic theories

In this chapter we present the framework our investigation is conducted in. Known as "general probabilistic theories" (GPTs) [61–66], it is general enough to accommodate essentially every physical theory, admitting probabilistic processes. The idea behind it is that a theory is defined by what an agent can do in a laboratory, and by the observations they collect, and the predictions they make.

GPTs come in two flavours: one based on convex geometry [62, 64, 81, 82], and the other, more general, based on the compositional structure of physical theories [67–74]. The theories described in the latter approach are often called *operational probabilistic theories* (OPTs). These two approaches are almost equivalent, but OPTs are able to describe also non-convex theories, arising e.g. from the lack of Causality.

For convex theories, one can translate concepts of one approach into the other, but the scope of the two approaches remains slightly different. The convex approach is more low-level: one starts from the state space of single systems, and builds composites from it. Its weak point is that one must specify all the details and constructions, and this can become cumbersome when one studies the composition of systems [120–122]. However, this approach is often inescapable if one is to deal with a concrete model. On the other hand, OPTs use the high-level language of circuits borrowed from category theory [75–80], and take composition as a primitive, rather than derive it from the structure of the state space. Its strong point is that it can be used to derive results about a theory without specifying its concrete details too much.

In this chapter and the rest of this thesis we will mainly adopt the OPT

variant: its focus on processes and their composition matches beautifully with the scope of thermodynamics, which is all about processes and transformations between states. Therefore, in the following, the term "GPT" will be used as a synonym of "OPT", or more precisely, of a general probabilistic theory treated in the OPT approach. Here we present the principles underpinning the OPT framework, where the ideas of process and composition play a central role. Even states are viewed as processes, specifically as preparation processes. The analysis of the operational structure of a physical theory is done by introducing a diagrammatic language, which will be used throughout this thesis. Then, in section 2.2, we insert the probabilistic ingredient: every theory must be able to provide probabilities of experimental outcomes.

In section 2.3 we introduce the axiom of Causality [67], which is often implicitly assumed in a lot of GPT literature. Causality will be a background assumption throughout this thesis. The choice of Causality as an axiom for a physical theory is motivated by several of its consequences, e.g. the lack of time loops or the no-signalling principle between different physical systems. Finally, this axiom is also appealing from a thermodynamic perspective, for it guarantees the ability to discard systems, and therefore to restrict ourselves to a subsystem of a larger thermodynamic system.

2.1 Events, tests, and the operational structure

As the name suggests, an operational probabilistic theory is made of two parts: the *operational* one and the *probabilistic* one. The operational part is the more fundamental: it describes how to build experiments in a laboratory by composing and connecting the associated devices. As such, it is the essential ingredient of every experimentally testable physical theory. The probabilistic part is built on top of that, and it deals with the predictive power of the theory, namely the ability to predict the likelihood of the various experimental observations.

In this section we introduce a formalism able to describe the operational structure of every physical theory [67,69,70,72,73], which has a nice graphical representation in terms of diagrams and circuits, taken from the graphical languages for symmetric monoidal categories [76–80].

2.1.1 Systems and tests

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

We denote systems by capital letters in Roman character: A, B, etc. There is also a particular system, the *trivial system*, which simply means "nothing", or the degrees of freedom the theory does not deal with. We will denote it by letter I. A device with the trivial system as input is simply a device with *no* input, and a device with the trivial system as output is simply a device with *no* output.

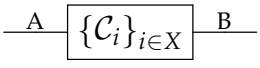
The application of a physical device can yield various outcomes. Each of them corresponds to a particular event that occurred in the laboratory, which can be identified by the experimenter by "reading" the device pointer. Therefore, we can give the following characterisation of tests.

Definition 2.1.1. A *test* with input system A and output system B is a collection of *events* $\{\mathcal{C}_i\}_{i\in X}$ that can occur in an experiment, labelled by the outcome i in some set X. X is called *outcome set.*

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

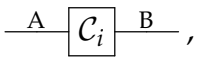
To clarify the role of outcome i better, we can regard it as what the experimenter actually sees when they perform their experiment (a sequence of digits, a spot in a photographic plate, the device pointer, etc.). The outcome set X is the set containing all the possible outcomes for a given test. In the following we will assume that all outcome sets are finite. This will simplify the later mathematical treatment, and will match a finite-dimensionality assumption we will make in section 2.2.

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



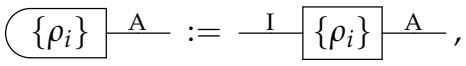
When there is no ambiguity, we will omit the outcome set X. If we want

to express that the specific event $\mathcal{C}_i$ has occurred, we will write

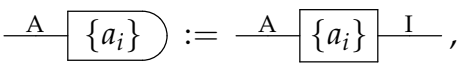


without braces.

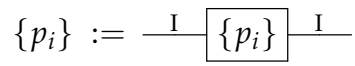
Whenever the trivial system I is involved, we omit the corresponding wire and letter. Specifically, when we have no physical input for our device—which means the trivial system as input—we have a *preparation-test* (a collection of *preparation-events*), which we represent as



namely with a rounded box on its left side. Intuitively, preparation-tests prepare a system in a particular "random state", although we will clarify this statement later. Similarly, when we have no physical output for our device—i.e. the trivial system as output—we have an *observation-test* (a collection of *observation-events*), which we represent as



namely with a rounded box on its right side. Intuitively, observation-tests destroy a system while acquiring some information from it, so they are related to demolition measurements. Finally, if we have a test $\{p_i\}_{i\in X}$ from the trivial system to itself, we omit both the wires and the box.



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

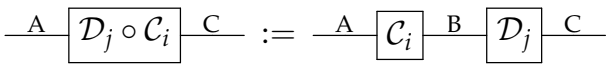
If a test is deterministic, we omit the braces and simply write $\mathcal{C}$ instead of $\{\mathcal{C}\}$. In a non-deterministic test, we cannot predict which particular outcome we will obtain. On the contrary, the outcome of a deterministic test is completely determined. Since we are not able to predict the outcome of non-deterministic tests, we set up a probabilistic structure that enables us to define probabilities for the various outcomes. We will address this issue in section 2.2, but first some other notions are needed.

2.1.2 Sequential and parallel composition

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

Definition 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 $X \times Y$.

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

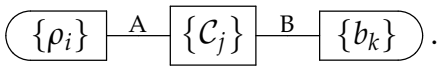


From this notation, and from its operational meaning, we immediately get that sequential composition is associative.

Sequential composition yields a natural ordering on tests. Indeed, some tests are performed first and other later. In graphical language this ordering goes from left to right: every box follows all the others on its left. However, we must not confuse this ordering with "temporal" or "causal" ordering. We will come back to this point in section 2.3.

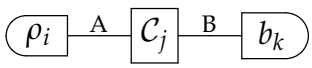
Now let us see an example of sequential composition of tests.

Example 2.1.4. Consider the diagram



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

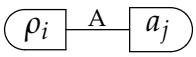
If we wish to express which events actually occurred, we write



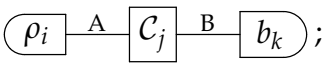
(2.1.1)

This means that the preparation-event ρ_i, the event $\mathcal{C}_j$, and the observation-event b_k occurred.

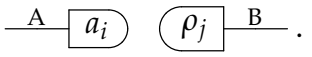
We will often make use of the following short-hand notations, inspired by quantum theory, to mean some common diagrams occurring in our analysis.

1. $(a_j|\rho_i) :=$ 

;

2. $(b_k|\mathcal{C}_j|\rho_i) :=$ 

;

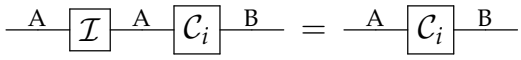
3. $|\rho_j)\,(a_i| :=$ 

.

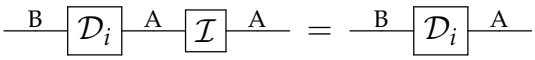
Let us now define the identity test.

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



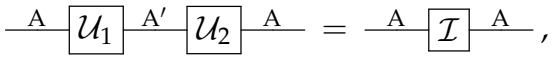
for every $\mathcal{D}_i$. According to this definition, it is clear that for every system A the identity test $\mathcal{I}_\mathrm{A}$ is unique.

Applying the identity test is just like doing nothing. For this reason we will often omit the box for the identity test, and write just a plain wire.

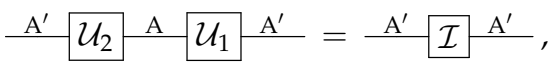
We sometimes want to "identify" similar systems, namely systems that behave exactly in the same way from an operational point of view, yet they are distinct. In quantum mechanics, for example, we can consider the polarisation of a photon and the spin of an electron. Although they are completely different physical systems, they are described by the same Hilbert space.[1]

[1] 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 A ≈ 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

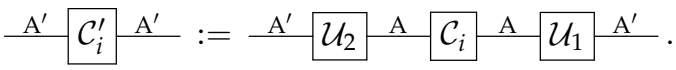


where $\mathcal{I}_{\mathrm{A}}$ is the identity test on A, and



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

If A ≈ 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. System composition has the following properties.

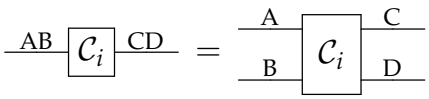
1. AI = IA = A for every system A, where I is the trivial system;
2. AB ≈ BA for all systems A and B;
3. A (BC) = (AB) C for all systems A, B, C.

These properties have a fairly intuitive meaning.

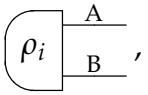
1. When we combine a system with "nothing", we still have the original system.
2. The composition of systems does not depend on the order we compose them.
3. This particular form of associativity allows us to write simply ABC, without parentheses. Again, the order of composition is irrelevant.

We represent composite systems diagrammatically as a collection of wires one under another. We will typically omit the wire for the trivial system.

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

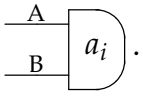


By property 2, it is completely irrelevant to write A or B on the upper input wire, and the same holds for every wire. For composite systems we depict preparation-events as



,

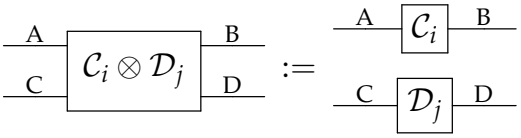
and observation-events as



.

Now we can define the parallel composition of tests.

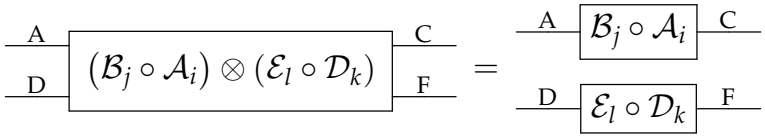
Definition 2.1.8. Let $\{\mathcal{C}_i\}_{i\in X}$ be a test from A to B, and let $\{\mathcal{D}_j\}_{j\in Y}$ be a test from C to D. The *parallel composition* $\{\mathcal{C}_i \otimes \mathcal{D}_j\}_{(i,j)\in X\times Y}$ (or tensor product) is a test from AC to BD with outcome set $X \times Y$, and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

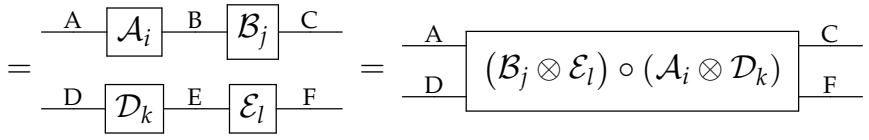


.

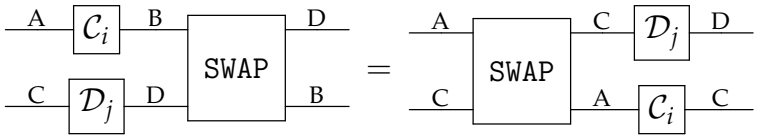
Again, from its operational meaning, it is immediate that parallel composition is associative. Note that this is captured by the graphical notation we are using.

We can combine parallel and sequential composition: 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





Let us analyse the properties of the deterministic test that intertwines system AB and BA. In practice, it swaps system A and system B, so we call it SWAP. Clearly swapping the systems twice yields the original system, thus $\texttt{SWAP}^{-1} = \texttt{SWAP}$. Moreover it swaps the events in a parallel composition.



Note that we can compose preparation-tests only in parallel; the same holds for observation-tests. We will often write sequential composition as a product: 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 $\mathcal{D}_j\mathcal{C}_i$.

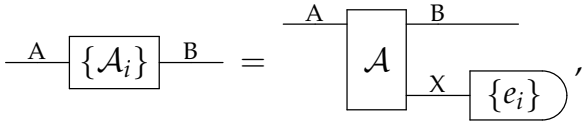
Now we can define operational theories.

Definition 2.1.9. An *operational theory* is given by a collection of systems, closed under composition, and a collection of tests, closed under sequential and parallel composition.

It is easy to see, from the properties presented above, that an operational theory is described by a strict symmetric monoidal category [67, 71, 72, 80, 123].

In the following we will assume that tests $\{\mathcal{A}_i\}$ from system A to system B are performed through a deterministic interaction between the systems and the measurement apparatus X, which is read by the observer with an observation-test.

Assumption 2.1.10 (Physicalisation of readout [71])**.** *Every test* $\{\mathcal{A}_i\}_{i\in X}$ *can be realised as follows:*



where $\mathcal{A}$ *is a deterministic test from* A *to* BX*, and* $\{e_i\}_{i\in X}$ *is an observation-test on* X*.*

2.2 The probabilistic structure

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

Definition 2.2.1. An *operational-probabilistic theory* (OPT) is an operational theory where, for every test $\{p_i\}_{i\in X}$ on the trivial system I, one has $p_i \in [0, 1]$ and $\sum_{i\in X} p_i = 1$.

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

This definition states that every event from I to itself can be interpreted as a probability. Consequently, we can associate a probability with every diagram with no external wires.

Example 2.2.2. Let us consider eq. (2.1.1) again. It is a diagram without external wires; 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 focus only on OPTs, namely on operational theories with a probabilistic structure.

Sometimes it happens that we obtain the same physical configuration with different experimental procedures. For instance, in quantum theory consider the mixed state $\rho = \frac{1}{2}\mathbf{1}$ of a qubit. This state can be prepared either by having no information on the state of the system, or by taking the partial trace of one of the Bell states. The issue is now how to distinguish different experimental preparations, or find out when they are equivalent.

Let us consider, for instance, preparation-events. If we compose a preparation-event with an observation-event, we get $p_{ij} = (a_j|\rho_i)$, the joint probability of having the preparation-event ρ_i and the observation-event a_j.

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}.

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

Similarly, if we have an observation-event 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}.

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

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

In this vein, we can introduce an equivalence relation between preparation-events (and similarly between observation-events). If ρ_i and σ_j are two preparation-events on system A, we say that they are *tomographically equivalent* (or *tomographically indistinguishable*), written as $\rho_i \sim \sigma_j$, 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 that they are *tomographically equivalent* (or *tomographically indistinguishable*), written as $a_i \sim b_j$, 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 2.2.3. Equivalence classes of tomographically indistinguishable preparation-events are called *states*. The set of states of system A is denoted as $\mathsf{St}(\mathrm{A})$.

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

Therefore, two states ρ_1 and ρ_2 of system A are equal if and only if $(a|\rho_1) = (a|\rho_2)$ for every effect $a \in \mathsf{Eff}(\mathrm{A})$. Similarly, two effects a_1 and a_2 of system A are equal if and only if $(a_1|\rho) = (a_2|\rho)$ for every state $\rho \in \mathsf{St}(\mathrm{A})$. The process of reconstructing a state (or an effect) from the statistics of measurements is called *tomography*.

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.

Example 2.2.4. The trivial system has a unique deterministic state and a unique deterministic effect: it is the number 1. All the other effects and states are elements of $[0, 1]$.

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

Example 2.2.5. In quantum mechanics we 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 $\operatorname{tr} \rho_i \leq 1$. This is essentially a random preparation: a state ρ_i is prepared with a probability given by $\operatorname{tr} \rho_i$. Therefore in quantum mechanics $\mathsf{St}(\mathrm{A})$ is the set of trace-class positive operators with trace less than or equal to one.

An effect is, instead, represented by a positive operator P, with $P \leq \mathbf{1}$, where $\mathbf{1}$ is the identity operator. Observation-tests are then POVMs. The pairing between states and effect is given by the trace: $(P|\rho) = \operatorname{tr} P\rho$. 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).

According to definition 2.2.3, states and effects are in fact real-valued functions. As a consequence we can take linear combinations of them with real coefficients; in other words they span real vector spaces. Let $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$ be the vector space spanned by states, and let $\mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$ be the vector space spanned by effects. These vector spaces can be finite- or infinite-dimensional. In our presentation, to avoid mathematical subtleties and simplify the treatment, we will assume that they are finite-dimensional. Clearly, $\mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$ is the dual vector space of $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$ and $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$ is the dual vector space of $\mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$. For finite-dimensional vector spaces, we have $\dim \mathsf{St}_{\mathbb{R}}(\mathrm{A}) = \dim \mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$.

We can extend the notion of tomography to the full vector space $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$: $\xi, \eta \in \mathsf{St}_{\mathbb{R}}(\mathrm{A})$ are equal if and only if, for every $a \in \mathsf{Eff}(\mathrm{A})$ (or for every $X \in \mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$) we have $(a|\xi) = (a|\eta)$ (or $(X|\xi) = (X|\eta)$). A similar fact holds for the vector space $\mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$. The fact that these vector spaces have finite dimension means that a finite set of effects (resp. states) is sufficient to do tomography on states (resp. effects).

Remark 2.2.6. Consider the states of a bipartite system AB. Clearly to do tomography one takes the effects of AB. Some of them will be of the

product form, i.e. $a \otimes b$, where $a \in \mathsf{Eff}(\mathrm{A})$ and $b \in \mathsf{Eff}(\mathrm{B})$, some others will not. If considering only product effects is enough to do tomography on all bipartite states, we say that the theory satisfies Local Tomography [61, 67, 70, 84, 86, 107, 124]. Quantum theory on complex Hilbert spaces satisfies Local Tomography, but quantum theory on real Hilbert spaces does *not* [106–109]. If a theory satisfies Local Tomography, we have $\mathsf{St}_{\mathbb{R}}(\mathrm{AB}) = \mathsf{St}_{\mathbb{R}}(\mathrm{A}) \otimes \mathsf{St}_{\mathbb{R}}(\mathrm{B})$ [67]. In this case, product states are enough to characterise all states of a composite system, because every state of the composite system can be written as a linear combination of product states. In other words, there are no "genuinely new" states arising when systems are composed. In the following we will *not* assume Local Tomography as an axiom, in fact we will provide concrete examples of theories that violate it.

We can also take linear combinations with non-negative coefficients, they are called *conical combinations*. Using conical combinations, states (resp. effects) span a convex cone, the cone of states $\mathsf{St}_+(\mathrm{A})$ (resp. the cone of effects $\mathsf{Eff}_+(\mathrm{A})$). These two cones K are *proper* by construction: it means that

1. $0 \in K$;
2. let $\xi \neq 0$; if $\xi \in \mathsf{St}_+(\mathrm{A})$, then $-\xi \notin \mathsf{St}_+(\mathrm{A})$;
3. for every vector $\xi \in V$, there exist $\xi_+, \xi_- \in K$ such that $\xi = \xi_+ - \xi_-$.

Here K denotes the cones $\mathsf{St}_+(\mathrm{A})$ and $\mathsf{Eff}_+(\mathrm{A})$, and V the corresponding vector spaces $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$ and $\mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$. Once we define the cone of states, we can consider the *dual cone* $\mathsf{St}_+^*(\mathrm{A})$, defined as the set of linear functionals X such that $(X|\xi) \geq 0$ for all $\xi \in \mathsf{St}_+(\mathrm{A})$. Clearly, the elements of the cone of effects are in the dual cone, because a conical combination of effects yields a non-negative number when applied to a state, so $\mathsf{Eff}_+(\mathrm{A}) \subseteq \mathsf{St}_+^*(\mathrm{A})$.

Definition 2.2.7 (No-restriction hypothesis [67])**.** We say that a theory is *non-restricted*, or that it satisfies the no-restriction hypothesis, if $\mathsf{Eff}_+(\mathrm{A}) = \mathsf{St}_+^*(\mathrm{A})$ for every system.

While this may look just a statement of mathematical interest, it has some important physical implications. Consider the subset of $\mathsf{St}_+^*(\mathrm{A})$ made of linear functionals f such that $(f|\rho) \in [0,1]$ for all states ρ. In a non-restricted theory, these elements f are also valid effects. In other

words, the no-restriction hypothesis states that every mathematically allowed effect is also a physical effect. Clearly the no-restriction hypothesis concerns more the mathematical structure of the theory than its operational one. Indeed, it is the duty of the physical theory to specify what objects are to be considered physical effects, even if they are admissible in principle, based on their mathematical properties. For this reason, the no-restriction hypothesis has been questioned various times on the basis of its lack of operational motivation [67,125,126]. Moreover, recently it has been shown that theories with almost quantum correlations [127] violate it [128].

Example 2.2.8. The trivial system of every theory has remarkable properties. We know that states are in $[0,1]$, therefore $\mathsf{St}_{\mathbb{R}}(\mathrm{I}) = \mathbb{R}$. The cone $\mathsf{St}_{+}(\mathrm{I})$ is the set of non-negative numbers. Similarly, $\mathsf{Eff}_{\mathbb{R}}(\mathrm{I}) = \mathbb{R}$, and $\mathsf{Eff}_{+}(\mathrm{I}) = \mathbb{R}_{\geq 0}$.

Example 2.2.9. Let us see what $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$ and $\mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$ are in finite-dimensional quantum theory, namely when the Hilbert space is finite-dimensional ($\mathcal{H} \approx \mathbb{C}^n$, for $n \geq 2$). $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$ is the vector space of hermitian matrices of order n. It is a real vector space with dimension n^2. $\mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$ is again the vector space of hermitian matrices of order n.

Instead, the cones $\mathsf{St}_{+}(\mathrm{A})$ and $\mathsf{Eff}_{+}(\mathrm{A})$ are both the convex cone of positive semidefinite matrices. In quantum theory, the no-restriction hypothesis is valid: $\mathsf{Eff}_{+}(\mathrm{A}) = \mathsf{St}_{+}^{*}(\mathrm{A})$.

Now it is time to move our attention to the equivalence classes of tomographically indistinguishable events for general tests.

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

$$\widehat{\mathcal{C}}_i : \rho_{\mathrm{A}} \mapsto \mathcal{C}_i \rho_{\mathrm{A}}, \tag{2.2.1}$$

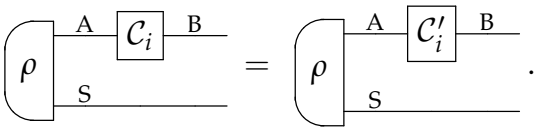
for every $\rho_{\mathrm{A}} \in \mathsf{St}(\mathrm{A})$. Note that $\mathcal{C}_i \rho_{\mathrm{A}}$ is a state of B. We want to check whether the linear extension of (2.2.1) is well defined. If $\xi \in \mathsf{St}_{\mathbb{R}}(\mathrm{A})$, we can express it as a linear combination of states, $\xi = \sum_j \lambda_j \rho_j$, where $\lambda_j \in \mathbb{R}$ for every j. The obvious linear extension of (2.2.1) is $\widehat{\mathcal{C}}_i \xi := \sum_j \lambda_j \widehat{\mathcal{C}}_i \rho_j$. The problem is that, in general, ξ does *not* have a unique expression as a linear combination of states. Suppose that $\xi = \sum_j \lambda_j \rho_j$ and $\xi = \sum_j \mu_j \sigma_j$,

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

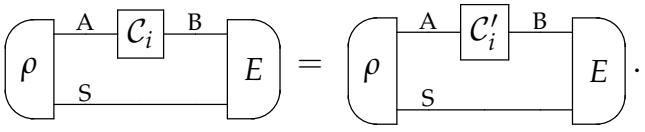
We have $\sum_j \lambda_j \rho_j = 0$ if and only if $\sum_j \lambda_j \left(a \middle| \rho_j\right) = 0$ for every effect $a \in \mathsf{Eff}(\mathrm{A})$. Let b be an arbitrary effect on B. Then $b\widehat{\mathcal{C}_i}$ is an effect on A, therefore $\sum_j \lambda_j \left(b \middle| \widehat{\mathcal{C}_i}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

Likewise, for every system S, the event $\mathcal{C}_i \otimes \mathcal{I}_{\mathrm{S}}$ from AS to BS will induce a linear operator from $\mathsf{St}_{\mathbb{R}}(\mathrm{AS})$ to $\mathsf{St}_{\mathbb{R}}(\mathrm{BS})$. Two events $\mathcal{C}_i$ and $\mathcal{C}'_i$ from A to B are *tomographically indistinguishable* if, for every system S, the linear operators associated with $\mathcal{C}_i \otimes \mathcal{I}_{\mathrm{S}}$ and $\mathcal{C}'_i \otimes \mathcal{I}_{\mathrm{S}}$ are the same.

In other words, for every system S, and every state $\rho \in \mathsf{St}(\mathrm{AS})$,



Recalling the definition of equal states, $\mathcal{C}_i$ and $\mathcal{C}'_i$ are tomographically indistinguishable if and only if, for every system S, every state $\rho \in \mathsf{St}(\mathrm{AS})$, and every effect $E \in \mathsf{St}(\mathrm{BS})$ one has



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

Definition 2.2.10. 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 2.2.11. One may wonder why we have given such a definition of tomographically indistinguishable events, involving an ancillary system S. The most obvious way of defining tomographic indistinguishability

would have been to say that $\mathcal{C}_i$ and $\mathcal{C}'_i$ are indistinguishable if $\mathcal{C}_i\rho = \mathcal{C}'_i\rho$ for every $\rho \in \mathsf{St}(\mathrm{A})$. Actually, this is not enough for OPTs. Indeed, Wootters provided a counterexample concerning quantum mechanics with real Hilbert space [108]. It can be shown that there exist two events that are locally indistinguishable, but if we add an ancillary system, they produce distinct output states. The condition $\mathcal{C}_i\rho = \mathcal{C}'_i\rho$ for every $\rho \in \mathsf{St}(\mathrm{A})$ is sufficient to identify indistinguishable events if the theory satisfies Local Tomography (see [67] for further details).

Again, we will assume that equivalence classes have been taken from the very beginning, so we will consider tests as collections of transformations. Transformations span a real vector space, denoted by $\mathsf{Transf}_{\mathbb{R}}(\mathrm{A},\mathrm{B})$.

Definition 2.2.12. A deterministic transformation $\mathcal{C} \in \mathsf{Transf}(\mathrm{A},\mathrm{B})$ is called *channel*.

We will denote the set of channels from A to B by $\mathsf{DetTransf}(\mathrm{A},\mathrm{B})$ (or by $\mathsf{DetTransf}(\mathrm{A})$ if $\mathrm{B} = \mathrm{A}$).

Among all possible channels, reversible ones are particularly important.

Definition 2.2.13. 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 $\mathcal{U}^{-1}\mathcal{U} = \mathcal{I}_{\mathrm{A}}$ and $\mathcal{U}\mathcal{U}^{-1} = \mathcal{I}_{\mathrm{B}}$.

Clearly, reversible channels on A form a group, denoted G_{A}. Now, we can rephrase the definition of operationally equivalent systems: two systems A and A′ are operationally equivalent if there exists a reversible channel from A to A′.

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

Example 2.2.14. A test in quantum theory from $\mathcal{H}_{\mathrm{A}}$ to $\mathcal{H}_{\mathrm{B}}$ is a collection of completely positive, trace non-increasing linear maps $\{\mathcal{C}_k\}$, called quantum operations [116, 117], such that $\sum_k \mathcal{C}_k$ is a trace-preserving map. Each quantum operation maps linear operators on $\mathcal{H}_{\mathrm{A}}$ into linear operators on $\mathcal{H}_{\mathrm{B}}$. A test is a quantum instrument, namely a collection of quantum operations [116].

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

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

2.2.1 Purity and coarse-graining

Even in an abstract probabilistic theory, it makes sense to define pure and mixed states, or, more generally, pure and non-pure transformations. The idea behind it is *coarse-graining*. Let us clarify this idea with the example of the roll of a die [129]. In this random experiment, there are some atomic events, which cannot be decomposed further: they are the numbers from 1 to 6. So, an atomic event is, for example, "the outcome of the roll is 2". However, we can consider the event "the outcome of the roll is odd". This event is the union of the atomic events relative to 1, 3, 5. We have just done 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 2.2.15. A test $\{\mathcal{B}_j\}_{j\in Y}$ is a *coarse-graining* of the test $\{\mathcal{A}_i\}_{i\in X}$ if there is a partition $\{X_j\}_{j\in Y}$ of X such that $\mathcal{B}_j = \sum_{i\in X_j} \mathcal{A}_i$. In this case, we say that $\{\mathcal{A}_i\}_{i\in X}$ is a *refinement* of $\{\mathcal{B}_j\}_{j\in Y}$.

As we can see, this definition gives a precise characterisation of what we mean by "joining together outcomes". A test that refines another extracts more information than the other. It is clear that if $\{\mathcal{B}_j\}_{i\in Y}$ is a coarse-graining of the test $\{\mathcal{A}_i\}_{i\in X}$, it has fewer outcomes. This concept is easily explained in fig. 2.1.

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 *all* the outcomes $i \in X$. Then we obtain the channel $\mathcal{C} = \sum_{i\in X} \mathcal{C}_i$ from A to B, which is called the channel associated with the test $\{\mathcal{C}_i\}_{i\in X}$. Similarly, we can obtain a deterministic state by summing all the states in a preparation-test; and we can get a deterministic effect by summing all the effects in an observation-test.

We can consider also refinements of single transformations.

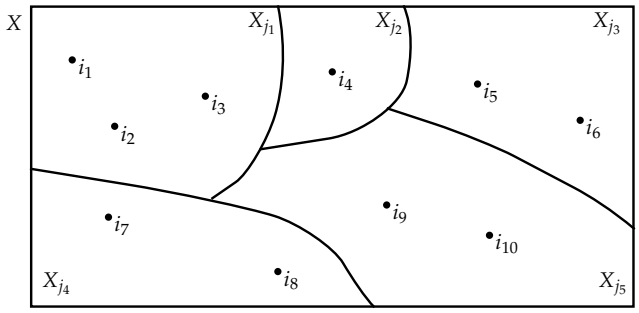


Figure 2.1: The outcome set X of the test $\{\mathcal{A}_i\}_{i\in X}$ has 10 outcomes. To perform a coarse-graining of it, we lump together some of its outcomes, relabelling them as a new outcome. For example, the outcomes i_1, i_2, and i_3 are relabelled as j_1. This gives rise to a partition $\{X_j\}_{j\in Y}$ of X. We associate a new transformation with each set in the partition, such that it is the sum of the transformations associated with the outcomes contained in that set. Thus $\mathcal{B}_{j_1} = \mathcal{A}_{i_1} + \mathcal{A}_{i_2} + \mathcal{A}_{i_3}$. The new test $\{\mathcal{B}_j\}_{j\in Y}$ has 5 outcomes.

Definition 2.2.16. 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 $\mathcal{C}$.

We can always obtain a refinement of a transformation $\mathcal{T}$ by taking a subset of a test, made of $\{p_i\mathcal{T}\}_{i\in X_0}$, with the property that $p_i \in (0,1]$ for every $i \in X_0$, and $\sum_{i\in X_0} p_i = 1$. Some transformations cannot be refined further, and they admit only trivial refinements of this form above.

Definition 2.2.17. A transformation $\mathcal{T}$ is *pure* if it has only trivial refinements.

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

Clearly, this definition applies also to states, which are particular transformations from the trivial system I to a system A. We will denote the set of pure states of system A by $\mathsf{PurSt}(\mathrm{A})$. The non-pure states are called *mixed*. In this way, a pure state represents maximal knowledge about the preparation of a system, whereas a mixed state expresses some lack of information. Similarly, we will denote the set of pure effects of system A by $\mathsf{PurEff}(\mathrm{A})$.

Let us see some examples in quantum theory.

Example 2.2.18. If we diagonalise a density operator $\rho = \sum_j p_j \left|\psi_j\right\rangle \left\langle\psi_j\right|$, each term $p_j \left|\psi_j\right\rangle \left\langle\psi_j\right|$ is a refinement[2] of ρ. More generally, a refinement of ρ is a state σ such that $\sigma \leq \rho$. This means that the support[3] of σ is contained in the support of ρ (see [130, appendix A.1] for a proof). A pure state is a density operator $\lambda \left|\psi\right\rangle \left\langle\psi\right|$, with $\lambda \in (0, 1]$, viz. proportional to a rank-one projector.

In quantum mechanics, we can associate Kraus operators $\{M_k\}$ with every quantum operation $\mathcal{C}$, such that $\mathcal{C}(\rho) = \sum_k M_k \rho M_k^\dagger$, for every state ρ [116,117]. A quantum operation is pure if and only it has only one Kraus operator.

Reversible channels are *not* pure in general, unless some axioms are imposed on the theory (see section 4.1).

Example 2.2.19. Consider finite-dimensional classical theory. Here states are vectors of non-negative numbers, whose entries sum to a number less than or equal to 1. In symbols, states are vectors $\mathbf{p} \in \mathbb{R}^d$, such that $p_i \geq 0$ for all $i = 1, \ldots, d$, and $\sum_{i=1}^d p_i \leq 1$. The cone of states is spanned by the d pure states δ_i, where δ_i denotes the vector with all zero entries, except the ith, which is 1. The identity channel can be written as $\mathcal{I} = \sum_{i=1}^d \left|\delta_i\right) \left(\delta_i^\dagger\right|$, where the $\delta_i^\dagger$'s are the effects such that $\left(\delta_i^\dagger \middle| \delta_j\right) = \delta_{ij}$, where δ_{ij} is Kronecker delta. Indeed, every element of the cone of states can be written as $\xi = \sum_{i=1}^d \lambda_i \delta_i$, where $\lambda_i \geq 0$, and

$$\mathcal{I}\xi = \sum_{i=1}^d \sum_{j=1}^d \lambda_j \left|\delta_i\right) \left(\delta_i^\dagger \middle| \delta_j\right) = \sum_{i=1}^d \lambda_i \delta_i = \xi.$$

Since classical theory satisfies Local Tomography, this is enough to conclude that $\sum_{i=1}^d \left|\delta_i\right) \left(\delta_i^\dagger\right|$ is the identity channel. This means that the identity channel is the coarse-graining of the test $\left\{\left|\delta_i\right) \left(\delta_i^\dagger\right|\right\}_{i=1}^d$. Therefore the identity channel in classical theory, albeit reversible, is *not* pure.

However, although not necessarily pure themselves, reversible channels send pure states into pure states [100]. They do not alter the "purity" of a state: they also map mixed states into mixed states.

[2] With a little abuse of terminology we also say that $\left|\psi_j\right\rangle \left\langle\psi_j\right|$ is a refinement of ρ.

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

Lemma 2.2.20. *Let $\mathcal{U}$ be a reversible channel from* A *to* B*. Then $\psi \in \mathsf{St}(\mathrm{A})$ is pure if and only if $\mathcal{U}\psi \in \mathsf{St}(\mathrm{B})$ is pure.*

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

$$\mathcal{U}\psi = \sum_i \rho_i \tag{2.2.2}$$

Let us show that each refinement ρ_i of $\mathcal{U}\psi$ is trivial, that is $\rho_i = p_i \mathcal{U}\psi$, for some $p_i \in (0,1]$, with $\sum_i p_i = 1$. By applying $\mathcal{U}^{-1}$ to both sides of eq. (2.2.2), we have $\psi = \sum_i \mathcal{U}^{-1}\rho_i$. Since ψ is pure, each refinement $\mathcal{U}^{-1}\rho_i$ is trivial, namely

$$\mathcal{U}^{-1}\rho_i = p_i \psi, \tag{2.2.3}$$

for some $p_i \in (0,1]$, with $\sum_i p_i = 1$. By applying $\mathcal{U}$ to both sides of eq. (2.2.3), we have $\rho_i = p_i \mathcal{U}\psi$. Since every refinement $\mathcal{U}\psi$ is trivial, $\mathcal{U}\psi$ is pure.

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

A similar statement holds also for effects: $b \in \mathsf{Eff}(\mathrm{B})$ is pure if and only if $b\mathcal{U} \in \mathsf{Eff}(\mathrm{A})$ is pure.

2.2.2 Norms for states, effects, and transformations

We can define a norm in the vector space of states $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$. It is defined as follows [67].

Definition 2.2.21. Let $\xi \in \mathsf{St}_{\mathbb{R}}(\mathrm{A})$. The *operational norm* of ξ is

$$\|\xi\| := \sup_{a \in \mathsf{Eff}(\mathrm{A})} (a|\xi) - \inf_{a \in \mathsf{Eff}(\mathrm{A})} (a|\xi)\,.$$

Let us show that it is indeed a norm. First of all, note that $\|\xi\| \geq 0$ because $\sup_{a \in \mathsf{Eff}(\mathrm{A})} (a|\xi) \geq \inf_{a \in \mathsf{Eff}(\mathrm{A})} (a|\xi)$. Then, let us show that $\|\xi\| = 0$ only if $\xi = 0$. If $\|\xi\| = 0$, then $\sup_{a \in \mathsf{Eff}(\mathrm{A})} (a|\xi) = \inf_{a \in \mathsf{Eff}(\mathrm{A})} (a|\xi)$. Now, we have that $\sup_{a \in \mathsf{Eff}(\mathrm{A})} (a|\xi) \geq 0$, and $\inf_{a \in \mathsf{Eff}(\mathrm{A})} (a|\xi) \leq 0$. Indeed, write ξ as $\xi = \xi_+ - \xi_-$, where $\xi_+, \xi_- \in \mathsf{St}_+(\mathrm{A})$. Then $(a|\xi) = (a|\xi_+) - (a|\xi_-)$, where $(a|\xi_+)$ and $(a|\xi_-)$ are both non-negative. As to the supremum, note that $(a|\xi) \geq (a|-\xi_-)$, therefore

$$\sup_{a \in \mathsf{Eff}(\mathrm{A})} (a|\xi) \geq \sup_{a \in \mathsf{Eff}(\mathrm{A})} (a|-\xi_-) = 0.$$

As to the infimum, note that $(a|\xi) \leq (a|\xi_+)$, whence

$$\inf_{a\in\mathsf{Eff}(\mathrm{A})} (a|\xi) \leq \inf_{a\in\mathsf{Eff}(\mathrm{A})} (a|\xi_+) = 0.$$

Then the only possibility of having $\sup_{a\in\mathsf{Eff}(\mathrm{A})} (a|\xi) = \inf_{a\in\mathsf{Eff}(\mathrm{A})} (a|\xi)$ is when

$$\sup_{a\in\mathsf{Eff}(\mathrm{A})} (a|\xi) = \inf_{a\in\mathsf{Eff}(\mathrm{A})} (a|\xi) = 0. \tag{2.2.4}$$

Since

$$\sup_{a\in\mathsf{Eff}(\mathrm{A})} (a|\xi) \geq (a|\xi) \geq \inf_{a\in\mathsf{Eff}(\mathrm{A})} (a|\xi)\,,$$

eq. (2.2.4) implies $(a|\xi) = 0$ for every effect a, so $\xi = 0$.

Let us prove that $\|\lambda\xi\| = |\lambda|\,\|\xi\|$ for every $\lambda \in \mathbb{R}$, and every ξ. Let $\lambda \geq 0$. Then $\sup_{a\in\mathsf{Eff}(\mathrm{A})} (a|\lambda\xi) = \lambda \sup_{a\in\mathsf{Eff}(\mathrm{A})} (a|\xi)$, and similarly $\inf_{a\in\mathsf{Eff}(\mathrm{A})} (a|\lambda\xi) = \lambda \inf_{a\in\mathsf{Eff}(\mathrm{A})} (a|\xi)$. Therefore

$$\|\lambda\xi\| = \lambda \sup_{a\in\mathsf{Eff}(\mathrm{A})} (a|\xi) - \lambda \inf_{a\in\mathsf{Eff}(\mathrm{A})} (a|\xi) = \lambda\,\|\xi\|\,.$$

Now, let $\lambda < 0$. In this case $\sup_{a\in\mathsf{Eff}(\mathrm{A})} (a|\lambda\xi) = \lambda \inf_{a\in\mathsf{Eff}(\mathrm{A})} (a|\xi)$, and similarly $\inf_{a\in\mathsf{Eff}(\mathrm{A})} (a|\lambda\xi) = \lambda \sup_{a\in\mathsf{Eff}(\mathrm{A})} (a|\xi)$. Hence

$$\|\lambda\xi\| = \lambda \inf_{a\in\mathsf{Eff}(\mathrm{A})} (a|\xi) - \lambda \sup_{a\in\mathsf{Eff}(\mathrm{A})} (a|\xi) = (-\lambda)\,\|\xi\|\,.$$

In conclusion $\|\lambda\xi\| = |\lambda|\,\|\xi\|$.

Finally, we need to prove the triangle inequality: $\|\xi + \eta\| \leq \|\xi\| + \|\eta\|$, for every ξ and η. We have

$$\|\xi + \eta\| = \sup_{a\in\mathsf{Eff}(\mathrm{A})} (a|\xi + \eta) - \inf_{a\in\mathsf{Eff}(\mathrm{A})} (a|\xi + \eta)\,. \tag{2.2.5}$$

Now

$$\sup_{a\in\mathsf{Eff}(\mathrm{A})} (a|\xi + \eta) \leq \sup_{a\in\mathsf{Eff}(\mathrm{A})} (a|\xi) + \sup_{a\in\mathsf{Eff}(\mathrm{A})} (a|\eta) \tag{2.2.6}$$

and

$$\inf_{a\in\mathsf{Eff}(\mathrm{A})} (a|\xi + \eta) \geq \inf_{a\in\mathsf{Eff}(\mathrm{A})} (a|\xi) + \inf_{a\in\mathsf{Eff}(\mathrm{A})} (a|\eta)\,, \tag{2.2.7}$$

as it is not hard to show. Thus, putting eqs. (2.2.5), (2.2.6), and (2.2.7) together, $\|\xi+\eta\| \leq \|\xi\| + \|\eta\|$. This shows that the operational norm is indeed a norm. In subsection 4.4.3 we will show that this norm is in fact the 1-norm (or the trace norm in quantum theory). For the trivial system, where $\mathsf{St}_{\mathbb{R}}(\mathrm{I}) = \mathbb{R}$, we have that $\|\xi\| = |\xi|$. Indeed, either $\xi \in \mathsf{St}_+(\mathrm{I})$ or $-\xi \in \mathsf{St}_+(\mathrm{I})$. Therefore

$$\|\xi\| = \sup_{k\in[0,1]} k\,|\xi| = |\xi|\,,$$

where we have used the fact that the effects of the trivial system are elements in $[0,1]$.

Clearly, for a state, $\|\rho\| \leq 1$, because effects yield probabilities when applied to a state. Therefore the set of states is bounded.

Definition 2.2.22. A state $\rho \in \mathsf{St}(\mathrm{A})$ is *normalised* if $\|\rho\| = 1$.

We will denote the set of normalised states of system A by $\mathsf{St}_1(\mathrm{A})$, and the set of normalised *pure* states by $\mathsf{PurSt}_1(\mathrm{A})$.

This norm is non-decreasing under the action of transformations [67, lemma 1].

Proposition 2.2.23. *For every vector $\xi \in \mathsf{St}_{\mathbb{R}}(\mathrm{A})$, $\|\mathcal{T}\xi\|_{\mathrm{B}} \leq \|\xi\|_{\mathrm{A}}$, where $\mathcal{T}$ is a transformation from* A *to* B*. If $\mathcal{T}$ is a reversible channel, then one has the equality.*

Proof. By definition

$$\|\mathcal{T}\xi\|_{\mathrm{B}} = \sup_{b\in\mathsf{Eff}(\mathrm{B})} (b|\mathcal{T}|\xi) - \inf_{b\in\mathsf{Eff}(\mathrm{B})} (b|\mathcal{T}|\xi)\,.$$

Now, $b\mathcal{T}$ is an effect of A, so $\sup_{b\in\mathsf{Eff}(\mathrm{B})} (b|\mathcal{T}|\xi) \leq \sup_{a\in\mathsf{Eff}(\mathrm{A})} (a|\xi)$, and $\inf_{b\in\mathsf{Eff}(\mathrm{B})} (b|\mathcal{T}|\xi) \geq \inf_{a\in\mathsf{Eff}(\mathrm{A})} (a|\xi)$. We conclude that

$$\|\mathcal{T}\xi\|_{\mathrm{B}} \leq \sup_{a\in\mathsf{Eff}(\mathrm{A})} (a|\xi) - \inf_{a\in\mathsf{Eff}(\mathrm{A})} (a|\xi) = \|\xi\|_{\mathrm{A}}\,.$$

If $\mathcal{T}$ is a reversible channel, then one has

$$\|\xi\|_{\mathrm{A}} = \left\|\mathcal{T}^{-1}\mathcal{T}\xi\right\|_{\mathrm{A}} \leq \|\mathcal{T}\xi\|_{\mathrm{B}}\,,$$

which means $\|\mathcal{T}\xi\|_{\mathrm{B}} = \|\xi\|_{\mathrm{A}}$. □

In particular, if we have $\xi = \rho - \sigma$, with ρ and σ two states, for every transformation

$$\|\mathcal{T}\rho - \mathcal{T}\sigma\| \leq \|\rho - \sigma\|,$$

thus recovering the same statement as in quantum theory.

The norm of states is the starting point to give the definition of a norm for the elements of the vector space of transformations.

Definition 2.2.24. Let $\Delta \in \mathsf{Transf}_{\mathbb{R}}(\mathrm{A}, \mathrm{B})$. The *norm* of Δ is

$$\|\Delta\| := \sup_{\mathrm{S}} \sup_{\rho \in \mathsf{St}(\mathrm{AS})} \|(\Delta \otimes \mathcal{I}_{\mathrm{S}})\,\rho\|.$$

Let us show that this is indeed a norm. First of all, note that, for every system S, and every state $\rho \in \mathsf{St}(\mathrm{AS})$,

$$\sup_{\mathrm{S}} \sup_{\rho \in \mathsf{St}(\mathrm{AS})} \|(\Delta \otimes \mathcal{I}_{\mathrm{S}})\,\rho\| \geq \|(\Delta \otimes \mathcal{I}_{\mathrm{S}})\,\rho\| \geq 0,$$

whence $\|\Delta\| \geq 0$. Moreover, the same inequality shows that if $\|\Delta\| = 0$, then $\|(\Delta \otimes \mathcal{I}_{\mathrm{S}})\,\rho\| = 0$, which in turn implies $(\Delta \otimes \mathcal{I}_{\mathrm{S}})\,\rho = 0$ for every system S, and every state $\rho \in \mathsf{St}(\mathrm{AS})$. We conclude that $\Delta = 0$.

Now, let us show that $\|\lambda\Delta\| = |\lambda|\,\|\Delta\|$, for every $\lambda \in \mathbb{R}$. Now,

$$\|\lambda\Delta\| = \sup_{\mathrm{S}} \sup_{\rho \in \mathsf{St}(\mathrm{AS})} \|(\lambda\Delta \otimes \mathcal{I}_{\mathrm{S}})\,\rho\| = \sup_{\mathrm{S}} \sup_{\rho \in \mathsf{St}(\mathrm{AS})} |\lambda|\,\|(\Delta \otimes \mathcal{I}_{\mathrm{S}})\,\rho\| =$$

$$= |\lambda| \sup_{\mathrm{S}} \sup_{\rho \in \mathsf{St}(\mathrm{AS})} \|(\Delta \otimes \mathcal{I}_{\mathrm{S}})\,\rho\| = |\lambda|\,\|\Delta\|.$$

Finally, we prove the triangle inequality: $\|\Delta + \Xi\| \leq \|\Delta\| + \|\Xi\|$, for every $\Delta, \Xi \in \mathsf{Transf}_{\mathbb{R}}(\mathrm{A}, \mathrm{B})$.

$$\|\Delta + \Xi\| = \sup_{\mathrm{S}} \sup_{\rho \in \mathsf{St}(\mathrm{AS})} \|[(\Delta + \Xi) \otimes \mathcal{I}_{\mathrm{S}}]\,\rho\| = \sup_{\mathrm{S}} \sup_{\rho \in \mathsf{St}(\mathrm{AS})} \|(\Delta \otimes \mathcal{I}_{\mathrm{S}})\,\rho + (\Xi \otimes \mathcal{I}_{\mathrm{S}})\,\rho\|.$$

Now, for every system S, and every state $\rho \in \mathsf{St}(\mathrm{AS})$, we have

$$\|(\Delta \otimes \mathcal{I}_{\mathrm{S}})\,\rho + (\Xi \otimes \mathcal{I}_{\mathrm{S}})\,\rho\| \leq \|(\Delta \otimes \mathcal{I}_{\mathrm{S}})\,\rho\| + \|(\Xi \otimes \mathcal{I}_{\mathrm{S}})\,\rho\| \leq$$

$$\leq \sup_{\mathrm{S}} \sup_{\rho \in \mathsf{St}(\mathrm{AS})} \|(\Delta \otimes \mathcal{I}_{\mathrm{S}})\,\rho\| + \sup_{\mathrm{S}} \sup_{\rho \in \mathsf{St}(\mathrm{AS})} \|(\Xi \otimes \mathcal{I}_{\mathrm{S}})\,\rho\|.$$

In conclusion, $\|\Delta + \Xi\| \leq \|\Delta\| + \|\Xi\|$, thus proving that we are dealing with an actual norm. Note that for a generic transformation $\mathcal{T} \in \mathsf{Transf}(\mathrm{A}, \mathrm{B})$ we have

$$\|\mathcal{T}\| = \sup_{\mathrm{S}} \sup_{\rho \in \mathsf{St}(\mathrm{AS})} \|(\mathcal{T} \otimes \mathcal{I}_{\mathrm{S}})\, \rho\| \leq 1,$$

because $(\mathcal{T} \otimes \mathcal{I}_{\mathrm{S}})\, \rho$ is a physical state. Therefore the set of transformations is bounded.

When $\Delta \in \mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$ the norm of Δ takes a simpler form [67, lemma 8].

Proposition 2.2.25. *Let $X \in \mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$. Then*

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

Note that this norm is the operator norm of the linear functional X on the set of states. Clearly, the set of effects $\mathsf{Eff}(\mathrm{A})$ of a generic system A is bounded too, for $\|a\| \leq 1$. The effects a such that $\|a\| = 1$ are called *normalised*, and their set for system A will be denoted by $\mathsf{Eff}_1(\mathrm{A})$. Similarly, the set of normalised pure effects will be denoted by $\mathsf{PurEff}_1(\mathrm{A})$.

2.2.3 Setting up a topology

The definition of tomography for states, effects, and transformations naturally yields a topology. The idea is that a sequence of states, effects, transformations converges to a limit if the states, effects, transformations in the sequence become tomographically indistinguishable from the limit. Let us clarify this idea separately for states, effects, and transformations.

We say that a sequence of states $\{\rho_n\}$ of system A converges to $\rho \in \mathsf{St}_{\mathbb{R}}(\mathrm{A})$ if, for every effect $a \in \mathsf{Eff}(\mathrm{A})$ we have

$$\lim_{n \to +\infty} (a|\rho_n) = (a|\rho)\,. \tag{2.2.8}$$

With this topology, the issue of the convergence of a sequence of states is turned into the convergence of a sequence of real numbers. Notice we wrote $\rho \in \mathsf{St}_{\mathbb{R}}(\mathrm{A})$, because in general the limit may not be a state, but it is just an element of the vector space of states. The limit is a state if the set of states is topologically closed. Physically this means that every vector

that can be arbitrarily well approximated by a sequence of states must be a state. This is fairly natural to assume, and we will do it.

We can extend the topology defined by eq. (2.2.8) to all vectors of $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$, for every system A: $\{\xi_n\}$ converges to ξ if

$$\lim_{n\to+\infty}(a|\xi_n) = (a|\xi)$$

for every effect $a \in \mathsf{Eff}(\mathrm{A})$.

Dually, we define a topology on the set of effects: a sequence $\{a_n\}$ of system A converges to $a \in \mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$ if, for every state $\rho \in \mathsf{St}(\mathrm{A})$, we have

$$\lim_{n\to+\infty}(a_n|\rho) = (a|\rho)\,.$$

Again, if for every convergent sequence of effects, the limit a is an effect too, the set of effects is closed for every system. We can extend this topology to all $\mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$: $\{X_n\}$ converges to X if, for every state ρ

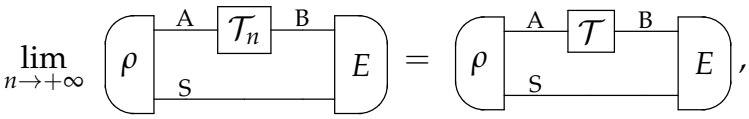
$$\lim_{n\to+\infty}(X_n|\rho) = (X|\rho)\,.$$

Finally, let us look at transformations. Recall that transformations are defined by their action on half of bipartite states. Therefore it is natural to say that a sequence $\{\mathcal{T}_n\}$ of transformations from A to B converges to $\mathcal{T} \in \mathsf{Transf}_{\mathbb{R}}(\mathrm{A},\mathrm{B})$ if, for every system S, one has

$$\lim_{n\to+\infty}(\mathcal{T}_n \otimes \mathcal{I}_{\mathrm{S}})\,\rho_{\mathrm{AS}} = (\mathcal{T} \otimes \mathcal{I}_{\mathrm{S}})\,\rho_{\mathrm{AS}}, \tag{2.2.9}$$

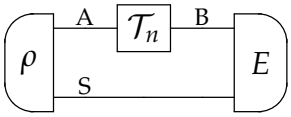
for every state $\rho \in \mathsf{St}(\mathrm{AS})$. In this way, we turn the convergence of a sequence of transformations into the convergence of a sequence of states, defined above. Again, if $\mathsf{Transf}(\mathrm{A},\mathrm{B})$ is closed, then every limit of every convergent sequence of transformations will be a transformation.

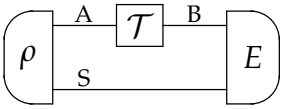
Recalling eq. (2.2.8), the condition of eq. (2.2.9) becomes



for every effect $E \in \mathsf{Eff}(\mathrm{BS})$. Note that this equation covers also the previous cases for states and effects: it is enough to take A (resp. B) to be the trivial system. Therefore we can give the following definition [72].

Definition 2.2.26 (Operational topology)**.** A sequence of transformations $\{\mathcal{T}_n\}$ from system A to system B converges to $\mathcal{T}$ if

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

$=$ 

,

for every system S, every state $\rho \in \mathsf{St}(\mathrm{AS})$, and every effect $E \in \mathsf{Eff}(\mathrm{BS})$.

Henceforth we will assume that all sets of states, effects, and transformations are topologically closed. The idea behind this is that a transformation (including a state or an effect) that can be arbitrarily well approximated by physical transformations, must be a physical transformation too. For the trivial system this is translated as follows.

Lemma 2.2.27. *The set* $\mathsf{St}(\mathrm{I})$ *is either* $\{0,1\}$ *or the whole interval* $[0,1]$.

Proof. If only 0 and 1 are allowed probabilities, there is nothing to prove. If instead the theory admits a probability $p \in (0,1)$, we have a test $\{p, 1-p\}$ on the trivial system, with which, repeating it several times and doing a suitable coarse-graining, we can approximate every element of $[0,1]$ arbitrarily well (see [67] for more details). Now, the closure in the operational topology for the trivial system coincides with the closure in the usual topology. Indeed $x_n \to x$ operationally if and only if $kx_n \to kx$ in the usual topology, for $k \in [0,1]$. Clearly, this is true if and only if $x_n \to x$ in the usual topology. Since the set of states is closed (in the usual topology), we have $\mathsf{St}(\mathrm{I}) = [0,1]$. □

In the former case, when $\mathsf{St}(\mathrm{I}) = \{0,1\}$, we say that the theory is *deterministic*.

Equivalent topologies for states, effects, and transformations.

We have seen that in $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$, $\mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$, and $\mathsf{Transf}_{\mathbb{R}}(\mathrm{A},\mathrm{B})$ we can introduce a norm, which induces another topology: we say that a sequence $\{\xi_n\}$ converges to ξ in norm if $\lim_{n\to+\infty} \|\xi_n - \xi\| = 0$. The topology of the norm is *stronger* than the operational topology defined above. This means that if a sequence converges in norm, it also converges in the operational topology. Let us prove this statement separately for states and transformations (which covers effects too).

Proposition 2.2.28. *Let $\{\xi_n\}$ be a sequence of elements of $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$. If $\{\xi_n\}$ converges to ξ in norm, it converges to ξ also operationally.*

Proof. Suppose ξ_n converges to ξ in norm, then $\lim_{n\to+\infty} \|\xi_n - \xi\| = 0$. Now to prove that ξ_n converges to ξ operationally, it is enough to prove that $\lim_{n\to+\infty} |(a|\xi_n - \xi)| = 0$ for every $a \in \mathsf{Eff}(\mathrm{A})$. Define $\eta_n := \xi_n - \xi$. Now, let us evaluate $(a|\eta_n)$. Clearly

$$(a|\eta_n) \leq \sup_{a\in\mathsf{Eff}(\mathrm{A})} (a|\eta_n)\,. \tag{2.2.10}$$

Recall that for every vector $x \in \mathsf{St}_{\mathbb{R}}(\mathrm{A})$, $\inf_{a\in\mathsf{Eff}(\mathrm{A})} (a|x) \leq 0$. Therefore we can add the term $-\inf_{a\in\mathsf{Eff}(\mathrm{A})} (a|\eta_n)$ to the right-hand side of eq. (2.2.10):

$$(a|\eta_n) \leq \sup_{a\in\mathsf{Eff}(\mathrm{A})} (a|\eta_n) - \inf_{a\in\mathsf{Eff}(\mathrm{A})} (a|\eta_n) = \|\eta_n\|\,,$$

where we have recognised the definition of the norm in $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$. We are done if we show that $|(a|\eta_n)| \leq \|\eta_n\|$. Note that

$$|(a|\eta_n)| = \begin{cases} (a|\eta_n) & \text{if } (a|\eta_n) \geq 0 \\ (a|-\eta_n) & \text{if } (a|\eta_n) < 0 \end{cases}.$$

Repeating the same argument, we get

$$(a|-\eta_n) \leq \|-\eta_n\| = \|\eta_n\|\,,$$

thus showing that $|(a|\eta_n)| \leq \|\eta_n\|$ for every $a \in \mathsf{Eff}(\mathrm{A})$. Since $\|\eta_n\| \to 0$, we have the thesis. □

Let us move to the corresponding statement for transformations.

Proposition 2.2.29. *Let $\{\Delta_n\}$ be a sequence of elements of $\mathsf{Transf}_{\mathbb{R}}(\mathrm{A},\mathrm{B})$. Then if $\{\Delta_n\}$ converges to Δ in norm, it converges to Δ also operationally.*

Proof. Suppose Δ_n converges to Δ in norm, then $\lim_{n\to+\infty} \|\Xi_n\| = 0$, where $\Xi_n := \Delta_n - \Delta$. Now Δ_n converges to Δ operationally, if for every system S, and every state $\rho \in \mathsf{St}(\mathrm{AS})$, $(\Delta_n \otimes \mathcal{I}_{\mathrm{S}})\,\rho$ converges to $(\Delta \otimes \mathcal{I}_{\mathrm{S}})\,\rho$ operationally, or in other words, if and only if $\lim_{n\to+\infty} (\Xi_n \otimes \mathcal{I}_{\mathrm{S}})\,\rho = 0$ operationally. Recalling the proof of proposition 2.2.28, for every effect $E \in \mathsf{Eff}(\mathrm{BS})$,

$$|(E|\Xi_n \otimes \mathcal{I}_{\mathrm{S}}|\rho)| \leq \|(\Xi_n \otimes \mathcal{I}_{\mathrm{S}})\,\rho\|\,.$$

Now

$$\|(\Xi_n \otimes \mathcal{I}_S)\,\rho\| \leq \sup_S \sup_{\rho \in \mathsf{St}(AS)} \|(\Xi_n \otimes \mathcal{I}_S)\,\rho\| = \|\Xi_n\|\,,$$

so $|(E|\Xi_n \otimes \mathcal{I}_S|\rho)| \leq \|\Xi_n\|$. Therefore if $\|\Xi_n\| \to 0$, it implies that $|(E|\Xi_n \otimes \mathcal{I}_S|\rho)| \to 0$, viz. $\lim_{n\to+\infty} (\Xi_n \otimes \mathcal{I}_S)\,\rho = 0$ operationally. This concludes the proof. □

We have assumed that the set of states, effects, and transformations are closed in the operational topology. Now we will make a stronger assumption.

Assumption 2.2.30. *The sets* $\mathsf{St}(A)$, $\mathsf{Eff}(A)$, *and* $\mathsf{Transf}(A,B)$ *are* closed *both in the operational and the norm topology, for all systems* A *and* B.

This assumption has far-reaching consequences. For example, the sets of states, effects, and transformations, being bounded and closed in the topology of the norm, are *compact* in this topology, as we are in finite dimension. This fact can be extended also to the operational topology. Indeed, an important consequence of this assumption is that the operational topology and the topology of the norm are *equivalent*: a sequence $\{\xi_n\}$ converges to ξ operationally if and only if it does it in norm. Let us prove the statement separately for states and transformations.

Proposition 2.2.31. *Let* $\{\xi_n\}$ *be a sequence of elements of* $\mathsf{St}_{\mathbb{R}}(A)$. *Then* ξ_n *converges to* ξ *operationally if and only if it converges to* ξ *in norm.*

Proof. We proved sufficiency in proposition 2.2.28. Let us prove necessity. Consider a sequence ξ_n that converges to ξ operationally. Then, for every $a \in \mathsf{Eff}(A)$, we have $\lim_{n\to+\infty} (a|\eta_n) = 0$, where $\eta_n := \xi_n - \xi$. Now, since the set of effects is compact (in the topology of the norm), and we are in finite dimension, the supremum $\sup_{a\in\mathsf{Eff}(A)} (a|\eta_n)$ of the linear function η_n is in fact a maximum, and achieved on the effect a^*. Similarly, the infimum $\inf_{a\in\mathsf{Eff}(A)} (a|\eta_n)$ is in fact a minimum, and achieved on the effect a_*. Then, by hypothesis, $\lim_{n\to+\infty} (a^*|\eta_n) = 0$ and $\lim_{n\to+\infty} (a_*|\eta_n) = 0$. Therefore

$$\lim_{n\to+\infty} \|\eta_n\| = \lim_{n\to+\infty} ((a^*|\eta_n) - (a_*|\eta_n)) = 0.$$

This proves that ξ_n that converges to ξ in norm too. □

This means that when we consider the conversion of states in the limit—e.g. in data compression or in the asymptotic conversion of states—we can choose either topology, according to its convenience in the problem we want to address.

Let us move to the similar statement for the topologies in the vector space of transformations, which covers the case of effects as well.

Proposition 2.2.32. *Let* $\{\Delta_n\}$ *be a sequence of elements of* $\mathsf{Transf}_{\mathbb{R}}(\mathrm{A},\mathrm{B})$. *Then* Δ_n *converges to* Δ *operationally if and only if it converges to* Δ *in norm.*

Proof. We proved sufficiency in proposition 2.2.29, let us show necessity. We have that Δ_n converges to Δ operationally if and only if, for every system S and every state $\rho \in \mathsf{St}(\mathrm{AS})$, we have that $(\Delta_n \otimes \mathcal{I}_{\mathrm{S}})\,\rho$ converges to $(\Delta \otimes \mathcal{I}_{\mathrm{S}})\,\rho$ operationally. By proposition 2.2.31, this means that

$$\lim_{n\to+\infty} \|(\Xi_n \otimes \mathcal{I}_{\mathrm{S}})\,\rho\| = 0,$$

where $\Xi_n := \Delta_n - \Delta$. Then in the definition of $\|\Xi_n\|$, by compactness, the supremum is achieved on system S^* and on the state $\rho^* \in \mathsf{St}(\mathrm{AS}^*)$. Therefore

$$\lim_{n\to+\infty} \|\Xi_n\| = \lim_{n\to+\infty} \|(\Xi_n \otimes \mathcal{I}_{\mathrm{S}^*})\,\rho^*\| = 0,$$

because, by hypothesis, $\lim_{n\to+\infty} \|(\Xi_n \otimes \mathcal{I}_{\mathrm{S}})\,\rho\| = 0$ for every system S and every $\rho \in \mathsf{St}(\mathrm{AS})$. This concludes the proof. □

One concludes that the two topologies are completely equivalent. Specifically, the statements about the compactness of the set of states, effects, and transformations are valid both in the operational and the norm topology.

We can use proposition 2.2.31 to prove a property of limits of sequences of transformations we will use in the following.

Lemma 2.2.33. *Let* $\{\mathcal{A}_n\}$ *be a sequence of transformations from* A *to* B *converging to* $\mathcal{A}$, *and let* $\{\mathcal{B}_n\}$ *be a sequence of transformations from* A *to* B *converging to* $\mathcal{B}$. *Then* $\{\mathcal{B}_n\mathcal{A}_n\}$ *converges to* $\mathcal{B}\mathcal{A}$.

Proof. We have that $\mathcal{A}_n$ converges to $\mathcal{A}$ if, for every S, $\{(\mathcal{A}_n \otimes \mathcal{I}_{\mathrm{S}})\,\rho_{\mathrm{AS}}\}$ converges to $(\mathcal{A} \otimes \mathcal{I}_{\mathrm{S}})\,\rho_{\mathrm{AS}}$. Similarly, $\{(\mathcal{B}_n \otimes \mathcal{I}_{\mathrm{S}})\,\rho_{\mathrm{AS}}\}$ converges to $(\mathcal{B} \otimes \mathcal{I}_{\mathrm{S}})\,\rho_{\mathrm{AS}}$. By proposition 2.2.31, this happens if and only if $\|(\mathcal{A}_n \otimes \mathcal{I}_{\mathrm{S}})\,\rho_{\mathrm{AS}} - (\mathcal{A} \otimes \mathcal{I}_{\mathrm{S}})\,\rho_{\mathrm{AS}}\| \to 0$, and $\|(\mathcal{B}_n \otimes \mathcal{I}_{\mathrm{S}})\,\rho_{\mathrm{AS}} - (\mathcal{B} \otimes \mathcal{I}_{\mathrm{S}})\,\rho_{\mathrm{AS}}\| \to 0$. Now we are ready to prove

that $\{\mathcal{B}_n\mathcal{A}_n\}$ converges to $\mathcal{BA}$, viz. that $\|(\mathcal{B}_n\mathcal{A}_n \otimes \mathcal{I}_S)\,\rho_{AS} - (\mathcal{BA} \otimes \mathcal{I}_S)\,\rho_{AS}\| \to 0$.

$$\|(\mathcal{B}_n\mathcal{A}_n \otimes \mathcal{I}_S)\,\rho_{AS} - (\mathcal{BA} \otimes \mathcal{I}_S)\,\rho_{AS}\| \leq \|\mathcal{B}_n\,(\mathcal{A}_n \otimes \mathcal{I}_S)\,\rho_{AS} - \mathcal{B}_n\,(\mathcal{A} \otimes \mathcal{I}_S)\,\rho_{AS}\| +$$

$$+ \|\mathcal{B}_n\,(\mathcal{A} \otimes \mathcal{I}_S)\,\rho_{AS} - \mathcal{B}\,(\mathcal{A} \otimes \mathcal{I}_S)\,\rho_{AS}\|$$

Now, $\|\mathcal{B}_n\,(\mathcal{A} \otimes \mathcal{I}_S)\,\rho_{AS} - \mathcal{B}\,(\mathcal{A} \otimes \mathcal{I}_S)\,\rho_{AS}\| \to 0$ because $\mathcal{B}_n \to \mathcal{B}$. Let us apply the monotonicity of the norm:

$$\|\mathcal{B}_n\,(\mathcal{A}_n \otimes \mathcal{I}_S)\,\rho_{AS} - \mathcal{B}_n\,(\mathcal{A} \otimes \mathcal{I}_S)\,\rho_{AS}\| \leq \|(\mathcal{A}_n \otimes \mathcal{I}_S)\,\rho_{AS} - (\mathcal{A} \otimes \mathcal{I}_S)\,\rho_{AS}\|\,.$$

Then $\|\mathcal{B}_n\,(\mathcal{A}_n \otimes \mathcal{I}_S)\,\rho_{AS} - \mathcal{B}_n\,(\mathcal{A} \otimes \mathcal{I}_S)\,\rho_{AS}\| \to 0$, because $\mathcal{A}_n \to \mathcal{A}$, by which we conclude that

$$\|(\mathcal{B}_n\mathcal{A}_n \otimes \mathcal{I}_S)\,\rho_{AS} - (\mathcal{BA} \otimes \mathcal{I}_S)\,\rho_{AS}\| \to 0.$$

□

2.3 Causality and its consequences

In this section we will examine how the direction in which "information flows" in a theory constrains the structure of the theory itself. In causal theories information propagates in the same order as the input-output order given by sequential composition. More poetically, in causal theories information propagates from the past to the future, and one is able to choose later experiments depending on the outcomes of present observations. Causality is a standard setting in most physical descriptions of Nature, and in the GPT literature it is often assumed implicitly, without even mentioning it directly. Causality will be the first axiom we impose in this thesis. Besides being reasonable from a physical point of view, Causality will bring a lot of interesting and important consequences to a theory, making it easier to tackle.

Let us begin with the formal definition of causal theory.

Definition 2.3.1 (Causality [67])**.** A theory is *causal* if for every preparation-test $\{\rho_i\}_{i\in X}$ and every observation-test $\{a_j\}_{j\in Y}$ on any system A, the probability $p_i := \sum_{j\in Y} (a_j|\rho_i)$ is *independent* of the observation-test $\{a_j\}_{j\in Y}$.

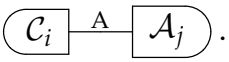
In other words, if $\{a_j\}_{j\in Y}$ and $\{b_k\}_{k\in Z}$ are two observation-tests, we have

$$\sum_{j\in Y}(a_j|\rho_i) = \sum_{k\in Z}(b_k|\rho_i)\,. \tag{2.3.1}$$

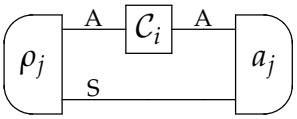
Loosely speaking, the preparation of the system does not depend on the choice of subsequent (or "future") measurements, a sort of no-signalling condition from the future. In this way, the direction in which information flows, as witnessed by marginal probabilities in definition 2.3.1, coincides with the ordering given by sequential composition. In general, this is not obvious, as the following example shows.

Example 2.3.2. Consider a theory where the states of a system are the quantum operations on that system. Specifically, deterministic states are quantum channels. Then we can consider the channels of this higher theory to be quantum "supermaps", which map quantum operations into quantum operations [131–136].

Let us consider a preparation of a state $\mathcal{C}_i$ followed by a measurement $\mathcal{A}_j$, which we represent in the higher theory 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 wire, in quantum theory such a diagram will look like



for some system S. Note that the effect $\mathcal{A}_j$ is split into two parts: one is before the quantum operation and the other is after, otherwise we could not have a diagram with no external wires. Therefore, in the theory in which states are quantum operations, the preparation of a state is influenced by a subsequent measurement, so information does not propagate in the same direction as sequential composition.

Definition 2.3.1 can be recast in an equivalent way [67, lemma 4], which is often more practical to work with.

Proposition 2.3.3. *A theory is causal if and only if for every system* A *there is a unique deterministic effect* u_{A}.

Proof. Necessity. Suppose, by contradiction, that there are two deterministic effects u and u' for system A. Deterministic effects are particular examples of observation-tests. Eq. (2.3.1) then states that $(u|\rho_i) = (u'|\rho_i)$ for every $\rho_i \in \mathrm{St}(\mathrm{A})$. This means that $u = u'$.

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

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

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

Example 2.3.4. We saw in example 2.2.5 that in quantum mechanics there is only one deterministic effect, the identity operator. Hence quantum mechanics is a causal theory.

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

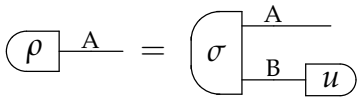
Let us see a straightforward corollary of uniqueness of the deterministic effect.

Corollary 2.3.5. *In a causal theory, if* u_{A} *and* u_{B} *are the deterministic effects of systems* A *and* B *respectively, then the deterministic effect for system* AB *is* $u_{\mathrm{A}} \otimes u_{\mathrm{B}}$.

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

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 marginalisation in probability theory, whence the name. In quantum mechanics, this operation is simply given by taking the partial trace over B.

Definition 2.3.6. The *marginal state* (*marginal* for short) ρ_{A} on system A of a bipartite state σ_{AB} is obtained by applying the deterministic effect to B:



For this reason, we will sometimes use the notation tr_{A} for the unique deterministic effect on A. Therefore, $\rho_{\mathrm{A}} = \mathrm{tr}_{\mathrm{B}}\sigma_{\mathrm{AB}}$.

In a causal theory, we have also useful characterisations of channels and tests [67, lemma 5].

Proposition 2.3.7. *Let $\mathcal{C} \in \mathsf{Transf}(\mathrm{A}, \mathrm{B})$. $\mathcal{C}$ is a channel if and only if $u_{\mathrm{B}}\mathcal{C} = u_{\mathrm{A}}$.*

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

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

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

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

Note that in quantum theory this is precisely the statement that a quantum operation is a quantum channel if and only if it is trace-preserving. This is even more obvious if we write u_{A} as tr_{A}.

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

$$\sum_{i\in X} u_{\mathrm{B}}\mathcal{C}_i = u_{\mathrm{A}}. \tag{2.3.2}$$

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

Suppose we have two parties sharing a bipartite state. In a causal theory it is impossible for a party to send a message to the other by acting locally on her own physical system and relying on correlations she shares with the other party. This form of instantaneous communication is called *signalling*. In more precise terms, in a causal theory it is not possible for a party to communicate the outcome of a local measurement on her system to the other without exchanging physical systems, classical communication included, as it is usually mediated by electromagnetic signals [67, theorem 1].

Theorem 2.3.8. *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, who share a bipartite state σ_{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 $p_{ij} := \operatorname{tr}_{\mathrm{AB}}\left(\mathcal{A}_i\otimes\mathcal{B}_j\right)\sigma_{\mathrm{AB}}$ and the marginal probabilities as $p_i^{(\mathrm{A})} := \sum_{j\in Y}\operatorname{tr}_{\mathrm{AB}}\left(\mathcal{A}_i\otimes\mathcal{B}_j\right)\sigma_{\mathrm{AB}}$ and $p_j^{(\mathrm{B})} := \sum_{i\in X}\operatorname{tr}_{\mathrm{AB}}\left(\mathcal{A}_i\otimes\mathcal{B}_j\right)\sigma_{\mathrm{AB}}$. Each party cannot acquire any information about the outcomes of the other based only on its marginal probability. Indeed, let us examine Alice's marginal probability $p_i^{(\mathrm{A})}$ better. Let ρ_{A} be the marginal state of σ_{AB} on system A.

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

$$= u_{\mathrm{A}}\mathcal{A}_i\otimes\operatorname{tr}_{\mathrm{B}}\sigma_{\mathrm{AB}} = \operatorname{tr}_{\mathrm{A}}\mathcal{A}_i\rho_{\mathrm{A}}$$

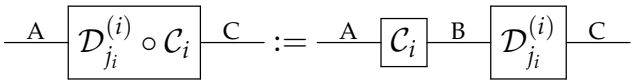
We see that Alice's marginal probability does not depend on the test performed by Bob at all, so she cannot get any information about the outcome

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

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

Definition 2.3.9. If $\{\mathcal{C}_i\}_{i\in X}$ is a test from A to B and, for every i, $\left\{\mathcal{D}^{(i)}_{j_i}\right\}_{j_i\in Y_i}$ is a test from B to C, then the *conditioned* (or *classically controlled*) *test* is a test from A to C with outcomes $(i, j_i) \in Z := \bigcup_{i\in X}\{i\}\times Y_i$, and events $\left\{\mathcal{D}^{(i)}_{j_i}\circ\mathcal{C}_i\right\}_{(i,j_i)\in Z}$.

The graphical representation is as usual.



Conditioning expresses the idea of choosing what to do at later steps using the classical information about outcomes obtained at previous steps. The test $\left\{\mathcal{D}^{(i)}_{j_i}\circ\mathcal{C}_i\right\}$ is well-defined thanks to Causality: it satisfies the necessary condition of eq. (2.3.2). Indeed

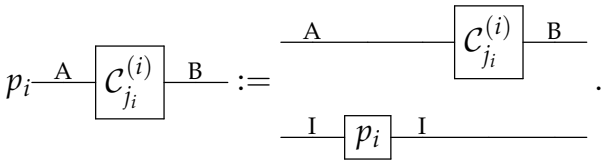
$$\sum_i\sum_{j_i} u\mathcal{D}^{(i)}_{j_i}\circ\mathcal{C}_i = \sum_i u\mathcal{C}_i = u.$$

A particular case of conditioning is randomisation.

Definition 2.3.10. If $\{p_i\}_{i\in X}$ is a set of probabilities[4] and, for every i, $\left\{\mathcal{C}^{(i)}_{j_i}\right\}_{j_i\in Y_i}$ is a test from A to B, we can construct the *randomised test* $\left\{p_i\mathcal{C}^{(i)}_{j_i}\right\}_{i\in X, j_i\in Y_i}$,

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

which is a test from A to B whose events are defined as



In a randomised test we are performing a classical random process and according to its outcome we apply a test $\left\{\mathcal{C}_{j_i}^{(i)}\right\}_{j_i \in Y_i}$, where i is the outcome of the classical random process. The existence of randomised tests tells us that a non-deterministic theory must be convex, thus recovering one of the assumptions of the convex approach to GPTs, but from a deeper principle, Causality.

Proposition 2.3.11. *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.*

Proof. Let $p \in [0,1]$. If the theory is non-deterministic, p is a state of the trivial system. Let $\{\mathcal{C}_i\}_{i\in X}$ and $\left\{\mathcal{D}_j\right\}_{j\in Y}$ be tests from A to B. By randomisation, we can consider the test $\{p\mathcal{C}_i\}_{i\in X} \cup \left\{(1-p)\,\mathcal{D}_j\right\}_{j\in Y}$. 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. □

Another important example of classically controlled tests are measure-and-prepare tests.

Definition 2.3.12. A test is *measure-and-prepare* if it is of the form $\{|\rho_i)\,(a_i|\}_{i\in X}$, where $\{a_i\}_{i\in X}$ is an observation-test, and ρ_i is a deterministic state for every $i \in X$.

A channel $\mathcal{C}$ is *measure-and-prepare* if it is the coarse-graining of a measure-and-prepare test:

$$\mathcal{C} = \sum_{i\in X} |\rho_i)\,(a_i|\,.$$

The idea behind a measure-and-prepare test is to perform an observation-test, and to prepare the deterministic state ρ_i if we get outcome i upon performing the observation-test.

The relationship between classical control and Causality is so tight that a theory where all classically controlled tests are possible is causal [67, lemma 7] (see also [137]).

Proposition 2.3.13. *A theory where every conditioned test is possible is causal.*

Proof. Suppose by contradiction that the theory is not causal, so there exist two deterministic effects $u \neq u'$ for a system A. Now, take a generic state $\rho \in \mathsf{St}(\mathrm{A})$. By definition, there exists a preparation-test $\{\rho_i\}$ such that $\rho_{i_0} = \rho$. Take the coarse-grained version $\{\rho_0, \rho_1\}$, where $\rho_0 := \rho_{i_0}$, and $\rho_1 := \sum_{i \neq i_0} \rho_i$. Now, consider the classically controlled test where one applies u if ρ_0 is prepared, and u' if ρ_1 is prepared. This is a valid test, given by $\{(u|\rho_0), (u'|\rho_1)\}$. This is a test on the trivial system, therefore

$$(u|\rho_0) + (u'|\rho_1) = 1.$$

Now, $\rho_0 + \rho_1$ is a deterministic state, and since u' is a deterministic effect

$$(u'|\rho_0) + (u'|\rho_1) = 1.$$

This implies $(u|\rho_0) = (u'|\rho_0)$. Since $\rho_0 = \rho$ is an arbitrary state of A, one has $u = u'$. The theory must be causal. □

2.3.1 States of causal theories

In causal theories, the norm of a state takes a particularly simple form.[5]

Proposition 2.3.14. *In a causal theory, for a state ρ we have $\|\rho\| = \operatorname{tr}\rho$.*

Proof. Clearly, for a state, we have $\|\rho\| = \sup_{a \in \mathsf{Eff}(\mathrm{A})} (a|\rho)$, therefore

$$\operatorname{tr}\rho \leq \sup_{a \in \mathsf{Eff}(\mathrm{A})} (a|\rho) = \|\rho\|,$$

because tr is a deterministic effect. Now, for every effect $a \in \mathsf{Eff}(\mathrm{A})$, we also have

$$(u|\rho) = (a|\rho) + (u - a|\rho) \geq (a|\rho),$$

whence

$$\|\rho\| = \sup_{a \in \mathsf{Eff}(\mathrm{A})} (a|\rho) \leq \operatorname{tr}\rho.$$

One concludes that $\|\rho\| = \operatorname{tr}\rho$. □

[5] This result clearly extends to all elements of $\mathsf{St}_+(\mathrm{A})$, which are just a non-negative rescaling of states.

For this reason, in a causal theory, a state is deterministic ($\mathrm{tr}\,\rho = 1$) if and only if it is normalised. The norm of a state is the probability of preparing that state in some preparation-test. Indeed, since tr is deterministic, the probability in $\mathrm{tr}\,\rho$ comes only from the randomness arising from the preparation of ρ.

Example 2.3.15. In quantum theory, we have

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

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

For every (non-zero) state ρ of a causal theory we can consider the normalised state

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

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, viz. a deterministic state. In a causal theory, every sub-normalised state ρ_i can be written as $\rho_i = p_i\overline{\rho}_i$, where $p_i = \|\rho_i\| \in [0,1]$ and $\overline{\rho}_i$ is a normalised state. Therefore every preparation-test $\{\rho_i\}_{i\in X}$ is a randomised test: we have a classical source of randomness $\{p_i\}_{i\in X}$, where $p_i = \|\rho_i\|$, and according to the classical outcome we prepare the deterministic state $\overline{\rho}_i$.

In conclusion, given the linearity of OPTs, in causal theories we can simply restrict ourselves to normalised states. In the following we will often drop the term "normalised" when talking about states, and when we write about the "state space" we will mean the set of normalised states $\mathsf{St}_1(\mathrm{A})$ of a system A. The state space of a non-deterministic causal theory is a convex set. Indeed, for any $p \in [0,1]$, consider the state $p\rho_0 + (1-p)\rho_1$, where ρ_0 and ρ_1 are two normalised states. $p\rho_0 + (1-p)\rho_1$ is still normalised:

$$\mathrm{tr}\left[p\rho_0 + (1-p)\rho_1\right] = p\,\mathrm{tr}\,\rho_0 + (1-p)\,\mathrm{tr}\,\rho_1 = p + 1 - p = 1.$$

Clearly, normalised pure states are the extreme points of the state space $\mathsf{St}_1(\mathrm{A})$. We have an interesting geometrical interpretation for the state space of any system A. It is the intersection of the hyperplane $\mathrm{tr}\,\xi = 1$ (for $\xi \in \mathsf{St}_{\mathbb{R}}(\mathrm{A})$) with the cone of states $\mathsf{St}_+(\mathrm{A})$, as illustrated in fig. 2.2.

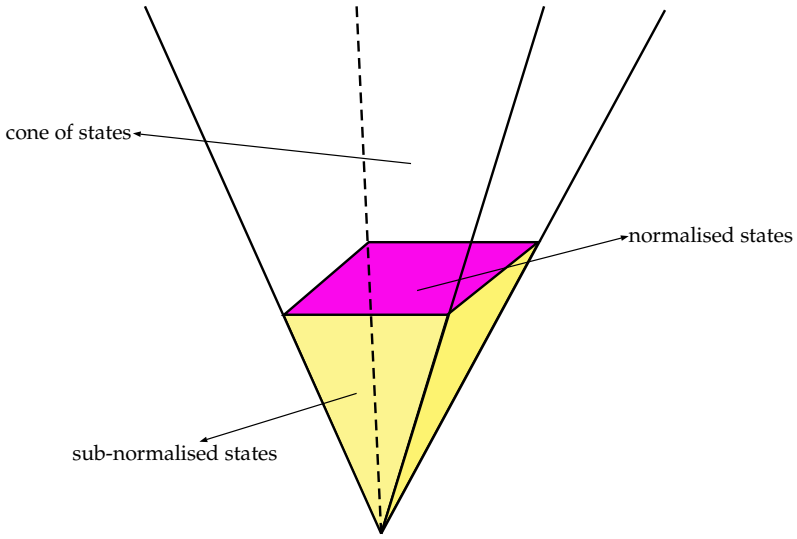


Figure 2.2: The cone of states of a causal theory. The set of normalised states (in purple) is given by the intersection of the hyperplane $\mathrm{tr}\,\xi = 1$, for ξ in $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$, with the cone of states $\mathsf{St}_{+}(\mathrm{A})$. Below that hyperplane are the sub-normalised states (in yellow). The coloured part in the cone of states is the set of states $\mathsf{St}(\mathrm{A})$. The white part corresponds to super-normalised elements of $\mathsf{St}_{+}(\mathrm{A})$, which are non-physical.

Convex combinations of normalised states do not have only a mathematical meaning, but can be also realised operationally. Suppose we have $\rho_p = p\rho_0 + (1-p)\,\rho_1$, where $\rho_0, \rho_1 \in \mathsf{St}_1(\mathrm{A})$. We can prepare ρ_p by using the following procedure.

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

2. If the outcome is i, then we prepare ρ_i. In this way, we realise the preparation-test $\{p_0\rho_0, p_1\rho_1\}$. Note that each state $p_i\rho_i$ is not normalised 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 $\rho_p = p\rho_0 + (1-p)\,\rho_1$.

A coarse-graining of normalised 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 it.

Definition 2.3.16. Let $\rho, \sigma \in \mathsf{St}_1(\mathrm{A})$. We say that σ is *contained* in ρ if ρ can be written as $\rho = p\sigma + (1-p)\,\tau$, where $p \in (0,1]$ and $\tau \in \mathsf{St}_1(\mathrm{A})$.

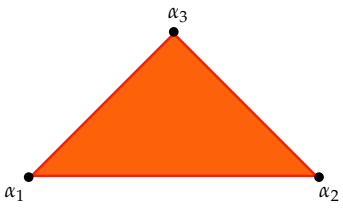
In other words, σ is contained in ρ if it arises in a convex decomposition of ρ. Clearly if ρ is pure, only ρ can be contained in it. At the other extreme we have internal states.

Definition 2.3.17. A state ω is *internal* if every (normalised) state ρ is contained in it.

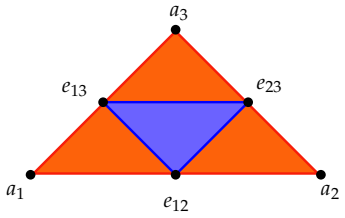
We will make use of the following definition too.

Definition 2.3.18. Two transformations $\mathcal{A}, \mathcal{A}' \in \mathsf{Transf}(\mathrm{A}, \mathrm{B})$ are *equal upon input* of ρ, written as $\mathcal{A} =_\rho \mathcal{A}'$ if $\mathcal{A}\sigma = \mathcal{A}'\sigma$ for every state σ contained in ρ.

We conclude this subsection with the important definition of perfectly distinguishable states, which are states that can be distinguished in a single shot. It means that if the state is not known, it can be identified with certainty.



(a) The state space of a restricted trit coincides with that of a classical trit.



(b) A cross-section of the effect cone of the classical trit (in orange) and of the restricted trit (in blue). The dual cone is the same in both cases (in orange). The restriction on effects is apparent.

Figure 2.3: The restricted trit

Definition 2.3.19. The (normalised) states $\{\rho_i\}_{i\in X}$ are *perfectly distinguishable* if there exists an observation-test $\{a_i\}_{i\in X}$, called the *perfectly distinguishing test*, such that

$$(a_i|\rho_j) = \delta_{ij}.$$

Perfectly distinguishable states might not exist in some systems of a theory, as shown in the following example [115].

Example 2.3.20. Start with the state space of the classical trit, represented in fig. 2.3a, with pure states α_1, α_2, α_3. They are perfectly distinguishable with observation-test $\{a_1, a_2, a_3\}$: $(a_i|\alpha_j) = \delta_{ij}$. Instead of allowing the full set of effects of classical theory, suppose that, for some reasons, the most fine-grained effects that are allowed are $e_{ij} = \frac{1}{2}(a_i + a_j)$, with $i \neq j$. A section of the dual cone (the same as the effect cone of classical theory), and of the effect cone of the restricted trit is represented in fig. 2.3b.

Since we have a smaller set of effects than the original classical trit, we must check what happens to the state space. Indeed it may happen that two states become tomographically indistinguishable because there are not enough effects to witness their difference. However, this is not the case of the restricted trit. The reason is that the effects e_{ij} are linearly independent, therefore they span exactly the same effect vector space as the effects a_i, which is what determines the tomographic power of a theory. Therefore the state space of the restricted trit coincides with that of the classical trit (cf. fig. 2.3a).

However, the restriction on the allowed effects has a dramatic consequence: there are *no* perfectly distinguishable pure states. The lack of perfectly distinguishable *pure* states determines the lack of perfectly distinguishable states. Indeed, if ρ_1 were perfectly distinguishable from ρ_2, all pure the states contained in ρ_1 would be perfectly distinguishable from the pure states contained in ρ_2. This is because if $(a|\rho) = 0$ (resp. $(a|\rho) = 1$) one has $a =_\rho 0$ (resp. $a =_\rho u$).

First of all, let us show that $\{\alpha_1, \alpha_2, \alpha_3\}$ are no longer perfectly distinguishable. Consider a generic effect $e = \lambda_{12}e_{12} + \lambda_{13}e_{13} + \lambda_{23}e_{23}$, where $\lambda_{ij} \geq 0$. This effect could yield 0 on α_2 and α_3 if and only if $\lambda_{12} = \lambda_{13} = \lambda_{23} = 0$, but this would be the zero effect, which cannot yield 1 on α_1. This means that the α_i's cannot be jointly perfectly distinguishable. Maybe we can still find a pair of α_i's that are perfectly distinguishable? The answer is again negative. To see it, take e.g. the pair $\{\alpha_1, \alpha_2\}$ (for the others the argument is the same). The only element in the effect cone that yields 1 on α_1 and 0 on α_2 is $2e_{13}$, but this is *not* a physical effect, because $u - 2e_{13} = a_2$, which is *not* an effect. In other words, $2e_{13}$ cannot exist in an observation-test of the form $\{2e_{13}, u - 2e_{13}\}$, but all effects must be part of some observation-test! In conclusion, the restricted trit has *no* perfectly distinguishable states.

The example above is based on a theory where the no-restriction hypothesis fails: the effect cone $\mathsf{Eff}_+(\mathrm{A})$ was strictly contained in the dual cone $\mathsf{St}_+^*(\mathrm{A})$. If the two cones coincide, we can always prove that at least two perfectly distinguishable pure states exist [115].

Proposition 2.3.21. *In an unrestricted theory, for every pure state ψ_1 there exists another pure state ψ_2 such that $\{\psi_1, \psi_2\}$ are perfectly distinguishable.*

Proof. Let ψ_1 be a pure state. The proof will consist of some steps. In the first step, let us prove that there exists a non-trivial element f of the dual cone $\mathsf{St}_+^*(\mathrm{A})$ such that $(f|\psi_1) = 0$. Note that being pure, ψ_1 lies in some supporting hyperplane through the origin of the cone $\mathsf{St}_+(\mathrm{A})$ [138]. Such a hyperplane must have equation $(f|x) = 0$ for all $x \in \mathsf{St}_{\mathbb{R}}(\mathrm{A})$, where f is some non-trivial linear functional on $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$, otherwise it would not pass through the origin (i.e. the null vector). Being a supporting hyperplane, we can choose f to be in the dual cone $\mathsf{St}_+^*(\mathrm{A})$ [138]. Thus we have found $f \in \mathsf{St}_+^*(\mathrm{A})$ such that $(f|\psi_1) = 0$.

Let us consider the maximum of f on the state space. Since f is continuous and the state space is compact, it achieves its maximum λ^* on some state ρ^*. Note that $\lambda^* > 0$, otherwise f would be the zero functional. Let us show that the maximum is attained on some pure state. If ρ^* is already a pure state, there is nothing to prove. If it is not, consider a refinement of ρ^* in terms of pure states, $\rho^* = \sum_i p_i \psi_i$, where $\{p_i\}$ is a probability distribution. Apply f to ρ^*:

$$\lambda^* = (f|\rho^*) = \sum_i p_i \, (f|\psi_i) \, .$$

Clearly $\lambda^* \leq \max_i (f|\psi_i)$, but being λ^* the maximum of f , in fact $\lambda^* = \max (f|\psi_i)$. This means that there is a pure state ψ_2, chosen among these ψ_i's, on which f attains its maximum.

Now consider the functional $a_2 := \frac{1}{\lambda^*} f$, which takes values in the interval $[0, 1]$ when applied to states. Specifically $(a_2|\psi_2) = 1$ and $(a_2|\psi_1) = 0$. By the no-restriction hypothesis, it is a valid effect, so we can construct the observation-test $\{a_1, a_2\}$, where $a_1 := u - a_2$, which distinguishes perfectly between ψ_1 and ψ_2. □

Even though the no-restriction hypothesis guarantees the existence of perfectly distinguishable states, we do not wish to assume it for its lack of operational motivation. Instead, in chapter 4 we will prove the existence of perfectly distinguishable states for a class of OPTs from first principles.

Finally we give the following definition.

Definition 2.3.22. A set $\{\rho_i\}_{i \in X}$ of perfectly distinguishable states is *maximal* if no state ρ_0 can be added such that $\{\rho_i\}_{i \in X} \cup \{\rho_0\}$ are still perfectly distinguishable.

If the ρ_i's are pure states, instead of writing "maximal set of perfectly distinguishable states", we will opt for the shorter terminology of "pure maximal set".

2.3.2 The group of reversible channels

Causality provides us with a new insight into the structure of reversible channels of a system A, and will enable us to prove some interesting properties. First of all, we show that the set of channels is compact [67, corollary 30].

Proposition 2.3.23. *For every pair of systems* A *and* B *the set of channels* $\mathsf{DetTransf}(\mathrm{A},\mathrm{B})$ *is compact.*

Proof. Let $\{\mathcal{C}_n\}$ be a sequence of channels that converges to some transformation $\mathcal{C} \in \mathsf{Transf}(\mathrm{A},\mathrm{B})$. Let us prove that $\mathcal{C}$ is a channel. We have $u_{\mathrm{B}}\mathcal{C}_n = u_{\mathrm{A}}$ for all $n \in \mathbb{N}$, because they are all channels. Then, for every state $\rho \in \mathsf{St}(\mathrm{A})$, we have

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

Therefore we conclude that $u_{\mathrm{B}}\mathcal{C} = u_{\mathrm{A}}$, which means that $\mathcal{C}$ is a channel, by proposition 2.3.7. We conclude that every convergent sequence of channels converges to a channel, by which we obtain the closure of $\mathsf{DetTransf}(\mathrm{A},\mathrm{B})$. $\mathsf{DetTransf}(\mathrm{A},\mathrm{B})$ is bounded because it is a subset of $\mathsf{Transf}(\mathrm{A},\mathrm{B})$, therefore it is compact in the norm topology. By the equivalence of the norm and operational topologies, we can say that $\mathsf{DetTransf}(\mathrm{A},\mathrm{B})$ is compact in both topologies. □

We can exploit the compactness of channels to prove that the group of reversible channels is compact as well [67, corollary 31].

Proposition 2.3.24. *For every system* A*, the group* G_{A} *of reversible channels is compact.*

Proof. Consider a convergent sequence $\{\mathcal{U}_n\}$ of reversible channels. By proposition 2.3.23, the limit $\mathcal{U}$ is a channel on A. Now consider the sequence $\{\mathcal{U}_n^{-1}\}$. Since the set of channels is compact, there exists a subsequence $\left\{\mathcal{U}_{n_k}^{-1}\right\}$ that converges to a channel $\mathcal{C}$. Now

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

and

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

where we have used lemma 2.2.33, and the fact that every subsequence of $\mathcal{U}_n$ converges to $\mathcal{U}$. This proves that G_{A} is closed. Recalling the fact that the set of channels is bounded, the group is compact (in both topologies). □

Finite groups are trivial examples of compact groups. Compact groups enjoy a remarkable mathematical property: they admit a finite Haar measure h, unique up to rescaling, namely a measure such that $h(G) < +\infty$, and that is invariant under left and right action of the group[6] [139]. In other words, if S is a Borel subset of G, one has

$$h(\mathcal{U}S) = h(S\mathcal{U}) = h(S),$$

for every $\mathcal{U} \in G$, where $\mathcal{U}S = \{\mathcal{U}\mathcal{V} : \mathcal{V} \in S\}$ and $S\mathcal{U} = \{\mathcal{V}\mathcal{U} : \mathcal{V} \in S\}$. Since h is a finite measure, it can be renormalised so that $h(G) = 1$, which we will always assume in the following. This measure will be one of the key ingredients in the construction of invariant states, an extremely important notion that will be used throughout this thesis.

Definition 2.3.25. A state $\chi \in \mathrm{St}(\mathrm{A})$ is called *invariant* if $\mathcal{U}\chi = \chi$ for every $\mathcal{U} \in G_{\mathrm{A}}$.

Since we are working in causal theories, in the following we will consider only *normalised* invariant states. Do they exist? The answer is affirmative.

Proposition 2.3.26. *In every causal theory there is at least one invariant state.*

Proof. Take a pure state ψ, and consider the state

$$\chi_\psi := \int_G \mathcal{V}\psi \, \mathrm{d}\mathcal{V},$$

where $\mathrm{d}\mathcal{V}$ is the Haar probability measure. χ_ψ is invariant, indeed for every reversible channel $\mathcal{U}$, we have

$$\mathcal{U}\chi_\psi = \int_G \mathcal{U}\mathcal{V}\psi \, \mathrm{d}\mathcal{V} = \int_G \mathcal{W}\psi \, \mathrm{d}\left(\mathcal{U}^{-1}\mathcal{W}\right) = \int_G \mathcal{W}\psi \, \mathrm{d}\mathcal{W} = \chi_\psi,$$

where we have set $\mathcal{W} := \mathcal{U}\mathcal{V}$, and we have exploited the invariance of the Haar probability measure. Therefore at least one invariant state always exists. □

[6] For a finite group, the Haar measure is simply the counting measure, that counts the number of elements in a subset of the group.

If G is a finite, the definition of χ_ψ becomes $\chi_\psi = \frac{1}{|G|}\sum_{i=1}^{|G|} \mathcal{V}_i\psi$.

In general, there will be more than one invariant state, because χ_ψ depends on the choice of ψ, and different ψ's may give rise to different invariant states χ_ψ. However, if ψ and ψ' are in the same orbit of the action of G, which means $\psi' = \mathcal{U}\psi$, they generate the same invariant state. Indeed,

$$\chi_{\psi'} = \int_G \mathcal{V}\psi' \,\mathrm{d}\mathcal{V} = \int_G \mathcal{V}\mathcal{U}\psi \,\mathrm{d}\mathcal{V} = \int_G \mathcal{W}\psi \,\mathrm{d}\left(\mathcal{W}\mathcal{U}^{-1}\right) = \int_G \mathcal{W}\psi \,\mathrm{d}\mathcal{W} = \chi_\psi,$$

setting $\mathcal{W} := \mathcal{V}\mathcal{U}$, and harnessing the invariance of the Haar probability measure. If there is only one orbit on the set of normalised pure states, the invariant state is unique. When there is only one orbit, the action is called *transitive*. In practice it means that for every pair of pure states ψ, ψ' there exists a reversible channel $\mathcal{U}$ such that $\psi' = \mathcal{U}\psi$.

Proposition 2.3.27. *A causal theory with transitive action has a unique invariant state, which is internal.*

Proof. We need to prove only the second part of the statement. Since the invariant state is unique, we have $\chi = \int_G \mathcal{V}\psi \,\mathrm{d}\mathcal{V}$ for every pure state ψ, therefore, by linearity

$$\chi = \int_G \mathcal{V}\rho \,\mathrm{d}\mathcal{V}, \tag{2.3.3}$$

for every state ρ. Now, since we are in finite dimension, by Carathéodory's theorem for convex geometry [140, 141], the integral (2.3.3) is a finite convex combination of reversible channels: $\chi = \sum_i p_i\mathcal{U}_i\rho$. Now, apply $\mathcal{U}_{i_0}^{-1}$ to χ, where $\mathcal{U}_{i_0}$ is one of the reversible channels in the convex combination. This yields

$$\chi = \mathcal{U}_{i_0}^{-1}\chi = p_{i_0}\rho + \sum_{i\neq i_0} p_i\mathcal{U}_{i_0}^{-1}\mathcal{U}_i\rho$$

for every ρ. This shows that every state is contained in χ, whence χ is internal. □

The uniqueness of the invariant state does *not* imply that the action is transitive, as shown in the following example.

Example 2.3.28. Consider theory with a system whose state space is a diamond, as illustrated in fig. 2.4. The pure states are its vertices ψ_1, ψ_2, ψ_3, and ψ_4. Assuming that all symmetries of the state space are allowed reversible channels, besides the identity we have the reflection across the

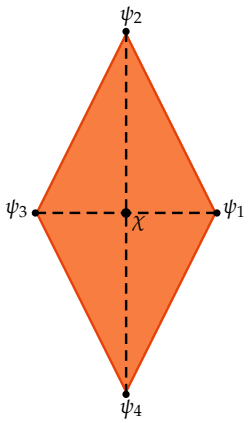


Figure 2.4: A state space where there is a unique invariant state, the centre χ of the diamond, but the action of reversible channels is not transitive on the pure states ψ_1, ψ_2, ψ_3, and ψ_4.

longer diagonal and the reflection across the shorter diagonal. There are two orbits on the set of pure states: $\{\psi_1, \psi_3\}$ and $\{\psi_2, \psi_4\}$, so the action is *not* transitive. Yet there is a unique invariant state χ, which is the centre of the diamond.

Another interesting property of theories with a transitive action of the group of reversible channels is that the set of pure states is compact.

Proposition 2.3.29. *In a causal theory with transitive action, the set of (normalised) pure states is compact.*

Proof. We only need to prove that the set of pure states is closed, being a subset of a compact set, the set of states. Let $\{\psi_n\}$ be a sequence of pure states converging to the state ψ. We have to prove that ψ is pure. Take a fixed pure state φ_0, then, by the transitivity of the action, there exists a sequence $\{\mathcal{U}_n\}$ of reversible channels such that $\psi_n = \mathcal{U}_n\varphi_0$. By proposition 2.3.24, there exists a subsequence $\{\mathcal{U}_{n_k}\}$ converging to the reversible channel $\mathcal{U}$. Hence

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

We now that reversible channels send pure states to pure states [100], so ψ is a pure state too. □

Chapter 3

Resource theories

In recent years, thermodynamics far from the thermodynamic limit has attracted remarkable attention [30–32], given the ever-increasing developments in the field of nanotechnology, and the actual experimental realisations of microscopic systems. In this new regime quantum effects become important, and must be taken into account in the physical description. To this end, quantum resource theories [33–35] have provided valuable tools, especially to study deterministic thermodynamics, viz. non-fluctuating processes. The power of resource theories stems from the fact that they describe thermodynamic transitions in a natural way: instead of specifying the actual details of the "mechanics" implementing a thermodynamic transition between two quantum states, they aim to study the convertibility criteria between states, namely *when* (and *not* how) a state can be transformed into another.

The ideas behind resource theories are pretty general and independent of the physical theory in which they are formulated [94–97]. The key concept is that the states of a physical system are to be viewed as resources. Then one studies the conversion between different resources by means of a restricted set of channels, which are singled out by the physical setting as those that easy to implement. For this reason, they are called *free operations*.

As to resources, some of them are very abundant or easy to get, therefore they are *free*; others, instead, may be extremely rare or hard to obtain, and these will be the most valuable. Thus, in the resource-theoretic framework a hierarchy of resources according to their value is naturally built in. This hierarchy is translated into mathematical terms as a partial preorder induced by free operations. In this preorder, a state is more valuable than

another if from the former we can reach a larger set of target states. In the light of this hierarchy, it is important to find functions that assign a value to resources in a way that it is consistent with the preorder. These are the *resource monotones* [33–35, 94, 142], and they can be very often identified with some thermodynamic potentials.

The subject of resource theories is extremely broad, both in quantum theory and in GPTs. In this chapter we give a brief presentation of the minimal necessary concepts to understand the contents and the spirit of the rest of the thesis. We will start with the definition of resource theory, and then we will introduce the resource preorder, establishing a hierarchy among the states of a GPT. Finally, we define resource monotones. We just mention that, besides the results presented here for resource theories in GPTs, we have studied how the thermodynamic limit emerges in the resource-theoretic approach [98, 99], but they are not reported here, for they have no direct consequences for the rest of the thesis.

3.1 Resource theories of states

In a great deal of situations, even in everyday life, one comes across resources, so often that we do not even question the meaning of this term. But what is actually a resource? To define a resource, one must first specify a task. In this setting, an object is a resource if it has some value with respect to performing that task. For example, a chemical product is a resource if it can be used to obtain something we want, for example it fertilises the soil.

Example 3.1.1. Ammonia has plenty of uses as a chemical product. In the process of producing ammonia, NH_3, we need hydrogen and nitrogen as resources, according to the reaction

$$N_2 + 3H_2 \to 2NH_3.$$

Regarding all the chemical species of the above reaction as resources, we understand an important feature of resources: they can be converted into one another.

As the example above highlights, chemistry is the prototypical resource theory, but there are plenty of examples even in physics, which is the

scope of our analysis. In particular, the advent of quantum information processing has provided very many examples of new tasks that can be performed only with access to quantum states, which, in our terminology, become quantum resources. Think, for example, of quantum teleportation [143] or dense-coding [144], which are only possible if one has access to entangled states. For this reason it is very natural to take quantum entanglement as a resource, and indeed it was the first quantum resource to be studied extensively [145].

In this presentation, we want to go beyond quantum theory, and to treat resources in general physical theories using the formalism of causal GPTs. Indeed, as shown in [94,95,98,99], many of the features of quantum resource theories [33–35] are in fact far more general in their scope. In the following we will always assume that our resources are states, like in most of concrete examples, but this need not be the case [35,94,136,146–148].

The idea behind a theory of resources is to identify which states are useless as resources, because they are abundant, easy to obtain, or of no use. These states are called *free states*. Clearly, all states of the trivial system, which are "non-physical states", are free, and represent a "void" resource. For a generic system A, we have a partition of $\mathsf{St}_1(\mathrm{A})$ into the set of free states F and its complement, which contains the true, costly, resources. As we highlighted above, the specification of F depends on the specific task we are considering. We can easily have a state free for one task, and a true resource for another.

We will assume the set of free states to be topologically closed: a state that can be arbitrarily well approximated by free states must be free too. Finally, is natural to assume that the tensor product of free states is another free state of the composite system. This is because if two resources have zero value if taken on their own, they have zero value also when taken together [94], therefore $F_{\mathrm{AB}} \supseteq F_{\mathrm{A}} \otimes F_{\mathrm{B}}$, where F_{S} denotes the set of free states of system S.

Clearly resources can be manipulated by an agent who has a task to perform, which means resources can be converted from one to another. However, the specific task one is considering constrains the allowed manipulation of resources: one is restricted to performing a subset of all transformations of the theory. These restricted transformations are usually regarded as being easy to implement, or realisable at no cost. They are called *free transformations*. Since we want the agent to be able to *choose* the operations they implement, it is customary to focus only on deterministic

transformations, i.e. channels, where there is no randomness involved. This is what we will do. For every input and output system A and B respectively, we will consider a partition of channels $\mathsf{DetTransf}(\mathrm{A},\mathrm{B})$ into a set of free channels $F_{\mathrm{A}\to\mathrm{B}}$ and a set of non-free channels. We will call $F_{\mathrm{A}\to\mathrm{B}}$ the set of *free operations*. For similar reason to the case of states, we take $F_{\mathrm{A}\to\mathrm{B}}$ to be topologically closed. Given the definition of states in OPTs, when A is the trivial system, $F_{\mathrm{A}\to\mathrm{B}}$ defines the free states of B. Therefore there is no need to deal with free operations and free states as separate objects [35, 94].

When B is the trivial system, we will always take the deterministic effect of A to be free. The reason why we do this will become clear in the following section.

Now, what are the properties of the free operations of a physical theory? They are summarised in the following requirements:

- the identity channel is free because doing nothing comes for free;
- the SWAP is a free operation;
- the sequential composition of free operations is a free operation;
- the parallel composition of free operations is a free operation.

These requirements are quite intuitive in their meaning. From these properties, it is obvious that the action of a free operation on a free state yields a free state, because it is the sequential composition of two free operations.

Now we are ready to collect all these remarks in the following definition of a resource theory.

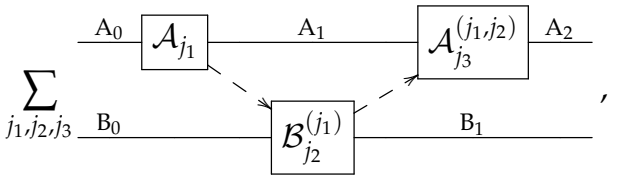
Definition 3.1.2. A *resource theory* on a causal GPT is the specification, for every pair of systems A and B, of a closed set of *free operations* $F_{\mathrm{A}\to\mathrm{B}} \subseteq \mathsf{DetTransf}(\mathrm{A},\mathrm{B})$. This set satisfies the following properties:

- $\mathcal{I} \in F_{\mathrm{A}\to\mathrm{A}}$;
- $\mathtt{SWAP} \in F_{\mathrm{AB}\to\mathrm{BA}}$;
- if $\mathcal{F}_1 \in F_{\mathrm{A}\to\mathrm{B}}$ and $\mathcal{F}_2 \in F_{\mathrm{B}\to\mathrm{C}}$, then $\mathcal{F}_2 \circ \mathcal{F}_1 \in F_{\mathrm{A}\to\mathrm{C}}$;
- if $\mathcal{F}_1 \in F_{\mathrm{A}\to\mathrm{B}}$ and $\mathcal{F}_2 \in F_{\mathrm{C}\to\mathrm{D}}$, then $\mathcal{F}_1 \otimes \mathcal{F}_2 \in F_{\mathrm{AC}\to\mathrm{BD}}$;
- $F_{\mathrm{A}\to\mathrm{I}} = \{u_{\mathrm{A}}\}$.

The interesting case, clearly, is when $F_{A\to B}$ is *strictly* contained in the set of all channels from A and B, because this means there is a true restriction on the allowed manipulations of the system. It is indeed from this restriction that the role of states as resources emerges, otherwise the resource theory is trivial.

From a more mathematical perspective, the properties satisfied by the set of free operations tell us that a resource theory specifies a strict symmetric monoidal subcategory of the underlying OPT on which it is defined [94, 95].

Example 3.1.3. The prototypical example of resource theory is the resource theory of quantum entanglement [145], where free states are separable states, and free operations are LOCC channels[1] [149–151], namely channels that can be decomposed as the action of local operations and classical communication exchanged between the two parties. Here is an example:



where dashed wires represent classical communication. Note that separable states are exactly the states that can be prepared with a LOCC protocol.

Alternatively, one can take as free operations the largest set of channels that send separable states to separable states. This is the set of separable operations [152].

We saw that free states are a special kind of free operations, therefore specifying free operations determines also free states: they are exactly those operations with trivial input. Yet, sometimes it is possible, or much easier, to specify only the free states. Can we recover a full resource theory just from the structure of free states? In general we will have a large freedom in the choice of free operations that respect the requirements of definition 3.1.2. The largest set of free operations compatible with the set of free states F is the set of channels from A to B that send the free states of A to free states of B:

$$F^{\max}_{A\to B} = \{\mathcal{C} \in \mathsf{DetTransf}\,(A, B) : \mathcal{C}f \in F_B,\ \text{where } f \in F_A\}\,, \tag{3.1.1}$$

[1] Recall that LOCC stands for "Local Operations and Classical Communication".

where F_S is the set of free states of system S. This is called the set of "resource-non-generating operations", and it is the largest set of free operations, because every other set of free operations satisfies the condition of eq. (3.1.1). However, this condition is not enough, because it says nothing about the behaviour of these operations under parallel composition. Indeed, from eq. (3.1.1) we know that the parallel composition of two free operations in $F^{\max}_{A\to B}$ preserves the product of free states $(\mathcal{C}\otimes\mathcal{C}')\,(f_A\otimes f_B)\in F_{AB}$, but the free states F_{AB} of the composite system AB may be also of a different form from product states. Therefore we take as free operations the subset of $F^{\max}_{A\to B}$ such that arbitrary tensor products of its elements preserve all free states of any composite system. These channels are also known as "completely resource-non-generating operations" [35].

3.2 The resource preorder

After giving the formal definition of resource theory, we want to examine its implications. One of the main applications of the resource-theoretic framework is to study state conversions when there is a restriction on the allowed channels. A physical example is thermodynamics, where one studies the thermodynamic convertibility of states under some external constraints, e.g. the fact that the temperature is constant or that no heat is exchanged in the thermodynamic transition. Thermodynamics also provides an example of a situation in which, to derive thermodynamic properties, such as relations between thermodynamic potentials, it is not important to specify the actual details of the transition between two states, but it is enough to know that such a transition is possible [153]. This is the typical situation arising when one considers functions of state, and this structure is beautifully captured by the preorder one can set up in a resource theory.

The preorder arises when one wishes to better specify the value of a resource. Indeed, in the previous section, we divided states into two classes: free and costly ones. This is a rough classification, because it is natural to expect that even among costly states there will be an internal hierarchy, with some of them closer to free states.

As it often happens experimentally, the processing of resources consumes or degrades them, even if such a processing is free. In this way, from a precious resource, we end up with resources that are more and

more useless. The idea is that a resource is more valuable than another if we can reach a larger set of resources by manipulating it. Therefore, the most natural way to define a hierarchy on resources is to take advantage of this idea about resource processing.

Definition 3.2.1. In a resource theory of states, we say that a state ρ is *more valuable* than a state[2] σ, and we write $\rho \succsim \sigma$, if there exists a free operation $\mathcal{F}$ such that $\mathcal{F}\rho = \sigma$.

We see that the hierarchy among resources is based on resource convertibility, and note that we do not need to specify the actual details of $\mathcal{F}$, but only that such an $\mathcal{F}$ exists. As already written, the relation $\succsim$ on the set of states is a preorder.

Proposition 3.2.2. *The relation $\succsim$ is a preorder*[3].

Proof. The relation $\succsim$ is reflexive. Indeed for any state ρ, $\rho \succsim \rho$ because the identity channel is a free operation. The relation is also transitive. Indeed, suppose we have $\rho \succsim \sigma$ and $\sigma \succsim \tau$. This means that there exist two free operations $\mathcal{F}_1$ and $\mathcal{F}_2$ such that $\mathcal{F}_1\rho = \sigma$ and $\mathcal{F}_2\sigma = \tau$. Taking the sequential composition, we have the free operation $\mathcal{F}_2\mathcal{F}_1$ such that $\mathcal{F}_2\mathcal{F}_1\rho = \tau$, which means $\rho \succsim \tau$. □

In general, however, we *cannot* conclude that, if $\rho \succsim \sigma$ and $\sigma \succsim \rho$, then $\rho = \sigma$. This only means that it is possible to convert ρ into σ with a free operation $\mathcal{F}$, and σ into ρ with a free operation $\mathcal{F}'$. Note that this does *not* even mean that $\mathcal{F}' = \mathcal{F}^{-1}$. Nevertheless, if $\rho \succsim \sigma$ and $\sigma \succsim \rho$, we can think of ρ and σ as *equivalent*, and we say that ρ is *as valuable as* σ. Indeed, it is straightforward to see that we can define an equivalence relation $\sim$, where $\rho \sim \sigma$ if $\rho \succsim \sigma$ and $\sigma \succsim \rho$. Taking the quotient of the set of states modulo $\sim$, the preorder $\succsim$ becomes a partial order $\succeq$ between equivalence classes.

Sometimes, given two states ρ and σ we cannot find neither a free operation converting ρ into σ, nor a free operation converting σ into ρ. In this case ρ and σ are "incomparable": we cannot establish which of the two is the more valuable. This means that the preorder on resources is only partial.

Let us prove that this preorder is compatible with the tensor product of states.

[2] Note that ρ and σ can be states of different systems.

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

Proposition 3.2.3. *Suppose $\rho \succsim \rho'$ and $\sigma \succsim \sigma'$, then one has $\rho \otimes \sigma \succsim \rho' \otimes \sigma'$.*

Proof. By hypothesis there exist two free operations $\mathcal{F}_1$ and $\mathcal{F}_2$ such that $\mathcal{F}_1\rho = \rho'$ and $\mathcal{F}_2\sigma = \sigma'$. Then $\mathcal{F}_1 \otimes \mathcal{F}_2$ is a free operation, and

$$(\mathcal{F}_1 \otimes \mathcal{F}_2)(\rho \otimes \sigma) = \rho' \otimes \sigma'.$$

This implies $\rho \otimes \sigma \succsim \rho' \otimes \sigma'$. □

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

Now we can prove that free states are indeed the least valuable states.

Proposition 3.2.4. *Let $f \in \mathsf{St}_1(\mathrm{A})$ be a free state. Then, for every state ρ of any system* B*, one has $\rho \succsim f$.*

Proof. Let us consider the channel $\mathcal{F} = |f)_{\mathrm{A}}(u|_{\mathrm{B}}$, which is a free operation because it is the sequential composition of free operations. Then for every $\rho \in \mathsf{St}_1(\mathrm{B})$, we have

$$\mathcal{F}\rho = f \operatorname{tr} \rho = f,$$

so $\rho \succsim f$. □

An immediate consequence is that all free states (even of different systems) are equivalent to each other. Indeed if f' is another free state, if we take $\rho = f'$, we have $f' \succsim f$. If instead we take $\rho = f$, proposition 3.2.4 tells us that $f \succsim f'$. In conclusion, $f \sim f'$.

Moreover, taking multiple copies of a free state does not increase its value. This is because the tensor product of free states is still a free state: we have $f \sim f^{\otimes n}$, for every $n \geq 1$. In summary, since there is no cost to prepare an arbitrary number of copies of free states, having many copies is just like having a single copy.

Remark 3.2.5. In the light of the resource preorder, we can understand why we required the deterministic effect, which destroys resources, to be a free operation. If f is a free resource of system A, there is a free process $\mathcal{F} : \mathrm{I} \to \mathrm{A}$ that prepares it from "nothing". According to our definition of the "more valuable" relation, this means that the void resource, i.e. "nothing", is more valuable than a real resource! This would be quite absurd if we did not impose that there is also a free process destroying f, thus implying that f is more valuable than "nothing". The conjunction of these two relations yields that a free state is equivalent to "nothing". This highlights how useless free states are.

We have seen that free states of system A are particular free operations, from the trivial system to A. However, if free operations are defined as having the same input and output system, the only free state is the state of the trivial system, or in other words, the theory does not have free states. In this case, can we identify a set of states that can be introduced "by hand" as free states of the theory? Proposition 3.2.4 states that free states are the minima of the preorder. We introduce the following definition, which characterises states that mimic some properties of free states, specifically the fact that they are the minima of the preorder.

Definition 3.2.6. A state $\rho \in \mathsf{St}_1(\mathrm{A})$ is *almost free* if for any state $\sigma \in \mathsf{St}_1(\mathrm{A})$ we have $\sigma \gtrsim \rho$.

We see that almost free states are all equivalent to each other: if ρ and ρ' are almost free states, by definition $\rho \gtrsim \rho'$ and $\rho' \gtrsim \rho$, so $\rho \sim \rho'$. Clearly free states are also almost free. If the theory has free states, almost free states are free too. Indeed, take σ to be a free state f, then we have $f \gtrsim \rho$. This implies that there is a free operation from I to A given by $\mathcal{F}f$, where $\mathcal{F}$ is a free operation converting f into ρ, such that $\rho = \mathcal{F}f$. Hence ρ is a free state too. Therefore, talking about almost free states is meaningful in theories where there are no free states.

Now, consider the almost free states that are stable under tensor product, by which we mean that the product of two almost free states is still an almost free state. These are the states that can be promoted "by hand" to free states of the theory. In conclusion, even if a theory has no free states, if it has almost free states stable under tensor product, we can introduce them as free states of the resource theory [35]: they are the minima of the resource preorder, and all equivalent to one another. The only difference is that we cannot prepare them with a free operation. We will use this fact in subsection 5.2.1, where we will add a free state "by hand" that cannot emerge from the definition of free operations.

3.3 Resource monotones

Having established a hierarchy among resources, it is sometimes useful to have a direct way of quantifying the value of a resource by assigning it a real number, its "price". This means translating the preorder on states

into the usual order on real numbers. To this end, we need real-valued functions that respect the preorder on resources.

Definition 3.3.1. A real-valued function $M : \mathsf{St}_1(\mathrm{A}) \to \mathbb{R}$ is a *resource monotone* (*monotone* for short) if $\rho \succsim \sigma$ implies $M(\rho) \geq M(\sigma)$.

In words, monotones assign a price to resources, consistent with their value. Specifically, if $\rho \sim \sigma$, then $M(\rho) = M(\sigma)$. Indeed, if $\rho \sim \sigma$, then $\rho \succsim \sigma$ and $\sigma \succsim \rho$, thus $M(\rho) \geq M(\sigma)$ and $M(\sigma) \geq M(\rho)$, whence one has $M(\rho) = M(\sigma)$.

A careful examination of the definition of monotones shows that they have some tricky subtleties. Indeed, it is not possible to translate the hierarchy of resources faithfully into the ordering of real numbers. The main reason for such a difficulty is that we can only establish a *partial preorder* among resources, whereas we have a *total order* on real numbers. Indeed, we can have two incomparable states ρ and σ, but if M is a monotone, we have either $M(\rho) \geq M(\sigma)$ or $M(\sigma) \geq M(\rho)$, because two real numbers can always be compared.

According to the definition of resource monotones, $\rho \succsim \sigma$ implies $M(\rho) \geq M(\sigma)$, but the converse implication in general does not hold. This means that the preorder $\succsim$ is more fundamental than the order induced by monotones. Indeed, if $M(\rho) \geq M(\sigma)$ we *cannot* conclude that $\rho \succsim \sigma$.

However, resource monotones are useful to detect *non-convertibility* of resources [94]. Recalling the definition, $M(\rho) < M(\sigma)$ means $\rho \not\succsim \sigma$, viz. there is no free operation converting ρ into σ. Similarly, if $M(\rho) = M(\sigma)$ we *cannot* conclude that $\rho \sim \sigma$. Indeed, a trivial monotone is a constant function that assigns the same value to all resources, irrespective of their place in the resource hierarchy. In this case, both equivalent and inequivalent resources have the same value.

To obtain a full equivalence between the preorder on resources and the ordering induced by monotones, we have to take more than one resource monotone. In this respect, a family of monotones $\{M_i\}_{i \in X}$ is said to be *complete* if we have $\rho \succsim \sigma$ if and only if $M_i(\rho) \geq M_i(\sigma)$ for every $i \in X$.

Proposition 3.3.2 ([94, proposition 5.2])**.** *Every resource theory admits a complete family of monotones.*

Proof. Take X to be state space of system A: $X = \mathsf{St}_1(\mathrm{A})$. Then let us label

monotones with a state τ. For every resource ρ, define $M_\tau(\rho)$ as

$$M_\tau(\rho) = \begin{cases} 1 & \text{if } \rho \succsim \tau \\ 0 & \text{if } \rho \not\succsim \tau \end{cases}.$$

Let us show that M_τ is a monotone for every state τ. Suppose $\rho \succsim \sigma$.

- If $\sigma \succsim \tau$, then by transitivity $\rho \succsim \tau$. In this case we have $M_\tau(\rho) = M_\tau(\sigma) = 1$, whence $M_\tau(\rho) \geq M_\tau(\sigma)$.
- If $\rho \not\succsim \tau$ and $\sigma \not\succsim \tau$, then $M_\tau(\rho) = M_\tau(\sigma) = 0$, whence $M_\tau(\rho) \geq M_\tau(\sigma)$.
- If $\rho \succsim \tau$ but $\sigma \not\succsim \tau$, $M_\tau(\rho) = 1$ and $M_\tau(\sigma) = 0$, and again $M_\tau(\rho) \geq M_\tau(\sigma)$.

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

Although we have managed to construct a complete family of resource monotones for every resource theory, such a family is not so practical, for it is indexed by the states themselves.

It is useful to classify resource monotones into some categories according to their behaviour under composition of resources [142]:

Additive $M(\rho \otimes \sigma) = M(\rho) + M(\sigma)$, for all states ρ and σ.

Partially additive $M(\rho^{\otimes n}) = nM(\rho)$ for every state ρ, and every $n \geq 1$.

Regularisable if $\lim_{n\to+\infty} \frac{1}{n} M(\rho^{\otimes n}) < +\infty$, and the limit exists for every state ρ.

Clearly

$$\text{Additive} \subseteq \text{Partially additive} \subseteq \text{Regularisable},$$

as it is straightforward to check. Later in section 5.4 we will encounter some examples of additive monotones. Regularisable monotones are fundamental when one wishes to study the thermodynamic limit [98, 99], because they can be extended to the regime where there are infinitely many

copies of a state (the thermodynamic limit). Indeed in this regime, the quantities that matter are densities, i.e. the value of the monotone per copy, and the condition $\lim_{n\to+\infty} \frac{1}{n} M\left(\rho^{\otimes n}\right) < +\infty$ means that it is possible to define a density in the thermodynamic limit, namely

$$M^{\infty}(\rho) := \lim_{n\to+\infty} \frac{1}{n} M\left(\rho^{\otimes n}\right).$$

M^{∞} is called the *regularisation* of M.

Proposition 3.3.3. *Given a regularisable monotone M, its regularisation M^{∞} is a resource monotone too.*

Proof. Suppose $\rho \succsim \sigma$, then $\rho^{\otimes n} \succsim \sigma^{\otimes n}$ by proposition 3.2.3. Therefore, since M is a monotone, $M\left(\rho^{\otimes n}\right) \geq M\left(\sigma^{\otimes n}\right)$. Then we get $\frac{1}{n} M\left(\rho^{\otimes n}\right) \geq \frac{1}{n} M\left(\sigma^{\otimes n}\right)$, and by taking the limit for $n \to +\infty$ of both sides we get $M^{\infty}(\rho) \geq M^{\infty}(\sigma)$. □

For this reason, it is meaningful to give the following definition.

Definition 3.3.4. A monotone M is *regularised* if there exists a monotone M', such that, for every ρ

$$M(\rho) = \lim_{n\to+\infty} \frac{1}{n} M'\left(\rho^{\otimes n}\right).$$

In words, a monotone is regularised if it is the regularisation of another (regularisable) monotone. Regularised monotones M are the meaningful monotones in the thermodynamic limit, because they are expressed as the density per particle of another monotone M'. Note that partially additive monotones are not only regularisable, but also regularised: it is enough to take $M' = M$. Indeed

$$M(\rho) = \lim_{n\to+\infty} \frac{1}{n} M\left(\rho^{\otimes n}\right) = \lim_{n\to+\infty} \frac{1}{n} \cdot nM(\rho).$$

Regularised monotones, which include partially additive monotones, exhibit a remarkable property.

Proposition 3.3.5. *Let M be a regularised monotone. If f is a free state, $M(f) = 0$. Moreover, for every state ρ, $M(\rho) \geq 0$.*

Proof. M is regularised, therefore there exists a monotone M' such that $M(f) = \lim_{n\to+\infty} \frac{1}{n} M'(f^{\otimes n})$. Since f is a free state, we have $f \sim f^{\otimes n}$, therefore $M'(f^{\otimes n}) = M'(f)$. Hence

$$M(f) = \lim_{n\to+\infty} \frac{1}{n} M'(f) = 0,$$

because $M'(f)$ is a constant. To conclude the proof, recall that, for every state ρ, $\rho \gtrsim f$. Being M a monotone, $M(\rho) \geq M(f) = 0$. □

We motivated the introduction of the preorder on states with the fact that some costly resources could be in fact close to free states. We model this idea of closeness by introducing a distance from a state to the set of free states.

$$d(\rho) := \inf_{f\in F_{\mathrm{A}}} \|\rho - f\|$$

Here ρ is a state of A, and f is a free state of A. Note that the infimum is achieved, so it is in fact a minimum, because we have assumed the set of free states to be topologically closed. This distance d is a resource monotone. Indeed, suppose $\rho \gtrsim \sigma$. This means there exists a free operation $\mathcal{F}$ such that $\mathcal{F}\rho = \sigma$. Then

$$d(\sigma) = \inf_{f\in F_{\mathrm{A}}} \|\sigma - f\| = \inf_{f\in F_{\mathrm{A}}} \|\mathcal{F}\rho - f\| \leq \inf_{f\in F_{\mathrm{A}}} \|\mathcal{F}\rho - \mathcal{F}f\|,$$

because $\{\mathcal{F}f\}$ is a subset of F_{A}. Now, the operational norm of a vector is non-increasing under a channel by proposition 2.2.23. Therefore

$$\|\mathcal{F}\rho - \mathcal{F}f\| = \|\mathcal{F}(\rho - f)\| \leq \|\rho - f\|.$$

Taking the infimum, we get $\inf_{f\in F_{\mathrm{A}}} \|\mathcal{F}\rho - \mathcal{F}f\| \leq \inf_{f\in F_{\mathrm{A}}} \|\rho - f\|$. In conclusion

$$d(\sigma) \leq \inf_{f\in F_{\mathrm{A}}} \|\rho - f\| = d(\rho).$$

This shows that d is indeed a monotone. We have presented this construction for the operational norm, but in fact it can be extended to any function $D(\rho, \sigma)$ that is decreasing under the action of channels [34, 35].

Chapter 4

Sharp theories with purification

In this chapter we introduce the theories we are going to study for the rest of the thesis. These theories can be roughly characterised as those admitting a level of description where all processes are pure and reversible, and all measurements are sharp. For these reasons, these theories are particularly appealing for the foundation of physics, and in particular thermodynamics. The key axiom defining them—Purification—underpins all dilation and extension theorems [67, 71], so important in quantum theory, and from a thermodynamic perspective it gives a formal guarantee that every observer can enlarge their system in order to deal with an isolated one, where information is maximal, and all evolutions are reversible. From a thermodynamic viewpoint, somehow Purification can be regarded as dual to Causality: if Causality allows one to go from larger to smaller systems, Purification enables us to do the opposite. Moreover, mixed states, so important for thermodynamics emerge in a different way: not only as ensembles, but also as marginals of pure states. This is particularly appealing for the foundations of thermodynamics, because one no longer needs to resort to fictitious and subjective ensembles, but mixed states arise because one is tracing out the degrees of freedom of the environment. A similar fact also holds for the issue of irreversibility.

Informally, sharp theories with purification are causal theories satisfying three additional axioms:

Purity Preservation The composition of two pure transformations is a pure transformation.

Pure Sharpness Every system has at least one pure sharp observable.

Purification Every state can be modelled as the marginal of a pure state. Such a modelling is unique up to local reversible transformations.

These axioms will be presented more formally in section 4.1, where their first consequences, in particular of Purification, will be examined. Sharp theories with purification enjoy some remarkable properties that, in some sense, make them close to quantum theory. In this chapter we will focus on those properties relevant to the thermodynamic analysis of chapter 5. One of the key features is a state-effect duality, by which with every normalised pure state we can associate a unique normalised pure effect, the *dagger* of the state, and vice versa [101, 110]. The second key result, which constitutes the high spot of this chapter, is the diagonalisation theorem, which states that in these theories every state can be *diagonalised* [101, 110], viz. written as a convex combination of perfectly distinguishable pure states, with unique coefficients [101]. The fact that with every state we can associate the probability distribution of its eigenvalues allows us to introduce entropic functions, which will be done in section 5.4. Furthermore, the two key properties of sharp theories with purification allow us to identify every element of the vector space of effects with a physical observable. This will prove of fundamental importance in section 5.6 to define thermal states, and to derive a lot of properties of a generalisation of Shannon-von Neumann entropy (section 5.5).

From these properties, others follow, but we do not report them in detail in this chapter because they are not so crucial for thermodynamics. We summarise them briefly here. A consequence of the state-effect duality is that sharp theories with purification are (strongly) self-dual [111], which enables us to extend the dagger to *all* transformations [111], not just to states and effects. The conjunction of strong self-duality with the state-effect duality implies that these theories satisfy the no-restriction hypothesis [67, 125]. Moreover, we can prove that there exists a pure projector on every face of the state space [111]. The conjunction of these two facts implies that sharp theories with purifications are Euclidean Jordan algebras [111, 120, 121, 154, 155], and that their interference is constrained at most to the second order [111]. However, note that not all Euclidean Jordan algebras are sharp theories with purification.

We conclude this chapter by presenting two new examples of sharp theories with purification. They are both theories constructed by imposing some superselection rules to quantum theory, and by defining system

composition in such a way that the three defining axioms are satisfied. The first example [105] is a theory where every non-trivial system is given by a pair of isomorphic quantum systems. This theory in section 5.8 will provide a counterexample in which majorisation is *not* sufficient to characterise certain thermodynamic transitions. The second example [101] is an extension of classical theory in which some systems look classical at the single-system level, but become entangled when composed. This shows that classical theory can be embedded and treated as a sub-theory of a sharp theory with purification, and that, ultimately, the results we obtain for sharp theories with purification can be extended to classical theory too, to provide an information-theoretic foundation of classical thermodynamics.

4.1 The axioms and their first consequences

In this section we present the axioms defining sharp theories with purification. These axioms are added on top of Causality, and will single out a class of theories where everything is pure and reversible at the fundamental level.

The first axiom to be added on top of Causality, Purity Preservation, states that no information can leak to the environment when two pure transformations are composed.

Axiom 4.1.1 (Purity Preservation [90])**.** *Sequential and parallel compositions of pure transformations yield pure transformations.*

We consider Purity Preservation as a fundamental requirement to do physics. Considering the theory as an algorithm to make deductions about physical processes, Purity Preservation ensures that, when presented with maximal information about two processes, the algorithm outputs maximal information about their composition [90]. Purity Preservation is very close to a slightly weaker axiom, Atomicity of Composition, introduced by D'Ariano in [156], and used in the axiomatisation of [68]. However Purity Preservation is stronger, in that it requires the preservation of purity also for *parallel* composition, and not just for sequential composition like in D'Ariano's original axiom. An immediate consequence of Purity Preservation is that the product of two pure states is pure, a fact usually

proved using Local Tomography [67]. Notably, quaternionic quantum theory fails this principle, for the product of two pure states is not pure in general [120, 121].

The second axiom, Pure Sharpness, guarantees that every system possesses at least one elementary property, in the sense of Piron [157]. Recall that here we are not assuming the no-restriction hypothesis, so the following axiom needs to be imposed.

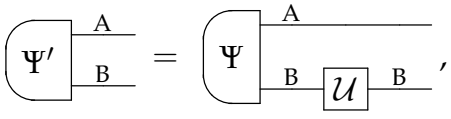
Axiom 4.1.2 (Pure Sharpness [110])**.** *For every system there exists at least one pure effect occurring with unit probability on some state.*

Pure Sharpness is reminiscent of the Sharpness axiom used in Hardy's 2011 axiomatisation [70, 73], which requires a one-to-one correspondence between pure states and effects that distinguish maximal sets of states. A similar axiom also appeared in works by Wilce [158–160], where he stipulates that for every pure effect there exists a *unique* state on which it occurs with probability 1.

The two axioms above are satisfied by both classical and quantum theory. Our last axiom, Purification, is precisely the one that characterises all physical theories admitting a fundamental level of description where all deterministic processes are pure and reversible. Essentially, Purification expresses a strengthened version of the principle of conservation of information [90, 129], demanding not only that information be conserved, but also that randomness can always be modelled as due to the presence of some inaccessible degree of freedom. In its simplest form, Purification is phrased as a requirement about *causal* theories, where the marginal of a bipartite state is defined in a canonical way. In this case, we say that a state $\rho \in \mathsf{St}_1(\mathrm{A})$ can be purified if there exists a pure state $\Psi \in \mathsf{PurSt}_1(\mathrm{AB})$ that has ρ as its marginal on system A. In this case, we call Ψ a *purification* of ρ and B a *purifying system*. The axiom is as follows.

Axiom 4.1.3 (Purification [67])**.** *Every state can be purified. Every two purifications of the same state, with the same purifying system, differ by a reversible channel on the purifying system.*

The second part of the axiom states that, if $\Psi, \Psi' \in \mathsf{PurSt}_1(\mathrm{AB})$ are such that $\mathrm{tr}_{\mathrm{B}}\Psi_{\mathrm{AB}} = \mathrm{tr}_{\mathrm{B}}\Psi'_{\mathrm{AB}}$, then



where $\mathcal{U}$ is a reversible channel on B.

Quantum theory, both on complex and real Hilbert spaces, satisfies Purification. Recently also Spekkens' toy model[1] [161] has been shown to satisfy Purification [162]. Other non-trivial examples are fermionic quantum theory [103, 104], and doubled quantum theory [105], presented in section 4.6. Remarkably, even classical theory can be regarded as a subtheory of a larger physical theory where Purification is satisfied (see section 4.7) [101].

Definition 4.1.4 (Sharp theories with purification)**.** A causal theory satisfying Purity Preservation, Pure Sharpness, and Purification will be called a *sharp theory with purification.*

In the rest of the section we will outline the first consequences of these axioms, especially of Purification.

4.1.1 First consequences

The easiest consequence of Purification is that reversible channels act transitively on the set of pure states [67].

Proposition 4.1.5 (Transitivity)**.** *For any pair of pure states ψ, ψ', there exists a reversible channel $\mathcal{U}$ that $\psi' = \mathcal{U}\psi$.*

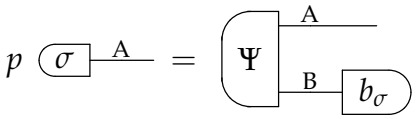
Proof. Every system A is a purifying system for the trivial system I. Then ψ and ψ' are two purifications of the same deterministic state of the trivial system (which is the number 1), therefore they differ by a reversible channel on the purifying system A, which means $\psi' = \mathcal{U}\psi$. □

As a consequence, every finite-dimensional system possesses a unique invariant state, which is an internal state (see proposition 2.3.27). Also, transitivity implies that the set of pure states is compact for every system (proposition 2.3.29). This is generally a non-trivial property—see [163] for a counterexample of a state space with a non-closed set of pure states.

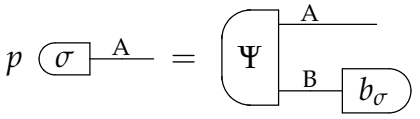
A crucial consequence of Purification is the *steering property*.

[1] The original, non-convex, version of it [161].

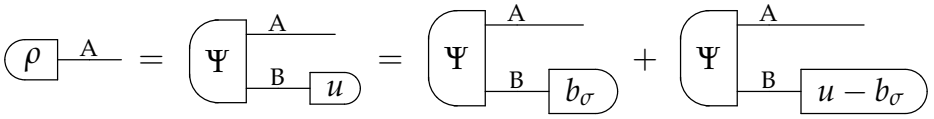
Theorem 4.1.6 (Pure Steering). *Let $\rho \in \mathsf{St}_1(\mathrm{A})$ and let $\Psi \in \mathsf{PurSt}_1(\mathrm{AB})$ be a purification of ρ. Then σ is contained in ρ if and only if there exist an effect b_σ on the purifying system* B *and a non-zero probability p such that*

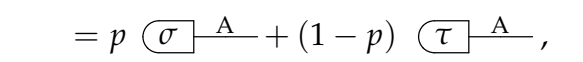


Proof. Sufficiency is easy. Consider the observation-test $\{b_\sigma, u - b_\sigma\}$ such that



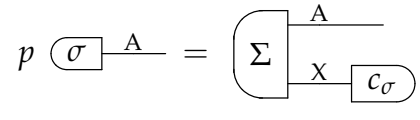
Then





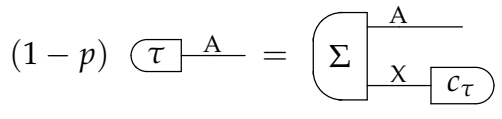
where τ is the state induced by applying the effect $u - b_\sigma$. This proves that σ is contained in ρ.

Conversely, if σ is contained in ρ, it means that $\rho = p\sigma + (1-p)\tau$, with $p \in (0,1)$. From an operational point of view, it means that there is a preparation-test $\{p\sigma, (1-p)\tau\}$, of which ρ is the coarse-graining. By Physicalisation of Readout, there exists a normalised bipartite state $\Sigma \in \mathsf{St}_1(\mathrm{AX})$, and an observation-test $\{c_\sigma, c_\tau\}$ on X such that



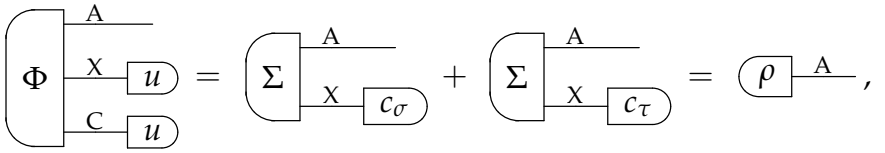
(4.1.1)

and

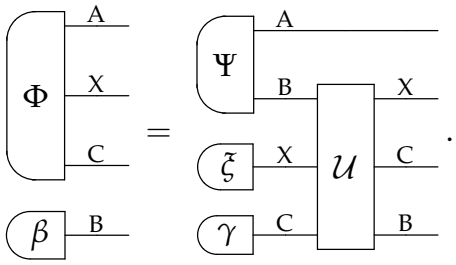


. (4.1.2)

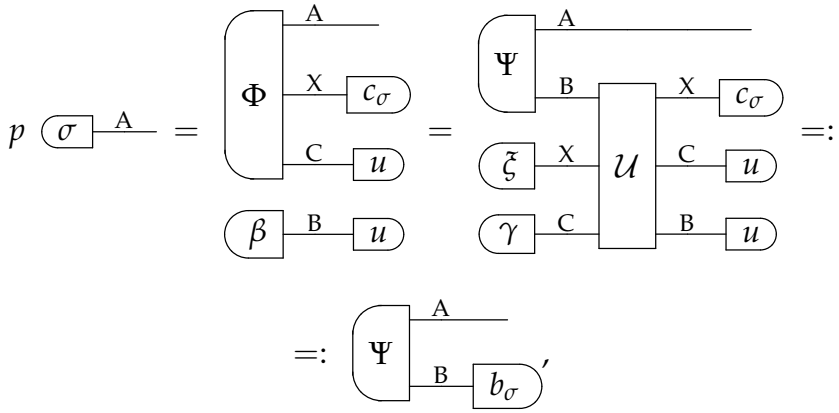
In general, Σ will not be pure, so let us take a purification $\Phi \in \mathsf{PurSt}_1(\mathrm{AXC})$ of Σ. Note that Φ is a purification of ρ too, indeed



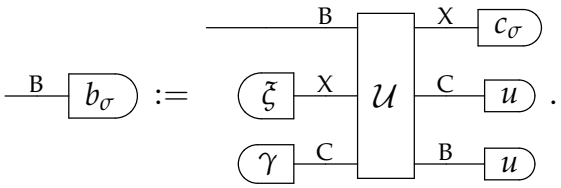
having used eqs. (4.1.1) and (4.1.2). Now, let us show that we can induce $p\sigma$ by applying a suitable effect on the purifying system of any purification $\Psi \in \mathsf{PurSt}_1(AB)$ of ρ. Now, take the pure states $\beta \in \mathsf{PurSt}_1(B)$, $\xi \in \mathsf{PurSt}_1(X)$, and $\gamma \in \mathsf{PurSt}_1(C)$. Then $\Phi \otimes \beta$ and $\Psi \otimes \xi \otimes \gamma$ are pure states (by Purity Preservation), and they are purifications of ρ with the same purifying system XCB (up to system swapping), so



Then



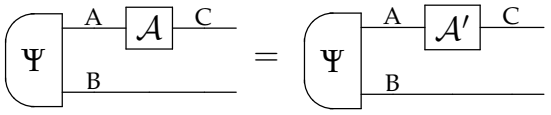
where



This proves the theorem. □

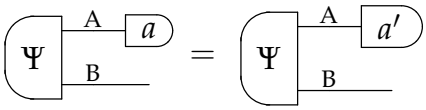
Purification also enables us to link equality upon input (as defined in definition 2.3.18) to equality on purifications [67, theorem 7], as an easy consequence of Pure Steering.

Proposition 4.1.7. *Let ρ be a state of system* A *and let* $\Psi \in \mathsf{St}_1(\mathrm{AB})$ *be a purification of* ρ*. Then, for every pair of transformations* $\mathcal{A}$ *and* $\mathcal{A}'$*, transforming* A *into* C*, if*



then $\mathcal{A} =_\rho \mathcal{A}'$*.*

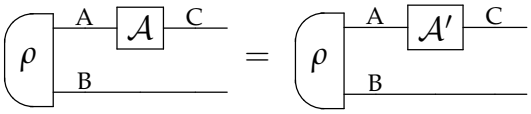
If system C *is trivial, then one has the full equivalence: for every pair of effects* a *and* a'



if and only if $a =_\rho a'$*.*

Pure Steering guarantees the existence of pure dynamically faithful states, in the following sense.

Definition 4.1.8. A state $\rho \in \mathsf{St}_1(\mathrm{AB})$ is *dynamically faithful* on system A if for every system C and for every pair of transformations $\mathcal{A}$ and $\mathcal{A}'$ transforming A into C



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

Thanks to Pure Steering, we have the following characterisation.

Proposition 4.1.9. *A pure state* Ψ_{AB} *is dynamically faithful on system* A *if and only if its marginal* ω_{A} *on* A *is internal.*

Proof. The proof is an adaptation of the arguments of [67, theorems 8, 9], which is valid even without invoking the Local Tomography axiom used therein [164]. □

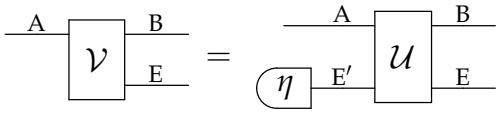
An indirect consequence of Pure Steering is a simple condition for a set of transformations to be a test (cf. [67, theorem 18]).

Proposition 4.1.10. *A set of transformations* $\{\mathcal{A}_i\}_{i=1}^n \subset \mathsf{Transf}(\mathrm{A}, \mathrm{B})$ *is a test if and only if* $\sum_{i=1}^n u_{\mathrm{B}} \mathcal{A}_i = u_{\mathrm{A}}$*. Specifically, a set of effects* $\{a_i\}_{i=1}^n$ *is an observation-test if and only if* $\sum_{i=1}^n a_i = u_{\mathrm{A}}$*.*

A proof of this proposition that does not make use of Local Tomography can be found in [130].

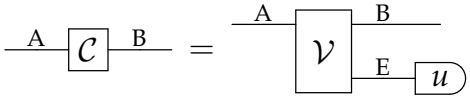
We will also make use of the following important consequences of Purification. The first states that every channel admits a dilation to a reversible channel [67, subsection 9 A].

Proposition 4.1.11. *For every channel* $\mathcal{C} \in \mathsf{Transf}(\mathrm{A}, \mathrm{B})$ *there is a reversible extension* $\mathcal{V} \in \mathsf{Transf}(\mathrm{A}, \mathrm{BE})$ *for some system* E*, namely a channel* $\mathcal{V}$ *of the form*

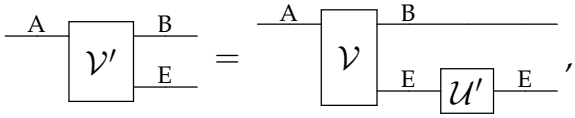


(4.1.3)

for some system E'*, some pure state* $\eta \in \mathsf{St}_1(\mathrm{E}')$*, where* $\mathcal{U}$ *is a reversible channel, such that*



Moreover, if $\mathcal{V}$ *and* $\mathcal{V}'$ *are two reversible extensions of the same channel, one has*

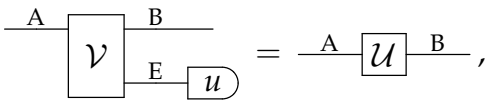


where $\mathcal{U}'$ *is a reversible channel on* E*.*

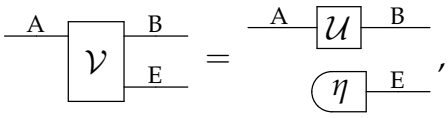
This is the form Purification takes for channels; indeed it is possible to prove that reversible channels are pure channels in theories with purification [67, corollary 28], so $\mathcal{V}$ is also a *pure* extension of the channel $\mathcal{C}$.

An immediate consequence is the following corollary.

Corollary 4.1.12. *Suppose* $\mathcal{V}$ *is a channel of the form* (4.1.3)*, and such that*

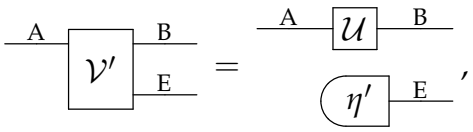


where $\mathcal{U}$ *is a reversible channel. Then*

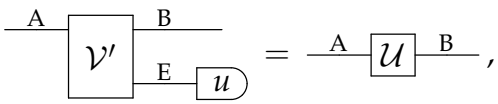


where η *is some pure state.*

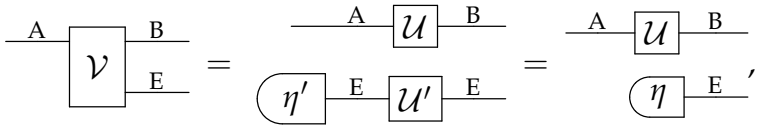
Proof. Consider the channel



for η' pure; which is of the form (4.1.3), where the joint reversible channel is $\mathcal{U} \otimes \mathcal{I}_E$. Clearly, we have



therefore, by proposition 4.1.11, there is a reversible channel $\mathcal{U}'$ on E such that



where $\eta = \mathcal{U}'\eta'$ is a pure state because η' is pure. □

Now we can move to study less immediate consequences of the axioms, which appeared in [101].

4.2 State-effect duality

In this section we derive the first important property of sharp theories with purification: a duality between normalised pure states and normalised pure effects. To do that, first we need some technical lemmas.

4.2.1 Technical lemmas

Pure Sharpness stipulates that for every system there is a pure effect occurring with probability 1 on some state. We can easily show that such a state must be pure [68, lemma 26, theorem 7].

Proposition 4.2.1. *Let a be a normalised pure effect. Then there exists a pure state α such that $(a|\alpha) = 1$. If ρ is another state such that $(a|\rho) = 1$, then $\rho = \alpha$.*

Combining the above result with our Pure Sharpness axiom, we derive the following proposition [110, proposition 9].

Proposition 4.2.2. *For every pure state $\alpha \in \mathsf{PurSt}_1(\mathrm{A})$ there exists at least one pure effect $a \in \mathsf{PurEff}(\mathrm{A})$ such that $(a|\alpha) = 1$.*

Proof. By Pure Sharpness, there exists at least one pure effect a_0 such that $(a_0|\alpha_0) = 1$, for some state α_0, where α_0 is pure. Now, for a generic pure state α, by transitivity, there is a reversible channel $\mathcal{U}$ such that $\alpha = \mathcal{U}\alpha_0$. Hence, the effect $a := a_0\mathcal{U}^{-1}$ is pure and $(a|\alpha) = 1$. □

In summary, for every normalised pure effect $a \in \mathsf{PurEff}_1(\mathrm{A})$, we can associate a *unique* pure state $\alpha \in \mathsf{PurSt}_1(\mathrm{A})$ with it such that $(a|\alpha) = 1$. A probabilistic model like this has been dubbed "sharp" by Wilce [158–160, 165]. Conversely, given a pure state α, there always exists at least one pure effect a such that $(a|\alpha) = 1$. This shows that there is a surjective correspondence between normalised pure effects and normalised pure states. We will show in a while that this correspondence is in fact bijective.

These results, presented here as propositions, were instead taken by Selby and Coecke as fundamental requirements for the definition of *test structure*, providing an operational characterisation of the Hermitian adjoint [166], used in a new reconstruction of quantum theory [137].

Probability balance of pure bipartite states

Given a normalised state $\rho \in \mathsf{St}_1(\mathrm{A})$, we define the probability p_* as the maximum probability that a pure state can have in a convex decomposition of ρ, namely[2]

$$p_* := \max_{\alpha \in \mathsf{PurSt}_1(\mathrm{A})} \{p : \rho = p\alpha + (1-p)\sigma, \sigma \in \mathsf{St}_1(\mathrm{A})\}.$$

We call p_* the *maximum eigenvalue* of ρ, and say that the *pure* state α is the corresponding *eigenstate*. The reason for this terminology will become clear once we prove our diagonalisation theorem.

A fundamental consequence of our axioms is that both marginals of a bipartite state have the same maximum eigenvalue.

Theorem 4.2.3. *Let Ψ be a pure bipartite state of system AB, let ρ_A and ρ_B be its marginals on systems A and B respectively. Then, ρ_A and ρ_B have the same*

[2] Note that the maximum is well-defined because the set of pure states is compact, thanks to transitivity.

maximum eigenvalue, namely

$$p_{*,\mathrm{A}} = p_{*,\mathrm{B}} =: p_*,$$

where $p_{*,\mathrm{A}}$ *and* $p_{*,\mathrm{B}}$ *are the maximum eigenvalues of* ρ_A *and* ρ_B *respectively.*

Moreover, when ρ_A *(or equivalently* ρ_B *) is decomposed as* $\rho_\mathrm{A} = p_*\alpha + (1-p_*)\,\sigma$ *for some* pure *state* α *and some state* σ*, the states* α *and* σ *are perfectly distinguishable with the observation-test* $\{a, u_\mathrm{A} - a\}$*, where* a *is any pure effect such that* $(a|\alpha) = 1$*.*

Proof. The fact that both marginals have the same maximum eigenvalue was proved in [110, theorem 2, corollary 1], and we will not report the proof here for the sake of brevity.

Now, write ρ_A as

$$\rho_\mathrm{A} = p_*\alpha + (1-p_*)\,\sigma, \tag{4.2.1}$$

where α is an eigenstate with maximum eigenvalue of ρ, and σ is possibly mixed. By [110, proposition 11], if a is a pure effect such that $(a|\alpha) = 1$ we have $(a|\rho_\mathrm{A}) = p_*$. Combining this equality with eq. (4.2.1) we finally obtain

$$p_* = (a|\rho_\mathrm{A}) = p_* + (1-p_*)\,(a|\sigma)\,,$$

which implies $(a|\sigma) = 0$ (unless $p_* = 1$, but in this case the state ρ_A is pure). Hence, α and σ are perfectly distinguishable with the test $\{a, u_\mathrm{A} - a\}$. □

Now we have managed to decompose every given state into a mixture of two perfectly distinguishable states. The probability balance has a lot of other consequences. The first is that every non-trivial system has at least two perfectly distinguishable pure states. To prove it, however, first of all, we must note that for the invariant state χ, due to its invariance under the action of reversible channels, every pure state is an eigenstate with maximum eigenvalue. Indeed, if χ is decomposed as

$$\chi = p_*\alpha + (1-p_*)\,\sigma,$$

it can also be decomposed as

$$\chi = p_*\mathcal{U}\alpha + (1-p_*)\,\mathcal{U}\sigma,$$

where $\mathcal{U}$ is a reversible channel. Owing to transitivity, every pure state α' can be obtained as $\mathcal{U}\alpha$ for some suitable reversible channel, meaning that every pure state is an eigenstate with maximum eigenvalue.

Corollary 4.2.4. *If* A $\neq$ I, *then every pure state of* A *is perfectly distinguishable from some other pure state.*

Proof. The proof is an application of theorem 4.2.3 to the case of the invariant state, and it has already appeared in [110, corollary 3]. Since for every pure state α, $\chi = p_* \alpha + (1 - p_*)\, \sigma$, α is perfectly distinguishable from σ, and from all the pure states contained in σ. □

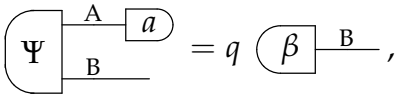
It is quite remarkable that the existence of perfectly distinguishable states pops out from the axioms, without being assumed from the start, or without relying on mathematical assumptions such as the no-restriction hypothesis (cf. proposition 2.3.21).

Another consequence of the probability balance is the following.

Corollary 4.2.5. *Let* ρ *be a mixed state of system* A. *Then, the following are equivalent:*

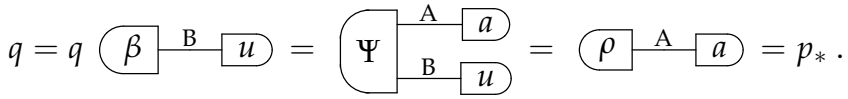
1. α *is an eigenstate of* ρ *with maximum eigenvalue* p_**;*
2. $(a|\rho) = p_*$ *for every pure effect* a *such that* $(a|\alpha) = 1$.

Proof. By [110, proposition 11], we already know that 1 $\Rightarrow$ 2. Let us prove the converse implication 2 $\Rightarrow$ 1. Suppose that $(a|\rho) = p_*$, where a is a pure effect such that $(a|\alpha) = 1$ on some pure state α. Let us show that α is an eigenstate of ρ with maximum eigenvalue p_*. Now, for every purification of ρ, say $\Psi \in \mathsf{PurSt}_1(\mathrm{AB})$, one has

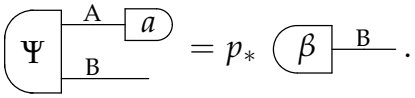


(4.2.2)

where β is a normalised state, pure by Purity Preservation. We have

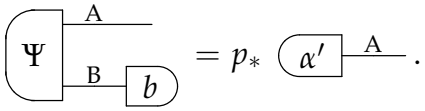


Hence eq. (4.2.2) becomes



(4.2.3)

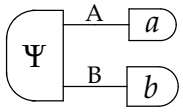
This condition implies that β is an eigenvector of the marginal state $\widetilde{\rho} = \mathrm{tr}_{\mathrm{A}}\Psi$. The implication $1 \Rightarrow 2$ guarantees that $(b|\widetilde{\rho}) = p_*$, for every pure effect b such that $(b|\beta) = 1$. The last condition implies an equation very similar to eq. (4.2.3):



(4.2.4)

for some pure state α'. Hence, α' is an eigenstate of ρ with eigenvalue p_*. To conclude the proof, it is enough to observe that $\alpha' = \alpha$. Indeed, combining eqs. (4.2.3) and (4.2.4) we have

$$(a|\alpha') = \frac{1}{p_*}$$



$$= (b|\beta) = 1,$$

which implies $\alpha' = \alpha$ by proposition 4.2.1. □

For every state we can also define another probability

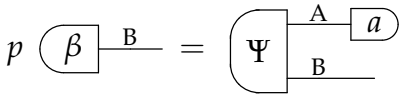
$$p^* := \sup_{a \in \mathsf{PurEff}(\mathrm{A})} (a|\rho)\,. \tag{4.2.5}$$

We will see soon that the supremum in the definition of p^* is in fact a maximum. Note that for pure states $p^* = 1$, thanks to proposition 4.2.2.

By corollary 4.2.5, for every state one has the bound $p_* \leq p^*$. We can in fact prove that one has the equality.

Proposition 4.2.6. *For every state $\rho \in \mathsf{St}_1(\mathrm{A})$ one has $p^* = p_*$.*

Proof. It is enough to show that $p^* \leq p_*$. Pick a pure effect $a \in \mathsf{PurEff}(\mathrm{A})$ such that $(a|\rho) \neq 0$. Such a pure effect exists because any pure effect a such that $(a|\alpha) = 1$, where α is an eigenstate of ρ with maximum eigenvalue p_*, has the property $(a|\rho) = p_* \neq 0$. Now consider a purification $\Psi \in \mathsf{PurSt}_1(\mathrm{AB})$ of ρ, and define the pure state β as



Note that p is non-vanishing because it is given by $p = (a|\rho)$. So β arises in a convex decomposition of the marginal of Ψ on B with probability p. By construction $p \leq p_*$, namely $(a|\rho) \leq p_*$. Taking the supremum over a, we finally obtain $p^* \leq p_*$, thus proving that $p^* = p_*$. □

As a consequence we have that p^* is achieved by applying any pure effect a such that $(a|\alpha) = 1$ to ρ, where α is an eigenstate of ρ with maximum eigenvalue. Therefore, the supremum in the definition of p^* in eq. (4.2.5) is in fact a maximum.

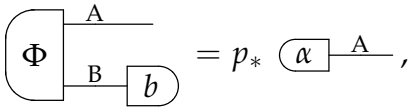
Probability balance for purifications of the invariant state

Recall that the invariant state $\chi \in \mathsf{St}_1(\mathrm{A})$ can be written as

$$\chi = p_* \alpha + (1 - p_*)\, \sigma$$

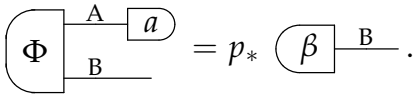
for every *pure* state $\alpha \in \mathsf{PurSt}_1(\mathrm{A})$, where σ is a suitable state. In addition, we have the following proposition.

Proposition 4.2.7. *Let $\Phi \in \mathsf{PurSt}_1(\mathrm{AB})$ be a purification of the invariant state χ_{A}. Then, for every pure state $\alpha \in \mathsf{PurSt}_1(\mathrm{A})$ there exists a pure effect $b \in \mathsf{PurEff}(\mathrm{B})$ such that*

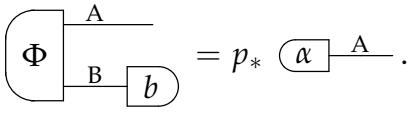


where p_* is the maximum eigenvalue of χ_{A}. As a consequence, every normalised pure effect $a \in \mathsf{PurEff}_1(\mathrm{A})$ satisfies the condition $(a|\chi_A) = p_*$.

Proof. We know that every pure state $\alpha \in \mathsf{PurSt}_1(\mathrm{A})$ is an eigenstate of χ_{A}. In the proof of corollary 4.2.5 we saw that, if a is a pure effect such that $(a|\alpha) = 1$, and $\Phi \in \mathsf{PurSt}_1(\mathrm{AB})$ is a purification of χ_{A}, we have



In the same proof we saw that, if b is a pure effect such that $(b|\beta) = 1$, we have



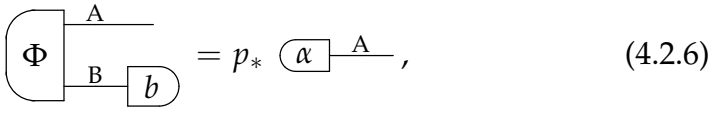
To conclude the proof, recall that there is a pure state α associated with every normalised pure effect a such that $(a|\alpha) = 1$, and that every pure state is an eigenstate of χ_{A} with maximum eigenvalue. Then by corollary 4.2.5, we have $(a|\chi_{\mathrm{A}}) = p_*$. □

4.2.2 State-effect duality

Using the results of the previous subsection, one can establish a one-to-one correspondence between normalised pure states and normalised pure effects. We refer to this correspondence as the *dagger* of states and effects.

Proposition 4.2.8. *For every pure state* $\alpha \in \mathsf{PurSt}_1(\mathrm{A})$ *there is a unique (normalised) pure effect* $a \in \mathsf{PurEff}_1(\mathrm{A})$ *such that* $(a|\alpha) = 1$.

Proof. The proof is essentially the same as in [68, theorem 8], even though we are assuming fewer axioms. Suppose that a and a' are two pure effects such that $(a|\alpha) = (a'|\alpha) = 1$. Then, let $\Phi \in \mathsf{PurSt}_1(\mathrm{AB})$ be a purification of the invariant state χ_{A}. By proposition 4.2.7, there exists a pure effect b such that

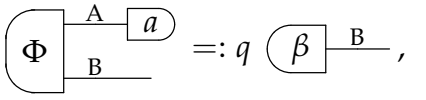


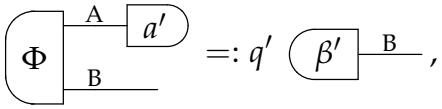
(4.2.6)

and the two effects a and a' must satisfy

$$(a|\chi_{\mathrm{A}}) = p_* = (a'|\chi_{\mathrm{A}})\,. \tag{4.2.7}$$

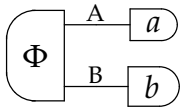
Now, let us define the pure states β and β' through the relations





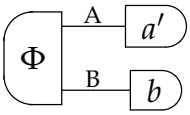
where q and q' are suitable probabilities. By applying the deterministic effect on both sides and using eq. (4.2.7) one obtains the equality $q = p_* = q'$. Hence, eqs. (4.2.6) and (4.2.7) lead to the equalities

$$(b|\beta) = \frac{1}{p_*}$$



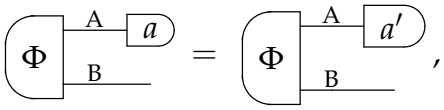
$$= (a|\alpha) = 1$$

$$(b|\beta') = \frac{1}{p_*}$$



$$= (a'|\alpha) = 1.$$

By proposition 4.2.1 we conclude that β and β' must be equal. Recalling the definitions of β and β', we then obtain the relation



which implies $a = a'$ because the state Φ is dynamically faithful on system A (proposition 4.1.9). □

Putting propositions 4.2.1, 4.2.2, and 4.2.8 together we finally obtain the desired result.

Theorem 4.2.9 (State-effect duality)**.** *For every system* A *there exists a bijective correspondence between normalised pure states and normalised pure effects, called the* dagger *and denoted by* $\dagger : \mathsf{PurSt}_1(\mathrm{A}) \to \mathsf{PurEff}_1(\mathrm{A})$. *The dagger satisfies the condition* $\left(\alpha^\dagger \middle| \alpha\right) = 1$, *for every* $\alpha \in \mathsf{PurSt}_1(\mathrm{A})$.

Therefore, for every normalised pure state α, $\alpha^\dagger$ denotes the associated pure effect such that $\left(\alpha^\dagger \middle| \alpha\right) = 1$. With a little abuse of notation we will denote the unique normalised pure state associated with a normalised pure effect a also with a dagger. For example, if a is a normalised pure effect, $a^\dagger$ is the unique normalised pure state such that $\left(a \middle| a^\dagger\right) = 1$.

Note that here this duality is limited only to states and effects, thus it is weaker than the dagger introduced in [166], however it was shown that it can be extended to all transformations too [111].

An easy corollary of the state-effect duality is the following (cf. [68, corollary 13]).

Corollary 4.2.10 (Transitivity on pure effects)**.** *For every pair of pure normalised effects* $a, a' \in \mathsf{PurEff}_1(\mathrm{A})$, *there exists a reversible channel* $\mathcal{U}$ *on* A *such that* $a' = a\mathcal{U}$.

In other words, reversible channels act transitively on the set of normalised pure effects too.

4.3 No disturbance without information

In sharp theories with purification, one can construct transformations that are "minimally-disturbing" [68, 167, 168]. Besides being important in their

own respect, these transformations will provide the crucial ingredient to prove the diagonalisation theorem in section 4.4.

The core result is the following proposition, from which important and useful corollaries follow. Note that a similar result, but under different axioms, was proved in [169].

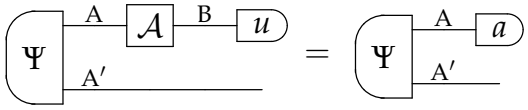
Proposition 4.3.1. *Let a be an effect such that $(a|\rho) = 1$, for some $\rho \in \mathsf{St}_1(\mathrm{A})$. Then there exists a* pure *transformation $\mathcal{T}$ on A such that $\mathcal{T} =_\rho \mathcal{I}$, where $\mathcal{I}$ is the identity, and $(u|\mathcal{T}|\sigma) \leq (a|\sigma)$, for every state $\sigma \in \mathsf{St}_1(\mathrm{A})$.*

Proof. The starting point of the proof is a result of [68], which guarantees that every normalised effect $a \in \mathsf{Eff}_1(\mathrm{A})$ can be written as

$$a = u_{\mathrm{B}}\mathcal{A}, \tag{4.3.1}$$

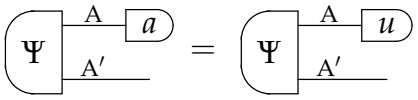
where $\mathcal{A}$ is a *pure* transformation from A to B, and B is a suitable system.

Now, let $\Psi \in \mathsf{PurSt}_1(\mathrm{AA}')$ be a purification of ρ. By eq. (4.3.1), we have



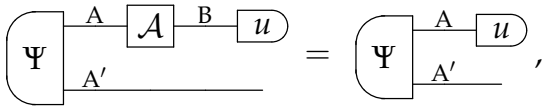
(4.3.2)

Now, since $(a|\rho) = 1$, we have $a =_\rho u_{\mathrm{A}}$. Hence, proposition 4.1.7 implies

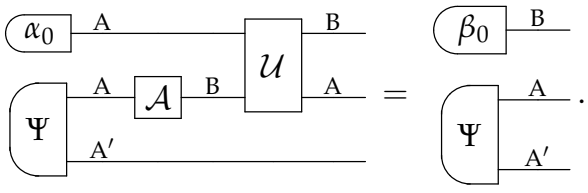


(4.3.3)

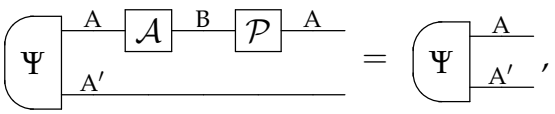
Combining eqs. (4.3.2) and (4.3.3), we obtain



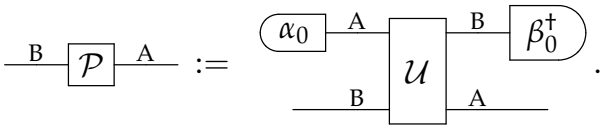
meaning that the two pure states $(\mathcal{A} \otimes \mathcal{I}_{\mathrm{A}'})\,\Psi$ and Ψ have the same marginal on system A'. By the uniqueness of purification, for fixed pure states $\alpha_0 \in \mathsf{PurSt}_1(\mathrm{A})$ and $\beta_0 \in \mathsf{PurSt}_1(\mathrm{B})$, there must exist a reversible channel $\mathcal{U}$ on AB, such that



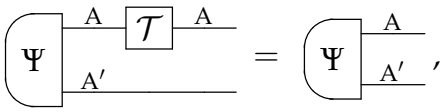
By applying $\beta_0^\dagger$ to both sides, we obtain



where $\mathcal{P}$ is the pure transformation defined as



Let us define the transformation $\mathcal{T} := \mathcal{P}\mathcal{A}$, which is pure by Purity Preservation. With this choice, we have



which implies $\mathcal{T} =_\rho \mathcal{I}$ by proposition 4.1.7. Finally, for all states $\sigma \in \mathsf{St}(\mathrm{A})$ we have the inequality

$$(u_{\mathrm{A}}|\mathcal{T}|\sigma) = (u_{\mathrm{A}}|\mathcal{P}\mathcal{A}|\sigma) \leq (u_{\mathrm{B}}|\mathcal{A}|\sigma) = (a|\sigma)\,.$$

Here, the inequality follows because all transformations are norm-non-increasing, and in this case we are dealing with the norm of $\mathcal{A}\sigma$ under the action of the transformation $\mathcal{P}$. The last equality follows from eq. (4.3.1). □

Note that the pure transformation $\mathcal{T}$ is non-disturbing on ρ because it acts as the identity on ρ and on all the states contained in it. In other words, whenever we have a (possibly mixed) effect occurring with unit probability on some state ρ, we can always find a transformation that does not disturb ρ (i.e. a non-disturbing non-demolition measurement). Being non-disturbing means that $\mathcal{T}$ occurs with unit probability on all the states contained in ρ. The other notable result of this proposition is that the probability of $\mathcal{T}$ occurring on a generic state σ is less than or equal to the probability of the original effect occurring on the same state.

The first consequence of proposition 4.3.1 is quite a technical result that is widely used in the rest of the chapter and of the thesis.

Corollary 4.3.2. *Let* $\{\alpha_i\}_{i=1}^n$ *be perfectly distinguishable* pure *states. Then one always has* $(\alpha_i^\dagger|\alpha_j) = \delta_{ij}$.

Proof. Clearly we need only to prove that, for every $i, j \in \{1, \dots, n\}$, whenever $j \neq i$, one has $(\alpha_i^\dagger|\alpha_j) = 0$. Let $\{a_i\}_{i=1}^n$ be the perfectly distinguishing test for the pure states $\{\alpha_i\}_{i=1}^n$. Since $(a_i|\alpha_i) = 1$, by proposition 4.3.1, for every $i \in \{1, \dots, n\}$ there exists a *pure* transformation $\mathcal{A}_i$ not disturbing α_i, namely

$$\mathcal{A}_i\alpha_i = \alpha_i. \tag{4.3.4}$$

Instead, for all $j \neq i$, by proposition 4.3.1 one has

$$(u|\mathcal{A}_i|\alpha_j) \leq (a_i|\alpha_j) = 0,$$

as $\{a_i\}_{i=1}^n$ is perfectly distinguishing. This implies that, for all $j \neq i$,

$$\mathcal{A}_i\alpha_j = 0. \tag{4.3.5}$$

Now, evaluating the expression $(\alpha_i^\dagger|\mathcal{A}_i|\alpha_i)$, and recalling eq. (4.3.4), we get

$$\left(\alpha_i^\dagger\middle|\mathcal{A}_i\middle|\alpha_i\right) = \left(\alpha_i^\dagger\middle|\alpha_i\right) = 1.$$

Since $\alpha_i^\dagger\mathcal{A}_i$ is a pure effect by Purity Preservation, and it occurs with unit probability on the state α_i, by the state-effect duality it must be

$$\alpha_i^\dagger\mathcal{A}_i = \alpha_i^\dagger. \tag{4.3.6}$$

Now, suppose $j \neq i$. By eqs. (4.3.6) and (4.3.5), one has

$$\left(\alpha_i^\dagger\middle|\alpha_j\right) = \left(\alpha_i^\dagger\middle|\mathcal{A}_i\middle|\alpha_j\right) = 0.$$

This concludes the proof. □

Note that, at this stage, this result does *not* mean that the effects $\{\alpha_i^\dagger\}_{i=1}^n$ make up an observation-test. This will be instead a consequence of corollary 4.3.4.

Using the existence of non-disturbing transformations we can also give a sufficient condition for the perfect distinguishability of states. The following condition, and especially its version for pure states, form the core of the proof of the diagonalisation theorem, and it is a more rigorous version of the construction used in [110].

Lemma 4.3.3. *Let* $\{\rho_i\}_{i=1}^n$ *be a set of normalised states. If there exists a set of effects*[3] $\{a_i\}_{i=1}^n$ *such that* $(a_i|\rho_i) = 1$ *for all* i*, and* $(a_i|\rho_j) = 0$ *for all* $j > i$*, then*

[3] Not necessarily an observation-test or a subset of an observation-test.

the states $\{\rho_i\}_{i=1}^n$ *are perfectly distinguishable.*

Proof. By hypothesis, the binary observation-test $\{a_i, u - a_i\}$ distinguishes perfectly between ρ_i and all the other states ρ_j with $j > i$. Equivalently, this observation-test distinguishes perfectly between ρ_i and the state $\widetilde{\rho}_i := \frac{1}{n-i}\sum_{j>i}\rho_j$. Specifically, $(u - a_i|\widetilde{\rho}_i) = 1$. Applying proposition 4.3.1, we can construct a pure transformation $\mathcal{A}_i^\perp$ such that $\mathcal{A}_i^\perp =_{\widetilde{\rho}_i} \mathcal{I}$, and, specifically,

$$\mathcal{A}_i^\perp \rho_j = \rho_j \tag{4.3.7}$$

for all $j > i$. Moreover, proposition 4.3.1 implies

$$\left(u\middle|\mathcal{A}_i^\perp\middle|\rho_i\right) \leq (u - a_i|\rho_i) = 0,$$

meaning that the transformation $\mathcal{A}_i^\perp$ never occurs on the state ρ_i. Let us define the effect $a_{i,0} := u - a_i - u\mathcal{A}_i^\perp$. Note that this effect is well-defined, because $(a_{i,0}|\sigma) \geq 0$, for all $\sigma \in \mathsf{St}_1(\mathrm{A})$. Indeed, by proposition 4.3.1, we have $\left(u\middle|\mathcal{A}_i^\perp\middle|\sigma\right) \leq (u - a_i|\sigma)$, for all $\sigma \in \mathsf{St}_1(\mathrm{A})$, whence $\left(u - a_i - u\mathcal{A}_i^\perp|\sigma\right) \geq 0$. Note that $a_{i,0}$ never occurs on the states ρ_k with $k \geq i$.

Now, define the transformations $\mathcal{A}_i = |\rho_i)(a_i|$ and $\mathcal{A}_{i,0} = |\rho_0)(a_{i,0}|$, where ρ_0 is a fixed normalised state. By proposition 4.1.10, the transformations $\{\mathcal{A}_i, \mathcal{A}_i^\perp, \mathcal{A}_{i,0}\}$ form a test. Summarising the above observations, the test satisfies the properties

$$\begin{cases} \mathcal{A}_i\rho_i = \rho_i & \\ \mathcal{A}_i\rho_j = 0 & \forall j > i \\ \mathcal{A}_i^\perp\rho_i = 0 & \\ \mathcal{A}_i^\perp\rho_j = \rho_j & \forall j > i \\ \mathcal{A}_{i,0}\rho_k = 0 & \forall k \geq i. \end{cases} . \tag{4.3.8}$$

By construction, the test distinguishes without error between the state ρ_i and all the states ρ_j with $j > i$, in such a way that the latter are not disturbed. Indeed, by construction $\mathcal{A}_i$ can only occur if the state is ρ_i, instead $\mathcal{A}_i^\perp$ never occurs on ρ_i, but it occurs with probability 1 if the state is any of the ρ_j's, with $j > i$, and it leaves them unchanged. Finally, $\mathcal{A}_{i,0}$ never occurs on the states ρ_k's with $k \geq i$, so it does not play a role in the discrimination process. Essentially, $\mathcal{A}_{i,0}$ only plays the role of making $\{\mathcal{A}_i, \mathcal{A}_i^\perp, \mathcal{A}_{i,0}\}$ a test.

Using the tests $\{\mathcal{A}_i, \mathcal{A}_i^\perp, \mathcal{A}_{i,0}\}$ it is easy to construct a protocol that distinguishes perfectly between the states $\{\rho_i\}_{i=1}^n$. The protocol works as follows: starting from $i = 1$ perform the test $\{\mathcal{A}_i, \mathcal{A}_i^\perp, \mathcal{A}_{i,0}\}$. If the transformation $\mathcal{A}_i$ takes place, then the state is ρ_i. If the transformation $\mathcal{A}_i^\perp$ takes place, then perform the test $\{\mathcal{A}_{i+1}, \mathcal{A}_{i+1}^\perp, \mathcal{A}_{i+1,0}\}$ (this can be done because $\mathcal{A}_i^\perp$ is non-disturbing). Using this protocol, every state in the set $\{\rho_i\}_{i=1}^n$ will be identified without error in at most n steps. Overall, the protocol is described by a test with $2n+1$ outcomes, corresponding to the transformations

$$\begin{aligned}
\mathcal{T}_1 &= \mathcal{A}_1 \\
\mathcal{T}_2 &= \mathcal{A}_2\mathcal{A}_1^\perp \\
&\vdots \\
\mathcal{T}_n &= \mathcal{A}_n\mathcal{A}_{n-1}^\perp \dots \mathcal{A}_1^\perp \\
\mathcal{T}_{n+1} &= \mathcal{A}_{1,0} \\
\mathcal{T}_{n+2} &= \mathcal{A}_{2,0}\mathcal{A}_1^\perp \\
&\vdots \\
\mathcal{T}_{2n} &= \mathcal{A}_{n,0}\mathcal{A}_{n-1}^\perp \dots \mathcal{A}_1^\perp \\
\mathcal{T}_{2n+1} &= \mathcal{A}_n^\perp \dots \mathcal{A}_1^\perp
\end{aligned}$$

To show that these transformations form a test, we use proposition 4.1.10: $\{\mathcal{T}_i\}_{i=1}^{2n+1}$ is a test if and only if $\sum_{i=1}^{2n+1} u\mathcal{T}_i = u$. An easy check shows that this is the case.

To complete the proof, we need to construct a perfectly distinguishing test $\{e_i\}_{i=1}^n$ for the states $\{\rho_i\}_{i=1}^n$. By discarding the output of the transformations $\{\mathcal{T}_i\}_{i=1}^{2n+1}$, we get an observation-test $\{t_i\}_{i=1}^{2n+1}$ with $2n+1$ outcomes and effects $t_i := u\mathcal{T}_i$. We claim that the observation-test

$$\{e_i\}_{i=1}^n = \{t_1, \dots, t_{n-1}, u - t_1 - \dots - t_{n-1}\} \tag{4.3.9}$$

is perfectly distinguishing for the states $\{\rho_i\}_{i=1}^n$. First of all, since $t_1, \dots, t_{n-1}$ coexist in a $(2n+1)$-outcome test, the effect $u - t_1 - \dots - t_{n-1}$ is well-defined. Now let us prove that the observation-test (4.3.9) perfectly distinguishes the states ρ_i's. We start from $t_1 = u\mathcal{A}_1$; by (4.3.8) we get

$$(t_1|\rho_j) = (u|\mathcal{A}_1|\rho_j) = \delta_{1j}\,(u|\rho_j) = \delta_{1j}.$$

Now, if $i > 1$,

$$t_i = u\mathcal{A}_i\mathcal{A}_{i-1}^{\perp}\ldots\mathcal{A}_1^{\perp}.$$

If we wish to calculate $\left(t_i\middle|\rho_j\right)$, by eq. (4.3.8), ρ_j is left invariant by all the $\mathcal{A}_k^{\perp}$ with $k < j$. If $i \neq j$, then

$$\left(t_i\middle|\rho_j\right) = \left(u\middle|\mathcal{A}_i\mathcal{A}_{i-1}^{\perp}\ldots\mathcal{A}_j^{\perp}\middle|\rho_j\right) = 0,$$

again by eq. (4.3.8). If, instead $j = i$,

$$\left(t_i\middle|\rho_i\right) = \left(u\middle|\mathcal{A}_i\middle|\rho_i\right) = \left(u\middle|\rho_i\right) = 1.$$

As a consequence of these results,

$$\left(u - t_1 - \ldots - t_{n-1}\middle|\rho_j\right) = \delta_{nj}.$$

We conclude that $\{e_i\}_{i=1}^n$ is really a perfectly distinguishing test, because we have $\left(e_i\middle|\rho_j\right) = \delta_{ij}$. □

In the case when the states are pure, and the effects in the statement of lemma 4.3.3 are the daggers of those pure states, we can prove something stronger.

Corollary 4.3.4. *Let $\{\alpha_i\}_{i=1}^n$ be a set of normalised* pure *states such that $\left(\alpha_i^{\dagger}\middle|\alpha_j\right) = 0$ for all $j > i$, then the states $\{\alpha_i\}_{i=1}^n$ are perfectly distinguishable, and the pure effects $\left\{\alpha_i^{\dagger}\right\}_{i=1}^n$ coexist in an observation-test, which distinguishes the states $\{\alpha_i\}_{i=1}^n$ perfectly.*

Proof. If we take $a_i := \alpha_i^{\dagger}$, by lemma 4.3.3, we know that the states $\{\alpha_i\}_{i=1}^n$ are perfectly distinguishable. Referring to the proof of lemma 4.3.3, note that, since $\mathcal{A}_i^{\perp}$ is pure, we have

$$\alpha_j^{\dagger}\mathcal{A}_i^{\perp} = \alpha_j^{\dagger} \qquad \forall j > i. \tag{4.3.10}$$

by a similar argument to the one in the proof of corollary 4.3.2. Indeed, the effect $\alpha_j^{\dagger}\mathcal{A}_i^{\perp}$ is pure by Purity Preservation, and satisfies

$$\left(\alpha_j^{\dagger}\middle|\mathcal{A}_i^{\perp}\middle|\alpha_j\right) = \left(\alpha_j^{\dagger}\middle|\alpha_j\right) = 1,$$

where we have used eq. (4.3.7). Let us construct the perfectly distinguishing observation-test like in the proof of lemma 4.3.3, by considering the effects $t_i = u\mathcal{A}_i$. One has, recalling that $\mathcal{A}_i = |\alpha_i)\,(\alpha_i^\dagger|$,

$$t_1 = u\mathcal{A}_1 = \alpha_1^\dagger$$

$$t_2 = u\mathcal{A}_2\mathcal{A}_1^\perp = \alpha_2^\dagger A_1^\perp = \alpha_2^\dagger$$

$$\vdots$$

$$t_n = u\mathcal{A}_n\mathcal{A}_{n-1}^\perp \dots \mathcal{A}_1^\perp = \alpha_n^\dagger,$$

having used eq. (4.3.10). This proves that the effects $\{\alpha_i^\dagger\}_{i=1}^n$ coexist in a $(2n+1)$-outcome observation-test. As a consequence, as shown in the proof of lemma 4.3.3, we have that

$$\left\{\alpha_1^\dagger, \dots, \alpha_{n-1}^\dagger, u - \alpha_1^\dagger - \dots - \alpha_{n-1}^\dagger\right\}$$

is perfectly distinguishing, and specifically $(\alpha_i^\dagger|\alpha_j) = \delta_{ij}$. □

As a consequence of this corollary, whenever some pure states are perfectly distinguishable, their daggers coexist in an observation-test that distinguishes them perfectly. Corollaries 4.3.2 and 4.3.4, taken together, state that a necessary and sufficient condition for the pure states $\{\alpha_i\}_{i=1}^n$ to be perfectly distinguishable is that their daggers satisfy $(\alpha_i^\dagger|\alpha_j) = \delta_{ij}$, for all $i, j \in \{1, \dots, n\}$.

4.4 Diagonalisation of states

This section represents the core of the whole chapter, for it introduces a key tool that will have plenty of consequences throughout this thesis: the diagonalisation of states. Since states are not density matrices, clearly diagonalisation here has a different meaning. To understand it, let us look at quantum theory from an operational angle. Note that in the diagonalisation of density matrices, a quantum state ρ of a d-dimensional Hilbert space is diagonalised as $\rho = \sum_{j=1}^d p_j\,|j\rangle\langle j|$, where $\{|j\rangle\}_{j=1}^d$ is an orthonormal basis, and $\{p_j\}_{j=1}^d$ is a probability distribution. Since $|j\rangle\langle j|$ represent

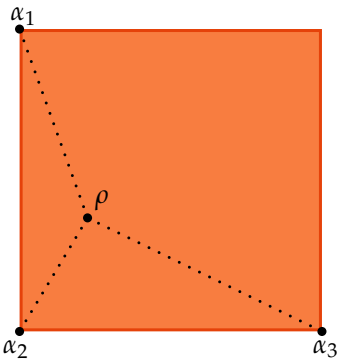


Figure 4.1: The state ρ is a non-trivial convex combination of the pure states α_1, α_2, and α_3, which are *not* jointly perfectly distinguishable. ρ cannot be diagonalised.

orthogonal pure states, we understand the operational meaning of diagonalisation in quantum theory: a state is diagonalised when it is written as a convex combination of perfectly distinguishable pure states. Indeed the p_j's—the eigenvalues of ρ—are the coefficients of a convex combination, and the pure states $\{|j\rangle\langle j|\}$ are distinguished perfectly by the projective measurement $\{|j\rangle\langle j|\}$.

Therefore it is natural to extend the definition of diagonalisation to GPTs as follows: a *diagonalisation* of ρ is a convex decomposition of ρ into perfectly distinguishable pure states. The probabilities in such a convex decomposition will be called the *eigenvalues* of ρ, and the perfectly distinguishable pure states the *eigenstates* [110].

Note that, while it is true that every state ρ in GPTs can be decomposed as a convex combination of pure states, the key point about diagonalisation is that every state should be written as a convex combination of *perfectly distinguishable* pure states. This is a non-trivial property, for example the square bit [62] does *not* satisfy it (see fig. 4.1).

Rather than postulate the diagonalisation of all states like in [87, 169, 170], here we *derive* the diagonalisation of states from the axioms of sharp theories with purification: Causality, Purity Preservation, Pure Sharpness, and Purification. This result already appeared in a preliminary form in [110], here we recast it in a more rigorous version, and we manage to prove the uniqueness of the eigenvalues of states from the axioms of sharp theories with purification, a fact that was only conjectured, and not proved

in [110].

We start with the actual diagonalisation theorem.

Theorem 4.4.1. *In a sharp theory with purification, every state of every (non-trivial) system can be diagonalised.*

The proof is a constructive procedure that returns a diagonalisation of ρ with the eigenvalues naturally listed in decreasing order, namely $p_i \geq p_{i+1}$ for every i. In particular, one has $p_1 = p_*$, which justifies why we called p_* the "maximum eigenvalue".

Proof. In order to diagonalise the state ρ, it is enough to proceed along the following steps:

1. Set $\rho_1 = \rho$ and $p_{*,0} = 0$.
2. For i starting from $i = 1$, decompose ρ_i as $\rho_i = p_{*,i}\alpha_i + (1 - p_{*,i})\,\sigma_i$, where $p_{*,i}$ is the maximum eigenvalue of ρ_i. Set $\rho_{i+1} = \sigma_i$, $p_i = p_{*,i}\prod_{j<i}(1 - p_{*,j})$. If $p_{*,i} = 1$, then terminate, otherwise continue to the step $i + 1$.

Recall that theorem 4.2.3 guarantees the condition $(\alpha_i^\dagger|\sigma_i) = 0$ at every step of the procedure. Since by construction every state α_j with $j > i$ is contained in σ_i, we also have $(\alpha_i^\dagger|\alpha_j) = 0$ for every $j > i$. Hence, corollary 4.3.4 implies that the states $\{\alpha_k\}_{k=1}^i$, generated by the first i iterations of the protocol, are perfectly distinguishable, for any i. For a finite-dimensional system, the procedure must terminate in a finite number of iterations. Once the procedure has been completed, the state ρ is decomposed as $\rho = \sum_{i=1}^r p_i\alpha_i$ where r is some positive (finite) integer, and $\{\alpha_i\}_{i=1}^r$ are perfectly distinguishable pure states. □

Later in subsection 4.4.2 we will show that the vector of the eigenvalues $\mathbf{p}$, also called the *spectrum* of ρ, is uniquely determined by the state ρ, which means that all diagonalisations of ρ have the same eigenvalues.

Before moving forward, it is important to note that the eigenvalues can be characterised as the outcome probabilities of a pure measurement performed on the system.

Corollary 4.4.2. *Let ρ be a generic state, diagonalised as $\rho = \sum_{i=1}^r p_i\alpha_i$. Then, one has $p_i = (\alpha_i^\dagger|\rho)$, for every $i \in \{1, \ldots, r\}$.*

Proof. Immediate from the combination of theorem 4.4.1 and corollary 4.3.2, because $(\alpha_i^\dagger|\alpha_j) = \delta_{ij}$. □

4.4.1 Diagonalisation of the invariant state

Let us examine the properties of the diagonalisations of the invariant states. In some respect they are special: first of all, all the eigenvalues are equal.

Proposition 4.4.3. *For every non-trivial system, there exists a (strictly) positive integer d such that*

1. *every diagonalisation of the invariant state consists of exactly d pure states;*
2. *the eigenvalues of the invariant states are all equal to $\frac{1}{d}$.*

Proof. Let $\chi = \sum_{i=1}^{r} p_i \alpha_i$ be a diagonalisation of the invariant state χ. By corollary 4.4.2, $p_i = (\alpha_i^\dagger|\chi)$, but by proposition 4.2.7 we have $(\alpha_i^\dagger|\chi) = p_*$, whence $p_i = p_*$ for every i. It follows that $p_* = \frac{1}{r}$. Now consider another diagonalisation of χ: $\chi = \sum_{i=1}^{r'} p'_i \alpha'_i$. Repeating the same argument, we conclude that $p'_i = p_* = \frac{1}{r'}$. This means that $r = r' =: d$. □

We will refer to d as the *dimension* of the system, for reasons that will become clear soon.

Let us show that the set of states $\{\alpha_i\}_{i=1}^{d}$ arising in any diagonalisation of the invariant state is *maximal*.

Proposition 4.4.4. *Let χ be written as a uniform mixture of pure states of the form $\chi = \frac{1}{d}\sum_{i=1}^{d} \alpha_i$, where d is the dimension of the system. Then*

1. *$\{\alpha_i\}_{i=1}^{d}$ is a maximal set of perfectly distinguishable pure states;*
2. *$\{\alpha_i^\dagger\}_{i=1}^{d}$ is a pure observation-test.*

Proof. Let us prove the two properties.

1. Suppose the invariant state is decomposed as $\chi = \frac{1}{d}\sum_{i=1}^{d} \alpha_i$. Then, by proposition 4.2.7 one has $(\alpha_i^\dagger|\chi) = \frac{1}{d}$, for any $i \in \{1, \ldots, d\}$, and consequently $(\alpha_i^\dagger|\alpha_j) = \delta_{ij}$. By corollary 4.3.4, the states $\{\alpha_i\}_{i=1}^{d}$ are perfectly distinguishable. Suppose by contradiction that this is *not* a maximal set; then we can add the pure state α_{d+1} so that $\{\alpha_i\}_{i=1}^{d+1}$ is a pure maximal set. If $\{a_i\}_{i=1}^{d+1}$ is the observation-test that distinguishes them, one must have $(a_{d+1}|\chi) = \frac{1}{d}\sum_{i=1}^{d} (a_{d+1}|\alpha_i) = 0$. But $(a_{d+1}|\chi) = 0$ implies $(a_{d+1}|\rho) = 0$ for every ρ, since every state is

contained in the invariant state, which is internal. This is in contradiction with the hypothesis $(a_{d+1}|\alpha_{d+1}) = 1$. Hence, we have proved that the set $\{\alpha_i\}_{i=1}^d$ is maximal.

2. Let us prove that $\{\alpha_i^\dagger\}_{i=1}^d$ is an observation-test, namely $\sum_{i=1}^d \alpha_i^\dagger = u$. By propositions 4.2.7 and 4.4.3, we have

$$\sum_{i=1}^d \left(\alpha_i^\dagger \middle| \chi\right) = \sum_{i=1}^d \frac{1}{d} = 1.$$

Since χ is internal, this means that $\sum_{i=1}^d \left(\alpha_i^\dagger|\rho\right) = 1$ for every normalised state $\rho \in \mathsf{St}_1(\mathrm{A})$, whence $\sum_{i=1}^d \alpha_i^\dagger$ is the deterministic effect u.

□

Propositions 4.4.3 and 4.4.4 imply that the invariant state is a uniform mixture of the states in a pure maximal set. Remarkably, the converse holds too: *every* pure maximal set, mixed with equal weights, yields the invariant state.

Proposition 4.4.5. *Let $\{\alpha_i\}_{i=1}^r$ be a pure maximal set. Then one has $r = d$ and $\chi = \frac{1}{d}\sum_{i=1}^d \alpha_i$.*

Proof. We know that every pure state is an eigenstate of χ with maximum eigenvalue. Specifically, we must have

$$\chi = \frac{1}{d}\alpha_1 + \frac{d-1}{d}\sigma_1$$

for a state σ_1 that is perfectly distinguishable from α_1 (cf. theorem 4.2.3). For every $n < r$, assume that the invariant state can be decomposed in the diagonalisation process as

$$\chi = \frac{1}{d}\left(\sum_{i=1}^n \alpha_i\right) + \frac{d-n}{d}\sigma_n, \tag{4.4.1}$$

where the states $\{\alpha_i\}_{i=1}^n \cup \{\sigma_n\}$ are perfectly distinguishable, and we prove that a decomposition of the same form can be found for $n+1$. To this purpose, we use the relations

$$(\alpha_i|\chi) = \frac{1}{d} \tag{4.4.2}$$

for all $i \in \{1, \dots, r\}$, following from proposition 4.2.7, and valid for all normalised pure effects, and

$$\left(\alpha_i^\dagger \middle| \alpha_j\right) = \delta_{ij} \tag{4.4.3}$$

for all $i, j \in \{1, \dots, r\}$, following from the assumption that the states $\{\alpha_i\}_{i=1}^r$ are perfectly distinguishable (cf. corollary 4.3.2). Eqs. (4.4.1), (4.4.2), and (4.4.3) yield the relation

$$\frac{1}{d} = \left(\alpha_{n+1}^\dagger \middle| \chi\right) = \frac{d-n}{d} \left(\alpha_{n+1}^\dagger \middle| \sigma_n\right),$$

or, equivalently,

$$\left(\alpha_{n+1}^\dagger \middle| \sigma_n\right) = \frac{1}{d-n}. \tag{4.4.4}$$

Hence, by proposition 4.2.6, the maximum eigenvalue of σ_n is greater than or equal to $\frac{1}{d-n}$. In fact, it must be equal to $\frac{1}{d-n}$, because otherwise the corresponding eigenstate α would lead to the contradiction, recalling eq. (4.4.1):

$$\frac{1}{d} = \left(\alpha^\dagger \middle| \chi\right) \geq \frac{d-n}{d} \left(\alpha^\dagger \middle| \sigma_n\right) > \frac{1}{d}.$$

Hence, eq. (4.4.4) and corollary 4.2.5 imply that α_{n+1} is an eigenstate of σ_n with maximum eigenvalue. Therefore, σ_n can be decomposed as $\sigma_n = \frac{1}{d-n}\alpha_{n+1} + \frac{d-n-1}{d-n}\sigma_{n+1}$, where the states α_{n+1} and σ_{n+1} are perfectly distinguishable. Inserting this relation into eq. (4.4.1) we obtain

$$\chi = \frac{1}{d}\left(\sum_{i=1}^{n+1} \alpha_i\right) + \frac{d-n-1}{d}\sigma_{n+1}.$$

Now, since the states $\{\alpha_i\}_{i=1}^n \cup \{\sigma_n\}$ are perfectly distinguishable, so are the states $\{\alpha_i\}_{i=1}^{n+1} \cup \{\sigma_{n+1}\}$. This proves the validity of eq. (4.4.1) for every $n \leq r$. To conclude the proof, consider eq. (4.4.1) for $n = r$. The condition that set $\{\alpha_i\}_{i=1}^r$ is maximal implies that the state σ_r should not arise in the decomposition, otherwise the pure states contained in σ_r would be perfectly distinguishable from the pure states $\{\alpha_i\}_{i=1}^r$, contradicting maximality. This is possible only if the corresponding probability is zero, namely only if one has $r = d$. □

In summary, the above proposition guarantees that all pure maximal sets of a system have the same cardinality, equal to d, and this is why we called d the dimension of the system. Moreover, any set of d perfectly distinguishable pure states is guaranteed to be maximal. As a consequence, every state can have at most d terms in its diagonalisations. Then, clearly every diagonalisation $\rho = \sum_{i=1}^{r} p_i \alpha_i$, where $p_i > 0$ for all $i \in \{1, \dots, r\}$, and $r \leq d$, can be rewritten as $\rho = \sum_{i=1}^{d} p_i \alpha_i$ by completing $\{\alpha_i\}_{i=1}^{r}$ to a maximal set $\{\alpha_i\}_{i=1}^{d}$, and taking some of the eigenvalues to be zero. This means that the spectrum of a state can always be taken to be a vector $\mathbf{p}$ with d entries. This will play an important role in the next chapter.

In other works [87, 110, 130, 169, 170], the above result and other properties of diagonalisations were derived from the Strong Symmetry axiom [87, 113], stating that all pure maximal sets are connected by reversible channels (see also subsection 5.8.1 for its thermodynamic implications in sharp theories with purification). Our result shows that the properties of diagonalisation can be derived from a very different set of axiom: Causality, Purity Preservation, Pure Sharpness, and Purification, and do not need Strong Symmetry, unlike in our previous works [110, 130].

The diagonalisation of the invariant state induces a one-to-one correspondence between maximal sets of perfectly distinguishable pure states and *pure sharp measurements* [168, 171], which can be characterised as follows.

Definition 4.4.6. An observation-test $\{a_i\}_{i=1}^{n}$ is a *pure sharp measurement* if every effect a_i is pure and normalised.

Under the validity of our axioms, every pure sharp measurement can be written as $\{\alpha_i^\dagger\}_{i=1}^{n}$, for some set of pure states $\{\alpha_i\}_{i=1}^{n}$ (cf. theorem 4.2.9).

Proposition 4.4.7. *For every pure maximal set $\{\alpha_i\}_{i=1}^{d}$, the effects $\{\alpha_i^\dagger\}_{i=1}^{d}$ form a pure sharp measurement. Conversely, for every pure sharp measurement $\{\alpha_i^\dagger\}_{i=1}^{n}$, the states $\{\alpha_i\}_{i=1}^{n}$ form a pure maximal set, and therefore $n = d$.*

Proof. Let $\{\alpha_i\}_{i=1}^{d}$ be a pure maximal set. By proposition 4.4.5, we know that $\frac{1}{d}\sum_{i=1}^{d} \alpha_i$ is a diagonalisation of the invariant state χ. Then, proposition 4.4.4 implies that $\{\alpha_i^\dagger\}_{i=1}^{d}$ is a pure sharp measurement.

Conversely, suppose that $\{a_i\}_{i=1}^{n}$ is a pure sharp measurement, then $a_i = \alpha_i^\dagger$ for some state α_i. By corollary 4.3.2, we know that $(\alpha_i^\dagger | \alpha_j) =$

δ_{ij}, and moreover $\{a_i\}_{i=1}^n$ is an observation-test, thus $\{a_i\}_{i=1}^n$ distinguishes the states $\{\alpha_i\}_{i=1}^n$ perfectly. The states $\{\alpha_i\}_{i=1}^n$ must form a pure maximal set. This can be proved by contradiction: suppose the set $\{\alpha_i\}_{i=1}^n$ is not maximal, and extend it to a maximal set $\{\alpha_i\}_{i=1}^d$. Then, by the first part of this proof we have that $\{\alpha_i^\dagger\}_{i=1}^d$ is an observation-test. By Causality, we then obtain

$$\sum_{i=1}^d \alpha_i^\dagger = u = \sum_{i=1}^n a_i = \sum_{i=1}^n \alpha_i^\dagger,$$

having used the equality $a_i = \alpha_i^\dagger$. In conclusion, we have obtained the relation $\sum_{i=n+1}^d \alpha_i^\dagger = 0$, which can be satisfied only if $n = d$. Hence, the states $\{\alpha_i\}_{i=1}^d$ form a pure maximal set. □

As a consequence, the product of two pure maximal sets is a pure maximal set for the composite system. This property was called "information locality" by Hardy [70, 73], and it has been recently shown, along with a weaker version of Purity Preservation, to play a major role in the emergence of local classical observers in GPTs [115]. In words, the dimension of a composite system AB is the product of the dimension of the components: $d_{\mathrm{AB}} = d_{\mathrm{A}} d_{\mathrm{B}}$.

Proposition 4.4.8 (Information locality)**.** *If* $\{\alpha_i\}_{i=1}^{d_{\mathrm{A}}}$ *is a pure maximal set for system* A *and* $\{\beta_j\}_{j=1}^{d_{\mathrm{B}}}$ *is a pure maximal set for system* B*, then*

$$\{\alpha_i \otimes \beta_j\}_{i\in\{1,\dots,d_{\mathrm{A}}\}, j\in\{1,\dots,d_{\mathrm{B}}\}}$$

is a pure maximal set for the composite system AB*.*

Proof. By proposition 4.4.7, $\{\alpha_i^\dagger\}_{i=1}^{d_{\mathrm{A}}}$ and $\left\{\beta_j^\dagger\right\}_{j=1}^{d_{\mathrm{B}}}$ are two observation-tests for systems A and B, respectively. Now, the product of two observation-tests is an observation-test (physically, corresponding to two measurements performed in parallel). Hence, the product $\left\{\alpha_i^\dagger \otimes \beta_j^\dagger\right\}_{i\in\{1,\dots,d_{\mathrm{A}}\}, j\in\{1,\dots,d_{\mathrm{B}}\}}$ is an observation-test on the composite system AB. Moreover, each effect $\alpha_i^\dagger \otimes \beta_j^\dagger$ is pure, due to Purity Preservation, and normalised. Using proposition 4.4.7 again, we obtain that $\{\alpha_i \otimes \beta_j\}_{i\in\{1,\dots,d_{\mathrm{A}}\}, j\in\{1,\dots,d_{\mathrm{B}}\}}$ is a pure maximal set. □

Diagonalisation of internal states

As a side remark, we show here that internal states have exactly d non-zero eigenvalues. In the quantum case, this amounts to saying that internal states are full-rank density matrices.

Proposition 4.4.9. *Every internal state ω has precisely d non-vanishing eigenvalues in every diagonalisation.*

Proof. Consider a complete state ω and one of its diagonalisations $\omega = \sum_{i=1}^{r} p_i \alpha_i$, where $r \leq d$, and the p_i's are non-vanishing, for all $i \in \{1, \dots, r\}$. Suppose by contradiction that $r < d$; this means that the states $\{\alpha_i\}_{i=1}^{r}$ do not form a pure maximal set, and therefore we can complete it by adding $d - r$ states $\{\alpha_i\}_{i=r+1}^{d}$. In this way we can rewrite the diagonalisation of ω as $\omega = \sum_{i=1}^{d} p_i \alpha_i$, where $p_i = 0$ for $i \in \{r+1, \dots, d\}$, and the states $\{\alpha_i\}_{i=1}^{d}$ are a pure maximal set. Take any α_i with $i \in \{r+1, \dots, d\}$; we have

$$0 = p_i = \left(\alpha_i^\dagger \middle| \omega\right).$$

On the other hand, ω is internal, therefore $\left(\alpha_i^\dagger \middle| \rho\right) = 0$ for all states $\rho \in \mathrm{St}_1(\mathrm{A})$, and $i \in \{r+1, \dots, d\}$. Hence $\left(\alpha_i^\dagger \middle| \alpha_i\right) = 0$, which is a contradiction. We conclude that $r = d$. □

Consequently, the pure states arising in every diagonalisation of ω form a maximal set.

The converse also holds, and the proof has already appeared in [68, corollary 19] (and does not make use of the stronger axioms assumed therein).

Proposition 4.4.10. *Let $\{\alpha_i\}_{i=1}^{d}$ be a maximal set of perfectly distinguishable pure states. Every convex combination of the α_i's with all* non-zero *coefficients yields a complete state.*

Proof. Let ω be a mixture of the pure states $\{\alpha_i\}_{i=1}^{d}$ with d non-zero probabilities $\omega = \sum_{i=1}^{d} p_i \alpha_i$, where $p_i > 0$. Consider the minimum eigenvalue $p_{\min} = \min_i \{p_i\}$. Then we can write $\omega = p_{\min}\chi + (1 - p_{\min})\,\sigma$, where σ is defined as

$$\sigma := \frac{1}{1 - p_{\min}} \sum_{i=1}^{d} \left(p_i - \frac{p_{\min}}{d}\right) \alpha_i,$$

and it is well-defined because $p_i \geq \frac{p_{\min}}{d}$. Since χ is contained in ρ, and χ is internal, we conclude that ρ is internal too. □

Double stochasticity of the transition matrices

Given two pure maximal sets, $\{\alpha_i\}_{i=1}^d$ and $\{\alpha'_i\}_{i=1}^d$, we call the matrix $T_{ij} = \left(\alpha_i^\dagger \middle| \alpha'_j\right)$ a *transition matrix*. With this definition, we have the following result [110, lemma 4].

Lemma 4.4.11. *In sharp theories with purification all transition matrices are doubly stochastic*[4].

We do not report the proof of this lemma, since in section 5.3 we will prove a stronger result (lemma 5.3.1), which will imply lemma 4.4.11. Moreover, the proof of lemma 5.3.1 will be virtually identical to the proof of lemma 4.4.11.

4.4.2 Uniqueness of the diagonalisation

Thanks to our axioms, the diagonalisation of a state is unique, up to the obvious freedom arising in the presence of degeneracy among the eigenvalues. This is a non-trivial consequence of the axioms: notably [170, 172] exhibited examples of GPTs where states can be diagonalised, but the same state can have more than one diagonalisation, with different spectra.

To take degeneracy into account, given a diagonalisation $\rho = \sum_{i=1}^r p_i \alpha_i$ of ρ, we define the *reduced spectrum* of ρ, as the set of the *distinct* eigenvalues of ρ, ordered in strictly decreasing order $\lambda_1 > \lambda_2 > \ldots > \lambda_s > 0$, and we rewrite the diagonalisation as

$$\rho = \sum_{k=1}^{s} \lambda_k \Pi_k,$$

where

$$\Pi_k := \sum_{i:p_i=\lambda_k} \alpha_i,$$

and the sum is over the α_i's arising in the given diagonalisation of ρ whose eigenvalue is λ_k. When expressed in this form, the diagonalisation is unique. Now we present the main theorem.

[4] Recall that a doubly stochastic matrix is a matrix with non-negative entries, in which every row and every column sum to 1.

Theorem 4.4.12. *Let $\rho = \sum_{k=1}^{s} \lambda_k \Pi_k$ and $\rho = \sum_{l=1}^{s'} \lambda'_l \Pi'_l$ be two diagonalisations of the same state. Then, one has $s = s'$, $\lambda_k = \lambda'_k$, $\Pi_k = \Pi'_k$, for all $k \in \{1, \dots, s\}$.*

Proof. Let the two diagonalisations be $\rho = \sum_i p_i \alpha_i$ and $\rho = \sum_j p'_j \alpha'_j$. First of all, let us prove that $\lambda_1 = \lambda'_1 = p_*$. This is a non-trivial statement to prove. Indeed, our diagonalisation algorithm of theorem 4.4.1 outputs the first eigenvalue to be the maximum eigenvalue p_*, but there might exist other diagonalisation algorithms yielding different eigenvalues, none of which equal to p_*. Let us define the degeneracies $d_1 = |\{i : p_i = \lambda_1\}|$ and $d'_1 = \left|\left\{j : p'_j = \lambda'_1\right\}\right|$, and assume $d_1 \geq d'_1$ without loss of generality. By definition, we have for $i \in \{1, \dots, d_1\}$

$$\lambda_1 = \left(\alpha_i^\dagger \middle| \rho\right) = \sum_j p'_j \left(\alpha_i^\dagger \middle| \alpha'_j\right) = \sum_j T_{ij} p'_j \leq \lambda'_1,$$

having used the fact that the transition matrix $T_{ij} = \left(\alpha_i^\dagger \middle| \alpha'_j\right)$ is doubly stochastic for $i, j \in \{1, \dots, d\}$, with d the dimension of the system. By a similar argument, with λ_1 and λ'_1 interchanged, finally we get the equality $\lambda_1 = \lambda'_1$. Since this applies to all diagonalisations of ρ, including those obtained with the algorithm of theorem 4.4.1, for which $\lambda_1 = p_*$, we conclude that $\lambda_1 = \lambda'_1 = p_*$.

The above relation implies the equality $\sum_{j=1}^{d'_1} \left(\alpha_i^\dagger \middle| \alpha'_j\right) = 1$, for every $i \in \{1, \dots, d_1\}$ or, equivalently, $(\alpha_i^\dagger | \chi'_1) = \frac{1}{d'_1}$ for all $i \in \{1, \dots, d_1\}$, where $\chi'_1 := \frac{1}{d'_1} \Pi'_1$. Note that $\frac{1}{d'_1}$ is the maximum eigenvalue of χ'_1 because the states $\left\{\alpha'_j\right\}_{j=1}^{d'_1}$ are perfectly distinguishable, and we have just proved that the maximum eigenvalue of a state arises in every diagonalisation. Thus corollary 4.2.5 implies that α_i is an eigenstate with maximum eigenvalue. In particular, choosing $i = 1$ we obtain the decomposition

$$\chi'_1 = \frac{1}{d'_1} \alpha_1 + \frac{d'_1 - 1}{d'_1} \sigma_1,$$

where σ_1 is a suitable state, perfectly distinguishable from α_1. We are now in the position to repeat the argument in the proof of proposition 4.4.5 for

the states $\{\alpha_i\}_{i=1}^{d_1}$, to find that $d_1 = d_1'$ and

$$\chi_1' = \frac{1}{d_1}\sum_{i=1}^{d_1}\alpha_1 = \frac{1}{d_1}\Pi_1.$$

Hence, we proved the equality $\Pi_1' = \Pi_1$. We can now define the state

$$\rho_2 := \frac{1}{1-d_1\lambda_1}\left(\rho - \lambda_1\Pi_1\right) = \frac{1}{1-d_1\lambda_1}\left(\sum_{k=2}^{s}\lambda_k\Pi_k\right) = \frac{1}{1-d_1\lambda_1}\left(\sum_{l=2}^{s'}\lambda_l'\Pi_l'\right).$$

Repeating the above argument, we can prove the equalities $\lambda_2 = \lambda_2'$ and $\Pi_2 = \Pi_2'$. Once all distinct eigenvalues have been scanned, the normalisation of the probability distribution implies the condition $s = s'$. □

A very close result was proved by Wilce in the framework of probabilistic models with conjugates and Jordan algebras [159, 160].

Theorem 4.4.12 shows that the diagonalisation is unique up to the choice of the eigenstates when we have degeneracy: only then do we have the freedom of choice of the eigenstates relative to degenerate eigenvalues, i.e. eigenvalues arising more than once in a diagonalisation. See [110] for another proof of the uniqueness of the eigenvalues based on majorisation (and also a further axiom).

Theorem 4.4.12 implies that we can associate a unique probability distribution with every state: its spectrum. The spectrum of a state will be the basis to define entropies and resource monotones for the thermodynamics of isolated systems.

4.4.3 Extending the diagonalisation to arbitrary vectors

The diagonalisation theorem, proved for normalised states, can be easily extended to arbitrary elements of the vector space $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$.

Proposition 4.4.13. *For every system* A *and for every vector* $\xi \in \mathsf{St}_{\mathbb{R}}(\mathrm{A})$ *there exist a unique set of d real numbers* $\{x_i\}_{i=1}^{d}$ *and a maximal set of perfectly distinguishable states* $\{\alpha_i\}_{i=1}^{d}$ *such that*

$$\xi = \sum_{i=1}^{d} x_i\alpha_i.$$

We omit the proof, which is the same as in [68, corollary 21]. Again the x_i's are called the eigenvalues of ξ, and the α_i's are called the eigenstates of ξ. A similar result was obtained also in [170] under different axioms.

Note that, since ξ is a generic vector of $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$, the eigenvalues are arbitrary real numbers; if instead ξ is in the cone $\mathsf{St}_{+}(\mathrm{A})$, the eigenvalues are non-negative real numbers.

Using this result, we can prove that the operational norm of a vector in $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$ coincides with the 1-norm of its spectrum $\mathbf{x}$ [111].

Proposition 4.4.14. *Let $\xi \in \mathsf{St}_{\mathbb{R}}(\mathrm{A})$ be diagonalised as $\xi = \sum_{i=1}^{d} x_i \alpha_i$. Then $\|\xi\| = \sum_{i=1}^{d} |x_i|$.*

Proof. Let us separate the terms with non-negative eigenvalues from the terms with negative eigenvalues, so that we can write $\xi = \xi_+ - \xi_-$, where $\xi_+ := \sum_{x_i \geq 0} x_i \alpha_i$, and $\xi_- = \sum_{x_i < 0} (-x_i)\, \alpha_i$. Clearly, $\xi_+, \xi_- \in \mathsf{St}_{+}(\mathrm{A})$. Recall the definition of the operational norm of a vector:

$$\|\xi\| = \sup_{a \in \mathsf{Eff}(\mathrm{A})} (a|\xi) - \inf_{a \in \mathsf{Eff}(\mathrm{A})} (a|\xi)\,.$$

In order to achieve the supremum of $(a|\xi)$ we must have $(a|\xi_-) = 0$. Moreover,

$$(a|\xi_+) = \sum_{x_i \geq 0} x_i\, (a|\alpha_i) \leq \sum_{x_i \geq 0} x_i$$

since $(a|\alpha_i) \leq 1$ for every i. The supremum of $(a|\xi_+)$ is achieved by $a = \sum_{x_i \geq 0} \alpha_i^{\dagger}$. Hence $\sup_a (a|\xi) = \sum_{x_i \geq 0} x_i$. By a similar argument, one shows that $\inf_a (a|\xi) = \sum_{x_i < 0} x_i$. Therefore

$$\|\xi\| = \sum_{x_i \geq 0} x_i + \sum_{x_i < 0} (-x_i) = \sum_{i=1}^{d} |x_i|\,.$$

□

For $p \geq 1$, the p-norm of a vector $\mathbf{x} \in \mathbb{R}^d$ is defined as $\|\mathbf{x}\|_p := \left(\sum_{i=1}^{d} |x_i|^p\right)^{\frac{1}{p}}$, thus we have $\|\xi\| = \|\mathbf{x}\|_1$, where $\mathbf{x}$ is the spectrum of ξ.

4.4.4 Extending the dagger map

Thanks to the diagonalisation theorems, the dagger map $\dagger : \mathsf{PurSt}_1(\mathrm{A}) \to \mathsf{PurEff}_1(\mathrm{A})$ can be extended to arbitrary vectors via the relation

$$\xi = \sum_{i=1}^{d} x_i \alpha_i \quad \longmapsto \quad \xi^\dagger := \sum_{i=1}^{d} x_i \alpha_i^\dagger.$$

Note that, since the diagonalisation is unique (up to degeneracy), the vector $\xi^\dagger$ is well-defined, i.e. it does not depend on the choice of the α_i's as long as they are eigenstates of ξ. Writing ξ like in theorem 4.4.12, $\xi = \sum_{k=1}^{s} \lambda_k \Pi_k$, to prove this fact, it suffices to show the following.

Proposition 4.4.15. *Let $\{\alpha_i\}_{i=1}^r$ and $\left\{\alpha'_j\right\}_{j=1}^r$ be two sets of perfectly distinguishable pure states. Then, if $\sum_{i=1}^r \alpha_i = \sum_{j=1}^r \alpha'_j$, one has $\sum_{i=1}^r \alpha_i^\dagger = \sum_{j=1}^r \alpha_j'^\dagger$.*

Proof. Let us extend $\{\alpha_i\}_{i=1}^r$ and $\left\{\alpha'_j\right\}_{j=1}^r$ to two pure maximal sets $\{\alpha_i\}_{i=1}^d$ and $\left\{\alpha'_j\right\}_{j=1}^d$. Then, the invariant state has the two diagonalisations $\chi = \frac{1}{d}\sum_{i=1}^d \alpha_i$ and $\chi = \frac{1}{d}\sum_{j=1}^d \alpha'_j$ (proposition 4.4.5). Using this fact and the condition $\sum_{i=1}^r \alpha_i = \sum_{j=1}^r \alpha'_j$, we obtain $\sum_{i=r+1}^d \alpha_i = \sum_{j=r+1}^d \alpha'_j$. Hence, the invariant state can be decomposed as

$$\chi = \frac{1}{d}\left(\sum_{j=1}^{r} \alpha'_j + \sum_{i=r+1}^{d} \alpha_i\right).$$

By proposition 4.4.4, this implies that the states $\left\{\alpha'_j\right\}_{j=1}^r \cup \{\alpha_i\}_{i=r+1}^d$ form a maximal set. Now, the correspondence between maximal sets and pure sharp measurements (proposition 4.4.7) implies that the effects $\left\{\alpha_j'^\dagger\right\}_{j=1}^r \cup \{\alpha_i^\dagger\}_{i=r+1}^d$ form a measurement. Causality yields

$$\sum_{j=1}^{r} \alpha_j'^\dagger + \sum_{i=r+1}^{d} \alpha_i^\dagger = u.$$

On the other hand, the normalisation of the measurement $\{\alpha_i^\dagger\}_{i=1}^d$ reads

$$\sum_{i=1}^{d} \alpha_i^\dagger = u.$$

Comparing the two equalities we obtain the desired relation $\sum_{i=1}^{r} \alpha_i^\dagger = \sum_{j=1}^{r} \alpha_j'^\dagger$. □

Similarly to what we did in section 4.2, with a little abuse of notation we will denote as † even the inverse map, from $\mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$ to $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$. Now we are ready to define observables, and to introduce a functional calculus on them.

4.4.5 Functional calculus on observables

Using Steering, one can convert the diagonalisation result for the elements of $\mathsf{St}_{\mathbb{R}}(\mathrm{A})$ into a diagonalisation result for the elements of $\mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$ (see [68, 164], and, for a different approach, [158–160, 169]).

Proposition 4.4.16. *For every finite-dimensional system* A *and for every vector* $X \in \mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$ *there exist a set of d real numbers* $\{x_i\}_{i=1}^d$*, and a pure maximal set of states* $\{\alpha_i\}_{i=1}^d$ *such that*

$$X = \sum_{i=1}^{d} x_i \alpha_i^\dagger.$$

This result allows us to give a concrete characterisation of the norm on the vector space of effects, along the same lines as we did for the norm on the vector space of states.

Proposition 4.4.17. *Let* $X \in \mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$ *be diagonalised as* $X = \sum_{i=1}^{d} x_i \alpha_i^\dagger$*. Then* $\|X\| = \max_i |x_i|$.

Proof. Recall the definition of the norm of a vector in $\mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$:

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

Now, we have

$$|(X|\rho)| = \left| \sum_{i=1}^{d} x_i \left(\alpha_i^\dagger \middle| \rho\right) \right| \leq \sum_{i=1}^{d} |x_i| \left(\alpha_i^\dagger \middle| \rho\right).$$

This is a sub-normalised "convex" combination of the $|x_i|$'s. Therefore

$$\sum_{i=1}^{d} |x_i| \left(\alpha_i^\dagger \middle| \rho\right) \leq \max_i |x_i| .$$

In conclusion $|(X|\rho)| \leq \max_i |x_i|$. The bound is achieved when ρ is the state α_{i_0} associated with $\max_i |x_i| =: x_{i_0}$. Therefore,

$$\|X\| = \sup_{\rho \in \mathsf{St}(\mathrm{A})} |(X|\rho)| = \max_i |x_i| .$$

□

In other words, the norm of a vector $X \in \mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$ is $\|X\| = \|\mathbf{x}\|_\infty$, where $\mathbf{x}$ is the spectrum of X, and $\|\mathbf{x}\|_\infty = \lim_{p\to+\infty} \|\mathbf{x}\|_p = \max_i |x_i|$.

Thanks to the diagonalisation theorem, all the elements of the vector space $\mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$ can be regarded as *observables*, in a similar sense to the use of the term in quantum theory. Indeed, given a diagonalisation $X = \sum_{i=1}^{d} x_i \alpha_i^\dagger$ one can think of the eigenvalues as the "values" associated with the outcomes of the sharp measurement $\{\alpha_i^\dagger\}_{i=1}^{d}$. In this way, one can interpret

$$\langle X \rangle_\rho := (X|\rho) = \sum_{i=1}^{d} x_i \left(\alpha_i^\dagger \middle| \rho\right)$$

as the *expectation value* of the observable X on ρ, because $\left(\alpha_i^\dagger \middle| \rho\right)$ are probabilities.

Like in quantum theory, the spectral theorem allows one to define a functional calculus on observables:[5] given an observable X and a function $f : \mathbb{R} \to \mathbb{R}$, one can define the observable

$$f(X) := \sum_{i=1}^{d} f(x_i)\, \alpha_i^\dagger .$$

Note that the observable $f(X)$ is well-defined, because the eigenvalues of X are unique, and as a consequence of proposition 4.4.15. In particular,

[5] See also [159, 160, 169] for a different approach.

one can choose the observable X to be the dagger of a state $\rho = \sum_{i=1}^{d} p_i \alpha_i$, thus obtaining

$$f\left(\rho^\dagger\right) = \sum_{i=1}^{d} f\left(p_i\right) \alpha_i^\dagger .$$

In the following chapter, we will take f to be the logarithm function on some base greater than 1, defining the "surprisal observable":

$$-\log_a \rho^\dagger = -\sum_{i=1}^{d} \left(\log_a p_i\right) \alpha_i^\dagger .$$

Therefore, Shannon-von Neumann entropy can be defined as the expectation value of the surprisal observable

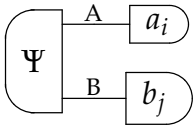
$$S\left(\rho\right) := \left(-\log_a \rho^\dagger \middle| \rho\right).$$

The functional calculus on observables will also be the basis to define the generalised relative entropy in section 5.5.

4.5 Schmidt decomposition

Using the diagonalisation theorem we can prove an operational version of the Schmidt decomposition of pure bipartite states [116,117]. The intuitive content of the Schmidt decomposition is that for every state of a bipartite system there exist two perfectly correlated pure observation-tests on the component systems, a similar situation to having conjugates [158–160]. More formally, this property is stated by the following theorem.

Theorem 4.5.1 (Schmidt decomposition)**.** *Let* Ψ *be a pure state of the composite system* AB*. Then, there exist a pure sharp measurement* $\{a_i\}_{i=1}^{d_\mathrm{A}}$ *on system* A*, and a pure sharp measurement* $\{b_j\}_{j=1}^{d_\mathrm{B}}$ *on system* B *such that*

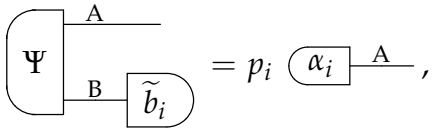


$$= p_i \delta_{ij} \qquad \forall i \in \{1, \ldots, r\}, \tag{4.5.1}$$

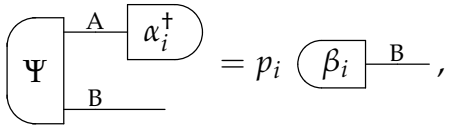
where $r \leq \min\{d_\mathrm{A}, d_\mathrm{B}\}$ *is a suitable integer, the* Schmidt rank*,* $\{p_i\}_{i=1}^{r}$ *is a probability distribution, with all non-vanishing elements.*

Moreover, one has the diagonalisations $\rho_{\mathrm{A}} = \sum_{i=1}^{r} p_i a_i^\dagger$ *and* $\rho_{\mathrm{B}} = \sum_{i=1}^{r} p_i b_i^\dagger$, *where* ρ_{A} *and* ρ_{B} *are the marginals of* Ψ *on systems* A *and* B *respectively.*

Proof. Let ρ_{A} be the marginal of Ψ on system A and let $\rho_{\mathrm{A}} = \sum_{i=1}^{r} p_i \alpha_i$ be a diagonalisation of ρ_{A}, where $p_i > 0$ for all $i \in \{1, \ldots, r\}$. By Pure Steering, there exists an observation-test on B, call it $\left\{\widetilde{b}_i\right\}_{i=1}^{r}$, such that

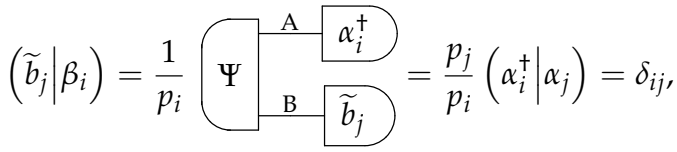


for every $i \in \{1, \ldots, r\}$. On the other hand, by corollary 4.4.2, the pure sharp measurement $\{a_i\}_{i=1}^{d_{\mathrm{A}}}$, where $d_{\mathrm{A}} \geq r$ and $a_i = \alpha_i^\dagger$ for $i \in \{1, \ldots, r\}$, induces pure states on system B, as follows

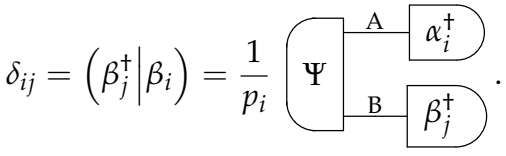


(4.5.2)

where each state β_i is pure and normalised, for every $i \in \{1, \ldots, r\}$. Note that the right-hand side vanishes if $i \in \{r+1, \ldots, d_{\mathrm{A}}\}$. Combining the two equations above, we obtain



for all $i, j \in \{1, \ldots, r\}$. Hence, the pure states $\{\beta_i\}_{i=1}^{r}$ are perfectly distinguishable. This means that, if ρ_{B} is the marginal of Ψ on system B, the pure sharp measurement $\{a_i\}_{i=1}^{d_{\mathrm{A}}}$ induces a diagonalisation of ρ_{B} in terms of the states $\{\beta_i\}_{i=1}^{r}$. Indeed $\sum_{i=1}^{r} p_i \beta_i = \rho_{\mathrm{B}}$ because $\sum_{i=1}^{d_{\mathrm{A}}} a_i = u$. By corollary 4.3.2, we know that the effects $\left\{\beta_i^\dagger\right\}_{i=1}^{r}$ are such that

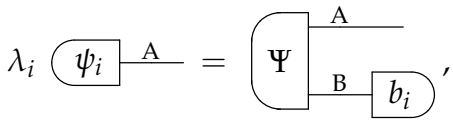


Hence, choosing $\{a_i\}_{i=1}^{d_A}$ and $\{b_j\}_{j=1}^{d_B}$ to be pure sharp measurements with $a_i := \alpha_i^\dagger$ and $b_i := \beta_i^\dagger$, for $i = \{1, \dots, r\}$, one obtains eq. (4.5.1). To conclude the proof, recall that $\rho_A = \sum_{i=1}^r p_i \alpha_i$ is a diagonalisation of ρ_A. Moreover, $\rho_B = \sum_{i=1}^r p_i \beta_i$ is a diagonalisation of ρ_B thanks to eq. (4.5.2). □

Theorem 4.5.1 guarantees that the diagonalisations of the two marginals of a pure bipartite state have the same non-vanishing eigenvalues. Moreover, it implies that we can induce the pure states in the diagonalisation of a state ρ by applying suitable normalised *pure* effects on the purifying system of *any* purification of ρ.

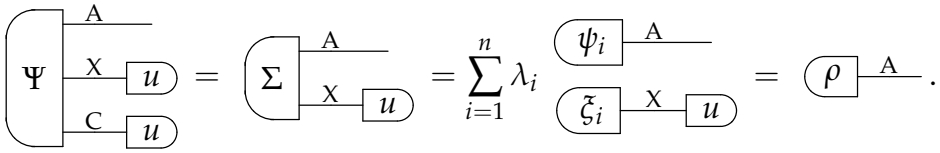
What about general convex decompositions of ρ into pure states that are not diagonalisations? The following corollary guarantees that they can always be induced by pure sharp measurements on the purifying system of *some* purification of ρ.

Corollary 4.5.2. *Let $\rho = \sum_{i=1}^n \lambda_i \psi_i$ be a convex decomposition of $\rho \in \mathsf{St}_1(A)$ into pure states, with $\lambda_i > 0$ for every $i \in \{1, \dots, n\}$. Then there exist a purification $\Psi \in \mathsf{PurSt}_1(AB)$ of ρ, with $d_B \geq n$, and a pure sharp measurement $\{b_j\}_{j=1}^{d_B}$ on B, such that*



for every $i \in \{1, \dots, n\}$.

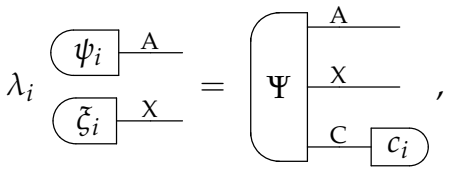
Proof. Consider a system X of dimension n, and let $\{\xi_i\}_{i=1}^n$ be a pure maximal set of X. Consider now the state $\Sigma \in \mathsf{St}_1(AX)$, given by $\Sigma := \sum_{i=1}^n \lambda_i \psi_i \otimes \xi_i$. This is a diagonalisation of Σ, for the states $\{\psi_i \otimes \xi_i\}_{i=1}^n$ are pure by Purity Preservation, and they are distinguished perfectly by the observation-test $\{u_A \otimes \xi_i^\dagger\}_{i=1}^n$. Now, let us consider a purification $\Psi \in \mathsf{PurSt}_1(AXC)$ of Σ. This is clearly a purification of ρ too, indeed



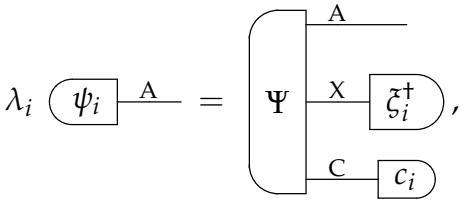
Now, by theorem 4.5.1 applied to the purification Ψ of Σ, there exists a pure sharp measurement[6] $\{c_k\}_{k=1}^{d_C}$ on C, that induces the pure states in the

[6] Clearly we have $d_C \geq n d_A$.

diagonalisation of Σ:



for $i \in \{1, \ldots, n\}$. Now, take the pure sharp measurement $\{\xi_i^\dagger\}_{i=1}^n$ on X, which yields



for $i \in \{1, \ldots, n\}$. To complete the proof, note that $\{\xi_i^\dagger \otimes c_k\}_{i=1,\, k=1}^{n\quad d_C}$ is still a pure sharp measurement $\{b_j\}$ on XC (by proposition 4.4.8), with $nd_C \geq n$ effects, where j runs on the pairs (i, k). Now it is enough to take the purifying system B to be XC, and to take $k = i$ in $\{\xi_i^\dagger \otimes c_k\}_{i=1,\, k=1}^{n\quad d_C}$. □

We will use this corollary in the proof of theorem 5.4.15.

4.6 Example: doubled quantum theory

In this section we present a new example of a sharp theory with purification [105], called "doubled quantum theory". This theory will provide a counterexample to thermodynamic convertibility and majorisation in section 5.8.

Consider a theory where every non-trivial system is the direct sum of two isomorphic quantum systems with Hilbert spaces $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Physically, we can think of the two Hilbert spaces as two superselection sectors. We associate each "doubled quantum system" with a pair of isomorphic Hilbert spaces $(\mathcal{H}_0, \mathcal{H}_1)$, with $\mathcal{H}_0 \approx \mathcal{H}_1$. We define the states of the doubled quantum system to be of the form

$$\rho = p\rho_0 \oplus (1-p)\,\rho_1 \tag{4.6.1}$$

where ρ_0 and ρ_1 are two density matrices in the two sectors and $p \in [0,1]$. The direct sum in eq. (4.6.1) means that there is no coherence between the two sectors.

Likewise, we define the effects to be all quantum effects of the form $e = e_0 \oplus e_1$, where e_0 and e_1 are two quantum effects in the two sectors. The allowed channels from the input system $(\mathcal{H}_0, \mathcal{H}_1)$ to the output system $(\mathcal{K}_0, \mathcal{K}_1)$ are the quantum channels (completely positive trace-preserving maps) that

1. send operators on $\mathcal{H}_0 \oplus \mathcal{H}_1$ to operators on $\mathcal{K}_0 \oplus \mathcal{K}_1$;

2. map block-diagonal operators to block-diagonal operators.

The set of allowed tests is defined as the set of quantum instruments $\{\mathcal{C}_j\}_{j \in X}$, where each quantum operation $\mathcal{C}_j$ respects the two conditions above for channels.

This means that in the allowed unitary channels $\mathcal{U}(\cdot) = U \cdot U^\dagger$, U must be of the form $U = (U_0 \oplus U_1) S^k$, where S is the unitary transformation that exchanges the two sectors (it exists because they are isomorphic), $k \in \{0,1\}$, and U_0 and U_1 are unitary transformations that act only on $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Therefore, if $k = 0$, there is no hopping of sector, and if $k = 1$ the two sectors are exchanged.

Doubled quantum theory satisfies Causality, Pure Sharpness, and Purity Preservation Causality is immediate: for every system, the only deterministic effect is the identity matrix. Pure Sharpness is also immediate: every rank-one projector is a pure sharp effect. As to Purity Preservation, note that the only pure transformations are quantum operations of the single-Kraus form $\mathcal{Q}(\cdot) = Q \cdot Q^\dagger$. Clearly, the composition of two single-Kraus operations (both in parallel and in sequence) is a single-Kraus operation. In other words, the composition of two pure transformations is pure.

4.6.1 Composite systems

To study if doubled quantum theory satisfies Purification, it is necessary to specify how systems compose in this theory.

The peculiarity of doubled quantum theory is the way systems are composed, which is *not* the intuitive way to compose systems with superselection rules. The product of two doubled quantum systems $(\mathcal{H}_0^{\mathrm{A}}, \mathcal{H}_1^{\mathrm{A}})$ and $(\mathcal{H}_0^{\mathrm{B}}, \mathcal{H}_1^{\mathrm{B}})$ is the doubled quantum system $(\mathcal{H}_0^{\mathrm{AB}}, \mathcal{H}_1^{\mathrm{AB}})$, with the two sectors defined by

$$\begin{cases} \mathcal{H}_0^{\mathrm{AB}} := (\mathcal{H}_0^{\mathrm{A}} \otimes \mathcal{H}_0^{\mathrm{B}}) \oplus (\mathcal{H}_1^{\mathrm{A}} \otimes \mathcal{H}_1^{\mathrm{B}}) \\ \mathcal{H}_1^{\mathrm{AB}} := (\mathcal{H}_0^{\mathrm{A}} \otimes \mathcal{H}_1^{\mathrm{B}}) \oplus (\mathcal{H}_1^{\mathrm{A}} \otimes \mathcal{H}_0^{\mathrm{B}}) \end{cases} . \tag{4.6.2}$$

Note that the direct sum inside each sector does *not* mean the presence of additional sectors. We illustrate this with an example.

Example 4.6.1. Consider the composite system of two doubled qubits, corresponding to $\mathcal{H}_0^{\mathrm{A}} \approx \mathcal{H}_1^{\mathrm{A}} \approx \mathcal{H}_0^{\mathrm{B}} \approx \mathcal{H}_1^{\mathrm{B}} \approx \mathbb{C}^2$. An example of state of the composite system is the pure state (the first index denotes the sector)

$$|\Psi\rangle = \frac{1}{\sqrt{2}} \left(|0,0\rangle_{\mathrm{A}} |0,0\rangle_{\mathrm{B}} + |1,0\rangle_{\mathrm{A}} |1,0\rangle_{\mathrm{B}} \right), \tag{4.6.3}$$

where $\{|0,0\rangle, |0,1\rangle\}$ is an orthonormal basis for $\mathcal{H}_0$ and $\{|1,0\rangle, |1,1\rangle\}$ is an orthonormal basis for $\mathcal{H}_1$. Thus we see that there *is* coherence allowed between $\mathcal{H}_0^{\mathrm{A}} \otimes \mathcal{H}_0^{\mathrm{B}}$ and $\mathcal{H}_1^{\mathrm{A}} \otimes \mathcal{H}_1^{\mathrm{B}}$. However, note that, when one of the two systems is traced out, the remaining local state has the block-diagonal form $\rho = \frac{1}{2} |0,0\rangle \langle 0,0| \oplus \frac{1}{2} |1,0\rangle \langle 1,0|$. This means that the coherence between the two terms in the state (4.6.3) is invisible at the single-system level.

From a physical point of view, doubled quantum theory can be thought of as ordinary quantum theory with a superselection rule on the *total* parity. Every system is split into two identical sectors of even and odd parity, respectively. When systems are composed, the sectors are grouped together based on the total parity, so that superpositions between subspaces with the same parity are allowed.

Here we summarise the basic operational features of doubled quantum theory concerning the composition of systems.

Doubled quantum theory violates Local Tomography An equivalent formulation of Local Tomography is that the dimension of the vector space spanned by the states of a composite system is equal to the product of

the dimensions of the vector spaces spanned by the states of the components [61,67,70], in formula $D_{\mathrm{AB}} = D_{\mathrm{A}} D_{\mathrm{B}}$, where D is the dimension of the vector space of states. This is because, if Local Tomography holds, one has $\mathrm{St}_{\mathbb{R}}(\mathrm{AB}) = \mathrm{St}_{\mathbb{R}}(\mathrm{A}) \otimes \mathrm{St}_{\mathbb{R}}(\mathrm{B})$ (cf. remark 2.2.6).

The equality $D_{\mathrm{AB}} = D_{\mathrm{A}} D_{\mathrm{B}}$ fails to hold in doubled quantum theory, where the dimension of the global vector space is strictly larger than the product of the dimensions of the individual vector spaces. To see it, note that the block-diagonal states of the form (4.6.1) span a vector space of dimension $D = 2d^2$, where d is the dimension of the Hilbert spaces $\mathcal{H}_0$ and $\mathcal{H}_1$. Given two systems A and B, the product of the individual dimensions is $D_{\mathrm{A}} D_{\mathrm{B}} = 2d_{\mathrm{A}}^2 \cdot 2d_{\mathrm{B}}^2 = (2d_{\mathrm{A}} d_{\mathrm{B}})^2$. On the other hand, each of the Hilbert spaces $\mathcal{H}_0^{\mathrm{AB}}$ and $\mathcal{H}_1^{\mathrm{AB}}$ in eq. (4.6.2) has dimension $d_{\mathrm{AB}} = 2d_{\mathrm{A}} d_{\mathrm{B}}$. Hence, the vector space spanned by the states of the composite system has dimension $D_{\mathrm{AB}} = 2d_{\mathrm{AB}}^2 = 2(2d_{\mathrm{A}} d_{\mathrm{B}})^2$, that is, twice the dimension of the vector space spanned by product states. This means that when systems are composed, genuinely new states arise, that cannot be reduced to states of the two components.

Doubled quantum theory satisfies Purification A generic state of a system $(\mathcal{H}_0, \mathcal{H}_1)$ can be diagonalised as

$$\rho = \left(\sum_{j=1}^{d} \lambda_j \left|\varphi_{j,0}\right\rangle \left\langle \varphi_{j,0}\right| \right) \oplus \left(\sum_{j=1}^{d} \mu_j \left|\psi_{j,1}\right\rangle \left\langle \psi_{j,1}\right| \right),$$

where $\left\{\left|\varphi_{j,0}\right\rangle\right\}_{j=1}^{d}$ is an orthonormal basis for $\mathcal{H}_0$ and $\left\{\left|\psi_{j,1}\right\rangle\right\}_{j=1}^{d}$ is an orthonormal basis for $\mathcal{H}_1$. The state can be purified e.g. by adding a copy of system $(\mathcal{H}_0, \mathcal{H}_1)$. Since the composite system has two superselection sectors, there will be two types of purification: purifications in the even subspace $\mathcal{H}_0^{\mathrm{AB}}$ and purifications in the odd subspace $\mathcal{H}_1^{\mathrm{AB}}$. A purification in the subspace $\mathcal{H}_0^{\mathrm{AB}}$ has the form

$$\left|\Psi_0\right\rangle = \left(\sum_{j=1}^{d} \sqrt{\lambda_j} \left|\varphi_{j,0}\right\rangle \left|\alpha_{j,0}\right\rangle \right) + \left(\sum_{j=1}^{d} \sqrt{\mu_j} \left|\psi_{j,1}\right\rangle \left|\beta_{j,1}\right\rangle \right),$$

where $\left\{\left|\alpha_{j,0}\right\rangle\right\}_{j=1}^{d}$ is an orthonormal basis for $\mathcal{H}_0$ and $\left\{\left|\beta_{j,1}\right\rangle\right\}_{j=1}^{d}$ is an orthonormal basis for $\mathcal{H}_1$. A purification in the subspace $\mathcal{H}_1^{\mathrm{AB}}$ will have the

form

$$|\Psi_1\rangle = \left(\sum_{j=1}^{d} \sqrt{\lambda_j}\, |\varphi_{j,0}\rangle \left|\alpha'_{j,1}\right\rangle\right) + \left(\sum_{j=1}^{d} \sqrt{\mu_j}\, |\psi_{j,1}\rangle \left|\beta'_{j,0}\right\rangle\right)$$

where $\left\{\left|\alpha'_{j,1}\right\rangle\right\}_{j=1}^{d}$ is an orthonormal basis for $\mathcal{H}_1$ and $\left\{\left|\beta'_{j,0}\right\rangle\right\}_{j=1}^{d}$ is an orthonormal basis for $\mathcal{H}_0$. Note that any two such purifications are equivalent under local unitary transformations: indeed, one has $|\Psi_1\rangle = (\mathbf{1} \otimes U)\,|\Psi_0\rangle$, where U is the unitary matrix defined by

$$U = \left(\sum_{j=1}^{d} \left|\alpha'_{j,1}\right\rangle \langle\alpha_{j,0}|\right) + \left(\sum_{j=1}^{d} \left|\beta'_{j,0}\right\rangle \langle\beta_{j,1}|\right).$$

The same arguments apply to purifications within the same sector and to purifications where the purifying system is not a copy of the original system. In summary, every state can be purified and every two purifications with the same purifying system are equivalent under local unitaries.

4.7 Example: extended classical theory

In this section we introduce another new example of a sharp theory with purification [101]. This example has a great importance because it shows that classical theory, which does *not* satisfy Purification, can be regarded as part of a larger theory obeying the Purification principle. Specifically, in this extended classical theory there are some systems that look entirely classical at the single-system level, but they compose in a different, coherent, way, so as to save the validity of Purification. This example therefore shows that all the results obtained above for sharp theories with purification (e.g. diagonalisation, etc.) can be carried over to classical theory, at least at the single-system level. Therefore classical theory is not excluded by our treatment.

Extended classical theory will include classical and non-classical systems, called *coherent dits* (or *codits* for short), in analogy with the similar notion in quantum Shannon theory [173]. The guiding idea is to entangle classical systems with each other, to provide the desired purifications, while at the same time to keep them classical at the single-system level.

In principle, we could have modelled classical theory as a sub-theory of quantum theory, but clearly in this case we would observe interference at the level of single systems, which we do not want.

4.7.1 Coherent composition of bits

To understand how the construction of extended classical theory works, let us illustrate the 2-dimensional case first. Recall that the state of a classical bit can be represented using the density matrix formalism as

$$\rho = p\,|0\rangle\langle 0| \oplus (1-p)\,|1\rangle\langle 1|\,, \tag{4.7.1}$$

where $p \in [0,1]$, and the direct sum sign is a reminder that the off-diagonal elements are forbidden. The composite system of two classical bits A and B is a 4-dimensional classical system, which is represented as a quantum system with 4 superselection sectors. In formulas, if $\left(\mathcal{H}_0^{\mathrm{A}}, \mathcal{H}_1^{\mathrm{A}}\right)$ represents the classical bit A, with its 2 superselection sectors $\mathcal{H}_0^{\mathrm{A}} = \mathrm{Span}\left\{|0\rangle_{\mathrm{A}}\right\}$ and $\mathcal{H}_1^{\mathrm{A}} = \mathrm{Span}\left\{|1\rangle_{\mathrm{A}}\right\}$, and $\left(\mathcal{H}_0^{\mathrm{B}}, \mathcal{H}_1^{\mathrm{B}}\right)$ is the classical system B, the composite system is $\left(\mathcal{H}_{00}^{\mathrm{AB}}, \mathcal{H}_{01}^{\mathrm{AB}}, \mathcal{H}_{10}^{\mathrm{AB}}, \mathcal{H}_{11}^{\mathrm{AB}}\right)$, with 4 sectors of dimension 1. Here the composition is the usual one in the presence of superselection sectors, namely

$$\begin{cases} \mathcal{H}_{00}^{\mathrm{AB}} := \mathcal{H}_0^{\mathrm{A}} \otimes \mathcal{H}_0^{\mathrm{B}} \\ \mathcal{H}_{01}^{\mathrm{AB}} := \mathcal{H}_0^{\mathrm{A}} \otimes \mathcal{H}_1^{\mathrm{B}} \\ \mathcal{H}_{10}^{\mathrm{AB}} := \mathcal{H}_1^{\mathrm{A}} \otimes \mathcal{H}_0^{\mathrm{B}} \\ \mathcal{H}_{11}^{\mathrm{AB}} := \mathcal{H}_1^{\mathrm{A}} \otimes \mathcal{H}_1^{\mathrm{B}} \end{cases}.$$

Clearly a state of this classical composite system is of the form

$$\rho_{\mathrm{AB}} = p_{00}\,|0\rangle\langle 0| \otimes |0\rangle\langle 0| \oplus p_{01}\,|0\rangle\langle 0| \otimes |1\rangle\langle 1| \oplus$$

$$\oplus p_{10}\,|1\rangle\langle 1| \otimes |0\rangle\langle 0| \oplus p_{11}\,|1\rangle\langle 1| \otimes |1\rangle\langle 1|$$

where $\{p_{ij}\}$, $i,j \in \{0,1\}$, is a probability distribution.

Now, to construct extended classical theory, let us consider a single 2-dimensional system with the same states as the classical bit; they are still of the form of eq. (4.7.1). We will change the way two classical bits compose, by imposing a superselection rule given by the *total* parity, in the same way we did for doubled quantum theory. Therefore the composition of two

classical bits $(\mathcal{H}_0^{\mathrm{A}}, \mathcal{H}_1^{\mathrm{A}})$ and $(\mathcal{H}_0^{\mathrm{B}}, \mathcal{H}_1^{\mathrm{B}})$, where $\mathcal{H}_0^{\mathrm{A}} \approx \mathcal{H}_1^{\mathrm{A}} \approx \mathcal{H}_0^{\mathrm{B}} \approx \mathcal{H}_1^{\mathrm{B}} \approx \mathbb{C}$ yields the system $(\mathcal{H}_0^{\mathrm{AB}}, \mathcal{H}_1^{\mathrm{AB}})$ where

$$\begin{cases} \mathcal{H}_0^{\mathrm{AB}} := (\mathcal{H}_0^{\mathrm{A}} \otimes \mathcal{H}_0^{\mathrm{B}}) \oplus (\mathcal{H}_1^{\mathrm{A}} \otimes \mathcal{H}_1^{\mathrm{B}}) \\ \mathcal{H}_1^{\mathrm{AB}} := (\mathcal{H}_0^{\mathrm{A}} \otimes \mathcal{H}_1^{\mathrm{B}}) \oplus (\mathcal{H}_1^{\mathrm{A}} \otimes \mathcal{H}_0^{\mathrm{B}}) \end{cases} . \tag{4.7.2}$$

Again, the direct sum inside each sector does *not* denote an internal superselection rule, so we allow coherence inside each sector, which is invisible at the single-system level. This fact will be true in all the direct sums we will write in this section about extended classical theory.

Note that eq. (4.7.2) gives exactly the same composition rule of doubled quantum theory, but restricted to 1-dimensional Hilbert spaces. In the composition rule, the Hilbert spaces are grouped together according to the residue classes modulo 2: if $k \in \{0, 1\}$, in $\mathcal{H}_k^{\mathrm{AB}}$ there is the direct sum of all terms whose indices sum to k modulo 2. Therefore

$$\mathcal{H}_0^{\mathrm{AB}} = \mathrm{Span}\left\{ |0\rangle_{\mathrm{A}} |0\rangle_{\mathrm{B}}, |1\rangle_{\mathrm{A}} |1\rangle_{\mathrm{B}} \right\};$$

and

$$\mathcal{H}_1^{\mathrm{AB}} = \mathrm{Span}\left\{ |0\rangle_{\mathrm{A}} |1\rangle_{\mathrm{B}}, |1\rangle_{\mathrm{A}} |0\rangle_{\mathrm{B}} \right\}.$$

Consequently, the pure states of the composite systems can be represented as unit vectors either of the form ($\alpha, \beta \in \mathbb{C}$)

$$|\Phi_0\rangle_{\mathrm{AB}} = \alpha |0\rangle_{\mathrm{A}} |0\rangle_{\mathrm{B}} + \beta |1\rangle |1\rangle,$$

or of the form

$$|\Phi_1\rangle_{\mathrm{AB}} = \alpha |0\rangle_{\mathrm{A}} |1\rangle_{\mathrm{B}} + \beta |1\rangle_{\mathrm{A}} |0\rangle_{\mathrm{B}}.$$

Note that in the composite system the only allowed states are those of the form $\rho = p\rho_0 \oplus (1-p)\rho_1$, where $p \in [0, 1]$, ρ_0 is a density matrix on $\mathcal{H}_0^{\mathrm{AB}}$, and ρ_1 is a density matrix on $\mathcal{H}_1^{\mathrm{AB}}$. The allowed effects are of the form $e = e_0 \oplus e_1$, where e_0 is a quantum effect on $\mathcal{H}_0^{\mathrm{AB}}$, and e_1 is a quantum effect on $\mathcal{H}_1^{\mathrm{AB}}$. The allowed channels on this systems are those quantum channels that respect the block-diagonal structure.

Remark 4.7.1. In the composition of the two classical bits we *cannot* have the full 4-dimensional quantum system

$$\mathrm{Span}\left\{ |0\rangle_{\mathrm{A}} |0\rangle_{\mathrm{B}}, |1\rangle_{\mathrm{A}} |1\rangle_{\mathrm{B}}, |0\rangle_{\mathrm{A}} |1\rangle_{\mathrm{B}}, |1\rangle_{\mathrm{A}} |0\rangle_{\mathrm{B}} \right\}.$$

Indeed, if this were the case, an allowed (pure) state of the composite system would be $\frac{1}{\sqrt{2}}\left(|0\rangle_{\mathrm{A}} + |1\rangle_{\mathrm{A}}\right)|0\rangle_{\mathrm{B}}$, which, when system B is traced out, would yield the forbidden state $\frac{1}{\sqrt{2}}\left(|0\rangle_{\mathrm{A}} + |1\rangle_{\mathrm{A}}\right)$. Therefore, to keep the state of A and B classical, we *need* the presence of the two superselection sectors $\mathcal{H}_0^{\mathrm{AB}}$ and $\mathcal{H}_1^{\mathrm{AB}}$.

With the above settings, it is easy to see that every state of a classical bit can be purified. For example, the generic bit state $\rho = p\,|0\rangle\langle 0| \oplus (1-p)\,|1\rangle\langle 1|$ has the purification

$$|\Psi_0\rangle_{\mathrm{AB}} = \sqrt{p}\,|0\rangle_{\mathrm{A}}\,|0\rangle_{\mathrm{B}} + \sqrt{1-p}\,|1\rangle_{\mathrm{A}}\,|1\rangle_{\mathrm{B}}\,.$$

In addition, it is possible to show that every two purifications of the same state differ by a local unitary operation on the purifying system. This includes, for example, the purification

$$|\Psi_1\rangle_{\mathrm{AB}} = \sqrt{p}\,|0\rangle_{\mathrm{A}}\,|1\rangle_{\mathrm{B}} + \sqrt{1-p}\,|1\rangle_{\mathrm{A}}\,|0\rangle_{\mathrm{B}}\,,$$

obtained from $|\Psi_0\rangle_{\mathrm{AB}}$ through the application of a bit flip on system B, which is an allowed transformation because it preserves the block-diagonal structure of system B.

Notice that the composition of two classical bits gives rise to a 4-dimensional system that is *not* classical. Therefore, in the theory we are constructing we will have classical and non-classical systems with the same dimension (e.g. a classical 4-dimensional system, and the coherent composition of 2 classical bits).

4.7.2 Coherent composition of dits

Let us generalise the results of the previous subsection to the coherent composition of two classical dits. A classical dit can be represented as a Hilbert space with d superselection sectors of dimension 1: $(\mathcal{H}_0, \ldots, \mathcal{H}_{d-1})$, where $\mathcal{H}_k = \mathrm{Span}\,\{|k\rangle\}$. The states of a classical dit are of the form

$$\rho = \bigoplus_{k=0}^{d-1} p_k\,|k\rangle\langle k|\,,$$

where $\{p_k\}_{k=0}^{d-1}$ is a probability distribution. Like for bits, the new composition rule for dits is based on residue classes (modulo d): the composite of two classical dits $\left(\mathcal{H}_0^{\mathrm{A}}, \ldots, \mathcal{H}_{d-1}^{\mathrm{A}}\right)$ and $\left(\mathcal{H}_0^{\mathrm{B}}, \ldots, \mathcal{H}_{d-1}^{\mathrm{B}}\right)$ is the system

$\left(\mathcal{H}_0^{\mathrm{AB}}, \ldots, \mathcal{H}_{d-1}^{\mathrm{AB}}\right)$ with d sectors of dimension d.

$$\begin{cases} \mathcal{H}_0^{\mathrm{AB}} := \left(\mathcal{H}_0^{\mathrm{A}} \otimes \mathcal{H}_0^{\mathrm{B}}\right) \oplus \left(\mathcal{H}_1^{\mathrm{A}} \otimes \mathcal{H}_{d-1}^{\mathrm{B}}\right) \oplus \ldots \oplus \left(\mathcal{H}_{d-1}^{\mathrm{A}} \otimes \mathcal{H}_1^{\mathrm{B}}\right) \\ \mathcal{H}_1^{\mathrm{AB}} := \left(\mathcal{H}_0^{\mathrm{A}} \otimes \mathcal{H}_1^{\mathrm{B}}\right) \oplus \left(\mathcal{H}_1^{\mathrm{A}} \otimes \mathcal{H}_0^{\mathrm{B}}\right) \oplus \ldots \oplus \left(\mathcal{H}_{d-1}^{\mathrm{A}} \otimes \mathcal{H}_2^{\mathrm{B}}\right) \\ \vdots \\ \mathcal{H}_{d-1}^{\mathrm{AB}} := \left(\mathcal{H}_0^{\mathrm{A}} \otimes \mathcal{H}_{d-1}^{\mathrm{B}}\right) \oplus \left(\mathcal{H}_1^{\mathrm{A}} \otimes \mathcal{H}_{d-2}^{\mathrm{B}}\right) \oplus \ldots \oplus \left(\mathcal{H}_{d-1}^{\mathrm{A}} \otimes \mathcal{H}_0^{\mathrm{B}}\right) \end{cases}.$$

Note that in $\mathcal{H}_k^{\mathrm{AB}}$, for $k \in \{0, \ldots, d-1\}$ there is a direct sum of all the terms $\mathcal{H}_j^{\mathrm{A}} \otimes \mathcal{H}_l^{\mathrm{B}}$ such that $j + l \equiv k \mod d$. In this way

$$\mathcal{H}_k^{\mathrm{AB}} = \mathrm{Span}\left\{ |j\rangle_{\mathrm{A}} |k - j \mod d\rangle_{\mathrm{B}} : j = 0, \ldots, d-1 \right\}.$$

The states of the composite system AB are therefore density matrices of the form

$$\rho_{\mathrm{AB}} = \bigoplus_{k=0}^{d-1} p_k \rho_k^{\mathrm{AB}}, \tag{4.7.3}$$

where ρ_k^{AB} is a density matrix of $\mathcal{H}_k^{\mathrm{AB}}$, and $\{p_k\}_{k=0}^{d-1}$ is a probability distribution. Again, the allowed effects are of the form $e = \bigoplus_{k=0}^{d-1} e_k$, where e_k is a quantum effect on $\mathcal{H}_k^{\mathrm{AB}}$. The allowed channels on this systems are those quantum channels that respect the block-diagonal structure.

The new composition of classical dits satisfies Purification It is easy to see that any state of a classical dit $\rho = \bigoplus_{j=0}^{d-1} p_j |j\rangle \langle j|$ can be purified. Specifically, we can find a purification in every sector of a composite system AB, where B is another classical dit. Indeed, for every sector k in AB, ρ can be purified as

$$|\Psi_k\rangle_{\mathrm{AB}} = \sum_{j=0}^{d-1} \sqrt{p_j} |j\rangle_{\mathrm{A}} |k - j \mod d\rangle_{\mathrm{B}}.$$

These purifications are all related to each other by a local unitary on B, which hops between the d sectors in B.

So far we have dealt with the composition of classical systems with the same dimension. Let us define the new, coherent, composition of a d_{A}-dimensional classical system A with a d_{B}-dimensional classical system B.

The standard way to compose system A, given by $\left(\mathcal{H}_0^{\mathrm{A}}, \ldots, \mathcal{H}_{d_{\mathrm{A}}-1}^{\mathrm{A}}\right)$, and system B, given by $\left(\mathcal{H}_0^{\mathrm{B}}, \ldots, \mathcal{H}_{d_{\mathrm{B}}-1}^{\mathrm{B}}\right)$, where each sector is 1-dimensional, is a system AB with $d_{\mathrm{A}} d_{\mathrm{B}}$ 1-dimensional sectors $\left(\mathcal{H}_{00}^{\mathrm{AB}}, \ldots, \mathcal{H}_{d_{\mathrm{A}}-1, d_{\mathrm{B}}-1}^{\mathrm{AB}}\right)$.

In this case, to define the new composition, we consider $\max\{d_{\mathrm{A}}, d_{\mathrm{B}}\}$ sectors, each of dimension $\min\{d_{\mathrm{A}}, d_{\mathrm{B}}\}$, so that altogether the Hilbert space for AB will have dimension $d_{\mathrm{A}} d_{\mathrm{B}}$. For concreteness, suppose $d_{\mathrm{A}} \leq d_{\mathrm{B}}$; then the composite system will be described by $\left(\mathcal{H}_0^{\mathrm{AB}}, \ldots, \mathcal{H}_{d_{\mathrm{B}}-1}^{\mathrm{AB}}\right)$, where

$$\begin{cases} \mathcal{H}_0^{\mathrm{AB}} := \left(\mathcal{H}_0^{\mathrm{A}} \otimes \mathcal{H}_0^{\mathrm{B}}\right) \oplus \left(\mathcal{H}_1^{\mathrm{A}} \otimes \mathcal{H}_{d_{\mathrm{B}}-1}^{\mathrm{B}}\right) \oplus \ldots \oplus \left(\mathcal{H}_{d_{\mathrm{A}}-1}^{\mathrm{A}} \otimes \mathcal{H}_{d_{\mathrm{B}}-d_{\mathrm{A}}+1}^{\mathrm{B}}\right) \\ \mathcal{H}_1^{\mathrm{AB}} := \left(\mathcal{H}_0^{\mathrm{A}} \otimes \mathcal{H}_1^{\mathrm{B}}\right) \oplus \left(\mathcal{H}_1^{\mathrm{A}} \otimes \mathcal{H}_0^{\mathrm{B}}\right) \oplus \ldots \oplus \left(\mathcal{H}_{d_{\mathrm{A}}-1}^{\mathrm{A}} \otimes \mathcal{H}_{d_{\mathrm{B}}-d_{\mathrm{A}}+2}^{\mathrm{B}}\right) \\ \vdots \\ \mathcal{H}_{d_{\mathrm{B}}-1}^{\mathrm{AB}} := \left(\mathcal{H}_0^{\mathrm{A}} \otimes \mathcal{H}_{d_{\mathrm{B}}-1}^{\mathrm{B}}\right) \oplus \left(\mathcal{H}_1^{\mathrm{A}} \otimes \mathcal{H}_{d_{\mathrm{B}}-2}^{\mathrm{B}}\right) \oplus \ldots \oplus \left(\mathcal{H}_{d_{\mathrm{A}}-1}^{\mathrm{A}} \otimes \mathcal{H}_{d_{\mathrm{B}}-d_{\mathrm{A}}}^{\mathrm{B}}\right) \end{cases}.$$

Again, in $\mathcal{H}_k^{\mathrm{AB}}$, for $k \in \{0, \ldots, d_{\mathrm{B}} - 1\}$ there is a direct sum of all the terms $\mathcal{H}_j^{\mathrm{A}} \otimes \mathcal{H}_l^{\mathrm{B}}$ such that $j + l \equiv k \mod d_{\mathrm{B}}$. In this way

$$\mathcal{H}_k^{\mathrm{AB}} = \mathrm{Span}\left\{|j\rangle_{\mathrm{A}} |k - j \mod d_{\mathrm{B}}\rangle_{\mathrm{B}} : j = 0, \ldots, d_{\mathrm{A}} - 1\right\}.$$

The states are still of the form (4.7.3), with $d = d_{\mathrm{B}}$.

4.7.3 The other composites

From the previous subsection we know that in extended classical theory, the generic system we have encountered so far is made of N superselection sectors $(\mathcal{H}_0, \ldots, \mathcal{H}_{N-1})$, each of which of dimension $n \leq N$. Note that this covers also the usual d-dimensional classical systems, for which $N = d$, and $n = 1$. To complete the theory, we must specify how these generic systems compose.

We *define* the composition of two general systems of extended classical theory to follow the same rules explained above for classical systems. More specifically, consider $(\mathcal{H}_0^{\mathrm{A}}, \ldots, \mathcal{H}_{N-1}^{\mathrm{A}})$, with sectors of dimension $n \leq N$, and $(\mathcal{H}_0^{\mathrm{B}}, \ldots, \mathcal{H}_{M-1}^{\mathrm{B}})$ with sectors of dimension $m \leq M$, and for concreteness suppose $N \leq M$. The composite system AB will have

$M = \max\{N, M\}$ sectors:

$$\begin{cases} \mathcal{H}_0^{\mathrm{AB}} := \left(\mathcal{H}_0^{\mathrm{A}} \otimes \mathcal{H}_0^{\mathrm{B}}\right) \oplus \left(\mathcal{H}_1^{\mathrm{A}} \otimes \mathcal{H}_{M-1}^{\mathrm{B}}\right) \oplus \ldots \oplus \left(\mathcal{H}_{N-1}^{\mathrm{A}} \otimes \mathcal{H}_{M-N+1}^{\mathrm{B}}\right) \\ \mathcal{H}_1^{\mathrm{AB}} := \left(\mathcal{H}_0^{\mathrm{A}} \otimes \mathcal{H}_1^{\mathrm{B}}\right) \oplus \left(\mathcal{H}_1^{\mathrm{A}} \otimes \mathcal{H}_0^{\mathrm{B}}\right) \oplus \ldots \oplus \left(\mathcal{H}_{N-1}^{\mathrm{A}} \otimes \mathcal{H}_{M-N+2}^{\mathrm{B}}\right) \\ \vdots \\ \mathcal{H}_{M-1}^{\mathrm{AB}} := \left(\mathcal{H}_0^{\mathrm{A}} \otimes \mathcal{H}_{M-1}^{\mathrm{B}}\right) \oplus \left(\mathcal{H}_1^{\mathrm{A}} \otimes \mathcal{H}_{M-2}^{\mathrm{B}}\right) \oplus \ldots \oplus \left(\mathcal{H}_{N-1}^{\mathrm{A}} \otimes \mathcal{H}_{M-N}^{\mathrm{B}}\right) \end{cases}, \tag{4.7.4}$$

where, again in $\mathcal{H}_k^{\mathrm{AB}}$ there is the direct sum of N terms $\mathcal{H}_j^{\mathrm{A}} \otimes \mathcal{H}_l^{\mathrm{B}}$ such that $j + l \equiv k \mod M$. We see that in this case each sector $\mathcal{H}_k^{\mathrm{AB}}$ in the composite system has dimension nmN. If we take $n = m = 1$, we recover the coherent composition law for classical systems. Note that, in general it is *not* true that $nmN \leq M$. Indeed, composing two systems arising from the coherent composition of dits, the resulting system has d sectors, each of which of dimension d^3. Therefore, in the most general system of extended classical theory there is no restriction on the dimension of sectors, and it is just a system with $N \geq 2$ isomorphic superselection sectors $(\mathcal{H}_0, \ldots, \mathcal{H}_{N-1})$. The same rule (4.7.4) still applies to these systems.

Generic states are of the form $\rho = \bigoplus_{k=0}^{N-1} p_k \rho_k$, where ρ_k is a density matrix of the sector $\mathcal{H}_k$, and $\{p_k\}_{k=0}^{N-1}$ is a probability distribution. Effects are of the form $e = \bigoplus_{k=0}^{N-1} e_k$, where e_k is a quantum effect on $\mathcal{H}_k$. Finally, all transformations between A and B are those quantum operations from A to B that preserve the block-diagonal structure.

Now we can finally show that extended classical theory is a sharp theory with purification. It is straightforward to show that the theory satisfies Causality, Purity Preservation and Pure Sharpness.

Extended classical theory satisfies Purification A state of $(\mathcal{H}_0^{\mathrm{A}}, \ldots, \mathcal{H}_{N-1}^{\mathrm{A}})$, with sectors of dimension n, can be diagonalised as

$$\rho = \bigoplus_{k=0}^{N-1} \left(\sum_{j=0}^{n-1} \lambda_{j,k} \left|\varphi_{j,k}\right\rangle \left\langle\varphi_{j,k}\right| \right),$$

where $\{\lambda_{j,k}\}$ is a probability distribution, and $\{\left|\varphi_{j,k}\right\rangle\}$ is an orthonormal basis of $\mathcal{H}_k^{\mathrm{A}}$, for every k. Then to obtain a purification of ρ, it is enough to take the same system as the purifying system B. Specifically a purification

in the sector $\mathcal{H}_l^{\mathrm{AB}}$ is given by

$$|\Psi_l\rangle_{\mathrm{AB}} = \sum_{k=0}^{N-1} \sum_{j=0}^{n-1} \sqrt{\lambda_{j,k}}\, |\varphi_{j,k}\rangle_{\mathrm{A}}\, |\alpha_{j,l-k}\rangle_{\mathrm{B}}$$

where $\{|\alpha_{j,l-k}\rangle\}_{j=0}^{n-1}$ is an orthonormal basis of sector $\mathcal{H}_{l-k}^{\mathrm{B}}$.[7] Since all superselection sectors are isomorphic, one can convert a purification on $\mathcal{H}_l^{\mathrm{AB}}$ into a purification on $\mathcal{H}_{l'}^{\mathrm{AB}}$ by a local hopping unitary on system B.

Extended classical theory violates Local Tomography We can show that Local Tomography fails in general. For a system $(\mathcal{H}_0^{\mathrm{A}}, \ldots, \mathcal{H}_{N-1}^{\mathrm{A}})$ with sectors of dimension n, the dimension of $\mathrm{St}_{\mathbb{R}}(\mathrm{A})$ is $D_{\mathrm{A}} = Nn^2$. Similarly for a system with M sectors of dimension m, we have $D_{\mathrm{B}} = Mm^2$. Now, by eq. (4.7.4) we have $D_{\mathrm{AB}} = MN^2m^2n^2$. Therefore

$$D_{\mathrm{AB}} = MN^2m^2n^2 > MNm^2n^2 = D_{\mathrm{A}}D_{\mathrm{B}},$$

which means that all composites in extended classical theory violate Local Tomography.

[7] Here, as above, $l - k$ is to be intended modulo N.

Chapter 5

Operational thermodynamics

After analysing the properties of sharp theories with purification in great detail, in this chapter finally we move to the actual study of thermodynamic properties of GPTs, with a special focus on sharp theories with purification. We mainly examine the simplest instance of thermodynamics, namely for systems with fixed energy, also known as *microcanonical thermodynamics*. Even this case will provide us with a lot of foundational insights. In accordance with recent thermodynamic results, we will use a resource-theoretic approach to microcanonical thermodynamics, which will allow us to extend its scope beyond classical and quantum theory. Here the relevant resource into play is the purity of states, therefore a resource-theoretic treatment of microcanonical thermodynamics involves setting up a resource theory of purity. It turns out that there are three fairly natural choices for such a resource theory, but in general they give rise to inequivalent notions of resources [105], unlike in quantum theory [37].

After a general treatment of microcanonical thermodynamics in causal GPTs [105] in sections 5.1 and 5.2, the axioms of sharp theories with purification are introduced in the thermodynamic analysis only from section 5.3. These theories will be the subject of the rest of the chapter, even when not specified explicitly. Majorisation will play a central role, for it gives a necessary criterion for the thermodynamic conversion of states under all three resource theories in sharp theories with purification. Requiring it to be sufficient too—or, in other words, that the three resource theories generate equivalent preorders—will result in a non-trivial constraint on the dynamics of the theory, called *unrestricted reversibility* [105].

Given the interplay between the three resource theories of purity and

majorisation, we use the diagonalisation of states in sharp theories with purification to define mixedness monotones as Schur-concave functions of the spectrum of a state [101]. We show that a large class of "spectral" monotones coincide with the "non-spectral" definitions already put forward in the literature [101, 163, 174, 175]. An important example of mixedness monotone is Shannon-von Neumann entropy, which exhibits some properties close to its quantum counterpart [101].

Moving beyond the setting of microcanonical thermodynamics, we use the entropic machinery of sharp theories with purification to define generalised Gibbs states using Jaynes' maximum entropy principle [16,17], thus introducing temperature in GPTs. We use these states to carry out an operational derivation of Landauer's principle, showing also how the bounds on energy dissipation can be overcome by using non-classical correlations, witnessed by negative conditional entropy [101].

5.1 The microcanonical framework

We start this chapter by examining the simplest example of thermodynamic situation: an isolated thermodynamic system. In the usual statistical mechanical treatment, this is described by the microcanonical ensemble [6, 7], extensively studied in classical and quantum theory. Here we want to extend the microcanonical description to arbitrary physical theories, to understand if all theories admit a sensible microcanonical ensemble. Being an isolated system, the energy of a microcanonical system is known.[1] This of course restricts the microstates of the system to a subset of the allowed microstates: a submanifold of the phase space in classical theory, and a subspace of the Hilbert space in quantum theory. One of the basic principles of statistical mechanics is that we can recover the thermodynamic properties of microcanonical systems from a suitable statistical mechanical state [7]. What state do we assign to isolated thermodynamic systems? The prescription comes from the *equal a priori probability postulate*.

Axiom 5.1.1 (Equal a priori probability postulate)**.** *The state of a system with a macroscopic constraint is given by the uniform mixture of all the microstates*

[1] Sometimes, for mathematical convenience, one tolerates a small uncertainty Δ on the energy [6].

compatible with that constraint.[2]

Justifying this principle is generally hard, and it usually involves arguments based on ergodic theory (see [15, chapter 15] for a review of the approach), at least in classical theory, or more recently, based on quantum entanglement [20, 21].

In the following we want to extend the microcanonical framework to arbitrary causal theories[3], without assuming any of the axioms of chapter 4. The microcanonical framework requires the extension of two ingredients to GPTs: the implementation of the macroscopic constraint, and the equal a priori probability postulate. As to the latter, we do not want to address the thorny issue of justifying it, which inevitably requires specifying a lot of details about the physical theory under examination, but rather to identify the conditions that allow one to formulate this principle in a theory more general than classical or quantum theory.

5.1.1 Theories of systems with constraints

We start by analysing how to define constrained systems in GPTs, and we will take inspiration from quantum theory. Note that constrained systems appear very often in physics and often as a result of conservation laws. Think e.g. of a system of particles confined in a fixed volume, or a system with conserved total angular momentum. In these scenarios, it is always meaningful to ask ourselves about the state of "minimum information" compatible with the macroscopic constraints, which will behave as a sort of generalised microcanonical state. Note that imposing constraints, at least for quantum and classical systems, leads to the definition of an *effective system* of smaller dimension than the original one. Let us see if this idea works in GPTs too, and if we can set up an effective OPT whose systems are these effective, constrained, systems.

To have a clearer picture, let us briefly review how the microcanonical constraint is usually imposed in quantum theory. In this case, we know the Hamiltonian H, and the specific value E of its spectrum. This restricts the allowed quantum states to (mixtures of) the eigenstates of H in the

[2] In the quantum case, some authors [7] formulate this postulate as the fact that the microcanonical state is an incoherent mixture of the basis states of the subspace identified by the macroscopic constraint.

[3] Here Causality is assumed for the sake of simplicity, but it is not necessary, cf. [105].

eigenspace associated with E. These are exactly the states ρ for which $P_E \rho P_E = \rho$, where P_E is the projector on the eigenspace with eigenvalue E. For example, the system S could be an electron in a hydrogen atom, in the absence of external fields. In general, the basis states of the electron are labelled as $|n, \ell, m_\ell, m_s\rangle$, where n, ℓ, m_ℓ, and m_s are the principal, orbital, magnetic, and spin quantum number respectively. Suppose we know the electron is in the ground state, corresponding to $n = 1$ and $\ell = m_\ell = 0$. In this case, the subspace associated with the ground state is a two-dimensional one, spanned by the "spin-up" and "spin-down" states $\left|n = 1, \ell = 0, m_\ell = 0, m_s = \frac{1}{2}\right\rangle$ and $\left|n = 1, \ell = 0, m_\ell = 0, m_s = -\frac{1}{2}\right\rangle$. This constrained system can be regarded as an *effective qubit*. Clearly, constraints different from energy can be treated in a similar way. Another example of an effective qubit is a single photon with wave vector $\mathbf{k}$: it is enough to restrict ourselves to the two-dimensional space spanned by the states $|\mathbf{k}, H, 1\rangle$ and $|\mathbf{k}, V, 1\rangle$, corresponding to vertical and horizontal polarisation respectively. In this case, we can see two constraints working together: a constraint on the wave vector, and a constraint on the energy of the field, namely the fact that we are dealing with a single photon.

The important feature of these examples of constrained systems is that they arise from a *linear* constraint $\mathcal{L}(\rho) = 0$ on the space of density matrices. For instance, in the microcanonical case described above, the linear map $\mathcal{L}$ is $\mathcal{L}(\cdot) = P_E(\cdot)P_E - \mathcal{I}(\cdot)$, $\mathcal{I}$ being the identity channel.

Motivated by the analysis of the quantum case, we define a *constrained* (or *effective*) *system* in GPTs as the original system S with some *linear* maps $\{\mathcal{L}_i\}_{i=1}^n$ implementing the constraints, where each $\mathcal{L}_i$ is an element of $\mathsf{Transf}_{\mathbb{R}}(\mathrm{S})$, the vector space spanned by physical transformations (cf. section 2.2):

$$\mathrm{A} := \left(\mathrm{S}, \{\mathcal{L}_i\}_{i=1}^n\right).$$

Clearly now we need to specify the set of states, effects, and transformations for the constrained system A. The states of A are clearly those that satisfy the constraints enforced by the linear maps $\{\mathcal{L}_i\}_{i=1}^n$:

$$\mathsf{St}(\mathrm{A}) = \{\rho \in \mathsf{St}(\mathrm{S}) : \mathcal{L}_i(\rho) = 0, i = 1, \ldots, n\}.$$

The transformations of the effective system A are those transformations of S that send states of A to states of A. The effects of A are just the effects of the original system S, restricted to the states in $\mathsf{St}(\mathrm{A})$, and possibly

identified if they become tomographically indistinguishable because of the reduced number of states, as shown in example 5.1.3.

Remark 5.1.2. In the classical case, a Hamiltonian is generally of the form $H(\mathbf{q},\mathbf{p}) = \frac{|\mathbf{p}|^2}{2m} + V(\mathbf{q})$. Thus imposing the microcanonical restriction $H(\mathbf{q},\mathbf{p}) = E$ leads to the constraint $\frac{|\mathbf{p}|^2}{2m} + V(\mathbf{q}) = E$ on $\mathbf{p}$ and $\mathbf{q}$, which is *not* linear. From this situation, one may be tempted to think that treating constraints as linear in all physical theories is not the right way. In fact, there is no contradiction to what we wrote above: the constraint $H(\mathbf{q},\mathbf{p}) = E$ is not linear on the phase space, i.e. the set of normalised pure states of classical theory. However, our constraints are applied to *all* states, not just the pure ones, so the right space on which to study the microcanonical restriction is the space of all probability density functions $\rho(\mathbf{q},\mathbf{p})$ on the phase space. In that space, the microcanonical constraint becomes that the probability density function $\rho(\mathbf{q},\mathbf{p})$ be supported on the submanifold $H(\mathbf{q},\mathbf{p}) = E$, which is a linear constraint on ρ.

Clearly, the formalism of constrained systems can be applied to more general situations than microcanonical thermodynamics. Example 5.1.3 shows that even classical sub-theories of a given GPT can be studied in the same way [115].

Example 5.1.3. Given a pure maximal set $\{\alpha_i\}_{i=1}^d$ of a system S of a causal theory, let $\boldsymbol{\alpha}$ be its convex hull: $\boldsymbol{\alpha} = \mathrm{Conv}\{\alpha_i : i = 1,\dots,d\}$, which represents the states of a classical sub-theory. Define the decoherence on $\boldsymbol{\alpha}$ as a channel $D_{\boldsymbol{\alpha}}$ such that $D_{\boldsymbol{\alpha}}\rho \in \boldsymbol{\alpha}$ for every state ρ, and $D_{\boldsymbol{\alpha}}\gamma = \gamma$ for every state γ in $\boldsymbol{\alpha}$ [115]. By this very definition, it is immediate to see that the classical states (i.e. the states in $\boldsymbol{\alpha}$) are exactly those for which $D_{\boldsymbol{\alpha}}\rho = \rho$. Hence the decoherence $D_{\boldsymbol{\alpha}}$ can be viewed as implementing a constraint on the states of a system, as explained above. What happens to the effects of this constrained subsystem? According to the recipe presented above, we simply restrict the effects of the original theory to the effective system $\boldsymbol{\alpha}$. Since now there are fewer states in $\boldsymbol{\alpha}$ than in the original theory, we must identify those effects that are no longer tomographically distinct on $\boldsymbol{\alpha}$. More precisely, we can introduce the following equivalence relation on the original set of effects $\mathsf{Eff}(\mathrm{S})$: $e \sim_{\boldsymbol{\alpha}} f$ if $(e|\gamma) = (f|\gamma)$ for every state γ in $\boldsymbol{\alpha}$. Therefore, the set of effects of the classical sub-theory is the set of equivalence classes $\mathsf{Eff}(\mathrm{S})/\boldsymbol{\alpha} := \mathsf{Eff}(\mathrm{S})/\sim_{\boldsymbol{\alpha}}$.

To complete our analysis, let us show that $\mathsf{Eff}\,(\mathrm{S})\,/\boldsymbol{\alpha}$ is actually the set of effects of some classical theory. Recall that in classical theory, every element in the cone of effects arises as a conical combination of the effects that distinguish the pure states. In our setting this means checking that every element of $\mathsf{Eff}_+\,(\mathrm{S})\,/\boldsymbol{\alpha}$ arises as a conical combination[4] of the equivalence classes $[a_i]$ of the effects that distinguish the pure states α_i in $\boldsymbol{\alpha}$. Consider a generic element ξ in $\mathsf{Eff}_+\,(\mathrm{S})$, and let us show that it is in the same equivalence class as $\xi' = \sum_{i=1}^{d} \lambda_i a_i$, where $\lambda_i = (\xi|\alpha_i)$ for all i. By linearity, to check the equivalence of two elements of $\mathsf{Eff}_+\,(\mathrm{S})$, it is enough to check that they produce the same numbers when applied to all pure states α_j. Now,

$$\left(\xi'\middle|\alpha_j\right) = \sum_{i=1}^{d} \lambda_i \left(a_i\middle|\alpha_j\right) = \lambda_j = \left(\xi\middle|\alpha_j\right)$$

This shows that the restricted effect cone $\mathsf{Eff}_+\,(\mathrm{S})\,/\boldsymbol{\alpha}$ of the sub-theory is actually a classical effect cone, generated by the effects that distinguish the pure states in $\boldsymbol{\alpha}$. This shows that the formalism of constrained systems works well also for the study of the emergence of classicality in GPTs.

So far, we have defined effective systems at the single-system level. In order to complete our picture, we need to define the composition of constrained systems too. Consider two effective systems $\mathrm{A} := \left(\mathrm{S_A}, \{\mathcal{L}_{\mathrm{A},i}\}_{i=1}^{n}\right)$ and $\mathrm{B} := \left(\mathrm{S_B}, \left\{\mathcal{L}_{\mathrm{B},j}\right\}_{j=1}^{m}\right)$. A natural way to define the effective composite system AB is to select the states of the unconstrained composite system $\mathrm{S_A S_B}$ that satisfy *both* constraints—i.e. to select the density matrices $\rho_{\mathrm{S_A S_B}}$ such that

$$\begin{cases} \left(\mathcal{L}_{\mathrm{A},i} \otimes \mathcal{I}_{\mathrm{S_B}}\right)\left(\rho_{\mathrm{S_A S_B}}\right) = 0 \\ \left(\mathcal{I}_{\mathrm{S_A}} \otimes \mathcal{L}_{\mathrm{B},j}\right)\left(\rho_{\mathrm{S_A S_B}}\right) = 0 \end{cases}, \tag{5.1.1}$$

for all $i \in \{1, \ldots, n\}$, and all $j \in \{1, \ldots, m\}$.

When the effective systems A and B result from an energy constraint, the effective system AB describes a system consisting of two parts, each with its own well-defined energy. In the quantum case, the constraints (5.1.1) amount to $\left(P_{E_\mathrm{A}} \otimes Q_{E_\mathrm{B}}\right) \rho_{\mathrm{S_A S_B}} \left(P_{E_\mathrm{A}} \otimes Q_{E_\mathrm{B}}\right) = \rho_{\mathrm{S_A S_B}}$, where E_A and E_B are the energies of the two local systems, and P_{E_A} and Q_{E_B} are the projectors on the corresponding eigenspaces. Note that this differs from the case

[4] Note that it is not hard to see that $\mathsf{Eff}_+\,(\mathrm{S})\,/\boldsymbol{\alpha}$ is still a cone, with the sum and the multiplication by a scalar inherited from $\mathsf{Eff}_+\,(\mathrm{S})$.

where in the composition of two microcanonical systems only a restriction on the *global* energy is imposed. In this latter case, the states satisfy the *weaker* condition $\Pi_{E_A+E_B}\rho_{S_AS_B}\Pi_{E_A+E_B} = \rho_{S_AS_B}$, where $\Pi_{E_A+E_B}$ is the projector on the eigenspace of $H_{S_A} + H_{S_B}$ with eigenvalue $E_A + E_B$.

The reason why we adopt eq. (5.1.1) as the rule of composition of effective systems is because we want to keep the systems A and B as independently constrained systems, which can be addressed separately on their own, even in a composite setting, still retaining their status of effective systems. This property is particularly important if we want to develop an operational theory of effective systems, so that we can treat them as actual physical systems, forgetting that they arise from constraints imposed on a physical system. In the case of the effective qubits arising from photon polarisation, this fact is particularly apparent. Consider two photons of different spatial modes, with wave vectors $\mathbf{k}_A$ and $\mathbf{k}_B$. If we put a constraint on the *total* energy of the two photons, one of the allowed states would be

$$|\Psi\rangle = \frac{1}{\sqrt{2}} \left(|\mathbf{k}_A, H, 2\rangle \, |\mathbf{k}_B, H, 0\rangle + |\mathbf{k}_A, H, 0\rangle \, |\mathbf{k}_B, H, 2\rangle \right),$$

where the third entry denotes the number of photons, but this cannot be interpreted as a state of two single photons, because the number of photons is undefined. Imposing the constraint (5.1.1) avoids this problem. For this reason, we will use the notation AB for effective systems defined by the constraint (5.1.1).

In summary, given a theory and a set of constraints composed as in eq. (5.1.1), one can build a new *effective theory*, which consists only of effective systems. In the microcanonical case, we can build an effective theory where every system has definite energy, and where all subsystems of a composite systems have definite energy too. For a given system A in such a theory, all the states in $\mathsf{St}(\mathrm{A})$ have the same energy by construction. Likewise, all the transformations in $\mathsf{Transf}(\mathrm{A})$ will be energy-preserving. For every pair of systems A and B, the composite system AB consists of two parts, each of which with its own, well-defined energy. The joint transformations in $\mathsf{Transf}(\mathrm{AB})$ will be interpreted as operations that preserve the energy of the first part *and* the energy of the second part.

The advantage of this effective description is that we need not specify the constraints: in principle, *every* linear constraint can fit into the framework. In this way, we can circumvent the thorny issue of defining the

notion of Hamiltonian in GPTs [176] (cf. section 5.6): in the effective description, we can simply regard each effective system as a system with trivial Hamiltonian, which assigns the same energy to all states of the system. Moreover, since effective system have the same operational structure of unconstrained ones, we can forget we are dealing with constrained systems, and treat them as ordinary ones. This is why we call them "effective systems". When dealing with microcanonical systems, one should always bear in mind that they arise from a constraint on the energy of the system, but for simplicity we will not mention this explicitly henceforth.

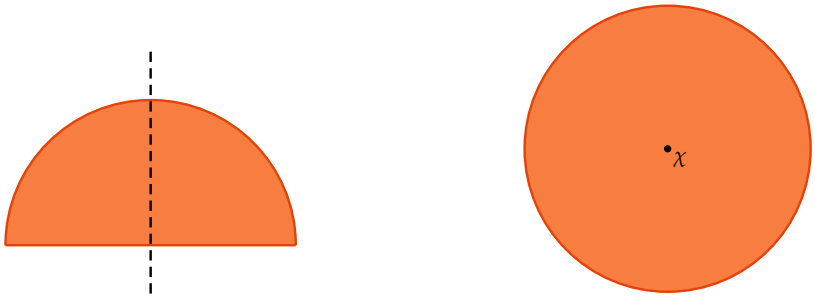
5.1.2 The principle of equal a priori probability

Let us now move to the equal a priori probability principle, according to which one should assign the same probability to all the microstates of the system compatible with a given macroscopic constraint. But what are the microstates? To understand it, let us see what happens in classical and quantum theory. In classical theory, they are the points in the phase space; in quantum theory they are vectors in the Hilbert space. In both cases they are normalised pure states. Therefore it is natural for us to define microstates as *normalised pure states*, representing those preparations of the system that are both deterministic and maximally fine-grained. Then the principle of equal a priori probability states that the system should be described by a uniform mixture of all deterministic pure states satisfying the constraint. For example, the microcanonical state of a finite-dimensional quantum system at energy E is described by the density matrix $\chi_E := \int_{S_E} p_E \, |\psi\rangle \langle\psi| \, \mathrm{d}\psi$, where S_E is the set of pure states in the eigenspace corresponding to the eigenvalue E, and $p_E \mathrm{d}\psi$ is the uniform probability measure over S_E. In the effective picture, where we call this subspace A, the microcanonical state is nothing but the *maximally mixed state* $\chi_{\mathrm{A}} := \int |\psi\rangle \langle\psi| \, \mathrm{d}\psi$ where $\mathrm{d}\psi$ is the uniform probability measure over the pure states of the system.

In GPTs the key problem is to define what we mean by "equal a priori probability", by which we could define the microcanonical state as

$$\chi := \int \psi \, \mathrm{d}\psi. \tag{5.1.2}$$

In quantum mechanics there is a canonical choice: the unitarily invariant probability measure on the pure states of the system. The natural exten-



(a) (b)

Figure 5.1: Two different state spaces. In fig. 5.1a, pure states form a half-circle. Owing to the limited symmetry of the state space, there is no canonical notion of equal a priori probability on the set of pure states. As such, there are a lot of invariant states: all the points of the state space on the symmetry axis. For the set in fig. 5.1b, pure states form a circle, and the notion of uniform probability distribution is uniquely defined. This means that there is a unique microcanonical state χ, which is the only invariant state under all the symmetries of the disk.

sion to GPTs is to consider the probability measures that are invariant under all reversible channels. The problem is, however, that, in general, there may be more than one invariant probability measure, as illustrated in the following example.

Example 5.1.4. Consider a system where the state space is a half-disk, like in fig. 5.1a. Pure states are the states on the half-circle, and they can be parametrised by a polar angle $\vartheta \in [0, \pi]$. Now, reversible channels must be symmetry transformations of the state space. For the half-disk, the only symmetry transformations are the identity and the reflection across the vertical symmetry axis (the black dashed line in fig. 5.1a). Hence, every probability density function that assigns the same value to the points ϑ and $\pi - \vartheta$ is guaranteed to be invariant under reversible channels. This means that the "equal a priori probability" is *not* unique.

The situation is different if the state space of the system is a full disk, as illustrated in fig. 5.1b, corresponding to a rebit, a two-level system in real quantum mechanics [106–109]. In this case, every rotation of the disk is (at least in principle) a reversible channel of the system. The invariant probability measure is unique and given by the probability density function

$p(\vartheta) = \frac{1}{2\pi}$.

This example shows that there exist probabilistic theories where the notion of "equal a priori probability" on pure states is not uniquely defined. The physical implications of this is that in general there is *not* a unique microcanonical state. Indeed, by eq. (5.1.2), $\mathrm{d}\psi$ is that invariant probability measure on pure states, therefore different choices of $\mathrm{d}\psi$ correspond to different microcanonical states. This is of course in contrast with the usual statistical mechanical treatment, which assumes that the macroscopic constraint is enough to determine the equilibrium state. In other words, there are no further constraints coming from additional conserved quantities that restrict the evolution of the system in the set of pure states. In order to formulate the principle of equal a priori probability, we introduce the following condition.

Condition 5.1.5. For every (finite-dimensional) system there exists a *unique* invariant probability measure on normalised pure states.

We saw that some systems do not admit a unique invariant probability measure, like the one in fig. 5.1a. However, are we sure that at least one invariant probability measure always exist? The answer is positive, and now we will show how to construct it. Recall that the group of reversible channels G is a compact group (cf. subsection 2.3.2), and that compact groups admit a finite Haar measure h, which can be renormalised so that $h(G) = 1$ (see subsection 2.3.2). Now, there is a (continuous) group action of G on normalised pure states $\cdot : G_{\mathrm{A}} \times \mathsf{PurSt}_1(\mathrm{A}) \to \mathsf{PurSt}_1(\mathrm{A})$, given by $\mathcal{U}\psi$, for every $\mathcal{U} \in G_{\mathrm{A}}$, and every $\psi \in \mathsf{PurSt}_1(\mathrm{A})$. The idea is to induce a probability measure on the set of normalised pure states from the Haar probability measure on G_{A}. To do this, let us *fix* a normalised pure state ψ_0 in the action of G_{A}, and consider the function $F_{\psi_0} : G_{\mathrm{A}} \to \mathsf{PurSt}_1(\mathrm{A})$ such that $\mathcal{U} \mapsto \mathcal{U}\psi_0$. Since F_{ψ_0} is continuous, we can induce a probability measure μ_{ψ_0} on $\mathsf{PurSt}_1(\mathrm{A})$ (called the "image measure" [139]) by setting

$$\mu_{\psi_0}(S) := h\left(F_{\psi_0}^{-1}(S)\right)$$

for every Borel subset of $\mathsf{PurSt}_1(\mathrm{A})$. Let us show that this probability measure μ_{ψ_0} is invariant under the action of G. We have $\mu_{\psi_0}(\mathcal{U}S) = h\left(F_{\psi_0}^{-1}(\mathcal{U}S)\right)$, where $F_{\psi_0}^{-1}(\mathcal{U}S) = \{\mathcal{V} \in G : \mathcal{V}\psi_0 \in \mathcal{U}S\}$. Now $F_{\psi_0}^{-1}(\mathcal{U}S) =$

$\mathcal{U}F_{\psi_0}^{-1}(S)$. To see it, take $\mathcal{V} \in G$ such that $\mathcal{V}\psi_0 \in \mathcal{U}S$. Now, this implies that $\mathcal{U}^{-1}\mathcal{V}\psi_0 \in S$, so $\mathcal{U}^{-1}\mathcal{V} \in F_{\psi_0}^{-1}(S)$, and clearly $\mathcal{V} \in F_{\psi_0}^{-1}(\mathcal{U}S)$ can be written as $\mathcal{U}$ times an element in $F_{\psi_0}^{-1}(S)$. Therefore $F_{\psi_0}^{-1}(\mathcal{U}S) \subseteq \mathcal{U}F_{\psi_0}^{-1}(S)$. To show the other inclusion, take $\mathcal{V} \in G$ such that $\mathcal{V}\psi_0 \in S$. Clearly $\mathcal{U}\mathcal{V}\psi_0 \in \mathcal{U}S$, so $F_{\psi_0}^{-1}(\mathcal{U}S) \supseteq \mathcal{U}F_{\psi_0}^{-1}(S)$, and therefore $F_{\psi_0}^{-1}(\mathcal{U}S) = \mathcal{U}F_{\psi_0}^{-1}(S)$. In conclusion

$$\mu_{\psi_0}(\mathcal{U}S) = h\left(\mathcal{U}F_{\psi_0}^{-1}(S)\right) = h\left(F_{\psi_0}^{-1}(S)\right) = \mu_{\psi_0}(S),$$

where we have used the invariance of the Haar probability measure h. This shows that the image probability measure μ_{ψ_0} on $\mathsf{PurSt}_1(\mathrm{A})$ is invariant under reversible channels. This means that there *always* exists an invariant probability measure on normalised pure states.

Condition 5.1.5 enforces the uniqueness of such a measure. By its very definition μ_{ψ_0} depends on the pure state ψ_0 we fix, so in general we expect that different choices of ψ_0 will result in different invariant probability measures on the set of normalised pure states. Let us examine this issue in greater detail. Consider now ψ_0' such that $\psi_0' = \mathcal{U}\psi_0$ for some $\mathcal{U} \in G$. What is the relationship between μ_{ψ_0} and $\mu_{\psi_0'}$? To answer the question, we need to understand how to write $F_{\psi_0'}^{-1}(S)$ in terms of $F_{\psi_0}^{-1}(S)$ for any Borel set S. We have $F_{\psi_0'}^{-1}(S) = F_{\psi_0}^{-1}(S)\mathcal{U}^{-1}$. To see it, consider $\mathcal{V} \in G$ such that $\mathcal{V}\psi_0' \in S$, which means $\mathcal{V}\mathcal{U}\psi_0 \in S$. Therefore $\mathcal{V}$ can be written as an element of $F_{\psi_0}^{-1}(S)$ (i.e. $\mathcal{V}\mathcal{U}$) times $\mathcal{U}^{-1}$. This shows $F_{\psi_0'}^{-1}(S) \subseteq F_{\psi_0}^{-1}(S)\mathcal{U}^{-1}$. To prove the other inclusion, consider $\mathcal{V} \in G$ such that $\mathcal{V}\psi_0 \in S$, and let us show that $\mathcal{V}\mathcal{U}^{-1}$ is such that $\mathcal{V}\mathcal{U}^{-1}\psi_0' \in S$. This is immediate, as $\psi_0' = \mathcal{U}\psi_0$. This proves that $F_{\psi_0'}^{-1}(S) \supseteq F_{\psi_0}^{-1}(S)\mathcal{U}^{-1}$, so $F_{\psi_0'}^{-1}(S) = F_{\psi_0}^{-1}(S)\mathcal{U}^{-1}$. Hence

$$\mu_{\mathcal{U}\psi_0}(S) = h\left(F_{\mathcal{U}\psi_0}^{-1}(S)\right) = h\left(F_{\psi_0}^{-1}(S)\mathcal{U}^{-1}\right) = h\left(F_{\psi_0}^{-1}(S)\right) = \mu_{\psi_0}(S),$$

where we have used the invariance properties of the Haar probability measure once more. This shows that pure states in the same orbit of the group action generate the same invariant probability measure. Therefore having a unique orbit (i.e. a transitive action) is a *sufficient* condition to have a unique invariant probability measure on normalised pure states (and hence a unique microcanonical state).

This condition is *necessary* too. To prove it, suppose there is more than one orbit, and take a pure state ψ_0 in an orbit, and another ψ_0' in a different one. Let us show that the induced probability measures are different. Take S to be the orbit $G\psi_0$ of ψ_0; since the action is continuous, this is a closed set, hence a Borel set. Then $F_{\psi_0}^{-1}(G\psi_0) = \{\mathcal{U} \in G : \mathcal{U}\psi_0 \in G\psi_0\} = G$, by definition of orbit of ψ_0. Hence $\mu_{\psi_0}(G\psi_0) = h(G) = 1$. On the other hand, for the same orbit $G\psi_0$, we have $F_{\psi_0'}^{-1}(G\psi_0) = \{\mathcal{U} \in G : \mathcal{U}\psi_0' \in G\psi_0\} = \varnothing$, because orbits are disjoint sets, so there are no elements in the orbits of ψ_0' also in the orbit of ψ_0. Hence $\mu_{\psi_0'}(G\psi_0) = h(\varnothing) = 0$. In conclusion, we have found a set to which the probability measures generated by ψ_0 and ψ_0' assign different values, therefore they are different measures.[5]

To summarise: an invariant probability measure on $\mathsf{PurSt}_1(\mathrm{A})$ always exists, and it is unique if and only if the action of reversible channels is transitive on $\mathsf{PurSt}_1(\mathrm{A})$. We collect all these remarks in the following theorem.

Theorem 5.1.6. *For every finite system* A*, the following are equivalent:*

1. *Condition 5.1.5 is satisfied.*
2. *For every pair of normalised pure states, there exists a reversible channel connecting them.*

Note that enforcing the uniqueness of the invariant probability measure on normalised pure states leads to a non-trivial requirement—transitivity—, which has appeared several times, either directly or indirectly, in various reconstructions of quantum theory [61, 68, 70, 83, 84, 86]. Theorem 5.1.6 provides one more motivation for this condition, this time from a completely different perspective, a thermodynamic one. The transitivity requirement already appeared in [100] as a requirement for a "canonical resource theory of purity" in the context of the duality between pure-state entanglement and purity in GPTs.

Note that transitivity means that, at least in principle, the dynamics allowed by the physical theory enable one to explore the entire set of nor-

[5] In fact, we can say even more: the probability measures associated with different orbits are *mutually singular* [139]: there exist two complementary subsets of $\mathsf{PurSt}_1(\mathrm{A})$, one with zero probability for one measure, and the other with zero probability for the other. This means that the two probability measures "live on disjoint sets".

malised pure states compatible with the macroscopic constraint on the energy, a fact that is often assumed or remarked in textbook presentations about the microcanonical ensemble [7].

We finish this subsection by noting that sharp theories with purification, having a transitive action (see section 4.1), satisfy condition 5.1.5. With these theories we are on the right track to have a sensible microcanonical thermodynamics.

5.1.3 The microcanonical state

Once we know that condition 5.1.5 holds in a physical theory, we can use eq. (5.1.2) to construct the microcanonical state, where now $\mathrm{d}\psi$ is the unique invariant probability measure over the normalised pure states of a system. The convexity of the state space guarantees that the microcanonical state is indeed a state. Moreover, since the state space is finite-dimensional, thanks to Carathéodory's theorem for convex geometry [140, 141], it is possible to replace the integral in eq. (5.1.2) with a *finite* sum. This means that the microcanonical state can (in principle) be generated by picking pure states at random from a finite set.

The following proposition highlights two important properties of the microcanonical state.

Proposition 5.1.7. *The microcanonical state χ satisfies the following two properties:*

1. *it is invariant under all reversible channels of the system;*
2. *it can be generated from any normalised pure state by a fixed random reversible dynamic.*

Proof. Let us prove the two properties.

1. For every reversible channel $\mathcal{U}$, one has

$$\mathcal{U}\chi = \int \mathcal{U}\psi \,\mathrm{d}\psi = \int \psi' \,\mathrm{d}\left(\mathcal{U}^{-1}\psi'\right) = \int \psi' \,\mathrm{d}\psi' = \chi, \tag{5.1.3}$$

where we have defined $\psi' := \mathcal{U}\psi$, and we have used the invariance of the probability measure $\mathrm{d}\psi$.

2. Consider the transformation $\mathcal{T} = \int_G \mathcal{U}\, \mathrm{d}\mathcal{U}$, where $\mathrm{d}\mathcal{U}$ is the Haar probability measure on the group of reversible channels. $\mathcal{T}$ is a channel because
$$u\mathcal{T} = \int_G u\mathcal{U}\, \mathrm{d}\mathcal{U} = u \int_G \mathrm{d}\mathcal{U} = u.$$
The channel $\mathcal{T}$ maps every normalised pure state ψ to the microcanonical state: indeed, one has
$$\mathcal{T}\psi = \int_G \mathcal{U}\psi\, \mathrm{d}\mathcal{U} = \chi,$$
like in eq. (5.1.3). Since we are working with finite-dimensional systems, the integral in the definition of $\mathcal{T}$ can be replaced by a *finite* convex combination of reversible channels, by Carathéodory's theorem. Clearly, this writing of $\mathcal{T}$ does *not* depend on the pure state ψ it is applied to. Therefore, there is a fixed convex combination of reversible channels—a random reversible dynamic—such that when it is applied to any pure state, it yields the microcanonical state.

□

Property 1 expresses the fact that the microcanonical state is an *equilibrium state*, in the sense that it does not evolve under any of the reversible dynamics compatible with the macroscopic constraint. Note that the notion of equilibrium here is different from the notion of thermal equilibrium, which refers to interactions with an external bath. This is rather a *dynamical equilibrium*: the probability assignments to pure states made in the definition of the microcanonical state are stable under all possible evolutions of the system.

By condition 5.1.5 we are dealing with GPTs with transitive action of reversible channels on pure states, and they have a unique invariant state (cf. proposition 2.3.27). Thus property 1 identifies microcanonical states with the invariant states of those GPTs: looking for the microcanonical state is equivalent to searching for the invariant state. In the following, given the overall thermodynamic character of this chapter we will stick to "microcanonical states" as terminology.

Example 5.1.8. In fig. (5.1b) we showed the state space of a rebit, a disk, which admits a unique invariant probability density function on pure states.

In this case the microcanonical state is the unique invariant state under all the symmetries of the disk, namely its centre.

In any sharp theory with purification, given any pure maximal set $\{\alpha_i\}_{i=1}^d$ the microcanonical state is the invariant state, diagonalised as

$$\chi = \frac{1}{d} \sum_{i=1}^{d} \alpha_i. \tag{5.1.4}$$

In quantum theory, the microcanonical state is the maximally mixed state $\frac{1}{d}\mathbf{1}$.

It is worth noting that eq. (5.1.4) is often taken as the *definition* of microcanonical state in quantum statistical mechanics [6, 7]. According to it, the microcanonical state is defined as the uniform mixture of orthonormal basis states of the eigenspace of the fixed energy. Proposition 4.4.5 shows that a similar definition is possible in every sharp theory with purification. One might be tempted to use eq. (5.1.4) to define the microcanonical state in *arbitrary* physical theories. However, the fact that the state resulting from the uniform mixture of the α_i's is independent of the choice of maximal set is not guaranteed to hold in every theory. Moreover, in general there is no relationship between the writing of eq. (5.1.4) and that of eq. (5.1.2). For this reason, we prefer to stick to eq. (5.1.2), and define the microcanonical state as the uniform mixture of *all* pure states with a given energy, rather than the uniform mixture of a particular pure maximal set. From a physical angle, the uniform mixture of all pure states represents the result of fully uncontrolled, but energy conserving fluctuations in the experimental setup. This describes the situation of a total lack of knowledge besides the knowledge of the value of the energy: not even the "energy eigenbasis"—the pure maximal set $\{\alpha_i\}_{i=1}^d$—is known.

Property 2 instead refers to the fact that the system can, at least in principle, be brought to equilibrium, so there exists a sort of thermalisation process. Physically, the fact that this process is a random reversible one can be interpreted as a situation where the experimenter has no full control on the preparation, but has control on the dynamics of the system, maybe through some classical control fields [100]. In this picture, property 2 guarantees that the experimenter can prepare the microcanonical state by drawing the parameters of their control fields at random. Further along this line, one can also imagine scenarios where the randomisation

occurs naturally as a result of fluctuations of the fields. Property 2 is important for the resource-theoretic approach (section 5.2), because it guarantees that the microcanonical state is "easy to prepare", therefore it can be rightfully considered a free state. Indeed we can reach the microcanonical state χ from *any* state ρ, even from mixed states. This follows from the linearity of $\mathcal{T}$, because every ρ can be written as a convex combination of pure states; therefore $\mathcal{T}\rho = \chi$.

5.1.4 Composition of microcanonical states

At the level of single systems, condition 5.1.5 guarantees the existence of a microcanonical state. But how does the microcanonical state behave under the composition of systems? It is important to answer this question because, from the operational point of view, it is natural to consider scenarios where the experimenter has more than one system at their disposal. This aspect is not so stressed in traditional textbook presentations of statistical mechanics, where the composition of microcanonical systems (done fixing only the global energy, therefore differently from us) is only mentioned to show that in the thermodynamic limit if a composite system is in the microcanonical state, so are its components [6, 7].

Here we want to take a closer look at the composition of microcanonical systems, according the rules of eq. (5.1.1). Composition is especially important in the context of resource theories, where it is natural to ask how resources interact when combined together. Think for example of the quantum resource theory of noisy operations [36, 37, 177]. There microcanonical states are treated as free. Since the experimenter can generate the microcanonical states χ_A and χ_B at no cost, then they can generate the product state $\chi_A \otimes \chi_B$ at no cost too.

If we insist that microcanonical states are the *only* free states in a resource-theoretic treatment of microcanonical thermodynamics, the product state $\chi_A \otimes \chi_B$ of two microcanonical states must be the microcanonical state χ_{AB} of the composite system AB. This is consistent with the intuitive interpretation of the microcanonical state as "the state of minimum information compatibly with the constraints". In other words, if one has minimum information on the parts of a system, then one has minimum information about the whole. This is indeed the case in quantum theory, where the product of two maximally mixed states is maximally mixed:

$\frac{1}{d_{\mathrm{A}}}\mathbf{1} \otimes \frac{1}{d_{\mathrm{B}}}\mathbf{1} = \frac{1}{d_{\mathrm{AB}}}\mathbf{1}$.

Recall that here we are dealing with effective systems, with their corresponding constraints. For energy constraints, the composite of two effective systems A and B is defined as a system consisting of two parts, each constrained to a specific value of the energy. Consistently with this interpretation, the microcanonical state of system AB is the "maximally mixed state" in the set of states with fixed local energies. It is therefore natural to require that minimum information about the parts should imply minimum information about the whole.

Condition 5.1.9. The microcanonical state of a composite system is the product of the microcanonical states of its components. In formula:

$$\chi_{\mathrm{A}} \otimes \chi_{\mathrm{B}} = \chi_{\mathrm{AB}}, \tag{5.1.5}$$

for every pair of effective systems A and B.

We call eq. (5.1.5) the *condition of informational equilibrium*. Note that, again, here we are not referring to thermal equilibrium between the two subsystems. This is clear from the fact that we do not allow an energy flow between the two systems A and B. Instead, we allow a flow of information, implemented by the joint dynamics of the composite system AB.

It is natural to ask which physical principles guarantee the condition of informational equilibrium. One such principle is Local Tomography, as shown in [67, 84]. However, Local Tomography is not necessary for informational equilibrium. Indeed quantum theory on real Hilbert spaces violates Local Tomography, but still satisfies the condition of informational equilibrium. As already stressed, in our analysis we will *not* assume Local Tomography in our set of physical principles. Nevertheless, our principles of sharp theories with purification guarantee the validity of the condition of informational equilibrium, which is the really important condition to set up a sensible theory of (microcanonical) thermodynamics.

Example 5.1.10. Sharp theories with purifications satisfy the condition of informational equilibrium [105]. To prove it, consider two systems A and B, and pick two pure maximal sets for A and B respectively, say $\{\alpha_i\}_{i=1}^{d_{\mathrm{A}}}$ and $\{\beta_j\}_{j=1}^{d_{\mathrm{B}}}$. Then, the product set $\{\alpha_i \otimes \beta_j\}_{i\in\{1,\dots,d_{\mathrm{A}}\}, j\in\{1,\dots,d_{\mathrm{B}}\}}$ is maximal

for the composite system AB, by proposition 4.4.8. Using the decomposition (5.1.4), we obtain

$$\chi_{\mathrm{AB}} = \frac{1}{d_{\mathrm{AB}}} \sum_{i=1}^{d_{\mathrm{A}}} \sum_{j=1}^{d_{\mathrm{B}}} \alpha_i \otimes \beta_j = \frac{1}{d_{\mathrm{A}} d_{\mathrm{B}}} \left(\sum_{i=1}^{d_{\mathrm{A}}} \alpha_i \right) \otimes \left(\sum_{i=1}^{d_{\mathrm{B}}} \beta_j \right) = \chi_{\mathrm{A}} \otimes \chi_{\mathrm{B}},$$

where we have used the information locality condition $d_{\mathrm{AB}} = d_{\mathrm{A}} d_{\mathrm{B}}$ (again proposition 4.4.8). In summary, sharp theories with purification satisfy our two conditions for the general microcanonical framework.

Now we are ready to extend the microcanonical framework from quantum and classical theory to general physical theories.

Definition 5.1.11. An OPT, interpreted as a theory of effective systems, is *microcanonical* if conditions 5.1.5 and 5.1.9 are satisfied.

Physically, a microcanonical theory is a theory where:

1. every system has a well-defined notion of uniform mixture of all pure states;

2. uniform mixtures are stable under parallel composition of systems.

Microcanonical theories provide the foundation for the definition of three important resource theories, analysed in the following sections.

5.2 Three microcanonical resource theories

In this section we study three different notions of state convertibility in microcanonical theories. We adopt the resource-theoretic framework of chapter 3, where one fixes a set of *free operations*, closed under sequential and parallel composition. As we saw in section 3.2, the basic question in the resource-theoretic framework is whether a given state ρ can be transformed into another state σ by means of free operations. This gives rise to a preorder between states, denoted as[6] $\rho \succeq_F \sigma$, where F is the set of free operations. In the following we define three resource theories for microcanonical thermodynamics.

[6]Here we adopt the notation $\succeq_F$ for the preorder, instead of the notation $\succsim$ of section 3.2, because of the analogy with the notation for majorisation, as we will become clear in section 5.3.

5.2.1 The RaRe resource theory

Our first resource theory takes random reversible channels [100, 178–181] as free operations.

Definition 5.2.1. A *random reversible (RaRe) channel* on system A is a channel $\mathcal{R}$ of the form $\mathcal{R} = \sum_i p_i \mathcal{U}_i$, where $\{p_i\}$ is a probability distribution and $\mathcal{U}_i$ is a reversible channel on system A for every i.

Physically, RaRe channels are the operations that can be implemented with limited control over the reversible dynamics of the system: if the dynamics are subject to random fluctuations, the lack of control on these fluctuations gives rise to a RaRe channel.

We already encountered a RaRe channel in proposition 5.1.7 as the channel enacting the "thermalisation" process, i.e. mapping every pure state to the microcanonical state. From this point of view, it is fairly natural to take RaRe channels as free operations. Moreover, from a more mathematical perspective, RaRe channels have all the properties required of free operations: the identity channel is RaRe, the sequential composition of two RaRe channels is a RaRe channel, and so is the parallel composition. We call the resulting resource theory the *RaRe resource theory* and we denote the corresponding preorder by $\succeq_{\text{RaRe}}$.

Note that, in principle, the RaRe resource theory can be formulated in *every* GPT, even those that do *not* satisfy conditions 5.1.5 and 5.1.9, because there is no microcanonical state explicitly involved in the definition of RaRe channels. Such generality, however, comes at a price: from the structure of RaRe channels we *cannot* infer free states by taking the input system of a free operation to be the trivial system. Indeed, RaRe channels have the same input and output system, so taking the input system to be trivial makes the channel itself trivial. Therefore, strictly speaking, the RaRe resource theory has no free states. As explained in section 3.2, if we want to add free states, we have to look for almost free states. Now, assume the theory on which RaRe channels are defined satisfies condition 5.1.5 and 5.1.9.

By proposition 5.1.7 we know that $\mathcal{T}\rho = \chi$ for every ρ, where $\mathcal{T}$ is the RaRe thermalisation channel, and χ the microcanonical state. This clearly means that $\rho \succeq_{\text{RaRe}} \chi$ for every ρ. Therefore all states equivalent to χ in the preorder $\succeq_{\text{RaRe}}$ will be the minima as resources. Now, for a state ρ to be equivalent to χ, it must be $\chi \succeq_{\text{RaRe}} \rho$. This means that there exists a

RaRe channel $\mathcal{R}$ such that $\mathcal{R}\chi = \rho$. Since χ is an invariant state, $\mathcal{R}\chi = \chi$, therefore there is only χ in its equivalence class. This means that there is a unique almost free state, the microcanonical state χ. It is stable under tensor product, so the construction presented in section 3.2 tells us that in microcanonical theories we can rightfully consider the microcanonical state as a free state, even though we cannot derive it from the structure of free operations [35].

On the other extreme we have pure states, which are the maximum resource. First of all, they are all equivalent resources thanks to transitivity: reversible channels are a special case of RaRe channels. Then take any pure state ψ, and any state ρ. ρ can always written as a (possibly trivial) convex combination of pure states ψ_i: $\rho = \sum_i p_i \psi_i$. Because of the transitive action of reversible channels on pure states, there exist some reversible channels $\mathcal{U}_i$ such that $\psi_i = \mathcal{U}_i \psi$, whence ρ can be written as

$$\rho = \sum_i p_i \mathcal{U}_i \psi.$$

This means that $\psi \succeq_{\text{RaRe}} \rho$ for any pure state ψ and any state ρ.

5.2.2 The noisy resource theory

Whilst the RaRe resource theory can be defined in every OPT, we now discuss a second resource theory that can only be defined in physical theories satisfying conditions 5.1.5 and 5.1.9. In this resource theory, free operations mimic a thermalisation process, and are generated by the following three processes:

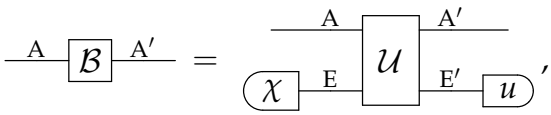
1. bringing in an ancillary system in the microcanonical state;

2. letting the target system and the "thermal" ancilla jointly evolve together;

3. removing the "thermal" ancilla.

More precisely, these operations, usually called "noisy" [36, 37, 177], are defined as follows.

Definition 5.2.2. A channel $\mathcal{B}$, from system A to system A′, is a *basic noisy operation* if it can be written as[7]



(5.2.1)

where E and E′ are suitable systems such that AE $\approx$ A′E′, and $\mathcal{U}$ is a reversible channel.

Note that, if we take A to be the trivial system, harnessing the invariance of the microcanonical state under reversible channels, we obtain that the microcanonical state is a free state from the very expression of basic noisy operations.

Definition 5.2.2, unlike RaRe channels, does not rely on the availability of external sources of randomness: all the randomness is accounted for in the preparation of the microcanonical state in the right-hand side of eq. (5.2.1). In other words, an external microcanonical state becomes the source of "thermalisation" for the target system, instead of the random fluctuations of reversible dynamics in the case of the RaRe resource theory.

Definition 5.2.2 uses the term "basic noisy operation" instead of the customary "noisy operation" to account for a mathematical subtlety which will arise in subsection 5.2.4: the set of basic noisy operations is generally *not* topologically closed. In quantum theory, for example, there exist counterexamples where the limit of a sequence of basic noisy operations is not a basic noisy operation [182]. It is then convenient to take the closure of the set of basic noisy operations.

Definition 5.2.3. A channel $\mathcal{N}$ is a *noisy operation* if it is the limit of a sequence of basic noisy operations $\{\mathcal{B}_n\}$.

The set of noisy operations satisfies all the requirements for being a set of free operations: the identity is a noisy operation, and the parallel and sequential composition of two noisy operations are noisy operations, thanks to the condition of informational equilibrium. The resource theory where the set of free operations is the set of noisy operations will be called the *noisy resource theory*. The corresponding preorder on states will be denoted by $\succeq_{\text{noisy}}$.

[7] Again, here Causality is assumed for convenience, it is possible to put forward a definition of noisy operations also in non-causal theories [105].

5.2.3 The unital resource theory

In the third resource theory, the set of free operations includes all the operations that transform microcanonical states into microcanonical states. This is the largest set of free operations compatible with χ being the free state, as described in chapter 3. These channels, called *unital*, are the most general operations that do not create resources out of free states. Mathematically, they are defined as follows.

Definition 5.2.4. A channel $\mathcal{D}$ from system A to system A′ is called *unital* if $\mathcal{D}\chi_{\mathrm{A}} = \chi_{\mathrm{A}'}$.

Unital channels are the operational generalisation of doubly stochastic matrices in classical probability theory [112, 183, 184].

Note that, by condition 5.1.9, all free states of a composite system are of the product form, therefore, focusing on resource-non-generating operations is enough, they are automatically completely resource-non-generating. If A is the trivial system, the preparation of the microcanonical state is a (free) unital channel, so we recover that the microcanonical state is free.

The set of unital channels enjoys all the properties required of a set of free operations: the identity is a unital channel, and thanks to the condition of informational equilibrium, the sequential and parallel composition of unital channels is a unital channel. The resource theory where free operations are unital channels will be called the *unital resource theory*. The corresponding preorder on states will be denoted by $\succeq_{\mathrm{unital}}$.

5.2.4 Inclusion relations

Let us highlight the relations between the three sets of free operations defined so far. First, RaRe channels are examples of unital channels. This is clear because every RaRe channel can be written as a mixture of reversible channels, each of which preserves the microcanonical state.

$$\mathcal{R}\chi = \sum_i p_i \mathcal{U}_i \chi = \sum_i p_i \chi = \chi$$

Hence, we have the inclusion

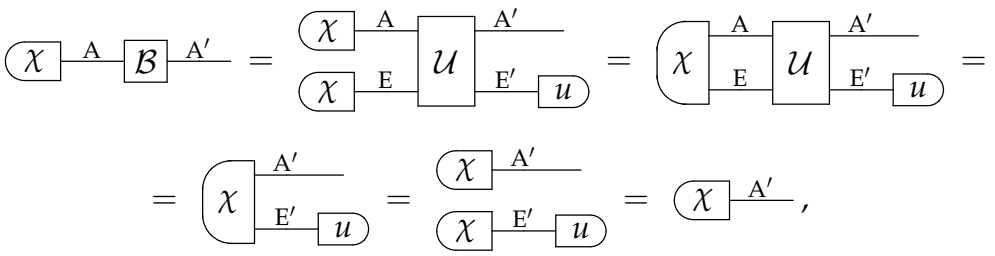
$$\mathsf{RaRe} \subseteq \mathsf{Unital}. \tag{5.2.2}$$

In classical probability theory, the inclusion is actually an equality, as a consequence of Birkhoff's theorem [112, 185]. Remarkably, in quantum theory there exist unital channels that are not random unitary, meaning that the inclusion (5.2.2) is generally strict. The simplest example is due to Landau and Streater [183].

We have seen that all RaRe channels are unital. Noisy operations are unital too.

Proposition 5.2.5. *Every noisy operation is unital.*

Proof. Suppose that $\mathcal{B}$ is a basic noisy operation, written as in eq. (5.2.1). Then, one has



having used the condition of informational equilibrium (5.1.5), the invariance of the microcanonical state χ_{AE} under reversible channels.[8] Hence, every basic noisy operation is unital. Taking the closure, one gets that all noisy operations are unital. □

In summary, one has the inclusion

$$\mathsf{Noisy} \subseteq \mathsf{Unital}. \tag{5.2.3}$$

The inclusion is strict in quantum theory, where Haagerup and Musat found examples of unital channels that cannot be realised as noisy operations [186].

It remains to understand the relation between RaRe channels and noisy operations. In quantum theory, the set of noisy operations (strictly) contains the set of RaRe channels as a proper subset [182]. In a generic theory, however, this inclusion relation need not hold.

[8] To be precise, here we have used the fact that $\mathcal{U}\chi_{\mathrm{AE}} = \chi_{\mathrm{A'E'}}$, taking advantage of the fact that $\mathrm{AE} \approx \mathrm{A'E'}$, so AE and $\mathrm{A'E'}$ can be treated for all practical purposes as though they were the same system.

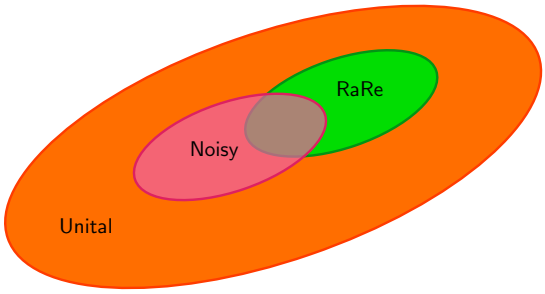
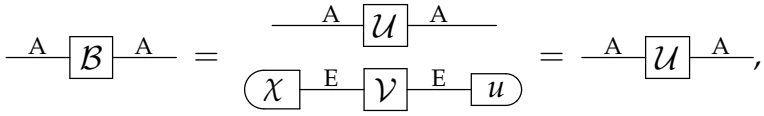


Figure 5.2: The most general inclusions between the three sets of free operations. At this stage we cannot say anything about the intersection between RaRe channels and noisy operations.

Example 5.2.6. As a counterexample, consider a theory where, besides the SWAP, only local reversible channels are allowed, like in PR boxes [187]. In this case, noisy operations on a given system A are just reversible channels. Indeed



where we have used the invariance of χ and the fact that it is a normalised state. In this theory, noisy operations are *strictly* contained in RaRe channels.

The inclusions (5.2.2) and (5.2.3) are the most general result one can derive from the definitions alone. The general situation is depicted in fig. 5.2. To go further, we need to introduce axioms, which we will do in the next subsection for sharp theories with purification.

Inclusion relations in sharp theories with purification

In sharp theories with purification, one can establish an inclusion between RaRe channels and noisy operations, the same we have in quantum theory. To obtain this result, we first restrict our attention to *rational* RaRe channels, i.e. RaRe channels of the form $\mathcal{R} = \sum_i p_i \mathcal{U}_i$ where each p_i is a rational number. With this definition, we have the following lemma.

Lemma 5.2.7. *In every sharp theory with purification, every rational RaRe channel is a basic noisy operation.*

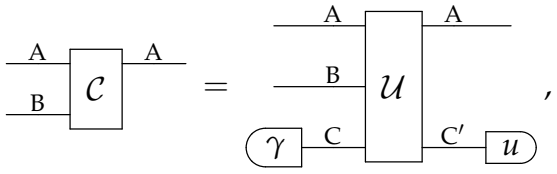
Proof. Let $\mathcal{R}$ be a rational RaRe channel, written as

$$\mathcal{R} = \sum_i \frac{n_i}{n} \mathcal{U}_i,$$

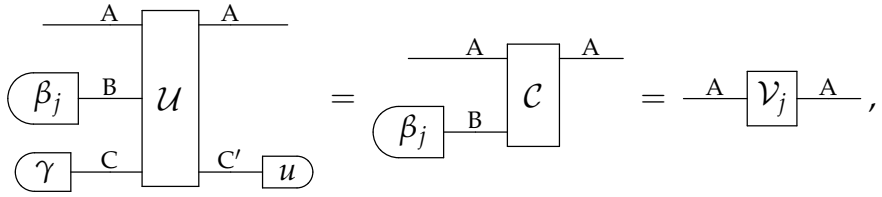
with $n_i \geq 0$ and $\sum_i n_i = n$. Let B be an n-dimensional system, and pick the pure maximal set $\{\beta_j\}_{j=1}^n$. Let $\mathcal{C}$ be the channel from AB to A, defined as

$$\mathcal{C} = \sum_{j=1}^n \mathcal{V}_j \otimes \beta_j^\dagger,$$

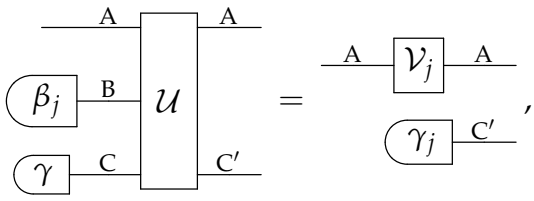
where $\{\mathcal{V}_j\}_{j=1}^n$ are n reversible channels on A, with n_1 of them equal to $\mathcal{U}_1$, n_2 equal to $\mathcal{U}_2$, and so on. Since the theory satisfies Purification, the channel $\mathcal{C}$ has a reversible dilation [67, 71], namely



where C and C′ are suitable systems, γ is a pure state, and $\mathcal{U}$ is a reversible channel. Now, by construction we have

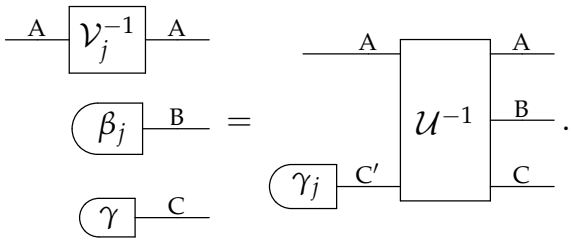


for every $j \in \{1, \ldots, n\}$. Since $\mathcal{V}_j$ is reversible, the joint map must factorise (corollary 4.1.12):



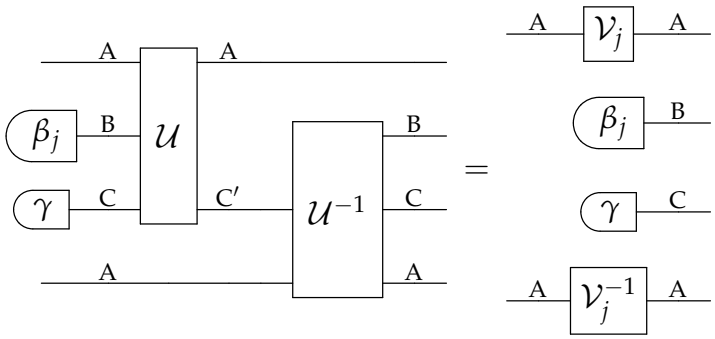
(5.2.4)

for some pure state γ_j of system C′. Composing both sides with $\mathcal{V}_j^{-1}$ on the left, and with $\mathcal{U}^{-1}$ on the right we obtain



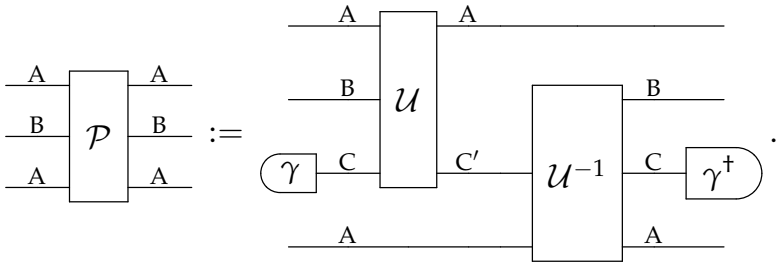
(5.2.5)

Now we can feed the expression for γ_j in terms of $\mathcal{U}$ from eq. (5.2.4) to the input of $\mathcal{U}^{-1}$ in eq. (5.2.5), getting

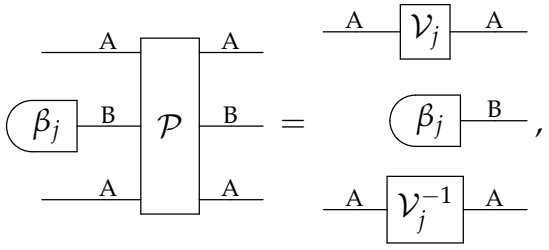


(5.2.6)

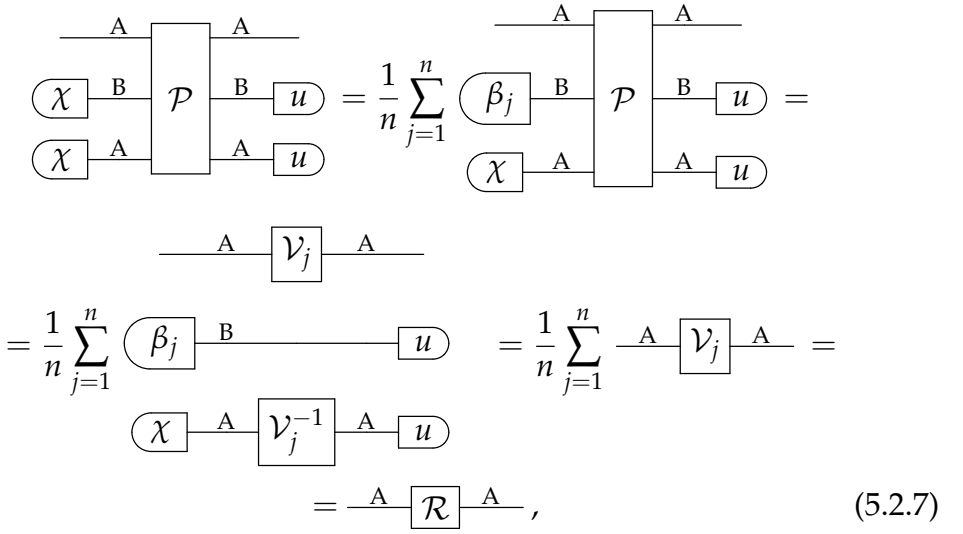
where most of the right-hand side comes from the left-hand side of eq. (5.2.5). At this point, we define the pure transformation



By eq. (5.2.6) $\mathcal{P}$ satisfies the relation

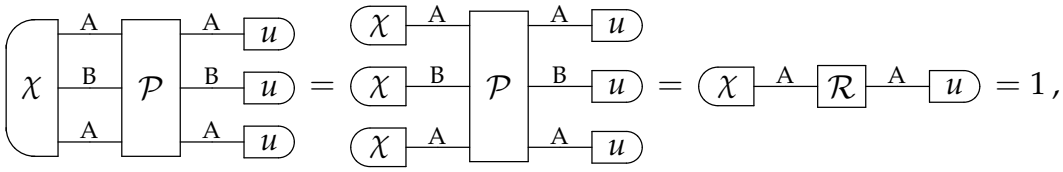


for all values of j. Using this relation and the expression of χ_{B} in terms of the β_x's, we can reconstruct $\mathcal{R}$ from $\mathcal{P}$:



(5.2.7)

where we have used the fact that $\sum_{j=1}^{n} \mathcal{V}_j = \sum_i n_i \mathcal{U}_i$. Finally, let us show that $\mathcal{P}$ is a channel. To this end, it is enough to show that $u\mathcal{P} = u$ (see proposition 2.3.7). This property is satisfied if and only if $(u|\mathcal{P}|\chi) = 1$, because every state lies in some convex decomposition of χ (see proposition 2.3.27). By the condition of informational equilibrium and eq. (5.2.7), we have



so $\mathcal{P}$ is a channel. Since every pure channel on a fixed system (here ABA) is reversible [67], $\mathcal{P}$ is reversible. Hence, eq. (5.2.7) shows that $\mathcal{R}$ is a basic noisy operation, with environment $\mathrm{E} = \mathrm{BA}$, and reversible channel $\mathcal{P}$. □

In quantum theory, this statement is quite immediate, as pointed out in [37]: a generic RaRe channel with rational probabilities $\left\{\frac{n_i}{n}\right\}_{i=1}^{r}$ and unitary gates $\{U_i\}_{i=1}^{r}$ can be realised as the basic noisy operation

$$\mathcal{B}(\rho) := \mathrm{tr}_{\mathrm{anc}}\left[U\left(\rho \otimes \frac{1}{n}\mathbf{1}\right)U^{\dagger}\right],$$

where $\mathrm{tr}_{\mathrm{anc}}$ is the partial trace on the n-dimensional system used as ancilla, and U is the control-unitary gate

$$U := \sum_{j=1}^{n} V_j \otimes |j\rangle \langle j| ,$$

$\{|j\rangle\}_{j=1}^{n}$ being an orthonormal basis for the ancillary system, and $\{V_j\}_{j=1}^{n}$ being a list of unitary gates, n_1 of which are equal to U_1, n_2 equal to U_2, and so on.

The proof of lemma 5.2.7 shows that the situation is in general more complicated in sharp theories with purification. The reason is that the simple construction of quantum theory based on control-unitary channels cannot be carried over. The analogue of the control-unitary U is a control-reversible transformation, which performs a reversible transformation on the target system depending on the state of a control system [114]. However, in section 5.8 we will show that not every sharp theory with purification admits control-reversible transformations. Specifically, the existence of control-reversible transformations is equivalent to a non-trivial property of the dynamics, which we will call "unrestricted reversibility" [105]. The non-trivial content of lemma 5.2.7 is that the inclusion of rational RaRe channels in noisy operations holds *in every sharp theory with purification*, without the need to assume unrestricted reversibility or the existence of control-reversible transformations.

Now, since rational RaRe channels are dense in the set of RaRe channels, and since the set of noisy operations is closed (see definition 5.2.3), we obtain the following theorem.

Theorem 5.2.8. *In every sharp theory with purification, RaRe channels are noisy operations.*

This theorem is important from a physical point of view, for it states that the "thermalisation" RaRe channel in proposition 5.1.7 can be regarded as a physical "thermalisation" process in which a system is put into contact with a "thermal bath" (the microcanonical state), and left there to "thermalise".

Note that the inclusion of RaRe channels in the set of noisy operations is generally strict: for example, in quantum theory there exist noisy operations that are not RaRe channels [182] . In summary, we have the inclu-

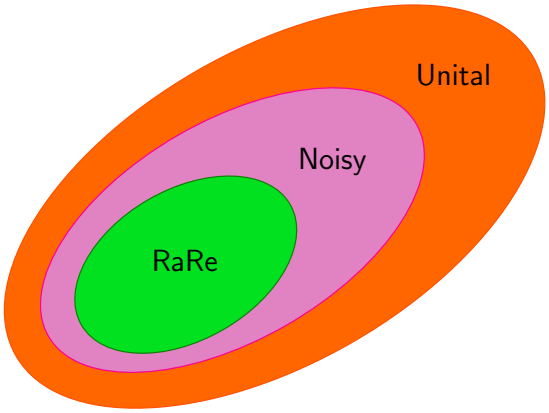


Figure 5.3: Inclusion relations between the three sets of free operations in sharp theories with purification.

sions

$$\mathsf{RaRe} \subseteq \mathsf{Noisy} \subseteq \mathsf{Unital}, \tag{5.2.8}$$

illustrated in fig. 5.3. These inclusions imply the relations

$$\rho \succeq_{\mathrm{RaRe}} \sigma \quad \Longrightarrow \quad \rho \succeq_{\mathrm{noisy}} \sigma \quad \Longrightarrow \quad \rho \succeq_{\mathrm{unital}} \sigma,$$

valid for every pair of states ρ and σ of the *same* system.[9] Note that the unital relation $\succeq_{\mathrm{unital}}$ is the weakest, i.e. the easiest to satisfy, whereas RaRe convertibility is the strongest. Starting from RaRe convertibility we can already characterise the states that are extremal in all the three preorders. In subsection 5.2.1 we already saw that $\rho \succeq_{\mathrm{RaRe}} \chi$ for every ρ. This clearly implies that $\rho \succeq_{\mathrm{noisy}} \chi$ and $\rho \succeq_{\mathrm{unital}} \chi$, so the microcanonical state is the resource with minimum value according all the three preorders. This should not surprise, since it is a free state (or can be regarded as such) for all the three resource theories. At the other extreme, we also saw that for every pure state ψ and every state ρ, $\psi \succeq_{\mathrm{RaRe}} \rho$, and that pure states are all equivalent to each other in the RaRe resource theory. By the inclusions (5.2.8), these properties carry over to the other two resource theories.

[9] Since RaRe channels have the same input and output system, we will restrict ourselves to state convertibility in the same system, so as to be able to compare the convertibility properties under *all three* resource theories.

5.3 Majorisation and unital channels

From this section on, we will always work in sharp theories with purification, even when we do not state it explicitly. We know that in these theories states can be diagonalised (see section 4.4), so we wonder whether their spectra (i.e. the vectors of their eigenvalues) play any role in determining the convertibility properties for the three resource theories, like in quantum theory [37]. Indeed, in general, determining when a state ρ can be converted into a state σ is a hard task, because it involves checking *all* possible free operations in order to find out if a transition is possible. Therefore a practical criterion that depends only on the two states is highly desirable. Here, we will explore the role of majorisation in this respect, which will provide us with a necessary and sufficient condition for the unital preorder.

In a broad sense, unital channels are the generalisation of doubly stochastic matrices, and in sharp theories with purification there is a more explicit connection.

Lemma 5.3.1. *Let $\mathcal{D}$ be a unital channel on system* A*, and let $\{\alpha_i\}_{i=1}^d$ and $\{\alpha'_i\}_{i=1}^d$ be two pure maximal sets of* A*. Then, the matrix D with entries $D_{ij} := \left(\alpha_i'^{\dagger}\middle|\mathcal{D}\middle|\alpha_j\right)$ is doubly stochastic.*

Proof. Every entry D_{ij} is a probability, so it is non-negative. Moreover, one has

$$\sum_{i=1}^{d} D_{ij} = \sum_{i=1}^{d} \left(\alpha_i'^{\dagger}\middle|\mathcal{D}\middle|\alpha_j\right) = (u|\mathcal{D}|\alpha_j) = \operatorname{tr}\alpha_j = 1,$$

for all $j = 1, \dots, d$, where we have used the fact that the effects $\{\alpha_i'^{\dagger}\}_{i=1}^d$ form an observation-test (proposition 4.4.7), and that $\mathcal{D}$ is a channel, so $u\mathcal{D} = u$. Finally, one has

$$\sum_{j=1}^{d} D_{ij} = \sum_{j=1}^{d} \left(\alpha_i'^{\dagger}\middle|\mathcal{D}\middle|\alpha_j\right) = d\left(\alpha_i'^{\dagger}\middle|\mathcal{D}\middle|\chi\right) = d\left(\alpha_i'^{\dagger}\middle|\chi\right) = d\cdot\frac{1}{d} = 1,$$

for all $i = 1, \dots, d$, where we have used proposition 4.4.5, and the fact that unital channels leave χ invariant. In conclusion we have shown that D is doubly-stochastic. □

This lemma is a generalisation of lemma 4.4.11, which can be obtained by taking $\mathcal{D}$ to be the identity channel $\mathcal{I}$.

Vice versa, every doubly stochastic matrix defines a unital channel.

Lemma 5.3.2. *Let D be a $d \times d$ doubly stochastic matrix, and let $\{\alpha_i\}_{i=1}^d$ and $\{\alpha'_i\}_{i=1}^d$ be any two pure maximal sets of system* A. *Then, the channel defined by $\mathcal{D} := \sum_{j=1}^d |\rho_j) \left(\alpha_j^\dagger\right|$, with $\rho_j := \sum_{i=1}^d D_{ij}\alpha'_i$, is unital.*

Proof. The transformation $\mathcal{D}$ is a measure-and-prepare channel: it can be implemented by performing the observation-test $\left\{\alpha_j^\dagger\right\}_{j=1}^d$ and by preparing the state ρ_j conditionally on outcome j. Moreover, one has

$$\mathcal{D}\chi = \sum_{j=1}^d \rho_j \left(\alpha_j^\dagger \middle| \chi\right) = \frac{1}{d}\sum_{j=1}^d\sum_{i=1}^d D_{ij}\alpha'_i = \frac{1}{d}\sum_{i=1}^d \alpha'_i = \chi,$$

the third equality following from the definition of doubly stochastic matrix, and the fourth from the diagonalisation of the state χ (proposition 4.4.5). □

Lemmas 5.3.1 and 5.3.2 establish a direct connection between unital channels and doubly stochastic matrices. Using this connection now we show that the ability to convert states in the unital resource theory is completely determined by a suitable majorisation criterion. Let us start by recalling the definition of majorisation [112].

Definition 5.3.3. Let $\mathbf{x}$ and $\mathbf{y}$ be two generic vectors in $\mathbb{R}^d$. One says that $\mathbf{x}$ *majorises* $\mathbf{y}$, denoted $\mathbf{x} \succeq \mathbf{y}$, if, when the entries of $\mathbf{x}$ and $\mathbf{y}$ are rearranged in decreasing order, one has

$$\sum_{i=1}^k x_i \geq \sum_{i=1}^k y_i \quad k = 1, \ldots, d-1$$

and

$$\sum_{i=1}^d x_i = \sum_{i=1}^d y_i.$$

Majorisation can be equivalently characterised in terms of doubly stochastic matrices: one has $\mathbf{x} \succeq \mathbf{y}$ if and only if $\mathbf{y} = D\mathbf{x}$, where D is a doubly stochastic matrix [112, 188]. The idea is that $\mathbf{y} \preceq \mathbf{x}$ if $\mathbf{y}$ is "more random" than $\mathbf{x}$.

In every sharp theory with purification, majorisation is a necessary and sufficient condition for convertibility under unital channels.

Theorem 5.3.4. *Let ρ and σ be normalised states, and let $\mathbf{p}$ and $\mathbf{q}$ be their spectra respectively. The state ρ can be converted into the state σ by a unital channel if and only if $\mathbf{p} \succeq \mathbf{q}$.*

Proof. Let $\rho = \sum_{j=1}^{d} p_j \alpha_j$ and $\sigma = \sum_{j=1}^{d} q_j \alpha'_j$ be diagonalisations of ρ and σ respectively. We first show that $\rho \succeq_{\text{unital}} \sigma$ implies $\mathbf{p} \succeq \mathbf{q}$. Since $\sigma = \mathcal{D}\rho$, with $\mathcal{D}$ unital channel, one has

$$\sum_{j=1}^{d} q_j \alpha'_j = \sum_{j=1}^{d} p_j \mathcal{D}\alpha_j$$

Applying $\alpha_i'^{\dagger}$ to both sides, we obtain

$$q_i = \sum_{j=1}^{d} p_j \left(\alpha_i'^{\dagger}\middle|\mathcal{D}\middle|\alpha_j\right) =: \sum_{j=1}^{d} D_{ij} p_j, \tag{5.3.1}$$

where we have set $D_{ij} := \left(\alpha_i'^{\dagger}\middle|\mathcal{D}\middle|\alpha_j\right)$. Now, the D_{ij}'s are the entries of a doubly stochastic matrix D by lemma 5.3.1. Hence, eq. (5.3.1) implies $\mathbf{p} \succeq \mathbf{q}$.

To prove sufficiency, suppose that $\mathbf{p} \succeq \mathbf{q}$ and let D be a doubly stochastic matrix such that $\mathbf{q} = D\mathbf{p}$. Define the measure-and-prepare channel $\mathcal{D} := \sum_{j=1}^{d} |\rho_j) \left(\alpha_j^{\dagger}\right|$, with $\rho_j := \sum_{i=1}^{d} D_{ij} \alpha'_i$. By construction, one has

$$\mathcal{D}\rho = \sum_{j=1}^{d} \rho_j \left(\alpha_j^{\dagger}\middle|\rho\right) = \sum_{i=1}^{d} \alpha'_i \sum_{j=1}^{d} D_{ij} p_j = \sum_{i=1}^{d} q_i \alpha'_i = \sigma.$$

Now, the channel $\mathcal{D}$ is unital by lemma 5.3.2. Hence, ρ can be converted into σ by a unital channel. □

Note that since RaRe channels and noisy operations are special cases of unital channels, majorisation is a *necessary* condition for convertibility in

the RaRe and noisy resource theories. We will examine when it is sufficient for RaRe channels in section 5.8. So far we have proved

$$\rho \succeq_{\text{RaRe}} \sigma \quad \Longrightarrow \quad \rho \succeq_{\text{noisy}} \sigma \quad \Longrightarrow \quad \rho \succeq_{\text{unital}} \sigma \quad \Longleftrightarrow \quad \mathbf{p} \succeq \mathbf{q}. \tag{5.3.2}$$

Now the relations $\psi \succeq_F \rho \succeq_F \chi$, where F denotes any of the three sets of free operations, can be easily understood in terms of majorisation. Indeed the spectrum of a pure state is $\begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}^T$, whereas the spectrum of the microcanonical state is $\begin{pmatrix} \frac{1}{d} & \dots & \frac{1}{d} \end{pmatrix}^T$, and it is straightforward to check that

$$\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \succeq \mathbf{p} \succeq \begin{pmatrix} \frac{1}{d} \\ \vdots \\ \frac{1}{d} \end{pmatrix}.$$

5.3.1 Operational characterisation of the eigenvalues

Besides giving a necessary and sufficient condition for the convertibility under unital channels, majorisation also provides an operational characterisation of the eigenvalues of a state: the eigenvalues are the least random probability distribution that can be generated by pure observation-tests. A similar result was also proved in GPTs satisfying other axioms in [169, 170, 172].

To prove the main result we need a lemma on the structure of pure observation-tests.

Lemma 5.3.5. *Let $\{a_i\}_{i=1}^n$ be a* pure *observation-test. Then, for every $i \in \{1, \dots, n\}$, $a_i = \lambda_i \alpha_i^\dagger$, for some pure state α_i, and $\lambda_i \in (0, 1]$. Moreover $\sum_{i=1}^n \lambda_i = d$, and $n \geq d$. One has $n = d$ if and only if $\{a_i\}_{i=1}^n$ is a pure sharp measurement.*

Proof. Since we can extend the diagonalisation to elements of $\mathsf{Eff}_{\mathbb{R}}(\mathrm{A})$, we can write $a_i = \lambda_i \alpha_i^\dagger$, and there is one term in the diagonalisation of a_i because it is pure. Being a_i an effect, it must be $\lambda_i \in (0, 1]$, because $\lambda_i = (a_i | \alpha_i)$. Now let us prove that $\sum_{i=1}^n \lambda_i = d$. By Causality, $\sum_{i=1}^n \lambda_i \alpha_i^\dagger = u$. Now consider

$$1 = \operatorname{tr} \chi = \sum_{i=1}^n \lambda_i \left(\alpha_i^\dagger \middle| \chi \right) = \sum_{i=1}^n \lambda_i \cdot \frac{1}{d},$$

whence $\sum_{i=1}^{n} \lambda_i = d$. Since $\lambda_i \leq 1$, we have

$$d = \sum_{i=1}^{n} \lambda_i \leq \sum_{i=1}^{n} 1 = n,$$

so $n \geq d$. Now let us prove that $n = d$ if and only if $\{a_i\}_{i=1}^{n}$ is a pure sharp measurement. Suppose $\{a_i\}_{i=1}^{n}$ is a pure sharp measurement, then $n = d$ by proposition 4.4.7. Conversely, suppose we know that $n = d$. Then in this case, the only possibility of having $\sum_{i=1}^{n} \lambda_i = d$ is when $\lambda_i = 1$ for every i. Therefore all the effects are normalised, and the observation-test can be rewritten as $\{\alpha_i^\dagger\}_{i=1}^{d}$ for some pure states $\{\alpha_i\}_{i=1}^{d}$. □

Now we are ready to prove the main result. It is useful to introduce the following notation: given two vectors $\mathbf{x} \in \mathbb{R}^n$, and $\mathbf{y} \in \mathbb{R}^m$, let us define

$$\mathbf{x} \oplus \mathbf{y} := \begin{pmatrix} \mathbf{x} \\ \hline \mathbf{y} \end{pmatrix}, \tag{5.3.3}$$

which is a vector of $\mathbb{R}^{n+m}$. This operation is nothing but appending $\mathbf{y}$ under $\mathbf{x}$ to create a larger vector.

Proposition 5.3.6. *Consider a* pure *observation-test* $\boldsymbol{a} = \{a_i\}_{i=1}^{n}$ *and state* ρ. *Let* $\mathbf{q}_{\boldsymbol{a}} \in \mathbb{R}^n$ *be the vector with entries* $q_{a,i} = (a_i|\rho)$ *for* $i \in \{1, \ldots, n\}$. *Then if* $\mathbf{p}$ *is the spectrum of* ρ, *define* $\widetilde{\mathbf{p}} := \mathbf{p} \oplus \mathbf{0}$, *where* $\mathbf{0}$ *is the* $n - d$ *dimensional null vector. Then* $\mathbf{q}_{\boldsymbol{a}} \preceq \widetilde{\mathbf{p}}$.

Note that, since by lemma 5.3.5 $n \geq d$, the vector $\mathbf{q}_{\boldsymbol{a}}$ and the spectrum $\mathbf{p}$ are not directly comparable in terms of majorisation, because they have different dimensions, $\mathbf{q}_{\boldsymbol{a}}$ being the larger. Therefore, to circumvent the problem, we make $\mathbf{p}$ larger by attaching $n - d$ 0 entries. In this way we get $\widetilde{\mathbf{p}}$, which is an n-dimensional vector like $\mathbf{q}_{\boldsymbol{a}}$. Now the let us see the proof, whose lines will be close to [174, lemma B.1] for quantum theory.

Proof. By lemma 5.3.5, for each $a_i \in \{a_i\}_{i=1}^{n}$, we have $a_i = \lambda_i \alpha_i^\dagger$, for some $0 < \lambda_i \leq 1$, and for some pure state α_i. Consider a diagonalisation of $\rho = \sum_{j=1}^{d} p_j \alpha'_j$. We have

$$q_{a,i} := (a_i|\rho) = \sum_{j=1}^{d} p_j \left(a_i \middle| \alpha'_j\right) = \sum_{j=1}^{d} \lambda_i p_j \left(\alpha_i^\dagger \middle| \alpha'_j\right).$$

Now, $M_{ij} := \lambda_i \left(\alpha_i^\dagger \middle| \alpha'_j\right)$ are the entries of an $n \times d$ matrix M such that $\mathbf{q_a} = M\mathbf{p}$. Clearly $M_{ij} \geq 0$ for all $i = 1, \dots n$, and $j = 1, \dots, d$. Calculating $\sum_{i=1}^n M_{ij}$, we have

$$\sum_{i=1}^{n} M_{ij} = \sum_{i=1}^{d} \left(\lambda_i \alpha_i^\dagger \middle| \alpha'_j\right) = \operatorname{tr} \alpha'_j = 1, \tag{5.3.4}$$

whence the column of the matrix M sum to 1. Now let us move to $\sum_{j=1}^d M_{ij}$.

$$\sum_{j=1}^{d} M_{ij} = \lambda_i \sum_{j=1}^{d} \left(\alpha_i^\dagger \middle| \alpha'_j\right) = \lambda_i d \left(\alpha_i^\dagger \middle| \chi\right) = \lambda_i d \cdot \frac{1}{d} = \lambda_i \leq 1 \tag{5.3.5}$$

If $n = d$ we are done, because in this case $\lambda_i = 1$ for every i (by lemma 5.3.5), M is doubly stochastic and $\mathbf{q_a} = M\mathbf{p}$, whence the thesis.

Now, suppose $n > d$; we wish to construct an $n \times n$ doubly stochastic matrix D from M, such that we can write $q_{a,i} = \sum_{j=1}^n D_{ij} \widetilde{p}_j$, where $\widetilde{\mathbf{p}} = \mathbf{p} \oplus \mathbf{0}$ is the vector of probabilities defined as

$$\widetilde{p}_j := \begin{cases} p_j & 1 \leq j \leq d \\ 0 & d+1 \leq j \leq n \end{cases}.$$

Let us define D as

$$D := \left(\underbrace{M}_{d \text{ columns}} \middle| \underbrace{\begin{matrix} \frac{1-\lambda}{n-d} & \dots & \frac{1-\lambda}{n-d} \end{matrix}}_{n-d \text{ columns}} \right),$$

where the last $n - d$ columns are all equal to each other, with their ith entry equal to $\frac{1-\lambda_i}{n-d}$.

Now, D is doubly stochastic. Indeed each entry is non-negative, because $\lambda_i \leq 1$ for all $i \in \{1, \dots, n\}$ and $n \geq d$. Furthermore,

$$\sum_{i=1}^{n} D_{ij} = \begin{cases} \sum_{i=1}^n M_{ij} & 1 \leq j \leq d \\ \frac{n - \sum_{i=1}^n \lambda_i}{n-d} & d+1 \leq j \leq n \end{cases} = 1$$

by eq. (5.3.4), and because $\sum_{i=1}^n \lambda_i = d$ (by lemma 5.3.5). Finally

$$\sum_{j=1}^{n} D_{ij} = \sum_{j=1}^{d} M_{ij} + \sum_{j=d+1}^{n} \frac{1-\lambda_i}{n-d} = 1,$$

having used eq. (5.3.5). Clearly now we have $q_{a,i} = \sum_{j=1}^{n} D_{ij}\widetilde{p}_j$, because, by construction of $\widetilde{\mathbf{p}}$ and D,

$$\mathbf{q}_a = \left(\begin{array}{c|ccc} M & \frac{1-\lambda}{n-d} & \dots & \frac{1-\lambda}{n-d} \end{array} \right) \left(\begin{array}{c} \mathbf{p} \\ \hline \mathbf{0} \end{array} \right).$$

Therefore $\mathbf{q}_a \preceq \widetilde{\mathbf{p}}$. □

5.4 Mixedness monotones

In the previous section we saw that the majorisation criterion determines whether a state is more resourceful than another in the unital resource theory, and it is a necessary condition in the other two. To be more quantitative, let us introduce monotones (cf. section 3.3). We noted that pure states are the most valuable states in all three resource theories (see subsection 5.2.4), so it is natural to call the three resource theories "resource theories of purity". A *purity monotone under the free operations* F for system A is a function $P : \mathsf{St}_1(\mathrm{A}) \to \mathbb{R}$ satisfying the condition $P(\rho) \geq P(\sigma)$ if $\rho \succeq_F \sigma$ [100].

In sharp theories with purification, thanks to theorem 5.3.4, unital purity monotones have a complete mathematical characterisation in terms of Schur-convex functions.

Definition 5.4.1. A function $f : \mathbb{R}^d \to \mathbb{R}$ is called *Schur-convex* if $\mathbf{x} \preceq \mathbf{y}$ implies $f(\mathbf{x}) \leq f(\mathbf{y})$.

These functions assign a greater real number to more ordered vectors. The following proposition is therefore quite natural.

Proposition 5.4.2. *A function on the state space $P : \mathsf{St}_1(\mathrm{A}) \to \mathbb{R}$ is a unital purity monotone if and only if $P(\rho) = f(\mathbf{p})$, where $\mathbf{p}$ is the spectrum of ρ and $f : \mathbb{R}^d \to \mathbb{R}$ is a Schur-convex function.*

Proof. Theorem 5.3.4 shows that the convertibility of states under unital channels is fully captured by their eigenvalues. Consequently, a unital monotone will be a function only of the eigenvalues of a state: there exists a function $f : \mathbb{R}^d \to \mathbb{R}$ such that $P(\rho) = f(\mathbf{p})$, for every normalised state ρ. Now, suppose that $\mathbf{p}$ and $\mathbf{q}$ are two probability distributions satisfying

$\mathbf{p} \succeq \mathbf{q}$. Then, theorem 5.3.4 implies that there is a unital channel transforming the state $\rho = \sum_{i=1}^{d} p_i \alpha_i$ into the state $\sigma = \sum_{i=1}^{d} q_i \alpha_i$, for any pure maximal set $\{\alpha_i\}_{i=1}^{d}$. As a result, we obtain the relation

$$f(\mathbf{p}) = P(\rho) \geq P(\sigma) = f(\mathbf{q}).$$

This means that f is Schur-convex.

Conversely, given a Schur-convex function f one can define a function P_f on the state space, as $P_f(\rho) := f(\mathbf{p})$, $\mathbf{p}$ being the spectrum of ρ. By theorem 5.3.4, if $\rho \succeq_{\text{unital}} \sigma$, then $\mathbf{p} \succeq \mathbf{q}$, where $\mathbf{p}$ and $\mathbf{q}$ are the spectra of the two states, respectively. Therefore

$$P_f(\rho) = f(\mathbf{p}) \geq f(\mathbf{q}) = P_f(\sigma),$$

where we have used the fact that f is Schur-convex. This shows that P_f is a unital purity monotone. □

Since majorisation is a necessary condition for convertibility in all the three resource theories, $P_f(\rho) = f(\mathbf{p})$, with $\mathbf{p}$ is the spectrum of ρ, and f Schur-convex, is a purity monotone in *all* the three resource theories. Hence these functions will simply be called "purity monotones", without the need to specify the resource theory. The fact that we have not proved the sufficiency of majorisation for RaRe channels means that there might exist RaRe purity monotones that are not generated by Schur-convex functions.

Remark 5.4.3. Note that the purity monotones arising from Schur-convex functions are invariant under reversible channels: $P_f(\rho) = P_f(\mathcal{U}\rho)$. This is because ρ and $\mathcal{U}\rho$ have the same spectrum. Indeed if $\rho = \sum_{i=1}^{d} p_i \alpha_i$ is a diagonalisation of ρ, then $\mathcal{U}\rho = \sum_{i=1}^{d} p_i \mathcal{U}\alpha_i$ is a diagonalisation of $\mathcal{U}\rho$, because the pure states $\mathcal{U}\alpha_i$ are distinguished by the observation-test $\{\alpha_i^\dagger \mathcal{U}^{-1}\}_{i=1}^{d}$. Then obviously ρ and $\mathcal{U}\rho$ have the same spectrum.

In thermodynamics the relevant quantities for isolated systems are entropies, which are often presented popularly as "measures of disorder". From this perspective, entropies are the opposite of purity monotones, which instead assign a higher value to ordered states. To comply with the standard thermodynamic treatment, we define *mixedness monotones*, and we will stick to them in the following presentation.

Definition 5.4.4. A *mixedness monotone* for system A is a function $M : \mathsf{St}_1(\mathrm{A}) \to \mathbb{R}$ such that $-M$ is a purity monotone.

In other words, if $\rho \succeq_F \sigma$, we have $M(\rho) \leq M(\sigma)$. In the same spirit, one defines Schur-concave functions.

Definition 5.4.5. A function $f : \mathbb{R}^d \to \mathbb{R}$ is called *Schur-concave* if $-f$ is Schur-convex.

This means that for a Schur-concave function, if $\mathbf{x} \preceq \mathbf{y}$ then $f(\mathbf{x}) \geq f(\mathbf{y})$.

Viewing proposition 5.4.2 in the light of mixedness monotones, given a Schur-concave function f, one can generate a mixedness monotone $M_f(\rho) := f(\mathbf{p})$, where ρ is a state, and $\mathbf{p}$ its spectrum. In the case of the unital resource theory, mixedness monotones are *all* generated by Schur-concave functions.

A slightly more restrictive notion is that of *generalised entropy*.

Definition 5.4.6. For every system A, let $M : \mathsf{St}_1(\mathrm{A}) \to \mathbb{R}$ be a mixedness monotone. We say that M is a *generalised entropy* if it is additive, that is

$$M(\rho_{\mathrm{A}} \otimes \sigma_{\mathrm{B}}) = M(\rho_{\mathrm{A}}) + M(\sigma_{\mathrm{B}})$$

for all $\rho_{\mathrm{A}} \in \mathsf{St}_1(\mathrm{A})$, and all $\sigma_{\mathrm{B}} \in \mathsf{St}_1(\mathrm{B})$.

By proposition 5.4.2, some generalised entropies can be obtained from particular Schur-concave functions.

Corollary 5.4.7. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a Schur-concave function for all d, satisfying the additivity property $f(\mathbf{p} \otimes \mathbf{q}) = f(\mathbf{p}) + f(\mathbf{q})$, where $\mathbf{p} \otimes \mathbf{q}$ denotes the Kronecker product. Then, the corresponding mixedness monotone $M_f(\rho) = f(\mathbf{p})$, where $\mathbf{p}$ is the spectrum of ρ, is a generalised entropy.*

Again, in the case of the unital resource theory, all generalised entropies are of the form of corollary 5.4.7. Let us see some examples.

Example 5.4.8. An important example of additive Schur-concave functions are Rényi entropies [189]

$$H_\alpha(\mathbf{p}) = \frac{1}{1-\alpha} \log_a \left(\sum_{i=1}^{d} p_i^\alpha \right),$$

with $a > 1$ and $\alpha \geq 0$, where $\mathbf{p}$ is a vector of probabilities[10]. If we take the limit $\alpha \to 1$ we recover *Shannon-von Neumann entropy* [190,191]:

$$H(\mathbf{p}) = -\sum_{i=1}^{d} p_i \log_a p_i = \lim_{\alpha \to 1} H_\alpha(\mathbf{p}).$$

Particularly important cases are when $\alpha = 0$, and $\alpha \to +\infty$. In this cases,

$$H_0(\mathbf{p}) = \log_a |\operatorname{supp} \mathbf{p}|,$$

where $|\operatorname{supp} \mathbf{p}|$ denotes the number of the non-vanishing entries of $\mathbf{p}$. Moreover,

$$H_\infty(\mathbf{p}) := \lim_{\alpha \to +\infty} H_\alpha(\mathbf{p}) = -\log_a p_{\max},$$

where $p_{\max}$ denotes the maximum entry of $\mathbf{p}$. Rényi entropies are decreasing in α, so

$$H_\infty(\mathbf{p}) \leq H_\alpha(\mathbf{p}) \leq H_0(\mathbf{p})$$

for $\alpha \geq 0$. For this reason, H_∞ is also known as the *min-entropy*, and H_0 as the *max-entropy*.

Now we can define the generalised Rényi entropies as $S_\alpha(\rho) := H_\alpha(\mathbf{p})$, for $\alpha \in [0, +\infty]$, where $\mathbf{p}$ is the spectrum of ρ. In particular $S(\rho) = H(\mathbf{p})$ is the generalised Shannon-von Neumann entropy. Note that one has the obvious bounds

$$0 \leq S_\alpha(\rho) \leq \log_a d,$$

for every state ρ, and every $\alpha \in [0, +\infty]$, where d is the dimension of the system. The lower bound is achieved by any pure state, and the upper bound by the microcanonical state χ.

5.4.1 Preparation and measurement monotones

In some cases, it is possible to connect generalised entropies, defined on the spectra of states, to measures of the minimum randomness we can extract from a state by pure measurements, or of the minimum randomness needed to prepare that state. To this end, it is useful to identify a particular class of Schur-concave functions, which we call *reducible*.

[10] This means a vector representing a probability distribution.

Definition 5.4.9. A Schur-concave function f is called *reducible* if for every vector $\mathbf{x} \in \mathbb{R}^d$ one has

$$f(\mathbf{x} \oplus \mathbf{0}) = f(\mathbf{x}),$$

where $\mathbf{0}$ is a null vector with any number of components.

Here we have used the same notation as in eq. (5.3.3) to denote appending some zero entries to a vector. In words, for a reducible Schur-concave function the zero entries of a vector do not matter, it only sees the non-vanishing ones. Examples of reducible Schur-concave functions are Rényi entropies (and hence Shannon-von Neumann entropy). Not all Schur-concave functions are reducible though, as shown in the following example.

Example 5.4.10. Given a vector of probabilities $\mathbf{p}$ of dimension d, consider the function $V(\mathbf{p}) = \frac{1}{d}\left(1 - \sum_{i=1}^{d} p_i^2\right)$, arising from the variance of $\mathbf{p}$. V is Schur-concave, but it is *not* reducible. Indeed, consider the vectors $\mathbf{p} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix}^T$, and $\widetilde{\mathbf{p}} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}^T$; we have $V(\mathbf{p}) = \frac{1}{4}$, but $V(\widetilde{\mathbf{p}}) = \frac{1}{6}$, whence V is not reducible.

It is worth noting that the marginals of a pure bipartite state have the same value for any reducible Schur-concave function. So if Ψ is a pure state of AB and ρ_{A} and ρ_{B} its marginals on A and B respectively, one has

$$M_f(\rho_{\mathrm{A}}) = M_f(\rho_{\mathrm{B}}),$$

for every reducible Schur-concave function f. This easily follows from the generalised Schmidt decomposition (theorem 4.5.1), which ensures that the marginals of a pure bipartite state have the same non-vanishing eigenvalues. These non-vanishing eigenvalues are the only ones seen by a reducible Schur-concave function.

In every sharp theory with purification, the mixedness monotones arising from reducible Schur-concave functions have a nice characterisation in terms of optimal measurements, or, dually, in terms of optimal ensemble decompositions. Let us start from two definitions.

Definition 5.4.11. Given a reducible Schur-concave function f, the *measurement monotone* M_f^{meas} of a state $\rho \in \mathrm{St}_1(\mathrm{A})$ is defined as

$$M_f^{\mathrm{meas}}(\rho) := \inf_{\boldsymbol{a}} f(\mathbf{q}),$$

where the infimum is taken over all pure observation-tests $\boldsymbol{a} := \{a_i\}$ of system A, and $q_i := (a_i|\rho)$.

Definition 5.4.12. The *preparation monotone* M_f^{prep} is defined as

$$M_f^{\text{prep}}(\rho) := \inf_{\sum_i \lambda_i \psi_i = \rho} f(\boldsymbol{\lambda}),$$

where the infimum is over all convex decompositions $\sum_i \lambda_i \psi_i$ of the state ρ in terms of pure states ψ_i.

In words, the measurement monotone M_f^{meas} is the smallest amount of mixedness (as measured by the function f) present in the probability distributions generated by pure observation-tests on ρ. Dually, the preparation monotone M_f^{prep} is the smallest amount of mixedness necessary to prepare ρ as an ensemble of pure states. Why do we have to take the infimum? To understand it, consider the following example for a preparation monotone.

Example 5.4.13. Consider a pure state ψ and a preparation monotone M_f^{prep}. Clearly, convex decompositions of ψ into pure states are all trivial, i.e. of the form $\sum_i \lambda_i \psi$. Comparing the probability distribution $\{\lambda_i\}$ with the extremal one $\boldsymbol{\delta}$, we have

$$\boldsymbol{\lambda} \preceq \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \boldsymbol{\delta}.$$

Therefore $f(\boldsymbol{\lambda}) \geq f(\boldsymbol{\delta})$; however this is a spurious effect, because the pure states involved in a convex realisation of ψ are exactly the same, and all equal to ψ! The presence of the infimum cancels these spurious effects, and we have $M_f^{\text{prep}}(\psi) = f(\boldsymbol{\delta})$.

Let us clarify why we need f to be reducible with an example from quantum theory.

Example 5.4.14. Consider a state $\rho = p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1|$ of a qutrit, where $\{|0\rangle, |1\rangle, |2\rangle\}$ is the computational basis, and $p \in (0,1)$. Clearly if we use the spectral measurement $\{|0\rangle\langle 0|, |1\rangle\langle 1|, |2\rangle\langle 2|\}$, in the evaluation of the measurement monotone we obtain the vector $\mathbf{q} = \begin{pmatrix} p & 1-p & 0 \end{pmatrix}^T$.

Now take the pure POVM $\left\{ |0\rangle\langle 0| , |1\rangle\langle 1| , \frac{1}{2}|2\rangle\langle 2| , \frac{1}{2}|2\rangle\langle 2| \right\}$, which gives rise to the vector $\mathbf{q}' = \begin{pmatrix} p & 1-p & 0 & 0 \end{pmatrix}^T$. Clearly there is no difference in the randomness one can extract from the former and the latter measurement when performed on ρ. However, if f is non-reducible, there is no guarantee that $f(\mathbf{q}) = f(\mathbf{q}')$.

Similarly, in the evaluation of the preparation monotone, the writing $\rho = p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1|$ gives rise to $\boldsymbol{\lambda} = \begin{pmatrix} p & 1-p \end{pmatrix}^T$. However, one can also write ρ as $\rho = p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1| + 0|2\rangle\langle 2|$, giving rise to $\boldsymbol{\lambda}' = \begin{pmatrix} p & 1-p & 0 \end{pmatrix}^T$. Again, there is no more randomness involved in the latter preparation than in the former, so we should have $f(\boldsymbol{\lambda}) = f(\boldsymbol{\lambda}')$.

The definitions of measurement and preparation monotones can be put forward in any causal GPT, since they involve very primitive elements of GPTs, such as the convex structure, and the fact that observation-tests yield probabilities when performed on states. Specifically [163, 174, 175] defined preparation and measurement Shannon entropy in GPTs.

Despite the name we used, in general GPTs, or even in microcanonical ones, M_f^{meas} and M_f^{prep} lack a clear interpretation as mixedness monotones, unless further assumptions are made on f (or on the theory). For instance, taking f concave makes M_f^{meas} a RaRe mixedness monotone [100]. However, in sharp theories with purification we find that M_f^{meas} and M_f^{prep} are actual mixedness monotones.

Theorem 5.4.15. *In every sharp theory with purification one has*

$$M_f^{\text{meas}}(\rho) = M_f^{\text{prep}}(\rho) = M_f(\rho),$$

for every reducible Schur-concave function f and for every state ρ.

Proof. Let us prove that M_f^{meas} coincides with M_f. Let $\rho = \sum_{i=1}^d p_i \alpha_i$ be a diagonalisation of $\rho \in \mathrm{St}_1(\mathrm{A})$. If we take the pure sharp measurement $\{\alpha_i^\dagger\}_{i=1}^d$, we have $(\alpha_i^\dagger|\rho) = p_i$. Hence,

$$M_f^{\text{meas}}(\rho) \leq f(\mathbf{p}) = M_f(\rho).$$

To prove the converse inequality, recall proposition 5.3.6: for every pure observation-test $\{a_i\}$, one has $\mathbf{q} \preceq \widetilde{\mathbf{p}}$, where $\mathbf{q}$ is the vector of probabilities

$q_i = (a_i|\rho)$ and $\widetilde{\mathbf{p}}$ is the vector of the eigenvalues of ρ (with additional zeros appended, if needed). Since f is Schur-concave, we have $f(\mathbf{q}) \geq f(\widetilde{\mathbf{p}})$ and, taking the infimum over all pure measurements

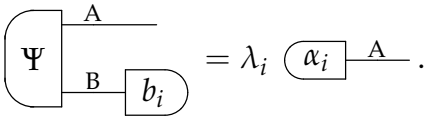
$$M_f^{\text{meas}}(\rho) \geq f(\widetilde{\mathbf{p}}) = f(\mathbf{p}) = M_f(\rho),$$

where we have used the fact that f is reducible. Summarising, we have obtained the equality $M_f^{\text{meas}} = M_f$.

We now prove the equality $M_f^{\text{prep}} = M_f$. By definition, we have

$$M_f^{\text{prep}}(\rho) \leq f(\mathbf{p}) = M_f(\rho),$$

because the diagonalisation is a special case of pure-state decomposition. The converse inequality follows from Pure Steering. By corollary 4.5.2, every convex decomposition of $\rho =: \rho_{\text{A}}$ into pure states can be induced by a pure sharp measurement applied to a suitable purification of ρ (which in general depends on the particular decomposition considered). Now take the decomposition $\rho_{\text{A}} = \sum_i \lambda_i \alpha_i$ of ρ_{A} into pure states, and let $\Psi \in \mathsf{PurSt}_1(\text{AB})$ be an associated purification as per corollary 4.5.2. Then there is a *pure* sharp measurement $\{b_i\}$ on system B that will induce the states[11] $\{\lambda_i \alpha_i\}$:



Discarding system A on both sides we obtain

$$(b_i|\rho_{\text{B}}) = \lambda_i,$$

where ρ_{B} is the marginal state on system B. In other words, $\boldsymbol{\lambda}$ is the vector of the outcome probabilities for the pure observation-test $\{b_i\}$. By definition of measurement monotone, we must have

$$f(\boldsymbol{\lambda}) \geq M_f^{\text{meas}}(\rho_{\text{B}}) = M_f(\rho_{\text{B}}). \tag{5.4.1}$$

Since f is reducible, for all the purifying systems used to induce pure-state decompositions of ρ, $M_f(\rho_{\text{B}}) = M_f(\rho_{\text{A}}) = M_f(\rho)$. Now, taking the

[11] Here we are allowing some of the λ_i's to be zero; this is why we are taking the same index i for the effects in the pure sharp measurement $\{b_i\}$, unlike in corollary 4.5.2.

infimum over all pure-state decompositions of $\rho_{\mathrm{A}} = \rho$ in eq. (5.4.1), we obtain the desired inequality

$$M_f^{\mathrm{prep}}(\rho) \geq M_f(\rho).$$

□

This result, in particular the equality between the measurement and preparation max-entropies S_0, was linked, in the presence of Strong Symmetry [87, 113], to the absence of higher-order interference, i.e. to the lack of irreducible behaviour in the interference pattern obtained from three (or more) slits [169]. Therefore, this points out that sharp theories with purification and Strong Symmetry (cf. also subsection 5.8.1) do not have higher-order interference. In fact, it was proved that Strong Symmetry is not necessary to prove the lack of higher-order interference [111].

5.5 Properties of Shannon-von Neumann entropy

The first property follows from theorem 5.4.15: since the measurement Shannon-von Neumann entropy was proved to be concave in [163, 174, 175], it follows that the "spectral" Shannon-von Neumann entropy we defined here is concave as too. This means that

$$S\left(\sum_i p_i \rho_i\right) \geq \sum_i p_i S(\rho_i),$$

where $\{p_i\}$ is a probability distribution.

As seen in subsection 4.4.5, Shannon-von Neumann entropy can be expressed as

$$S(\rho) = \left(-\log_a \rho^{\dagger} \middle| \rho\right),$$

meaning that $S(\rho)$ is the expectation value of the *surprisal observable* $-\log_a \rho^{\dagger}$. This alternative formulation is useful because it suggests a generalisation of the relative entropy to sharp theories with purification.

Definition. Let ρ and σ be two normalised states. The *relative entropy* of ρ to σ is

$$S(\rho \parallel \sigma) := \left(\log_a \rho^{\dagger} - \log_a \sigma^{\dagger} \middle| \rho\right).$$

The key property of the relative entropy is Klein's inequality (cf. also [130, 169].

Lemma 5.5.1 (Klein's inequality)**.** *Let ρ and σ be two normalised states. One has $S(\rho \parallel \sigma) \geq 0$ and $S(\rho \parallel \sigma) = 0$ if and only if $\rho = \sigma$.*

Proof. The proof follows similar lines to the quantum case (see e.g. [116]). Let $\rho = \sum_{i=1}^{d} p_i \alpha_i$ and $\sigma = \sum_{i=1}^{d} q_i \alpha_i'$ be diagonalisations of ρ and σ. Now, let us compute $S(\rho \parallel \sigma)$ explicitly. Assume that all the eigenvalues of ρ and σ are non-zero, for the result in the general case can be obtained by using the continuity of the logarithm, and by taking limits suitably. Hence,

$$\left(\log \rho^{\dagger} \middle| \rho\right) = \sum_{i=1}^{d} p_i \log_a p_i,$$

and

$$\left(\log_a \sigma^{\dagger} \middle| \rho\right) = \sum_{i,j=1}^{d} \left(\alpha_i'^{\dagger} \middle| \alpha_j\right) p_j \log_a q_i = \sum_{i,j=1}^{d} T_{ij} p_j \log_a q_i,$$

where $T_{ij} := \left(\alpha_i'^{\dagger} \middle| \alpha_j\right)$ are the entries of a doubly stochastic matrix (lemma 4.4.11). Then

$$S(\rho \parallel \sigma) = \sum_{j=1}^{d} p_j \left(\log_a p_j - \sum_{i=1}^{d} T_{ij} \log_a q_i \right) \geq \sum_{j=1}^{d} p_j \left(\log_a p_j - \log_a r_j \right), \tag{5.5.1}$$

having used the concavity of the logarithm, and having set $r_j := \sum_{i=1}^{d} T_{ij} q_i$. The right-hand side of the last inequality is the classical relative entropy $D(\mathbf{p} \parallel \mathbf{r})$. Since $D(\mathbf{p} \parallel \mathbf{r})$ is always non-negative, we obtain the bound

$$S(\rho \parallel \sigma) \geq D(\mathbf{p} \parallel \mathbf{r}) \geq 0.$$

The fact that $S(\rho \parallel \sigma) = 0$ if $\rho = \sigma$ is obvious from the very definition of relative entropy. Let us prove the converse. Since the classical relative entropy vanishes if and only if $\mathbf{p} = \mathbf{r}$, the condition $S(\rho \parallel \sigma) = 0$ implies $p_j = \sum_{i=1}^{d} T_{ij} q_i$, for all $i \in \{1, \ldots, d\}$. Inserting this equality into eq. (5.5.1) we obtain the relation

$$0 = \sum_{j=1}^{d} p_j \left[\log \left(\sum_{i=1}^{d} T_{ij} q_i \right) - \sum_{i=1}^{d} T_{ij} \log q_i \right].$$

Since the logarithm is a strictly concave function, the equality implies that T is a permutation matrix. Hence, we have $T_{ij} = \delta_{i,\pi(j)}$, for a suitable permutation π. Recalling the definition of T, we obtain

$$T_{ij} = \left(\alpha_i'^{\dagger} \middle| \alpha_j\right) = \delta_{i,\pi(j)},$$

which in turn implies $\alpha_i = \alpha'_{\pi(i)}$, for all $i \in \{1, \dots, d\}$ due to the state-effect duality. In conclusion, we have obtained

$$\rho = \sum_{i=1}^{d} p_i \alpha_i = \sum_{i=1}^{d} q_{\pi(i)} \alpha'_{\pi(i)} = \sigma.$$

□

Like in quantum theory, this version of Klein's inequality allows one to prove a number of important properties. The easiest application is the subadditivity of Shannon-von Neumann entropy.

Proposition 5.5.2 (Subadditivity)**.** *Let ρ_{AB} be a bipartite state of system* AB*, and let ρ_{A} and ρ_{B} be its marginals on system* A *and* B *respectively. Shannon-von Neumann is* subadditive*, namely*

$$S(\rho_{\mathrm{AB}}) \leq S(\rho_{\mathrm{A}}) + S(\rho_{\mathrm{B}}).$$

The equality holds if and only if ρ_{AB} is a product state.

The proof follows from the application of Klein's inequality to the states $\rho := \rho_{\mathrm{A}} \otimes \rho_{\mathrm{B}}$ and $\sigma := \rho_{\mathrm{AB}}$. The subadditivity of the entropy guarantees that the *mutual information*, defined as

$$I(\mathrm{A};\mathrm{B})_{\rho_{\mathrm{AB}}} := S(\rho_{\mathrm{A}}) + S(\rho_{\mathrm{B}}) - S(\rho_{\mathrm{AB}})$$

is a non-negative quantity, and vanishes if and only if ρ_{AB} is a product state. Therefore, the mutual information can be used as a measure of correlations. On the other hand, the *conditional entropy*

$$S(\mathrm{A}|\mathrm{B}) := S(\rho_{\mathrm{AB}}) - S(\rho_{\mathrm{B}})$$

can be negative, thanks to Purification, because AB can be in a pure state, $S(\rho_{\mathrm{AB}}) = 0$, while ρ_{B} can be mixed, so $S(\rho_{\mathrm{B}}) > 0$. Like in quantum theory, the negativity of conditional entropy can be exploited for novel thermodynamic protocols [192], as we will see in subsection 5.7.1.

Another consequence of Klein's inequality is the *triangle inequality*.

Proposition 5.5.3 (Triangle inequality)**.** *For every bipartite state ρ_{AB} one has*

$$S(\rho_{AB}) \geq |S(\rho_A) - S(\rho_B)|,$$

where ρ_A and ρ_B are the marginals of ρ_{AB} on A *and* B *respectively.*

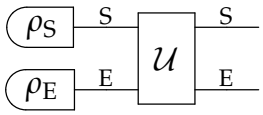
Its proof is the same as in the quantum case (see e.g. [116]), and we will omit it for brevity. Combining subadditivity and the triangle inequality, one obtains the bound

$$|S(\rho_A) - S(\rho_B)| \leq S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B),$$

valid in all sharp theories with purification.

5.5.1 The second law lemma

Using the mathematical properties of Shannon-von Neumann entropy we can prove a physical one that arises naturally when considering a system evolving jointly with its environment. Assume the system and the surrounding environment are uncorrelated at the initial time. Consistently with Purification, here we assume that, by suitably enlarging the environment, the interaction can be modelled by a reversible channel $\mathcal{U}$. We denote the initial states of the system and the environment by ρ_S and ρ_E respectively, so the initial state of the composite system is $\rho_{SE} = \rho_S \otimes \rho_E$. Primed states will denote the states after the interaction.



The result of the interaction is typically to create correlations between the system and the environment, thus increasing the mutual information from the initial zero value to a final non-zero value. The creation of correlations can be equivalently phrased as an increase of the sum of the system and environment entropies. Indeed, the positivity of the mutual information gives the bound

$$0 \leq I(S;E)_{\rho'_{SE}} = S(\rho'_S) + S(\rho'_E) - S(\rho'_{SE}) =$$

$$= S(\rho'_S) + S(\rho'_E) - S(\rho_{SE}) = S(\rho'_S) + S(\rho'_E) - S(\rho_S) - S(\rho_E),$$

the second equality coming from the fact that reversible channels do not change the entropy (see remark 5.4.3), so $S\left(\rho'_{\mathrm{SE}}\right) = S\left(\rho_{\mathrm{SE}}\right)$. The resulting bound

$$S\left(\rho'_{\mathrm{S}}\right) + S\left(\rho'_{\mathrm{E}}\right) \geq S\left(\rho_{\mathrm{S}}\right) + S\left(\rho_{\mathrm{E}}\right) \tag{5.5.2}$$

is sometimes regarded as an elementary instance of the second law of thermodynamics [117]. It is important, however, not to confuse the sum of the entropies $S\left(\rho'_{\mathrm{S}}\right) + S\left(\rho'_{\mathrm{E}}\right)$ with the total entropy $S\left(\rho'_{\mathrm{SE}}\right)$, which remains *unchanged* due to the reversibility of the global evolution. The best reading of eq. (5.5.2) is probably that a decrease in the entropy of the system must be accompanied by an increase of the entropy of the environment. Following Reeb and Wolf [193] we will refer to eq. (5.5.2) as the *second law lemma*.

Operationally, the second law lemma is the statement that uncorrelated systems can only become more correlated as a result of reversible interactions. The interesting part of it is that "correlations" here are measured in terms of entropies: the existence of an entropic measure of correlations is a non-trivial consequence of the axioms.

5.6 Generalised Gibbs states

After analysing microcanonical thermodynamics in GPTs, and in particular in sharp theories with purification, it is time to move beyond it, and begin studying the role of temperature. In classical and quantum statistical mechanics, we know that temperature, and the associated equilibrium state, the *canonical ensemble*, are tightly related to the Hamiltonian of the system. Here we encounter the first difficulty: defining the Hamiltonian in GPTs is a rather thorny issue [176]. To circumvent the problem, instead of determining which observable is the Hamiltonian, let us instead focus on the fact that the Hamiltonian is some observable, and recall that in subsection 4.4.5, we learnt a few things about observables in sharp theories with purification. Now, suppose we are given an observable H, which we may think of as the Hamiltonian of the system. However, for the following derivation, this identification is not at all necessary, and our treatment provides an immediate way to see generalisations of the canonical ensemble beyond the Hamiltonian case even in quantum theory.

In the following, since Shannon-von Neumann entropy can be seen as the expectation value of the surprisal observable, we will make use of it

to derive the form of thermal states following Jaynes' maximum entropy principle [16, 17]

Now, suppose that the only information we have about the state of system A is the expectation value of a certain observable H (e.g. the Hamiltonian). Which state should we assign to the system? The maximum entropy principle posits that, among the states with the given expectation value, we should choose the one that maximises Shannon-von Neumann entropy, namely the state $\rho_{\max}$ such that

$$\rho_{\max} = \arg\max \left\{ S(\rho) : \langle H \rangle_\rho = E \right\}. \tag{5.6.1}$$

Before proceeding, let us make a brief remark about the choice of the base a for the logarithm in the definition of Shannon-von Neumann entropy. The argument in the following can be carried out using any base $a > 1$ for the logarithm in the definition of entropy. However, if we want the entropy to have a thermodynamic meaning, we must recover thermodynamic predictions, such as the correct calculation of entropy differences for the ideal gas. An argument by von Neumann [190], later extended to GPTs [169, 172], shows that one must take the natural logarithm. With this in mind, let us determine the state that maximises Shannon-von Neumann entropy in sharp theories with purification, subject to the constraint $\langle H \rangle = E$. Like in quantum theory, it turns out that there is a one-parameter family of states: *Gibbs states*. They are of the form

$$\gamma_\beta := \frac{\mathrm{e}^{-\beta H^\dagger}}{\operatorname{tr} \mathrm{e}^{-\beta H^\dagger}}, \tag{5.6.2}$$

where $\beta \in [-\infty, +\infty]$, and the value of the parameter β is a function of E. The expression in eq. (5.6.2) means in fact

$$\gamma_\beta = \frac{1}{Z} \sum_{i=1}^{d} \mathrm{e}^{-\beta E_i} \varphi_i,$$

with $Z := \sum_{i=1}^{d} \mathrm{e}^{-\beta E_i}$, where the E_i's are the eigenvalues of H, and each φ_i is a pure state such that $(H|\varphi_i) = E_i$, namely the corresponding eigenstate. How can we check that the Gibbs states are exactly the solution to the maximisation problem in eq. (5.6.1)? Instead of solving it directly, let us show that the entropy of Gibbs states is higher than of any other state

with the same expectation value E for H. To this end, let us calculate the expectation value of H on a Gibbs state:

$$E(\beta) := \langle H \rangle_{\gamma_\beta} = -\frac{\mathrm{d}}{\mathrm{d}\beta} \ln Z.$$

This can assume all values between $E_{\min}$ and $E_{\max}$ (the minimum and maximum eigenvalue of H). Now, in the maximisation problem (5.6.1), E is fixed, rather than a function of β. It is instead β to be a function of E. In other words, we have to invert the function $E(\beta)$. Now, if $E_{\min} < E_{\max}$, namely H is not fully degenerate, the function $E(\beta)$ is invertible for all $\beta \in [-\infty, +\infty]$ [193, lemma 9]. If H is fully degenerate, the Gibbs state is the microcanonical state $\chi = \frac{1}{d} \sum_{i=1}^{d} \varphi_i$, for every β, a case we treated in the previous sections. Therefore, it is not restrictive to assume $E_{\min} < E_{\max}$. In this case, there is an inverse function $\beta(E)$, and now we are ready to prove that the Gibbs state $\gamma_{\beta(E)}$ is the maximum entropy state with expectation value E. The proof is based on an argument by Preskill [117]. First, note that the entropy of a Gibbs state is

$$S\left(\gamma_{\beta(E)}\right) = \beta(E) E + \ln Z. \tag{5.6.3}$$

Then use Klein's inequality between any ρ with $\langle H \rangle_\rho = E$, and $\gamma_{\beta(E)}$.

$$0 \leq S\left(\rho \parallel \gamma_{\beta(E)}\right) = \left(\ln \rho^\dagger - \ln \gamma_{\beta(E)}^\dagger \middle| \rho\right) = \left(\ln \rho^\dagger + \beta(E) H + u \ln Z \middle| \rho\right) =$$

$$= -S(\rho) + \beta(E) E + \ln Z = -S(\rho) + S\left(\gamma_{\beta(E)}\right),$$

where we have used the fact that $S(\rho) = -\left(\ln \rho^\dagger | \rho\right)$, that $E = \langle H \rangle_\rho = (H|\rho)$, and eq. (5.6.3). This yields the bound

$$S(\rho) \leq S\left(\gamma_{\beta(E)}\right),$$

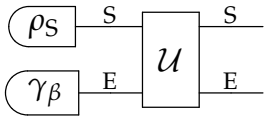
for every ρ such that $\langle H \rangle_\rho = E$.

Motivated by this characterisation, we regard the Gibbs state $\gamma_{\beta(E)}$ as the equilibrium state of a system with fixed expectation value E of the observable H. In the following we will focus on the case where H is the "energy of the system". In this case, we will write the parameter β as

$\beta = \frac{1}{kT}$, where k is Boltzmann's constant, and T is interpreted as the "temperature". Consistently, we will regard γ_β as the "equilibrium state at temperature T". Since the system is finite-dimensional, in principle we allow *negative* temperatures [194,195], corresponding to a situation in which higher energy levels are favoured. However, the interpretation of negative β's as sensible inverse *thermodynamic* temperatures has been recently questioned [196], so in the following treatment we will always assume temperatures to be non-negative.

5.7 An operational derivation of Landauer's principle

The entropic tools constructed from the axioms allow us to prove an operational version of Landauer's principle, based on a recent argument by Reeb and Wolf [193]. The scenario considered here is that of a system S that interacts reversibly with an environment E, initially in the equilibrium state at temperature T, i.e. in a Gibbs state. Like in subsection 5.5.1, primed states and quantities denote the states and the quantities after the interaction.



In this context, Landauer's principle amounts to the statement that a decrease in the entropy of the system must be accompanied by an increase in the expected energy of the environment. More formally, we have the following theorem, exactly equal to the result by Reeb and Wolf [193], but where quantum entropies are replaced by our definition of Shannon-von Neumann entropy [101].

Theorem 5.7.1. *Suppose that the system and the environment are initially in the product state $\rho_{SE} = \rho_S \otimes \gamma_{E,\beta}$ where $\gamma_{E,\beta}$ is Gibbs state at inverse temperature β. Let H_E be the energy observable of the environment. After a reversible interaction $\mathcal{U}$, the system and the environment will satisfy the equality*

$$\langle H_E \rangle' - \langle H_E \rangle = kT \left(S(\rho_S) - S(\rho'_S) + I(S;E)_{\rho'_{SE}} + S(\rho'_E \parallel \gamma_{E,\beta}) \right), \tag{5.7.1}$$

where $\langle H_{\mathrm{E}} \rangle = \left(H_{\mathrm{E}} \middle| \gamma_{\mathrm{E},\beta}\right)$ *and* $\langle H_{\mathrm{E}} \rangle' = \left(H_{\mathrm{E}} \middle| \rho'_{\mathrm{E}}\right)$ *are the expectation values of the environment energy at the initial and final times respectively.*

The proof is identical to the quantum case [193], and we do not report it for brevity. The key point is, again, Klein's inequality, which implies that the last two terms in the right-hand side of eq. (5.7.1) are always non-negative, and therefore one has the bound

$$\langle H_{\mathrm{E}} \rangle' - \langle H_{\mathrm{E}} \rangle \geq kT \left(S\left(\rho_{\mathrm{S}}\right) - S\left(\rho'_{\mathrm{S}}\right) \right), \tag{5.7.2}$$

stating that it is impossible to reduce the entropy of the system without heating up the environment. Furthermore, the equality condition in Klein's inequality implies that the lower bound (5.7.2) is attained if and only if

1. the system and the environment remain uncorrelated after the interaction;

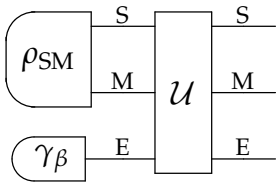
2. the environment remains in the equilibrium state.

Note that in this case $\langle H_{\mathrm{E}} \rangle' = \langle H_{\mathrm{E}} \rangle$, and there is in fact no dissipation. This is because eq. (5.7.1) prescribes $S\left(\rho_{\mathrm{S}}\right) = S\left(\rho'_{\mathrm{S}}\right)$, so the entropy of the state cannot decrease, and no real erasure takes place here.

5.7.1 The role of memories and negative conditional entropy

Combining the settings of [192] and [193], we can explore the thermodynamic meaning of *negative* conditional entropy in sharp theories with purification. We will see that non-classical correlations, witnessed by negative conditional entropy, can be harnessed to overcome Landauer's bound (5.7.2): we can reduce the entropy of the system without heating up the environment.

Suppose there is a system M, the *memory*, which contains some information about the system S because it is correlated with it. Now we can consider the case where the composite system SME is in the state $\rho_{\mathrm{SM}} \otimes \gamma_{\mathrm{E},\beta}$, with $\gamma_{\mathrm{E},\beta}$ a Gibbs state.



We will also assume that the overall reversible evolution does not increase the entropy of the memory, viz. $S\left(\rho'_{\mathrm{M}}\right) \leq S\left(\rho_{\mathrm{M}}\right)$ [192]. It is immediate to see that the second law lemma (subsection 5.5.1) takes the form

$$S\left(\rho'_{\mathrm{SM}}\right) - S\left(\rho_{\mathrm{SM}}\right) + S\left(\rho'_{\mathrm{E}}\right) - S\left(\rho_{\mathrm{E}}\right) = I\left(\mathrm{SM};\mathrm{E}\right)_{\rho'_{\mathrm{SME}}}$$

which allows one to reformulate Landauer's principle as follows [193].

Proposition 5.7.2. *Suppose that the system, the memory, and the environment are initially in the state $\rho_{\mathrm{SME}} = \rho_{\mathrm{SM}} \otimes \gamma_{\mathrm{E},\beta}$ with $\gamma_{\mathrm{E},\beta}$ a Gibbs state. After a reversible interaction $\mathcal{U}$ such that $S\left(\rho'_{\mathrm{M}}\right) \leq S\left(\rho_{\mathrm{M}}\right)$, the system, the memory, and the environment satisfy the equality*

$$\langle H_{\mathrm{E}}\rangle' - \langle H_{\mathrm{E}}\rangle = kT\left(S\left(\mathrm{S}|\mathrm{M}\right)_{\rho_{\mathrm{SM}}} - S\left(\mathrm{S}|\mathrm{M}\right)_{\rho'_{\mathrm{SM}}} + \right.$$

$$\left. + S\left(\rho_{\mathrm{M}}\right) - S\left(\rho'_{\mathrm{M}}\right) + I\left(\mathrm{SM};\mathrm{E}\right)_{\rho'_{\mathrm{SME}}} + S\left(\rho'_{\mathrm{E}} \parallel \gamma_{\mathrm{E},\beta}\right)\right).$$

Given that $S\left(\rho'_{\mathrm{M}}\right) \leq S\left(\rho_{\mathrm{M}}\right)$, this means again that

$$\langle H_{\mathrm{E}}\rangle' - \langle H_{\mathrm{E}}\rangle \geq kT\left(S\left(\mathrm{S}|\mathrm{M}\right)_{\rho_{\mathrm{SM}}} - S\left(\mathrm{S}|\mathrm{M}\right)_{\rho'_{\mathrm{SM}}}\right) \tag{5.7.3}$$

Comparing this with eq. (5.7.2), we notice the presence of conditional entropies which may be *negative*. In the particular case of SM in a *pure* state, $S\left(\mathrm{S}|\mathrm{M}\right)_{\rho_{\mathrm{SM}}} = -S\left(\rho_{\mathrm{S}}\right)$ due to Schmidt decomposition (theorem 4.5.1). Then eq. (5.7.3) reads

$$\langle H_{\mathrm{E}}\rangle' - \langle H_{\mathrm{E}}\rangle \geq kT\left(S\left(\rho'_{\mathrm{S}}\right) - S\left(\rho_{\mathrm{S}}\right)\right), \tag{5.7.4}$$

where the roles of the initial and final states are swapped with respect to eq. (5.7.2). Now we will show that this swapping allows us to perform the erasure of a mixed state of S towards a fixed pure state of S at *no* thermodynamic cost, thus overcoming Landauer's bound (5.7.2).

Suppose SM is initially in a *pure entangled* state $\Psi \in \mathsf{PurSt}_1\left(\mathrm{SM}\right)$; in this case ρ_{S} is mixed [100], and we have $S\left(\rho_{\mathrm{M}}\right) = S\left(\rho_{\mathrm{S}}\right) > 0$. Suppose we want to erase ρ_{S} to a fixed pure state ψ_0 of S. Now, let us consider the joint reversible evolution of SME to be $\mathcal{U}_{\mathrm{SM}} \otimes \mathcal{I}_{\mathrm{E}}$, where $\mathcal{U}_{\mathrm{SM}}$ is the reversible channel mapping Ψ to $\psi_0 \otimes \varphi_0$, where φ_0 is some pure state of the memory M ($\mathcal{U}_{\mathrm{SM}}$ exists thanks to transitivity). Clearly this reversible evolution respects the hypotheses of proposition 5.7.2 because $0 = S\left(\rho'_{\mathrm{M}}\right) < S\left(\rho_{\mathrm{M}}\right)$,

so it performs the erasure of ρ_S to ψ_0. Let us evaluate its thermodynamic cost $\langle H_E \rangle' - \langle H_E \rangle$. Since initially the environment is uncorrelated in the state $\gamma_{E,\beta}$, and the evolution is $\mathcal{U}_{SM} \otimes \mathcal{I}_E$, we have $\rho'_E = \gamma_{E,\beta}$, so the erasure occurs at zero thermodynamic cost. Note that eq. (5.7.4) is satisfied, indeed its left-hand side vanishes, while its right-hand side is negative and equal to $-kTS(\rho_S)$.

Again, pure-state entanglement, captured by the negativity of the conditional entropy, is a resource even in sharp theories with purification, in that it allows us to overcome Landauer's principle, and to perform erasure at no thermodynamic cost (cf. the quantum case in [192]).

5.8 Sufficiency of majorisation and unrestricted reversibility

After the above digression about entropies in sharp theories with purification, and their role in determining the equilibrium state when the expectation value of an observable is known, it is time to go back to the three resource theories of section 5.2, and to deal with the last question we left unanswered: is majorisation sufficient for the convertibility under RaRe channels? If not, when is it so?

The answer is *negative* [105]. First, let us try to give an intuitive reason for why it is so, which will guide us in the search for an actual counterexample. Note that majorisation is merely a condition on the spectra of states, and carries no information about the dynamics allowed by the theory. Instead, RaRe convertibility is all about the dynamics: if a theory does not have enough reversible dynamics, a state could majorise another without a RaRe channel transforming the former into the latter. So, a priori majorisation and RaRe convertibility might not related, and it is instead surprising that in quantum theory majorisation is sufficient to characterise the RaRe preorder. If we want to look for a counterexample, we need to focus on theories where there is a "restriction" on the allowed reversible channels. For this reason, let us focus on doubled quantum theory, presented in section 4.6, where reversible channels are constrained by the parity superselection rule.

Consider the following states of a doubled qubit (the first index de-

notes the sector $\mathcal{H}_0$ or $\mathcal{H}_1$):

$$\rho = \frac{1}{2}\left(|0,0\rangle\langle 0,0| + |0,1\rangle\langle 0,1|\right) \tag{5.8.1}$$

and

$$\sigma = \frac{1}{2}|0,0\rangle\langle 0,0| \oplus \frac{1}{2}|1,0\rangle\langle 1,0|, \tag{5.8.2}$$

where $\{|0,0\rangle, |0,1\rangle\}$ is an orthonormal basis for $\mathcal{H}_0$ and $\{|1,0\rangle, |1,1\rangle\}$ an orthonormal basis for $\mathcal{H}_1$. The key point here is that the state ρ is fully contained in one sector (the even parity sector), while the state σ is a mixture of two states in two different sectors.

The two states have the same spectrum, and therefore they are equivalent in terms of majorisation of their spectra. However, there is no RaRe channel transforming one state into the other. To see this, we use the following lemmas.

Lemma 5.8.1. *If any two states ρ and σ are equivalent under RaRe channels, there exists a reversible channel $\mathcal{U}$ such that $\sigma = \mathcal{U}\rho$.*

The proof can be found in [181].

Lemma 5.8.2. *No unitary matrices U in doubled quantum theory are such that $\sigma = U\rho U^\dagger$, where ρ and σ the states of eqs.* (5.8.1) *and* (5.8.2).

Proof. The proof is by contradiction. Suppose that one has $\sigma = U\rho U^\dagger$, for some unitary matrix U. Then, define the vectors $|\varphi_0\rangle := U|0,0\rangle$ and $|\varphi_1\rangle := U|0,1\rangle$. With this definition, we have

$$U\rho U^\dagger = \frac{1}{2}\left(|\varphi_0\rangle\langle\varphi_0| + |\varphi_1\rangle\langle\varphi_1|\right).$$

Now, $U\rho U^\dagger$ must be an allowed state in doubled quantum theory. This means that there are only two possibilities: either $|\varphi_0\rangle$ and $|\varphi_1\rangle$ belong to the same sector, or they do not. But σ is a mixture of states in both sectors. Hence, $|\varphi_0\rangle$ and $|\varphi_1\rangle$ must belong to different sectors, if the relation $U\rho U^\dagger = \sigma$ is to hold. At this point, there are only two possibilities: either

$$\begin{cases} U|0,0\rangle = |0,0\rangle \\ U|0,1\rangle = |1,0\rangle \end{cases}$$

or

$$\begin{cases} U|0,0\rangle = |1,0\rangle \\ U|0,1\rangle = |0,0\rangle \end{cases}.$$

However, neither of these conditions can be satisfied by a unitary matrix in doubled quantum theory: every unitary matrix satisfying either condition would map the valid state $|0,+\rangle = \frac{1}{\sqrt{2}}(|0,0\rangle + |0,1\rangle)$ into the *invalid* state $\frac{1}{\sqrt{2}}(|0,0\rangle + |1,0\rangle)$, which is forbidden by the parity superselection rule. □

Since unitary channels are the only reversible channels in doubled quantum theory, we conclude that no RaRe channel can convert ρ into σ (and vice versa), despite being equivalent in terms of majorisation.[12] Summarising: in general, in sharp theories with purification, majorisation is *not* sufficient for the convertibility via RaRe channels. Clearly, this means that in general, in sharp theories with purification, the three resource theories give rise to inequivalent preorders on resources.

If instead majorisation is sufficient for RaRe convertibility, the three preorders become equivalent: this is the missing step to close the chain of implications (5.3.2). In fact, the sufficiency of majorisation is equivalent to the equivalence of the three resource preorders.

5.8.1 Unrestricted reversibility

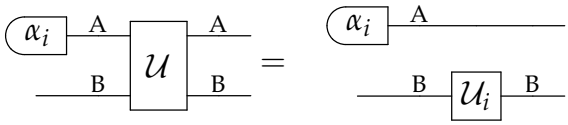
The condition for the equivalence of the three resource theories can be expressed in three, mutually equivalent ways, corresponding to three axioms independently introduced by different authors.

Axiom 5.8.3 (Permutability [70,73])**.** *Every permutation of every pure maximal set can be implemented by a reversible channel.*

Axiom 5.8.4 (Strong Symmetry [87, 113])**.** *For every two pure maximal sets, there exists a reversible channel that converts the states in one set into the states in the other.*

[12] Note that a very similar counterexample can be constructed in extended classical theory using the coherent composition of classical bits. However, in this theory, at the level of single classical systems, majorisation is still sufficient for RaRe convertibility.

Axiom 5.8.5 (Reversible Controllability [114]). *For every pair of systems* A *and* B, *every pure maximal set* $\{\alpha_i\}_{i=1}^d$ *of system* A *and every set of reversible channels* $\{\mathcal{U}_i\}_{i=1}^d$ *on system* B, *not necessarily distinct, there exists a reversible channel* $\mathcal{U}$ *on the composite system* AB *such that*



for every $i \in \{1, \ldots, d\}$.

This last axiom states the possibility of implementing control-reversible transformations in a globally reversible fashion.

Permutability, Strong Symmetry, and Reversible Controllability are logically distinct requirements. For example, Strong Symmetry implies Permutability, but the converse is not true in general, as shown in the following example.

Example 5.8.6. Consider the square bit [62]. Here the state space is a square, and the pure states are its vertices. The group of reversible channels is the symmetry group of the square, which is the dihedral group D_4. Every pair of vertices is a set of perfectly distinguishable pure states. Fig. 5.4 shows the situation for the pure states

$$\alpha_1 = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} \qquad \alpha_2 = \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} \qquad \alpha_3 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix},$$

where the third component gives the normalisation. The pure observation-test $\{a_1, a_2\}$, where

$$a_1 = \frac{1}{2}\begin{pmatrix} 0 & 1 & 1 \end{pmatrix} \qquad a_2 = \frac{1}{2}\begin{pmatrix} 0 & -1 & 1 \end{pmatrix},$$

is the perfectly distinguishing test for the two sets $\{\alpha_1, \alpha_2\}$ and $\{\alpha_1, \alpha_3\}$.

Now, since every set of perfectly distinguishable pure states has two elements, the only non-trivial permutation of the elements of such a set is the transposition. This permutation can be implemented by considering the reflection across the axis of the segment connecting the two points. Hence the square bit satisfies Permutability. On the other hand, the square

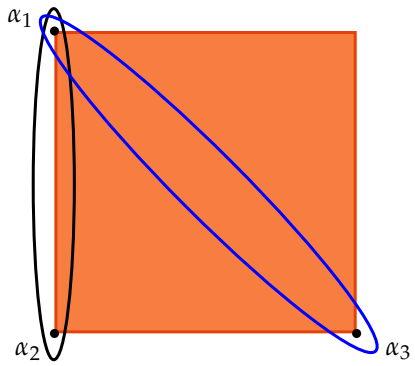


Figure 5.4: Normalised states of the square bit. The two sets $\{\alpha_1, \alpha_2\}$ (circled in black) and $\{\alpha_1, \alpha_3\}$ (circled in blue) consist of perfectly distinguishable pure states. Permutability holds, because every permutation of every pair of perfectly distinguishable pure states can be implemented by a reversible channel, corresponding to a symmetry of the square. However, no reversible channel can transform α_2 into α_3 while leaving α_1 unchanged. Hence, Strong Symmetry cannot hold for the square bit.

bit does *not* satisfy Strong Symmetry. Consider the two maximal sets $\{\alpha_1, \alpha_2\}$ and $\{\alpha_1, \alpha_3\}$. There are no reversible channels mapping the former to the latter because no symmetries of the square map a side to a diagonal.

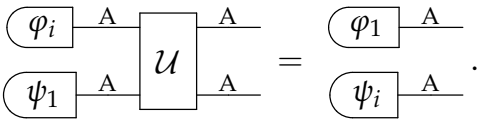
Although different in general, Permutability, Strong Symmetry, and Reversible Controllability become equivalent in sharp theories with purification.

Proposition 5.8.7. *In every sharp theory with purification, Permutability, Strong Symmetry, and Reversible Controllability are equivalent requirements.*

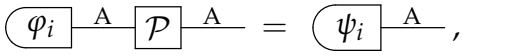
Proof. The implication "Strong Symmetry ⇒ Permutability" follows immediately from the definitions. The implication "Strong Symmetry ⇒ Reversible Controllability" was proved by Lee and Selby [114] using Causality, Purification, and the property that the product of two pure states is pure, which is guaranteed by Purity Preservation. Hence, we only need to prove the implications "Permutability ⇒ Strong Symmetry" and "Reversible Controllability ⇒ Strong Symmetry".

Let us prove that Permutability implies Strong Symmetry. The first part of the proof is similar to [70, theorem 30]. Consider two maximal

sets of perfectly distinguishable pure states $\{\varphi_i\}_{i=1}^d$ and $\{\psi_i\}_{i=1}^d$. Assuming Permutability, we will show that there exists a reversible channel $\mathcal{U}$ such that $\psi_i = \mathcal{U}\varphi_i$, for all $i = 1, \dots, d$. First of all, note that the states $\{\varphi_i \otimes \psi_j\}$ are pure (by Purity Preservation) and perfectly distinguishable. Then Permutability implies there exists a reversible channel $\mathcal{U}$ such that for all $i = 1, \dots, d$ [73]

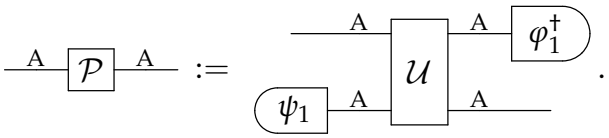


Applying the pure effect $\varphi_1^\dagger$ to both sides of the equation we obtain



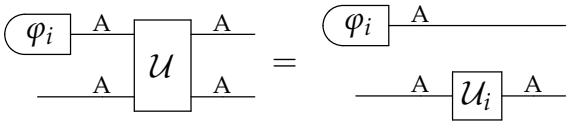
(5.8.3)

with



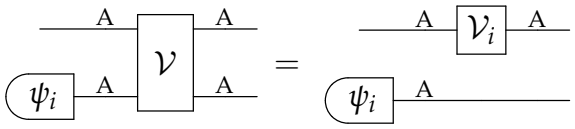
By construction, $\mathcal{P}$ is pure (by Purity Preservation, and because reversible channels are pure), and occurs with probability 1 on all the states $\{\varphi_i\}_{i=1}^d$. Moreover, the diagonalisation $\chi = \frac{1}{d}\sum_{i=1}^d \varphi_i$ implies that $\mathcal{P}$ occurs with probability 1 on every state because $(u|\mathcal{P}|\chi) = 1$, and χ is internal. Since $\mathcal{P}$ is a pure deterministic transformation on A, it must be reversible [67]. Hence, eq. (5.8.3) proves that the states $\{\varphi_i\}_{i=1}^d$ can be reversibly transformed into the states $\{\psi_i\}_{i=1}^d$. In short, Permutability implies Strong Symmetry.

Let us prove now that Reversible Controllability implies Strong Symmetry. Let $\{\varphi_i\}_{i=1}^d$ and $\{\psi_i\}_{i=1}^d$ be two pure maximal sets of a generic system A. Since reversible channels act transitively on pure states, for every $i \in \{1, \dots, d\}$, one can find a reversible channel $\mathcal{U}_i$ that maps ψ_1 into ψ_i, in formula $\mathcal{U}_i\psi_1 = \psi_i$. Moreover, Reversible Controllability implies that we can find a reversible channel $\mathcal{U}$ such that



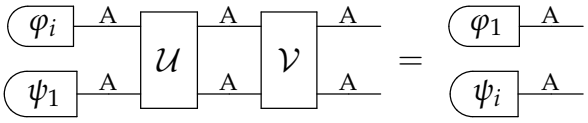
(5.8.4)

for every $i \in \{1, \dots, d\}$. Likewise, for every $i \in \{1, \dots, d\}$, one can always find a reversible channel $\mathcal{V}_i$ that transforms φ_i into φ_1, in formula $\mathcal{V}_i \varphi_i = \varphi_1$. And again, one can find a reversible channel $\mathcal{V}$ such that



(5.8.5)

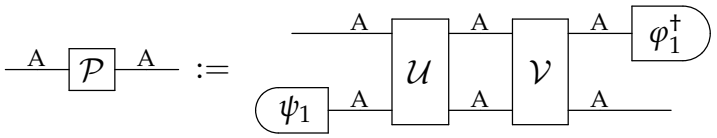
for every $i \in \{1, \dots, d\}$. Combining eqs. (5.8.4) and (5.8.5) with the definition of $\mathcal{U}_i$ and $\mathcal{V}_i$, we obtain



for every i. Hence, one has

, (5.8.6)

with



.

By the same argument used in the first part of the proof, we conclude that $\mathcal{P}$ is a reversible channel. Hence, eq. (5.8.6) implies that the set $\{\varphi_i\}_{i=1}^d$ can be reversibly converted into the set $\{\psi_i\}_{i=1}^d$. In short, Reversible Controllability implies Strong Symmetry. □

Since Permutability, Strong Symmetry, and Reversible Controllability are equivalent in the present context, we conflate them into a single notion.

Definition 5.8.8. A sharp theory with purification has *unrestricted reversibility* if the theory satisfies Permutability, or Strong Symmetry, or Reversible Controllability.

The fact that three desirable properties of GPTs become equivalent under our axioms gives further evidence that the axioms of sharp theories with purification capture an important structure of physical theories.

Examples of sharp theories with purification and unrestricted reversibility are quantum theory on real and complex Hilbert space.

5.8.2 When the three resource theories are equivalent

Now we characterise exactly when the RaRe, Noisy, and Unital Resource theories are equivalent in terms of state convertibility. Owing to the inclusions RaRe $\subseteq$ Noisy $\subseteq$ Unital, a sufficient condition for the equivalence is that the convertibility under unital channels implies the convertibility under RaRe channels, or in other words, that majorisation is sufficient for RaRe convertibility. The characterisation is as follows.

Theorem 5.8.9. *In every sharp theory with purification, the following statements are equivalent:*

1. *the RaRe, noisy, and unital Resource theories are equivalent in terms of state convertibility;*

2. *the theory has unrestricted reversibility.*

Proof. To prove the implication 2 $\Rightarrow$ 1, it is enough to show that unrestricted reversibility implies that majorisation is sufficient for the RaRe preorder [110]. Consider two states ρ and σ, diagonalised as $\rho = \sum_{i=1}^{d} p_i \alpha_i$ and $\sigma = \sum_{i=1}^{d} q_i \alpha'_i$. Suppose $\mathbf{p} \succeq \mathbf{q}$, then $\mathbf{q} = D\mathbf{p}$ for some doubly stochastic matrix D. By Birkhoff's theorem [112, 185] D can be written as a convex combination of permutation matrices $D = \sum_k \lambda_k \Pi_k$, where the Π_k's are permutation matrices, and $\{\lambda_k\}$ is a probability distribution. Therefore $\mathbf{q} = \sum_k \lambda_k \Pi_k \mathbf{p}$; specifically, this means that $q_i = \sum_k \lambda_k \sum_{j=1}^{d} (\Pi_k)_{ij} p_j$ for every $i \in \{1, \dots, d\}$. Therefore, we have

$$\sigma = \sum_{i=1}^{d} q_i \alpha'_i = \sum_{i=1}^{d} \sum_k \lambda_k \sum_{j=1}^{d} (\Pi_k)_{ij} p_j \alpha'_i = \sum_k \lambda_k \sum_{j=1}^{d} p_j \sum_{i=1}^{d} (\Pi_k)_{ij} \alpha'_i \tag{5.8.7}$$

Now, $\sum_{i=1}^{d} (\Pi_k)_{ij} \alpha'_i$ is a pure state, given by $\alpha'_{\pi_k(j)}$, for a suitable permutation $\pi_k \in S_d$, the symmetric group with d elements. By unrestricted reversibility, the permutation π_k is implemented by a reversible channel $\mathcal{V}_k$. Moreover, by unrestricted reversibility there exists a reversible channel $\mathcal{U}$ such that $\mathcal{U}\alpha_j = \alpha'_j$ for every $j \in \{1, \dots, d\}$. Defining $\mathcal{U}_k = \mathcal{V}_k \mathcal{U}$, we then have

$$\mathcal{U}_k \alpha_j = \mathcal{V}_k \alpha'_j = \alpha'_{\pi_k(j)} = \sum_{i=1}^{d} (\Pi_k)_{ij} \alpha'_i,$$

which, combined with eq. (5.8.7), yields

$$\sigma = \sum_k \lambda_k \sum_{j=1}^{d} p_j \mathcal{U}_k \alpha_j = \sum_k \lambda_k \mathcal{U}_k \rho.$$

Hence, $\rho \succeq_{\text{RaRe}} \sigma$.

To prove the implication $1 \Rightarrow 2$, we show that condition 1 implies the validity of Strong Symmetry. Let $\{\alpha_i\}_{i=1}^d$ and $\{\alpha'_i\}_{i=1}^d$ be two pure maximal sets, and let $\{p_i\}_{i=1}^d$ be a probability distribution, with $p_1 > p_2 > \ldots > p_d > 0$. Consider the two states ρ and σ defined by $\rho = \sum_{i=1}^d p_i \alpha_i$, and $\sigma = \sum_{i=1}^d p_i \alpha'_i$. Since the two states ρ and σ have the same eigenvalues, the majorisation criterion guarantees that ρ can be converted into σ, and σ into ρ, by a unital channel (theorem 5.3.4). Now, our hypothesis is that convertibility under unital channels implies convertibility under RaRe channels. The mutual convertibility of ρ and σ under RaRe channels implies that there exists a reversible channel $\mathcal{U}$ such that $\sigma = \mathcal{U}\rho$ (lemma 5.8.1). Applying the effect $\alpha_1'^{\dagger}$ to both sides of the equality $\sigma = \mathcal{U}\rho$, we obtain

$$p_1 = \left(\alpha_1'^{\dagger} \middle| \sigma\right) = \sum_{j=1}^{d} p_j \left(\alpha_1'^{\dagger} \middle| \mathcal{U} \middle| \alpha_j\right) = \sum_{j=1}^{d} D_{1j} p_j \leq p_1,$$

having used the fact that $D_{ij} := \left(\alpha_i'^{\dagger} \middle| \mathcal{U} \middle| \alpha_j\right)$ are the entries of a doubly stochastic matrix (lemma 5.3.1). The above condition is satisfied only if $\left(\alpha_1'^{\dagger} \middle| \mathcal{U} \middle| \alpha_1\right) = 1$. By the state-effect duality (theorem 4.2.9), this condition is equivalent to the condition

$$\mathcal{U}\alpha_1 = \alpha'_1. \tag{5.8.8}$$

Now, decompose the states ρ and σ as $\rho = p_1 \alpha_1 + (1 - p_1)\, \rho_1$ and $\sigma = p_1 \alpha'_1 + (1 - p_1)\, \sigma_1$, where

$$\rho_1 := \frac{\sum_{i=2}^d p_i \alpha_i}{\sum_{i=2}^d p_i}$$

and

$$\sigma_1 := \frac{\sum_{i=2}^d p_i \alpha'_i}{\sum_{i=2}^d p_i}.$$

Combining eq. (5.8.8) with the equality $\mathcal{U}\rho = \sigma$, we obtain the condition $\mathcal{U}\rho_1 = \sigma_1$. Applying to ρ_1 and σ_1 the same argument we used for ρ and σ, we obtain the equality $\mathcal{U}\alpha_2 = \alpha'_2$. Iterating the procedure $d-1$ times, we finally obtain the equality $\mathcal{U}\alpha_i = \alpha'_i$ for every $i \in \{1, \dots, d\}$. Hence, every two pure maximal sets are connected by a reversible channel. □

Theorem 5.8.9 gives necessary and sufficient conditions for the equivalence of the three resource theories of microcanonical thermodynamics. In addition, it provides a thermodynamic motivation for the condition of unrestricted reversibility: the equivalence of the three resource theories of purity. Again, thermodynamics constrains the underlying structure of a physical theory. We see that from this argument we can rule out doubled quantum theory and extended classical theory: in those these theories, majorisation is *not* sufficient, and they do *not* satisfy unrestricted reversibility.

The results of this section can be summed up in the following theorem:

Theorem 5.8.10. *In every sharp theory with purification and unrestricted reversibility, the following are equivalent*

1. $\rho \succeq_{\text{RaRe}} \sigma$
2. $\rho \succeq_{\text{noisy}} \sigma$
3. $\rho \succeq_{\text{unital}} \sigma$
4. $\mathbf{p} \succeq \mathbf{q}$

for arbitrary normalised states ρ and σ, where $\mathbf{p}$ and $\mathbf{q}$ are the spectra of ρ and σ, respectively.

Proof. The implications $1 \Rightarrow 2$ and $2 \Rightarrow 3$ follow from the inclusions (5.2.8). The implication $3 \Rightarrow 4$ follows from theorem 5.3.4. The implication $4 \Rightarrow 1$ follows from theorem 5.8.9. □

Theorem 5.8.9 tells us that the RaRe, noisy, and unital resource theories are all equivalent in terms of state convertibility. It is important to stress that the equivalence holds despite the fact that the three sets of operations are generally different. Since the preorders $\succeq_{\text{RaRe}}$, $\succeq_{\text{noisy}}$, and $\succeq_{\text{unital}}$ coincide, we can say that the RaRe, noisy, and unital resource theories define the same notion of resource, which one may rightfully call "*purity*". Accordingly, we will talk about "the resource theory of purity" without specifying the set of free operations.

An important consequence of the equivalence is that the RaRe, noisy, and unital resource theories have the same quantitative measures of resourcefulness.

Proposition 5.8.11. *Let* $P : \mathsf{St}_1(\mathrm{A}) \to \mathbb{R}$ *be a real-valued function on the state space of system* A. *If* P *is a monotone under one of the sets* RaRe, Noisy *and* Unital, *then it is a monotone under all the other sets.*

In the light of proposition 5.4.2, this means that *all* purity monotones in *all three* resource theories are of the form $P(\rho) = f(\mathbf{p})$, where f is a Schur-convex function, and $\mathbf{p}$ the spectrum of ρ. Dually, all mixedness monotones are of the form $M(\rho) = f(\mathbf{p})$, where this time f is a Schur-concave function.

5.9 Entanglement-thermodynamics duality

We conclude the chapter by showing that sharp theories with purification and unrestricted reversibility exhibit a fundamental duality between the resource theory of purity and the resource theory of pure bipartite entanglement [100], where free operations are LOCC channels [149–151]. The content of the duality is that a pure bipartite state is more entangled than another if and only if the marginal states of the latter are purer than the marginal states of the former. More formally, the duality can be stated as follows [100].

Definition 5.9.1. A theory satisfies the *entanglement-thermodynamics duality* if, for every pair of systems A and B, and every pair of pure states $\Phi, \Psi \in \mathsf{PurSt}_1(\mathrm{AB})$ the following are equivalent:

1. Ψ can be converted into Φ by local operations and classical communication[13];
2. the marginal of Φ on system A can be converted into the marginal of Ψ on system A by a RaRe channel;
3. the marginal of Φ on system B can be converted into the marginal of Ψ on system B by a RaRe channel.

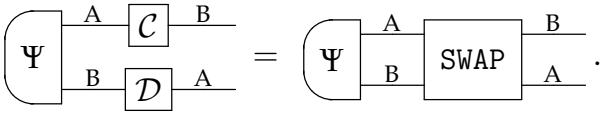
[13] Note that classical communication between the two parties can be easily modelled as an example of classical control, allowed by Causality: Bob chooses what to do based on the classical outcome of a previous test performed by Alice.

The duality can be illustrated by the diagrams

$$\begin{array}{ccc} \Psi & \xrightarrow{\text{LOCC}} & \Phi \\ {\scriptstyle \text{tr}_\text{B}}\downarrow & & \downarrow{\scriptstyle \text{tr}_\text{B}} \\ \rho_\text{A} & \xleftarrow{\text{RaRe}} & \rho'_\text{A} \end{array} \qquad \begin{array}{ccc} \Psi & \xrightarrow{\text{LOCC}} & \Phi \\ {\scriptstyle \text{tr}_\text{A}}\downarrow & & \downarrow{\scriptstyle \text{tr}_\text{A}} \\ \rho_\text{B} & \xleftarrow{\text{RaRe}} & \rho'_\text{B} \end{array}.$$

Our earlier work [100] showed that the entanglement-thermodynamics duality can be proved from four axioms: Causality, Purity Preservation, Purification, and Local Exchangeability—the last defined as follows.

Axiom 5.9.2 (Local Exchangeability [100])**.** *For every pair of systems* A *and* B*, and for every pure state* $\Psi \in \mathsf{PurSt}_1(\text{AB})$ *there exist two channels* $\mathcal{C} \in \mathsf{Transf}(\text{A},\text{B})$ *and* $\mathcal{D} \in \mathsf{Transf}(\text{B},\text{A})$*, which in general depend on* Ψ*, such that*



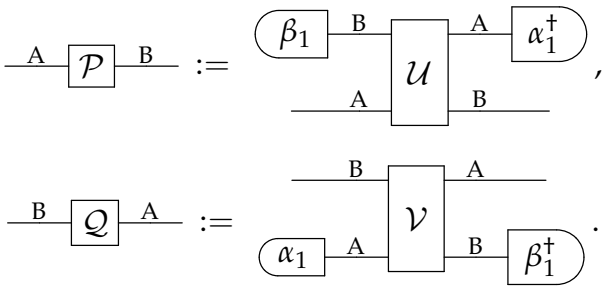
Since Causality, Purity Preservation, and Purification are already assumed among our axioms, proving the entanglement-thermodynamics duality is reduced to proving the validity of Local Exchangeability.

Proposition 5.9.3. *Every sharp theory with purification and unrestricted reversibility satisfies Local Exchangeability.*

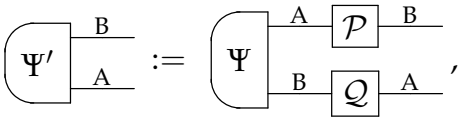
Proof. Let $\Psi \in \mathsf{PurSt}_1(\text{AB})$ be a generic pure state and let ρ_A and ρ_B be its marginal states, diagonalised as $\rho_\text{A} = \sum_{i=1}^{r} p_i \alpha_i$ and $\rho_\text{B} = \sum_{i=1}^{r} p_i \beta_i$, where $p_i > 0$ for all $i = 1, \dots, r$, and $r \leq \min\{d_\text{A}, d_\text{B}\}$. Here we are invoking Schmidt decomposition (theorem 4.5.1), by which the marginals of a pure bipartite state have the same spectrum up to vanishing elements. Now, we extend the set of eigenstates of ρ_A and ρ_B to two pure maximal sets. Without loss of generality assume $d_\text{A} \leq d_\text{B}$. By Permutability, there must exist a reversible channel $\mathcal{U} \in \mathsf{Transf}(\text{BA},\text{AB})$ such that[14] $\mathcal{U}(\beta_1 \otimes \alpha_i) = \alpha_1 \otimes \beta_i$ for every $i \in \{1, \dots, d_\text{A}\}$. Similarly, there must exist a reversible channel $\mathcal{V} \in \mathsf{Transf}(\text{BA},\text{AB})$ such that $\mathcal{V}(\beta_i \otimes \alpha_1) = \alpha_i \otimes \beta_1$ for every

[14] Strictly speaking BA $\neq$ AB, whereas Permutability refers to states of the same system. This can be easily accommodated by inserting the SWAP channel suitably.

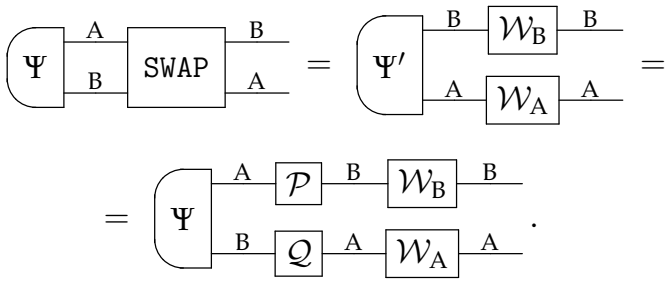
$i \in \{1, \dots, d_{\mathrm{A}}\}$ At this point, we define the pure transformations



and the pure state



where the purity of $\mathcal{P}$, $\mathcal{Q}$, and Ψ' follows from Purity Preservation. Like in the proof of proposition 5.8.7, we can prove that $\mathcal{P}$ and $\mathcal{Q}$ are in fact channels, so $u_{\mathrm{B}}\mathcal{P} = u_{\mathrm{A}}$ and $u_{\mathrm{A}}\mathcal{Q} = u_{\mathrm{B}}$. Hence Ψ' and $\mathtt{SWAP}\Psi$ have the same marginals. Then, the uniqueness of purification applied to both systems A and B (viewed as purifying systems of one another) implies that there exist two reversible channels $\mathcal{W}_{\mathrm{A}}$ and $\mathcal{W}_{\mathrm{B}}$ such that



Hence, we have shown that there exist two local *pure* channels $\mathcal{C} := \mathcal{W}_{\mathrm{B}}\mathcal{P}$ and $\mathcal{D} := \mathcal{W}_{\mathrm{A}}\mathcal{Q}$ that reproduce the action of the swap channel on the state Ψ. □

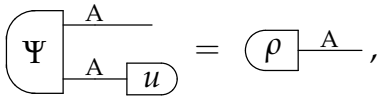
Note that Local Exchangeability is implemented by *pure* channels in sharp theories with purification and unrestricted reversibility.

To sum up, every sharp theory with purification and unrestricted reversibility satisfies the entanglement-thermodynamics duality. As a consequence of the duality, mixedness monotones, characterised at the end

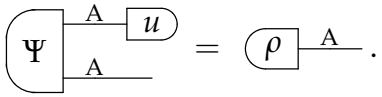
of the previous section, are in one-to-one correspondence with measures of pure bipartite entanglement. For example, Shannon-von Neumann entropy of the marginals of a pure bipartite state can be regarded as the *entanglement entropy* [145,197,198], an entropic measure of entanglement that is playing an increasingly important role in quantum field theory [199,200] and condensed matter [201].

In passing, we also mention that the validity of Local Exchangeability implies that every state admits a *symmetric purification* [100, theorem 3], in the following sense.

Definition 5.9.4. Let ρ be a state of system A and let Ψ be a pure state of AA. We say that Ψ is a *symmetric purification* of ρ if



and



With the above notation, we have the following.

Proposition 5.9.5. *In every sharp theory with purification and unrestricted reversibility, every state admits a symmetric purification.*

Chapter 6

Conclusions and outlook

In this DPhil thesis, the first doctoral thesis on this topic, we studied the foundations of thermodynamics and statistical mechanics using the tool of general probabilistic theories. In particular, we focused on microcanonical thermodynamics, describing systems where the energy is known and fixed. For the first time, we extended the microcanonical framework to arbitrary physical theories, beyond the known and well-studied cases of classical and quantum theory. Being a general paradigm, one expects thermodynamics to be valid in all physical theories, yet we discovered that, in order to have a physically meaningful microcanonical thermodynamics, we need to impose two requirements on the underlying theory. These two requirements are:

1. for every system there is a unique invariant probability measure;

2. the product of microcanonical states is still a microcanonical state.

Requirement 1 implies that for every system, once the energy is fixed, the microcanonical state is uniquely determined. This is because that invariant probability measure is used to define the microcanonical state. Requirement 2 expresses the stability of the equilibrium state under parallel composition, and it is similar in spirit to the notion of *complete passivity* for quantum thermal states [202–204]. In other words, there is no "activation process" when composing equilibrium states that could bring us out of equilibrium (and therefore extract work).

Specifically, requirement 1 is fully equivalent to one of the conditions (transitivity) that appeared in several reconstructions of quantum theory

from first principles: the fact that for every pair of pure states there exists a reversible channel connecting them [61, 68, 70, 83, 84, 86]. Therefore, our results offer a new perspective on transitivity, this time from thermodynamics: transitivity is a necessary condition to have a well-posed microcanonical thermodynamics. Since thermodynamics is not considered a fundamental theory, but rather emergent from an underlying theory through the paradigm of statistical mechanics, we usually expect that it is the underlying theory to constrain the thermodynamic behaviour. Instead, in our case we found the opposite: the reasonable thermodynamic desiderata of requirements 1 and 2 constrain the underlying theory.

For theories satisfying requirements 1 and 2 we set up a resource-theoretic treatment, where the microcanonical state was taken as the only free state, and we chose three different sets of free operations, with similar definitions to quantum theory:

1. random reversible (RaRe) channels;
2. noisy operations;
3. unital channels.

We studied these three resource theories in a class of physical theories, sharp theories with purification, which we propose as an axiomatic foundation for statistical mechanics. The fundamental feature of these theories is that they admit a level of description where all states are pure, and all evolutions reversible. By this we mean that every mixed state can be written as the marginal of a pure state of a larger system, and that every channel can be written as discarding one of the outputs of a bipartite reversible channel. Therefore, in these theories "impurity", partial information (viz. probabilistic mixtures), and irreversibility arise because of discarding. There is no need for the presence of an external agent who assigns probabilities subjectively or performs some coarse-graining. For this reason, sharp theories with purification are particularly suitable as candidate theories to solve the remaining tension between the pure and deterministic character of the fundamental dynamics, and the subjectivity of statistical ensembles. In these theories, mixed states, in particular thermal states, and their associated probabilities arise because in the thermodynamic description one is tracing out some degrees of freedom, typically those of the environment. This opens the way to the derivation of equilibrium states from typicality arguments based on entanglement [20, 21, 205, 206], since it is not

hard to see that every sharp theory with purification *must* have entangled states [100].

With these properties, sharp theories with purification appear to be reasonably close to quantum theory. Indeed the convex examples known so far [101, 103–109] are variations on quantum theory, obtained by imposing superselection rules [101, 103–105], or by considering real, instead of complex, amplitudes [106–109]. In fact, it is possible to prove that all sharp theories with purification are Euclidean Jordan algebras, and that they can exhibit at most second-order interference [87, 111, 207–209]. Besides these examples, we showed that, quite surprisingly, even classical theory can be extended to a sharp theory with purification. In this extension, at the level of single systems, classical systems look perfectly classical, and have all the properties of classical systems. What changes is the way they compose, because we need to have entangled states in composite systems: the composition of two classical systems is no longer a classical system.

After introducing sharp theories with purification, we studied their properties in relation to thermodynamics. The first is a state-effect duality, by which with every normalised pure state we can associate a unique normalised pure effect (and vice versa) that occurs with unit probability on that pure state. In [111] this was proved to be the stepping stone for the definition of the dagger of all transformations. The second important property is that all states can be diagonalised, i.e. written as a convex combination of perfectly distinguishable pure states, with unique coefficients, the eigenvalues of the state. Finally, sharp theories with purification satisfy requirements 1 and 2, so they admit a well-defined microcanonical thermodynamics. In these theories it is therefore possible to introduce the three resource theories.

We showed that the sets of free operations obey the same inclusion relations as in quantum theory: the set of RaRe channels is included in the set of noisy operations, which is in turn included in the set of unital channels. In addition, the convertibility of states under unital channels is fully described by majorisation on the spectra of states, thanks to the diagonalisation theorem. This allowed us to find mixedness monotones aplenty: they are Schur-concave functions on the spectrum of states. We were able to prove that, for a large class of them, their definition on the spectrum coincides with two definitions given in the GPT literature [163, 174, 175], based on pure measurements and pure preparations respectively.

Among mixedness monotones, we focused on Shannon-von Neumann

entropy, and we used it to define the generalised Gibbs states through Jaynes' maximum entropy principle [16,17]: we fixed the expectation value of an energy observable, and determined the state that maximises the Shannon-von Neumann entropy with that constraint, promoting it to an equilibrium state. This was the only part of the thesis where we departed from microcanonical thermodynamics and we explored the role of temperature. This was instrumental in proving an operational version of Landauer's principle in sharp theories with purification, where we linked the reduction of entropy in the system to the heat dissipated into the environment, following the approach by Reeb and Wolf [193].

We showed that, if we want the three resource theories to be equivalent and define the same preorder on states, the axioms of sharp theories with purification are not enough, and we *must* add a further principle, unrestricted reversibility, expressing the richness of the reversible dynamics of the theory. Again, from a thermodynamic requirement we derived a constraint on the underlying theory. This constraint restricts the set of allowed theories even further. For instance, doubled quantum theory and the coherent composites of extended classical theory are ruled out, and we get even closer to quantum theory. This could be an indication that quantum theory is eventually the only theory supporting a physically sensible thermodynamics.

From unrestricted reversibility, we proved that the three resource theories obey a duality with the resource theory of entanglement [100]. This connects the entanglement of pure bipartite states with the purity of their marginals: a pure state is more entangled than another if and only if the marginals of the latter are purer than the marginals of the former. In this way entanglement becomes a fertile ground for the foundations of thermodynamics, at least in the microcanonical setting.

The results of this thesis are only the surface of a deep operational structure, where thermodynamic and information-theoretic features are interwoven at the level of fundamental principles. The work initiated here clearly still has a lot of potentialities for further exploration. For instance, we still lack an operational derivation of strong subadditivity of Shannon-von Neumann entropy [210], and of the monotonicity of the relative entropy under the action of channels [211]. The proof of these results is notoriously difficult even in ordinary quantum theory, but the motivation is extremely strong, for they are the key to the derivation of the second law of thermodynamics [117] and of its quantum generalisations [41], not to

mention the consequences for information processing. An operational derivation of these results will shed a new light on quantum theory too, highlighting the principles leading to strong subadditivity, which at present are hidden behind the technical character of the existing proofs.

Another area of future research is the completion of the characterisation of microcanonical thermodynamics in sharp theories with purification. The aspects related to the thermodynamic limit are currently under investigation [99], but it is worth exploring other sides, such as catalysis, and the role of entropies in the single-shot work extraction [37, 212].

Clearly, the whole area of canonical thermodynamics, viz. for systems at a fixed *temperature*, rather than fixed energy, in general probabilistic theories is still largely unexplored. In this thesis, we briefly touched on it with the derivation of the generalised Gibbs states using the maximum entropy principle. However, we did not study the possible resource theories one can introduce in this setting [38–40, 42], nor did we study the existence of the canonical state in full generality, viz. without assuming the axioms of sharp theories with purification. To this end, a promising way to derive the canonical state is to harness the idea of complete passivity [202, 203], and enforce it in arbitrary physical theories. Alternatively, we can pursue the derivation of equilibrium states (including the canonical ones) in sharp theories with purification based on typicality and entanglement, along the lines of [20, 205, 206]. Results in this direction would bring further evidence that sharp theories with purification provide the appropriate ground for the construction of a well-founded statistical mechanics.

As we noted, the essence of sharp theories with purification is entanglement and the possibility of purifying states and transformations. What about classical theory? Is classical thermodynamics well founded? The answer to this question is not obvious nor easy. Clearly, our experience and the physical results, both theoretical and experimental, tell us that this is the case. Although we do it all the time, and it works very well, in classical theory there is no principle that *formally* justifies enlarging an open system to recover the isolated picture. The missing principle we are looking for is Purification [67], which is quintessentially thermodynamic. Indeed, when in thermodynamics we model an open system as part of an isolated system, where the other part has been discarded or neglected, it is precisely what happens in a theory where Purification is at work. There, all non-reversible channels can be seen as a reversible evolution in a larger system (the isolated system, where evolution is assumed to be reversible),

part of which has been discarded. Instead in classical theory, whatever is mixed and irreversible, it stays so, irrespective of how much we enlarge the system. This is precisely due to the lack of entanglement. Then why does thermodynamics work in classical theory?

A possible answer might be that classical thermodynamics works because Nature is ultimately quantum, where Purification holds. Therefore the formal underpinnings of classical thermodynamics may be found in quantum thermodynamics, of which it is a sub-theory. Another possible answer might come from our results: classical theory admits an extension to a sharp theory with purification with actual classical systems among its systems. This extended classical theory offers a new possibility for the foundations of classical statistical mechanics, allowing one to view classical ensembles, at least from a formal point of view, as arising from joint pure entangled states. At the same time, it allows us to export the results and the proof techniques of sharp theories with purification to classical theory. This motivates the following conjecture.

Conjecture. *Every theory with a "well-behaved" thermodynamics can be extended to a sharp theory with purification.*

As we currently lack a formal definition of "well-behaved" thermodynamics, our conjecture is not a mathematical statement for the time being, but rather an open research programme. Addressing this programme directly will be rather hard, and will mean rigorously formulating a set of desiderata about thermodynamics, from which to *derive* the requirements that the underlying physical theory should meet. An example of thermodynamic desiderata is provided by Lieb and Yngvason's axioms [213], recently revisited from a quantum information perspective [214,215], which capture the fundamental structures underpinning the second law of thermodynamics. Connecting general probabilistic theories with Lieb and Yngvason's desiderata is a promising route to approach our conjecture, and produce new results in the axiomatic foundations of thermodynamics.

Acknowledgements

I am really grateful to a great number of people in these three years as a DPhil student. They have been great years, in terms of scientific and personal growth, and I can truly say I have learnt a lot during this time. All the people I have met, in some way or another, have taught me something, therefore they all deserve a thank you. Here I will mention those who deserve a special acknowledgement. I am rather sure to forget somebody, so I will apologise in advance to those I forgot to mention.

My first and biggest thanks goes to my supervisor Prof. Jon Barrett, a brilliant and sharp scientist, always keen to discuss topics in the foundations of quantum theory and, more generally, of physics. He has guided me in my DPhil research, and he has granted me the freedom to pursue my independent lines of research, always available to discuss my results and to provide feedback or interesting insights. Without his support and guidance, all this research would not have been possible.

Another big thanks goes to my collaborator and former supervisor Prof. Giulio Chiribella, with whom I did most of the research presented here. It has been a great pleasure to work and have exciting scientific discussions with him. I have learnt really a lot from him, and he has always provided very good advice and scientific support, as well as excellent guidance in the research project of this thesis.

I really want to thank my examiners, Prof. Sam Staton and Dr. Markus Müller, for the careful reading of my thesis and the insightful and stimulating comments they made, which led to a fruitful scientific discussion during my viva.

Then Prof. Bob Coecke deserves a special mention, for co-leading the Quantum Group at the Department of Computer Science, and for creating such a nice, relaxed, and friendly atmosphere. This is the ideal environment to foster new ideas and collaborations. I have had the opportunity to

collaborate with him various times, on a scientific article, on an MSc thesis supervision, and teaching the classes for his course. They all were very nice experiences.

Of all the other people in the Quantum Group, Dr. Matty Hoban deserves a special thanks. He has helped me a lot in several ways, sharing his knowledge and expertise, or simply providing guidance and encouragement in the less bright times one inevitably encounters in one's DPhil. I am indebted to him for his patience and support.

In fact all the Quantum Group deserve a thank you, starting from the faculty members, including Prof. Samson Abramsky and Dr. Jamie Vicary, to postdocs, Dr. Nathan Walk, Dr. Ognyan Oreshkov, Dr. Niel de Beaudrap, Dr. Matt Pusey, Dr. Stefano Gogioso, down to my fellow DPhil students Robin Lorenz, Subhayan Roy Moulik, Vojtěch Havlíček, Marietta Stasinou, Fatimah Ahmadi, David Reutter, Fabrizio Genovese. I want also to mention some former fellow DPhil students, who have successfully graduated: Dr. Ciarán Lee, Dr. John Selby, and Dr. John-Mark Allen. It has been a pleasure to work with them, and I have learnt a lot from them all.

Being at the University of Oxford I have had the opportunity to interact with brilliant researchers outside the my group. I would like to mention Dr. David Jennings, and Prof. Harvey Brown, for the interesting and inspiring discussions I have had with them.

I would like to thank all my collaborators, from Oxford and other institutions: Prof. Giulio Chiribella, Prof. Bob Coecke, Prof. Howard Barnum, Prof. Jonathan Oppenheim, Prof. Paweł Horodecki, Dr. Ciarán Lee, Dr. John Selby, Dr. Stefano Gogioso, Dr. Lídia del Rio, Dr. Philippe Faist, Dr. Roberto Salazar, Dr. Jarek Jorbicz, and Carlo Sparaciari. A particular thanks goes to Prof. Jonathan Oppenheim and Prof. Paweł Horodecki, who invited me to visit their institutions. Prof. Paweł Horodecki deserves a special mention, for being always so nice and friendly, available and willing to help, despite being very busy as one of the world-leading experts in quantum information theory. It has been a sheer pleasure to work with him and learn from his brilliant ideas.

I wish to thank also Perimeter Institute, especially Prof. Lucien Hardy and Prof. Rob Spekkens, for hosting me during my visits there; and the Paris Centre for Quantum Computing, especially Dr. Alexei Grinbaum, for inviting me to a QuPa meeting in May 2018. I had a great and fruitful time in these visits, and I benefited a lot from inspiring conversations.

I want to conclude the academic part of my acknowledgements by mentioning Prof. Paolo Ciatti, Prof. Pieralberto Marchetti, Prof. Kurt Lechner, Prof. Antonio Saggion, Prof. Paolo Villoresi at the University of Padua, and Dr. Francesco Sorge for their support, and for passing their passion for theoretical physics, quantum mechanics, thermodynamics, and the foundations of physics on to me.

I feel very grateful for the help and support from many (old and new) friends during my time at Oxford. First of all, I would like to thank Aris Filos-Ratsikas, who has been an amazing friend in these three years; I had a great time with him, even in less bright moments, and he has always given me very good advice. I wish to thank some other friends for their time together, it would be too long to thank each of them separately, therefore I will list them here: Stefano Steffenoni, Enzo Arnosti, Alex Kavvos, Philip Lazos, Ninad Rajagopal, Stefano Gogioso, Marietta Stasinou, Vanessa Monteventi, Rosalba García Millán, Sheng Peng, George Vere, Gabriele Damiani, and Stefania Monterisi. Thanks for making these DPhil years better and lighter.

Clearly, all this would not have been possible without the love and support of my Mum and Dad, to whom this thesis is dedicated. They have always supported me and my passions throughout my life, and have always encouraged me to work hard and do my best to achieve my goals. They have been my guidance and inspiration.

To conclude, I want to thank Miriam Saccon for her love, support, and patience during my DPhil and the writing of this thesis; without her this would have been a much harder time. Despite not being her area of expertise, she did a very good proofreading of earlier drafts of this thesis, and I am very grateful to her for this.

Finally, I acknowledge the support of the EPSRC doctoral training grant 1652538 "Information-theoretic foundations of quantum thermodynamics", and of the Oxford-Google DeepMind graduate scholarship.

Bibliography

[1] Maxwell J C. Illustrations of the dynamical theory of gases. Part I. On the motions and collisions of perfectly elastic spheres. Phil. Mag., 1860, 19:19–32.

[2] Maxwell J C. Illustrations of the dynamical theory of gases. Part II. On the process of diffusion of two or more kinds of moving particles among one another. Phil. Mag., 1860, 20:21–37.

[3] Boltzmann L. Lectures on Gas Theory, reprint 2011 ed. Dover Books on Physics. Mineola, New York: Dover Publications, 1896.

[4] Gibbs J W. Elementary Principles in Statistical Mechanics, reprint 2014 ed. Dover Books on Physics. Mineola, New York: Dover Publications, 1902.

[5] Callen H B. Thermodynamics and an Introduction to Thermostatistics, 2nd ed. Hoboken, NJ: Wiley, 1985.

[6] Huang K. Statistical Mechanics, 2nd ed. Hoboken, NJ: Wiley, 2000.

[7] Kardar M. Statistical Physics of Particles. Cambridge: Cambridge University Press, 2007.

[8] Kardar M. Statistical Physics of Fields. Cambridge: Cambridge University Press, 2007.

[9] Anderson M H, Ensher J R, Matthews M R, et al. Observation of Bose-Einstein Condensation in a Dilute Atomic Vapor. Science, 1995, 269(5221):198–201.

[10] London F. The λ-Phenomenon of Liquid Helium and the Bose-Einstein Degeneracy. Nature, 1938, 141:643–644.

[11] Birkhoff G D. Proof of the Ergodic Theorem. Proc. Natl. Acad. Sci., 1931, 17(12):656–660.

[12] von Neumann, J. Proof of the Quasi-Ergodic Hypothesis. Proc. Natl. Acad. Sci., 1932, 18(1):70–82.

[13] Khinchin A I. Mathematical Foundations of Statistical Mechanics, reprint 2013 ed. Dover Books on Mathematics. Mineola, New York: Dover Publications, 1949.

[14] Krylov N S. Works on the Foundations of Statistical Physics, reprint 2014 ed. Princeton: Princeton University Press, 1980.

[15] Fasano A, Marmi S. Analytical mechanics: an introduction. Oxford: Oxford University Press, 2006.

[16] Jaynes E T. Information Theory and Statistical Mechanics. Phys. Rev., 1957, 106:620–630.

[17] Jaynes E T. Information Theory and Statistical Mechanics. II. Phys. Rev., 1957, 108:171–190.

[18] Deffner S, Zurek W H. Foundations of statistical mechanics from symmetries of entanglement. New J. Phys., 2016, 18(6):063013.

[19] Schrödinger E. Discussion of Probability Relations between Separated Systems. Math. Proc. Cambridge Phil. Soc., 1935, 31:555–563.

[20] Popescu S, Short A J, Winter A. Entanglement and the foundations of statistical mechanics. Nat. Phys., 2006, 2(11):754–758.

[21] Goldstein S, Lebowitz J L, Tumulka R, et al. Canonical Typicality. Phys. Rev. Lett., 2006, 96:050403.

[22] Bocchieri P, Loinger A. Ergodic Foundation of Quantum Statistical Mechanics. Phys. Rev., 1959, 114:948–951.

[23] Lloyd S. Black Holes, Demons, and the Loss of Coherence: [PhD Dissertation]. Rockfeller University, 1988.

[24] Lubkin E, Lubkin T. Average quantal behavior and thermodynamic isolation. Int. J. Theor. Phys., 1993, 32(6):933–943.

[25] Gemmer J, Otte A, Mahler G. Quantum Approach to a Derivation of the Second Law of Thermodynamics. Phys. Rev. Lett., 2001, 86:1927–1930.

[26] Gemmer J, Michel M, Mahler G. Quantum Thermodynamics: Emergence of Thermodynamic Behavior Within Composite Quantum Systems, volume 784 of *Lecture Notes in Physics*. Heidelberg: Springer Verlag, 2009.

[27] Müller M P, Gross D, Eisert J. Concentration of Measure for Quantum States with a Fixed Expectation Value. Commun. Math. Phys., 2011, 303(3):785–824.

[28] Brandão F G S L, Cramer M. Equivalence of Statistical Mechanical Ensembles for Non-Critical Quantum Systems. arXiv:1502.03263 [quant-ph], 2015.

[29] Ithier G, Benaych-Georges F. Thermalisation of a quantum system from first principles. arXiv:1510.04352 [quant-ph], 2015.

[30] Goold J, Huber M, Riera A, et al. The role of quantum information in thermodynamics—a topical review. J. Phys. A, 2016, 49(14):143001.

[31] Millen J, Xuereb A. Perspective on quantum thermodynamics. New J. Phys., 2016, 18(1):011002.

[32] Vinjanampathy S, Anders J. Quantum thermodynamics. Contemp. Phys., 2016, 57:1–35.

[33] Horodecki M, Oppenheim J. (Quantumness in the context of) resource theories. Int. J. Mod. Phys. B, 2013, 27(01n03):1345019.

[34] Brandão F G S L, Gour G. Reversible Framework for Quantum Resource Theories. Phys. Rev. Lett., 2015, 115:070503.

[35] Chitambar E, Gour G. Quantum Resource Theories. arXiv:1806.06107 [quant-ph], 2018.

[36] Horodecki M, Horodecki P, Oppenheim J. Reversible transformations from pure to mixed states and the unique measure of information. Phys. Rev. A, 2003, 67:062104.

[37] Gour G, Müller M P, Narasimhachar V, et al. The resource theory of informational nonequilibrium in thermodynamics. Phys. Rep., 2015, 583:1–58.

[38] Janzing D, Wocjan P, Zeier R, et al. Thermodynamic Cost of Reliability and Low Temperatures: Tightening Landauer's Principle and the Second Law. Int. J. Theor. Phys., 2000, 39(12):2717–2753.

[39] Horodecki M, Oppenheim J. Fundamental limitations for quantum and nanoscale thermodynamics. Nat. Commun., 2013, 4:2059.

[40] Brandão F G S L, Horodecki M, Oppenheim J, et al. Resource Theory of Quantum States Out of Thermal Equilibrium. Phys. Rev. Lett., 2013, 111:250404.

[41] Brandão F G S L, Horodecki M, Ng N, et al. The second laws of quantum thermodynamics. Proc. Natl. Acad. Sci., 2015, 112(11):3275–3279.

[42] Faist P, Oppenheim J, Renner R. Gibbs-preserving maps outperform thermal operations in the quantum regime. New J. Phys., 2015, 17(4):043003.

[43] Ćwikliński P, Studziński M, Horodecki M, et al. Limitations on the Evolution of Quantum Coherences: Towards Fully Quantum Second Laws of Thermodynamics. Phys. Rev. Lett., 2015, 115:210403.

[44] Lostaglio M, Jennings D, Rudolph T. Description of quantum coherence in thermodynamic processes requires constraints beyond free energy. Nat. Commun., 2015, 6:6383.

[45] Lostaglio M, Korzekwa K, Jennings D, et al. Quantum Coherence, Time-Translation Symmetry, and Thermodynamics. Phys. Rev. X, 2015, 5:021001.

[46] Yunger Halpern N, Renes J M. Beyond heat baths: Generalized resource theories for small-scale thermodynamics. Phys. Rev. E, 2016, 93:022126.

[47] Yunger Halpern N, Faist P, Oppenheim J, et al. Microcanonical and resource-theoretic derivations of the thermal state of a quantum system with noncommuting charges. Nat. Commun., 2016, 7:12051.

[48] Guryanova Y, Popescu S, Short A J, et al. Thermodynamics of quantum systems with multiple conserved quantities. Nat. Commun., 2016, 7:12049.

[49] Lostaglio M, Jennings D, Rudolph T. Thermodynamic resource theories, non-commutativity and maximum entropy principles. New J. Phys., 2017, 19(4):043008.

[50] Masanes L, Oppenheim J. A general derivation and quantification of the third law of thermodynamics. Nat. Commun., 2017, 8:14538.

[51] Wilming H, Gallego R. Third Law of Thermodynamics as a Single Inequality. Phys. Rev. X, 2017, 7:041033.

[52] Sparaciari C, Oppenheim J, Fritz T. Resource theory for work and heat. Phys. Rev. A, 2017, 96:052112.

[53] Scharlau J, Müller M P. Quantum Horn's lemma, finite heat baths, and the third law of thermodynamics. Quantum, 2018, 2:54.

[54] Boes P, Wilming H, Eisert J, et al. Statistical ensembles without typicality. Nat. Commun., 2018, 9:1022.

[55] Sparaciari C, del Rio L, Scandolo C M, et al. The first law of general quantum resource theories. arXiv:1806.04937 [quant-ph], 2018.

[56] Leff H, Rex A F. Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing. Boca Raton, FL: CRC Press, 2002.

[57] Szilard L. On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings. Behav. Sci., 1964, 9(4):301–310.

[58] Landauer R. Irreversibility and Heat Generation in the Computing Process. IBM J. Res. Dev., 1961, 5(3):183–191.

[59] Bennett C H. The thermodynamics of computation—a review. Int. J. Theor. Phys., 1982, 21(12):905–940.

[60] Bennett C H. Notes on Landauer's principle, reversible computation, and Maxwell's Demon. Stud. Hist. Philos. Sci. B, 2003, 34(3):501–510.

[61] Hardy L. Quantum Theory From Five Reasonable Axioms. arXiv quant-ph/0101012, 2001.

[62] Barrett J. Information processing in generalized probabilistic theories. Phys. Rev. A, 2007, 75:032304.

[63] Barnum H, Barrett J, Leifer M, et al. Generalized No-Broadcasting Theorem. Phys. Rev. Lett., 2007, 99:240501.

[64] Janotta P, Hinrichsen H. Generalized probability theories: what determines the structure of quantum theory? J. Phys. A, 2014, 47(32):323001.

[65] Chiribella G, Spekkens R W, editors. Quantum Theory: Informational Foundations and Foils, volume 181 of *Fundamental Theories of Physics*. Dordrecht: Springer Netherlands, 2016.

[66] Barnum H, Wilce A. Post-Classical Probability Theory. Dordrecht: Springer Netherlands, 2016, pages 367–420.

[67] Chiribella G, D'Ariano G M, Perinotti P. Probabilistic theories with purification. Phys. Rev. A, 2010, 81:062348.

[68] Chiribella G, D'Ariano G M, Perinotti P. Informational derivation of quantum theory. Phys. Rev. A, 2011, 84:012311.

[69] Hardy L. Foliable Operational Structures for General Probabilistic Theories. Cambridge: Cambridge University Press, 2011, pages 409–442.

[70] Hardy L. Reformulating and reconstructing quantum theory. arXiv:1104.2066 [quant-ph], 2011.

[71] Chiribella G. Dilation of states and processes in operational-probabilistic theories. in: Coecke B, Hasuo I, Panangaden P, editors, Proceedings of the 11th workshop on Quantum Physics and Logic, Kyoto, Japan, 4-6th June 2014, 2014, 1–14.

[72] Chiribella G, D'Ariano G M, Perinotti P. Quantum from Principles. Dordrecht: Springer Netherlands, 2016, pages 171–221.

[73] Hardy L. Reconstructing Quantum Theory. Dordrecht: Springer Netherlands, 2016, pages 223–248.

[74] D'Ariano G M, Chiribella G, Perinotti P. Quantum Theory from First Principles: An Informational Approach. Cambridge: Cambridge University Press, 2017.

[75] Abramsky S, Coecke B. A categorical semantics of quantum protocols. in: Proceedings of the 19th Annual IEEE Symposium on Logic in Computer Science, 2004, 415–425.

[76] Coecke B. Kindergarten Quantum Mechanics: Lecture Notes. AIP Conference Proceedings, 2006, 810(1):81–98.

[77] Coecke B. Quantum picturalism. Contemp. Phys., 2010, 51:59–83.

[78] Selinger P. A Survey of Graphical Languages for Monoidal Categories. Berlin, Heidelberg: Springer, 2011, pages 289–355.

[79] Coecke B, Duncan R, Kissinger A, et al. Generalised Compositional Theories and Diagrammatic Reasoning. Dordrecht: Springer Netherlands, 2016, pages 309–366.

[80] Coecke B, Kissinger A. Picturing Quantum Processes: A First Course in Quantum Theory and Diagrammatic Reasoning. Cambridge: Cambridge University Press, 2017.

[81] Barnum H, Wilce A. Information Processing in Convex Operational Theories. Electronic Notes in Theoretical Computer Science, 2011, 270(1):3–15. Proceedings of the Joint 5th International Workshop on Quantum Physics and Logic and 4th Workshop on Developments in Computational Models (QPL/DCM 2008).

[82] Wilce A. Formalism and Interpretation in Quantum Theory. Found. Phys., 2010, 40(4):434–462.

[83] Dakić B, Brukner Č. Quantum Theory and Beyond: Is Entanglement Special? Cambridge: Cambridge University Press, 2011, pages 365–392.

[84] Masanes L, Müller M P. A derivation of quantum theory from physical requirements. New J. Phys., 2011, 13(6):063001.

[85] Wilce A. Four and a Half Axioms for Finite-Dimensional Quantum Probability. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pages 281–298.

[86] Masanes L, Müller M P, Augusiak R, et al. Existence of an information unit as a postulate of quantum theory. Proc. Natl. Acad. Sci., 2013, 110(41):16373–16377.

[87] Barnum H, Müller M P, Ududec C. Higher-order interference and single-system postulates characterizing quantum theory. New J. Phys., 2014, 16(12):123029.

[88] Dakić B, Brukner Č. The Classical Limit of a Physical Theory and the Dimensionality of Space. Dordrecht: Springer Netherlands, 2016, pages 249–282.

[89] Müller M P, Masanes L. Information-Theoretic Postulates for Quantum Theory. Dordrecht: Springer Netherlands, 2016, pages 139–170.

[90] Chiribella G, Scandolo C M. Conservation of information and the foundations of quantum mechanics. EPJ Web Conf., 2015, 95:03003.

[91] Pikovski I, Vanner M R, Aspelmeyer M, et al. Probing Planck-scale physics with quantum optics. Nat. Phys., 2006, 8(5):393–397.

[92] Bartlett S D, Rudolph T, Spekkens R W. Reconstruction of Gaussian quantum mechanics from Liouville mechanics with an epistemic restriction. Phys. Rev. A, 2012, 86:012103.

[93] Spekkens R W. Quasi-Quantization: Classical Statistical Theories with an Epistemic Restriction. Dordrecht: Springer Netherlands, 2016, pages 83–135.

[94] Coecke B, Fritz T, Spekkens R W. A mathematical theory of resources. Inform. Comput., 2016, 250:59–86. Quantum Physics and Logic.

[95] Fritz T. Resource convertibility and ordered commutative monoids. Math. Structures Comput. Sci., 2015, pages 1–89.

[96] del Rio L, Krämer L, Renner R. Resource theories of knowledge. arXiv:1511.08818 [quant-ph], 2015.

[97] Krämer L, del Rio L. Currencies in resource theories. arXiv:1605.01064 [quant-ph], 2016.

[98] East R. The thermodynamic limit in the resource-theoretic framework: [Master Thesis]. UK: University of Oxford, 2017.

[99] East R, Scandolo C M. The thermodynamic limit in the resource-theoretic framework for general physical theories. in preparation.

[100] Chiribella G, Scandolo C M. Entanglement and thermodynamics in general probabilistic theories. New J. Phys., 2015, 17(10):103027.

[101] Chiribella G, Scandolo C M. Entanglement as an axiomatic foundation for statistical mechanics. arXiv:1608.04459 [quant-ph], 2016.

[102] Kitaev A, Mayers D, Preskill J. Superselection rules and quantum protocols. Phys. Rev. A, 2004, 69:052326.

[103] D'Ariano G M, Manessi F, Perinotti P, et al. Fermionic computation is non-local tomographic and violates monogamy of entanglement. Europhys. Lett., 2014, 107(2):20009.

[104] D'Ariano G M, Manessi F, Perinotti P, et al. The Feynman problem and fermionic entanglement: Fermionic theory versus qubit theory. Int. J. Mod. Phys. A, 2014, 29(17):1430025.

[105] Chiribella G, Scandolo C M. Microcanonical thermodynamics in general physical theories. New J. Phys., 2017, 19(12):123043.

[106] Stückelberg E C. Quantum theory in real Hilbert space. Helv. Phys. Acta, 1960, 33:458.

[107] Araki H. On a characterization of the state space of quantum mechanics. Commun. Math. Phys., 1980, 75(1):1–24.

[108] Wootters W K. Local accessibility of quantum states. in: Zurek W H, editor, Proceedings of Complexity, entropy and the physics of information, pages 39–46. Westview Press, 1990.

[109] Hardy L, Wootters W K. Limited Holism and Real-Vector-Space Quantum Theory. Found. Phys., 2012, 42(3):454–473.

[110] Chiribella G, Scandolo C M. Operational axioms for diagonalizing states. in: Heunen C, Selinger P, Vicary J, editors, Proceedings of the 12th International Workshop on Quantum Physics and Logic, Oxford, U.K., July 15-17 2015, 2015, 96–115.

[111] Barnum H, Lee C M, Scandolo C M, et al. Ruling out Higher-Order Interference from Purity Principles. Entropy, 2017, 19(6):253.

[112] Marshall A W, Olkin I, Arnold B C. Inequalities: Theory of Majorization and Its Applications. Springer Series in Statistics. New York: Springer, 2011.

[113] Müller M P, Ududec C. Structure of Reversible Computation Determines the Self-Duality of Quantum Theory. Phys. Rev. Lett., 2012, 108:130401.

[114] Lee C M, Selby J H. Generalised phase kick-back: the structure of computational algorithms from physical principles. New J. Phys., 2016, 18(3):033023.

[115] Scandolo C M, Salazar R, Korbicz J K, et al. Is it possible to be objective in every physical theory? arXiv:1805.12126 [quant-ph], 2018.

[116] Nielsen M A, Chuang I L. Quantum Computation and Quantum Information, 10th Anniversary ed. Cambridge: Cambridge University Press, 2011.

[117] Preskill J. Lecture notes on quantum computation. http://www.theory.caltech.edu/people/preskill/ph229.

[118] Wilde M M. Quantum Information Theory, 2nd ed. Cambridge: Cambridge University Press, 2017.

[119] Watrous J. The Theory of Quantum Information. Cambridge: Cambridge University Press, 2018.

[120] Barnum H, Graydon M A, Wilce A. Some Nearly Quantum Theories. in: Heunen C, Selinger P, Vicary J, editors, Proceedings of the

12th International Workshop on Quantum Physics and Logic, Oxford, U.K., July 15-17 2015, 2015, 59-70.

[121] Barnum H, Graydon M A, Wilce A. Composites and Categories of Euclidean Jordan Algebras. arXiv:1606.09331 [quant-ph], 2016.

[122] Galley T D, Masanes L. Any modification of the Born rule leads to a violation of the purification and local tomography principles. arXiv:1801.06414 [quant-ph], 2018.

[123] Coecke B, Paquette É. Categories for the Practising Physicist. Berlin, Heidelberg: Springer, 2011, pages 173–286.

[124] Bergia S, Cannata F, Cornia A, et al. On the actual measurability of the density matrix of a decaying system by means of measurements on the decay products. Found. Phys., 1980, 10(9):723–730.

[125] Janotta P, Lal R. Generalized probabilistic theories without the no-restriction hypothesis. Phys. Rev. A, 2013, 87:052131.

[126] Janotta P, Lal R. Non-locality in theories without the no-restriction hypothesis. in: Coecke B, Hoban M J, editors, Proceedings of the 10th International Workshop on Quantum Physics and Logic, Castelldefels (Barcelona), Spain, 17-19th July 2013, 2014, 84-89.

[127] Navascués M, Guryanova Y, Hoban M J, et al. Almost quantum correlations. Nat. Commun., 2015, 6:6288.

[128] Sainz A B, Guryanova Y, Acín A, et al. Almost-Quantum Correlations Violate the No-Restriction Hypothesis. Phys. Rev. Lett., 2018, 120:200402.

[129] Chiribella G, D'Ariano G M, Perinotti P. Quantum Theory, Namely the Pure and Reversible Theory of Information. Entropy, 2012, 14(10):1877–1893.

[130] Scandolo C M. Entanglement and thermodynamics in general probabilistic theories: [Master Thesis]. Padua, Italy: Università degli Studi di Padova, 2014.

[131] Chiribella G, D'Ariano G M, Perinotti P. Quantum Circuit Architecture. Phys. Rev. Lett., 2008, 101:060401.

[132] Chiribella G, D'Ariano G M, Perinotti P. Transforming quantum operations: Quantum supermaps. EPL, 2008, 83(3):30004.

[133] Chiribella G, D'Ariano G M, Perinotti P. Theoretical framework for quantum networks. Phys. Rev. A, 2009, 80:022339.

[134] Perinotti P. Causal Structures and the Classification of Higher Order Quantum Computations. Cham: Springer International Publishing, 2017, pages 103–127.

[135] Bisio A, Perinotti P. Axiomatic theory of Higher-Order Quantum Computation. arXiv:1806.09554 [quant-ph], 2018.

[136] Gour G. Comparison of Quantum Channels with Superchannels. arXiv:1808.02607 [quant-ph], 2018.

[137] Selby J H, Scandolo C M, Coecke B. Reconstructing quantum theory from diagrammatic postulates. arXiv:1802.00367 [quant-ph], 2018.

[138] Boyd S, Vandenberghe L. Convex Optimization. Cambridge: Cambridge University Press, 2004.

[139] Folland G B. Real Analysis: Modern Techniques and Their Applications, 2nd ed. Pure and Applied Mathematics. Hoboken, NJ: Wiley, 2013.

[140] Carathéodory C. Über den Variabilitätsbereich der Koeffizienten von Potenzreihen, die gegebene Werte nicht annehmen. Mathematische Annalen, 1907, 64(1):95–115.

[141] Steinitz E. Bedingt konvergente Reihen und konvexe Systeme. J. Reine Angew. Math., 1913, 143(143):128–175.

[142] Horodecki M. Entanglement Measures. Quantum Inf. Comput., 2001, 1(1):3–26.

[143] Bennett C H, Brassard G, Crépeau C, et al. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. Phys. Rev. Lett., 1993, 70:1895–1899.

[144] Bennett C H, Wiesner S J. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. Phys. Rev. Lett., 1992, 69:2881–2884.

[145] Horodecki R, Horodecki P, Horodecki M, et al. Quantum entanglement. Rev. Mod. Phys., 2009, 81:865–942.

[146] Ben Dana K, García Díaz M, Mejatty M, et al. Resource theory of coherence: Beyond states. Phys. Rev. A, 2017, 95:062327.

[147] Faist P, Berta M, Brandão F G S L. Thermodynamic Capacity of Quantum Processes. arXiv:1807.05610 [quant-ph], 2018.

[148] Gour G, Wilde M M. Entropy of a quantum channel. arXiv:1808.06980 [quant-ph], 2018.

[149] Bennett C H, Brassard G, Popescu S, et al. Purification of Noisy Entanglement and Faithful Teleportation via Noisy Channels. Phys. Rev. Lett., 1996, 76:722–725.

[150] Bennett C H, DiVincenzo D P, Smolin J A, et al. Mixed-state entanglement and quantum error correction. Phys. Rev. A, 1996, 54:3824–3851.

[151] Lo H K, Popescu S. Concentrating entanglement by local actions: Beyond mean values. Phys. Rev. A, 2001, 63:022301.

[152] Bennett C H, DiVincenzo D P, Fuchs C A, et al. Quantum nonlocality without entanglement. Phys. Rev. A, 1999, 59:1070–1091.

[153] Giles R. Mathematical Foundations of Thermodynamics. International series of monographs in pure and applied mathematics. Amsterdam: Elsevier Science, 2016.

[154] Alfsen E M, Shultz F W. State spaces of Jordan algebras. Acta Math., 1978, 140:441–487.

[155] Alfsen E M, Shultz F W. Geometry of state spaces of operator algebras. Mathematics: Theory & Applications. Basel: Birkhäuser, 2003.

[156] D'Ariano G M. Probabilistic theories: What is special about Quantum Mechanics? Cambridge: Cambridge University Press, 2010, pages 85–126.

[157] Piron C. Foundations of quantum physics. Mathematical Physics Monograph Series. San Francisco: Benjamin-Cummings Publishing Company, 1976.

[158] Wilce A. Conjugates, Filters and Quantum Mechanics. arXiv:1206.2897 [quant-ph], 2012.

[159] Wilce A. A Royal Road to Quantum Theory (or Thereabouts). arXiv:1606.09306 [quant-ph], 2016.

[160] Wilce A. A Royal Road to Quantum Theory (or Thereabouts), Extended Abstract. in: Duncan R, Heunen C, editors, Proceedings of the 13th International Conference on Quantum Physics and Logic, Glasgow, Scotland, 6-10 June 2016, 2017, 245-254.

[161] Spekkens R W. Evidence for the epistemic view of quantum states: A toy theory. Phys. Rev. A, 2007, 75:032110.

[162] Disilvestro L, Markham D. Quantum protocols within Spekkens' toy model. Phys. Rev. A, 2017, 95:052324.

[163] Barnum H, Barrett J, Orloff Clark L, et al. Entropy and information causality in general probabilistic theories. New J. Phys., 2010, 12(3):033024.

[164] Chiribella G. private communication.

[165] Wilce A. Formalism and Interpretation in quantum Theory. Found. Phys., 2010, 40(4):434–462.

[166] Selby J H, Coecke B. A Diagrammatic Derivation of the Hermitian Adjoint. Found. Phys., 2017, 47(9):1191–1207.

[167] Pfister C, Wehner S. An information-theoretic principle implies that any discrete physical theory is classical. Nat. Commun., 2013, 4:1851.

[168] Chiribella G, Yuan X. Bridging the gap between general probabilistic theories and the device-independent framework for nonlocality and contextuality. Inform. Comput., 2016, 250:15–49.

[169] Krumm M, Barnum H, Barrett J, et al. Thermodynamics and the structure of quantum theory. New J. Phys., 2017, 19(4):043025.

[170] Barnum H, Barrett J, Krumm M, et al. Entropy, majorization and thermodynamics in general probabilistic theories. in: Heunen C, Selinger P, Vicary J, editors, Proceedings of the 12th International Workshop on Quantum Physics and Logic, Oxford, U.K., July 15-17 2015, 2015, 43–58.

[171] Chiribella G, Yuan X. Measurement sharpness cuts nonlocality and contextuality in every physical theory. arXiv:1404.3348 [quant-ph], 2014.

[172] Krumm M. Thermodynamics and the Structure of Quantum Theory as a Generalized Probabilistic Theory. arXiv:1508.03299 [quant-ph], 2015. Master thesis.

[173] Harrow A. Coherent Communication of Classical Messages. Phys. Rev. Lett., 2004, 92:097902.

[174] Short A J, Wehner S. Entropy in general physical theories. New J. Phys., 2010, 12(3):033023.

[175] Kimura G, Nuida K, Imai H. Distinguishability measures and entropies for general probabilistic theories. Rep. Math. Phys., 2010, 66(2):175–206.

[176] Branford D, Dahlsten O C O, Garner A J P. On Defining the Hamiltonian Beyond Quantum Theory. Found. Phys., 2018, 48(8):982–1006.

[177] Horodecki M, Horodecki K, Horodecki P, et al. Local Information as a Resource in Distributed Quantum Systems. Phys. Rev. Lett., 2003, 90:100402.

[178] Uhlmann A. Sätze über Dichtematrizen. Wiss. Z. Karl-Marx Univ. Leipzig, 1971, 20:633–637.

[179] Uhlmann A. Endlich-dimensionale Dichtematrizen I. Wiss. Z. Karl-Marx Univ. Leipzig., 1972, 21:421–452.

[180] Uhlmann A. Endlich-dimensionale Dichtematrizen II. Wiss. Z. Karl-Marx Univ. Leipzig, 1973, 22:139–177.

[181] Müller M P, Masanes L. Three-dimensionality of space and the quantum bit: an information-theoretic approach. New J. Phys., 2013, 15(5):053040.

[182] Shor P W. Structure of Unital Maps and the Asymptotic Quantum Birkhoff Conjecture. presentation, 2010.

[183] Landau L J, Streater R F. On Birkhoff's theorem for doubly stochastic completely positive maps of matrix algebras. Linear Algebra Appl., 1993, 193:107–127.

[184] Mendl C B, Wolf M M. Unital Quantum Channels — Convex Structure and Revivals of Birkhoff's Theorem. Commun. Math. Phys., 2009, 289(3):1057–1086.

[185] Birkhoff G. Tres observaciones sobre el algebra lineal. Univ. Nac. Tucumán Rev. Ser. A, 1946, 5:147–151.

[186] Haagerup U, Musat M. Factorization and Dilation Problems for Completely Positive Maps on von Neumann Algebras. Commun. Math. Phys., 2011, 303(2):555–594.

[187] Gross D, Müller M P, Colbeck R, et al. All Reversible Dynamics in Maximally Nonlocal Theories are Trivial. Phys. Rev. Lett., 2010, 104:080402.

[188] Hardy G H, Littlewood J E, Pólya G. Some simple inequalities satisfied by convex functions. Messenger Math., 1929, 58(145–152):310.

[189] Rényi A. On Measures of Entropy and Information. in: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics, Berkeley: University of California Press, 1961, 547–561.

[190] von Neumann J. Mathematical foundations of quantum mechanics. Princeton: Princeton University Press, 1955.

[191] Shannon C E. A Mathematical Theory of Communication. Bell Syst. Tech. J., 1948, 27(3):379–423.

[192] del Rio L, Åberg J, Renner R, et al. The thermodynamic meaning of negative entropy. Nature, 2011, 474:61–63.

[193] Reeb D, Wolf M M. An improved Landauer principle with finite-size corrections. New J. Phys., 2014, 16(10):103011.

[194] Ramsey N F. Thermodynamics and Statistical Mechanics at Negative Absolute Temperatures. Phys. Rev., 1956, 103:20–28.

[195] Landsberg P T. Heat engines and heat pumps at positive and negative absolute temperatures. J. Phys. A, 1977, 10(10):1773.

[196] Dunkel J, Hilbert S. Consistent thermostatistics forbids negative absolute temperatures. Nat. Phys., 2014, 10(1):67–72.

[197] Plenio M B, Virmani S. An introduction to entanglement measures. Quant. Inf. Comp., 2007, 7(1):1–51.

[198] Janzing D. Entropy of Entanglement. Berlin, Heidelberg: Springer, 2009, pages 205–209.

[199] Ryu S, Takayanagi T. Holographic Derivation of Entanglement Entropy from the anti-de Sitter Space/Conformal Field Theory Correspondence. Phys. Rev. Lett., 2006, 96:181602.

[200] Nishioka T, Ryu S, Takayanagi T. Holographic entanglement entropy: an overview. J. Phys. A, 2009, 42(50):504008.

[201] Eisert J, Cramer M, Plenio M B. *Colloquium*: Area laws for the entanglement entropy. Rev. Mod. Phys., 2010, 82:277–306.

[202] Pusz W, Woronowicz S L. Passive states and KMS states for general quantum systems. Commun. Math. Phys., 1978, 58(3):273–290.

[203] Lenard A. Thermodynamical proof of the Gibbs formula for elementary quantum systems. J. Stat. Phys., 1978, 19(6):575–586.

[204] Sparaciari C, Jennings D, Oppenheim J. Energetic instability of passive states in thermodynamics. Nat. Commun., 2017, 8:1895.

[205] Müller M P, Dahlsten O C O, Vedral V. Unifying Typical Entanglement and Coin Tossing: on Randomization in Probabilistic Theories. Commun. Math. Phys., 2012, 316(2):441–487.

[206] Müller M P, Oppenheim J, Dahlsten O C O. The black hole information problem beyond quantum theory. J. High Energy Phys., 2012, 2012(9).

[207] Sorkin R D. Quantum mechanics as quantum measure theory. Mod. Phys. Lett. A, 1994, 9(33):3119–3127.

[208] Sorkin R D. Quantum Classical Correspondence: The 4th Drexel Symposium on Quantum Nonintegrability. Boston: International Press, 1997, chapter Quantum Measure Theory and its Interpretation, pages 229–251.

[209] Lee C M, Selby J H. Higher-Order Interference in Extensions of Quantum Theory. Found. Phys., 2017, 47(1):89–112.

[210] Lieb E H, Ruskai M B. Proof of the strong subadditivity of quantum-mechanical entropy. J. Math. Phys., 1973, 14(12):1938–1941.

[211] Petz D. Monotonicity of quantum relative entropy revisited. Rev. Math. Phys., 2003, 15(1):79–91.

[212] Dahlsten O C O, Renner R, Rieper E, et al. Inadequacy of von Neumann entropy for characterizing extractable work. New J. Phys., 2011, 13(5):053015.

[213] Lieb E H, Yngvason J. The physics and mathematics of the second law of thermodynamics. Phys. Rep., 1999, 310(1):1–96.

[214] Weilenmann M, Krämer L, Faist P, et al. Axiomatic Relation between Thermodynamic and Information-Theoretic Entropies. Phys. Rev. Lett., 2016, 117:260601.

[215] Weilenmann M, Krämer Gabriel L, Faist P, et al. Smooth entropy in axiomatic thermodynamics. arXiv:1807.07583 [quant-ph], 2018.