COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

Non-classical effects on generalized quantum channels

Dissertation

April 2019   Martin Plávala

COMENIUS UNIVERSITY, BRATISLAVA

FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

Non-classical effects on generalized quantum channels

Dissertation

Field:	9-1-9 Applied mathematics
Advisor:	Mgr. Anna Jenčová, DrSc.

April 2019   Martin Plávala

Abstract

In this review we introduce the framework of general probabilistic theories. Our aim is to present the standard definitions and results as they are used in the current research but we aim to provide more in-depth explanations than the ones we can find in a usual research article. We also prove several standard results in a self-contained way that are meant to provide the reader with consistent introduction to the framework.

keywords: general probabilistic theories, channels, incompatibility

Abstrakt

V práci zavedieme pojem všeobecnej probabilistickej teórie. Naším cieľom je prezentovať štandardné definície a výsledky tak, ako sú v súčasnosti používané, ale chceme poskytnúť obsažnejšie vysvetlenia ako tie, ktoré sa zväčša uvádzajú vo vedeckých článkoch. Rovnako dokážeme niekoľko výsledkov, ktoré majú čitateľovi poskytnúť konzistentný úvod do všeobecných probabilistických teórii.

kľúčové slová: všeobecné probabilistické teórie, kanály, nekompatibilita

Contents

List of Symbols

$\mathbb{N}$ natural numbers

$\mathbb{R}$ real numbers

$\mathbb{R}^+$ non-negative real numbers

conv   convex hull

span   linear hull

int   interior of a set

ext   extreme points of convex set or pure states of state space

id   identity map

K state space

$A(K)$ vector space of affine functions $K \to \mathbb{R}$

$A(K)^+$ cone of positive affine functions $K \to \mathbb{R}^+$

$E(K)$ effect algebra, i.e. the set affine functions $K \to [0,1]$

$A(K)^*$ vector space of linear functionals $A(K) \to \mathbb{R}$

$A(K)^{*+}$ cone of positive linear functionals $A(K) \to \mathbb{R}$

$\mathfrak{S}(E)$ states on the effect algebra E

$K_A \dot\otimes K_B$ minimal tensor product of state space K_A and K_B

$K_A \tilde\otimes K_B$ real tensor product of state space K_A and K_B

$K_A \hat\otimes K_B$ maximal tensor product of state space K_A and K_B

$\mathfrak{C}(K_A \to K_B)$ channels that map state space K_A to state space K_B

$\mathcal{H}$ finite-dimensional complex Hilbert space

$\dim(\mathcal{H})$ dimension of the Hilbert space $\mathcal{H}$

$\mathbb{1}$ the identity matrix

$B_h(\mathcal{H})$ self-adjoint operators on $\mathcal{H}$

$\mathrm{Tr}(A)$ trace of the operator $A \in B_h(\mathcal{H})$

$\mathfrak{D}_{\mathcal{H}}$ states on $\mathcal{H}$, $\rho \in \mathfrak{D}_{\mathcal{H}}$ if and only if $\rho \geq 0$, $\mathrm{Tr}(\rho) = 1$

$\mathcal{E}(\mathcal{H})$ effects on $\mathcal{H}$, $E \in \mathcal{E}(\mathcal{H})$ if and only if $0 \leq E \leq \mathbb{1}$

S_n simplex with n vertexes

S square state space

Introduction

The aim of this review is to introduce the framework of general probabilistic theories in a well motivated and self-contained way and to present the standard results and constructions. We expect the reader to have some familiarity with basic notions from linear algebra, convex geometry and quantum information theory. In case it would be needed, we recommend the reader to [1, 2] to find more information on linear algebra and convex analysis, to [3] to find more information on tensor products, to [4, 5] to find more information on topology and to [6–9] to find more information on quantum information theory. Linear programming and cone programming is a useful tool; although we will not use it in this review, we would like to point the reader to [10, 11] as these tools were recently used to obtain several interesting results [12–15].

This review is organized as follows: in Sec. 1 we introduce the framework of general probabilistic theories and the axioms on which we are going to build the framework of general probabilistic theories. We also discuss some of the limitations of our axioms and we even point out some of the assumptions that are usually swept under the rug. In Sec. 2 we introduce all of the basic definitions and results. We describe the convex geometry of the state space, the effect algebra and the duality between them. We also introduce some examples of probabilistic theories. In Sec. 3 we introduce products of state spaces and effect algebras. We begin with the possible notions of tensor products, but we also describe the direct products and direct convex sums. In Sec. 4 we introduce transformations of state spaces. The main paradigm of our approach is to see everything as a channel, even measurements and instruments corresponding to a measurement. In Sec. 5 we define compatibility of channels and we prove some of the main results and in Sec. 6 we introduce the phenomenons of steering and Bell non-locality. The last sections of the review are brief and their aim is only to introduce the discussed topics and phenomenons.

1 Introduction to probabilistic theories

Let's build a new theory. How do we do that? And if we succeed in building the new theory, how do we compare the new theory to the ones we already know? As strange as it may seem, these questions are not entirely hypothetical as at the moment questions such as quantum supremacy or post quantum cryptography are gaining importance. In order to start comparing theories or to say what properties of a given theory specify it among others we clearly need a unified framework, where all theories would be comparable on an equal setting. One of such frameworks is known as general probabilistic theories (or GPTs for short), which is a geometrical framework that heavily relies on the ideas of convexity and probabilistic mixtures. The settings of GPTs is great if one wishes to get a deeper understanding of some of the non-classical effects that we observe in quantum theory and the framework is general enough to include both finite dimensional classical and quantum theory.

The basic premise on which we are going to build the framework of GPTs and on which many other frameworks are built is the idea that there are states and there are measurements. As straightforward as it may seem, one has to remember that this basic premise is already something that we assume on the grounds that we have never seen anything else. A measurement is a procedure (or an equivalence class of procedures) that give us some information about the system. A state of a system is a mathematical object (or an equivalence class of objects) that fully describes the outcomes of all possible measurements that we can perform on said system. Here we are assuming that we can distinguish measurements from states, i.e. that we can tell what are measurements and what are states. As strange as it may seem in more general settings one may only have two sets of objects and if we decide that one of the sets are states then the other are measurements.

Before we proceed we are going to set up our naming convention. Since we can safely assume that states form a set, we are going to refer to this set as state space. In a similar manner we are going to referring to the set of measurements as to the set that includes all possible measurements that we can perform.

As first we are going to assume that we can prepare every state from the state space and that every state that we can prepare is included in the state space. This is quite straightforward as it is equivalent to saying that if something never happens then we do not have to deal with it and if something can happen then we do have to deal with it. We are implicitly assuming similar thing about measurements as if there is some procedure that yields information about the system we are assuming there is at least theoretical way to perform the given procedure. The assumption that we can perform every mathematically well defined measurement is called No-Restriction hypothesis and we will revisit this assumption in Subsec. 2.5

The next assumption is that randomness exists. It is usually stated that we can always flip a coin and generate some random data, but that implicitly assumes that a random coin exists; something that is not trivial for e.g. classical computers. We are going to argue that in many theories we do not have to assume the existence of randomness as whenever we will have a measurement in the theory which has non-deterministic outcomes then we can use said measurement to generate randomness; this is for example the case for quantum theory. The main aspect of randomness that we are going to use is that for any two states we can toss a random coin and based on the random outcome we can prepare either the first state or the second state. We are going to refer to this mixing using randomness as probabilistic mixtures and we are going to assume that we can perform the same with measurements. Moreover we are going to assume that an outcome of a measurement performed on a probabilistic mixture of two states is going to be a probabilistic mixture (with the same probabilities) of the measurement outcomes of the given states and that a probabilistic mixture of the measurements is going to give us a new measurement such that the outcome is again going to be a probabilistic mixture of the the outcomes of the two original measurements (with the same probabilities). We are going to formalize the idea of probabilistic mixtures using convex combinations.

We are going to assume that the set of states is bounded. This is because we are going to be interested in mapping the set of states to probabilities. If the set of states would not be bounded then we could not map the direction of recession to probabilities without violating the operational interpretation of convex combinations. Another possible approach is to not assume boundedness

of the set of state but later introduce relation of equivalence given as follows: two states are equivalent if the results of all measurements on the given two states are the same. Clearly we would like to factor the set of state with respect to this relation of equivalence as if we can not distinguish two states then we can as well assume that they represent the same state. One can see that this would also lead to obtaining a bounded set of states.

The last operational assumption we are going to assume is completeness, that is we are going to assume that if we can prepare a state that is arbitrary close to some other fixed state, then we can also prepare the limit state. The one thing that we have to specify is what it means that we can prepare a state arbitrary close to another one and we are going to interpret this as that the outcome of any measurement would be arbitrary close. This is a usable definition up to the point where one realizes that now we have to specify what it means that outcomes of two measurements are close to each other; we are going to address this problem soon.

There still are several assumptions that we are taking for granted. We are assuming that we know what information (as an outcome of a measurement) is and how to measure distances between measurement outcomes. This might seem obvious as what would be the point of a measurement if we couldn't interpret the information it gives in a reasonable sense yet one may argue that the way we do it does not have to be unique. Similarly, we are going to assume that probabilities are described by real numbers, which corresponds to our standard intuition of probabilities as relative frequencies, but arguments using relative frequencies seem rather strange when describing single shot experiments such as the probability of one photon passing through a polarization filter or the probability of a given candidate wining an election. It is an open question whether we can extend the notion of probability beyond our standard understanding, what would the generalization give us and whether it is actually necessary or not.

Apart from the all aforementioned assumption we are going to assume that both the state space and set of measurements will have a description using finite dimensional vector spaces. We are going to assume that the number of dimensions is finite as this will simplify quite a few of our calculations and in some cases it will be necessary.

2 Basic definitions

In this section we are going to define the state space, the effect algebra, the duality between them and present examples of state spaces and effect algebras. We are going to start by defining state space and then we will construct the effect algebra.

2.1 Definition of the state space

Definition 1. Let V denote a finite dimensional real vector space with the standard Euclidean topology. We say that $K \subset V$ is a state space if, according to the assumptions from Sec. 1, K is convex, closed and bounded.

We will consistently use K to denote a state space, unless stated otherwise. The first result is almost immediate.

Proposition 1. *Every state space is compact.*

Proof. It is known fact that every bounded subset of a real, finite dimensional vector space is completely bounded, which implies it is compact [5, Theorem 27.3]. $\square$

There are two types of states: states that are convex combinations of other states and states that are pure, i.e. they can not be expressed as convex combinations of other states. This is an important aspect as the pure states are more fundamental than the mixed states in the sense that we can see the mixed states as randomizations of pure states. This motivates the following definition.

Definition 2. We say that $x \in K$ is a pure state if for $\lambda \in [0,1] \subset \mathbb{R}$ and $y, z \in K$ we have that $x = \lambda y + (1 - \lambda)z$ implies $y = z$.

Pure states are also referred to as extreme points in the contexts of general theory of convex sets. Still one may argue that not all mixed states are equal as some may be mixed more than other. To characterize this property we introduce the notion of face.

Definition 3. Let $F \subset K$ be a convex set such that if for $\lambda \in (0,1)$ and $x, y \in K$ we have $\lambda x + (1 - \lambda)y \in F$ then also $x, y \in F$.

One can see that a pure state is a face that contains only one point. There are plethora of other results concerning the extreme points and faces of compact convex sets, see e.g. [1, Section 18]. We are going to present the following two.

Theorem 1 (Carathéodory). *Let $K \subset V$ be a state space and let $B \subset K$ be a set such that $K = \text{conv}(B)$, where* conv *denotes the convex hull. Then any $x \in K$ can be expressed as a convex combination of at most $\dim(V) + 1$ points from B.*

Proof. See [1, Theorem 17.1]. $\square$

Theorem 2. *Let K be a state space and let $\text{ext}(K)$ be the set of pure states of K then $K = \text{conv}(\text{ext}(K))$.*

Proof. See [1, Theorem 18.5]. $\square$

There is a special types of state spaces: polytopes. Polytopes are somehow simpler to deal with than a general state space and often it is simpler to prove results for polytopes.

Definition 4. We say that a state space K is a polytope if it has finitely many pure states, i.e. $\text{ext}(K)$ contains finitely many points.

Another special type of state space is a strictly convex state space.

Definition 5. We say that a state space K is strictly convex if for any $x, y \in K$ and $\lambda \in (0,1)$ the state $\lambda x + (1 - \lambda)y$ belongs to the interior of K, i.e. the smallest face containing $\lambda x + (1 - \lambda)y$ is K.

2.2 Definition of the effect algebra

Having defined a state space, we would want to define measurements, but for the time being we will abstain from doing it. For now, we will only define a more restricted notion of effect. Effects are maps that assign probabilities to states and as we will see in Subsec. 4.4 effects are the building blocks of measurements. For a given theory the set of effects will be called effect algebra, the name will be explained in Subsec. 2.3. In the following we will rely on the notion of affinity. We say that a map Φ is affine if for $x, y \in K$ and $\lambda \in [0, 1]$ we have $\Phi(\lambda x + (1 - \lambda)y) = \lambda \Phi(x) + (1 - \lambda)\Phi(y)$. Clearly affinity of the maps that we will work with follows from our probabilistic interpretation of convex combinations. Since we have assumed to describe probabilities by real numbers we are going to be interested in affine maps that assign real numbers to states.

Definition 6. $A(K)$ will denote the set of all affine maps $f : K \to \mathbb{R}$. $A(K)$ has a natural order given for $f, g \in A(K)$ as $f \leq g$ if and only if for all $x \in K$ we have $f(x) \leq g(x)$ and a norm $\| \cdot \|_{\sup}$ given as

$$\|f\|_{\sup} = \sup_{x \in K} |f(x)|.$$

Proposition 2. $A(K)$ is a partially ordered normed vector space.

Proof. The fact that linear combination of affine maps is again affine is obvious; it is also straightforward to see that $\| \cdot \|_{\sup}$ is a norm. At first note that since K is compact, then $\sup_{x \in K} |f(x)|$ is always finite. For $f, g \in A(K)$ and $x \in K$ we have that $f(x) = 0$ for all $x \in K$ implies $f = 0$, for $\alpha \in \mathbb{R}$ we have $(\alpha f)(x) = \alpha f(x)$ and $\sup_{x \in K} |(f + g)(x)| \leq \sup_{x \in K} |f(x)| + \sup_{x \in K} |g(x)|$. Also note that we can replace the supremum with maximum as K is closed.

Let $f, g, h \in A(K)$ and $x \in K$, then $f \leq f$ as $f(x) \leq f(x)$, i.e. $\leq$ is reflexive. Assume that $f \leq g$ and $g \leq f$, then we have $f(x) \leq g(x)$ and $g(x) \leq f(x)$ which implies $f(x) = g(x)$ for all $x \in K$ and $f = g$ so $\leq$ is antisymmetric. Assume that $f \leq g$ and $g \leq h$, then we have $f(x) \leq g(x) \leq h(x)$ for all $x \in K$ so $f \leq h$ which shows that $\leq$ is transitive. $\square$

We are also going to write $f \geq g$ whenever $g \leq f$. A well known fact is that the partial order over a vector space is equivalent to picking a cone inside the vector space. A cone P is a subset of a vector space such that for every $\lambda \in \mathbb{R}^+$, we have $v \in P$ if and only if $\lambda v \in P$, where $\mathbb{R}^+$ denotes the set of non-negative real numbers.

Definition 7. $A(K)^+$ will denote the cone of positive elements of $A(K)$, i.e.

$$A(K)^+ = \{f \in A(K) : f \geq 0\}.$$

A somewhat special role is played by the constant functions which will be denoted by the value they attain, e.g. $1 \in A(K)$ is a function such that $1(x) = 1$ for all $x \in K$.

We say that a cone P is pointed when $v \in P$ and $-v \in P$ implies $v = 0$ and we say that P is generating when for $P \subset W$ where W is a vector space and every $w \in W$ we have $v_1, v_2 \in P$ such that $w = v_1 - v_2$.

Proposition 3. $A(K)^+$ is a convex closed pointed and generating cone.

Proof. It is straightforward that $A(K)^+$ is convex. Assume that $\{f_n\}_{n=1}^{\infty}$ is a sequence of positive functions that converges to f, then for every $x \in K$ we have that $\{f_n(x)\}_{n=1}^{\infty}$ converges to $f(x)$ so $f_n(x) \geq 0$ for all $n \in \mathbb{N}$ implies $f(x) \geq 0$ and we get $f \geq 0$. To show that $A(K)^+$ is pointed assume that $f \geq 0$ and $f \leq 0$ at the same time, then we have $f(x) \geq 0$ and $f(x) \leq 0$ for all $x \in K$ which implies $f = 0$.

The proof that $A(K)^+$ is generating is actually somewhat interesting and useful to remember. Let $f \in A(K)$ and let $m = \min_{x \in K} f(x)$ then we have $f = (f - m) + m$ where $f - m \geq 0$ as $\min_{x \in K}(f - m)(x) = 0$. The interesting part is that we can always choose the negative part of f to be a constant function. Of course this is just a consequence of $1 \in \text{int}(A(K)^+)$ where int denotes the interior. $\square$

We will proceed with defining a dual object to the state space: the effect algebra. Effect algebra will be the set of all maps that assign probabilities to states and it will be important later on for the concept of measurement. We will also revisit the notion of effect algebra in Subsec. 2.3 where will compare our definition to a more abstract one as well as explain why it is called an algebra.

Definition 8. Let K be a state space then the effect algebra over K is denoted as $E(K)$ and it is defined as
$$E(K) = \{f \in A(K) : 0 \le f \le 1\}.$$
We call $f \in E(K)$ an effect. Moreover we will refer to the function $1 \in E(K)$ as unit effect.

Note that effect algebra is again a compact convex set, it has extreme points and the results of the Thm. 1 and Thm. 2 apply to it as well. The following is a simple yet important result.

Proposition 4. Let $f \in E(K)$, then also $1 - f \in E(K)$.

Proof. $1 - f \ge 0$ follows from $f \le 1$ and $1 - f \le 1$ follows from $f \ge 0$. $\qquad\square$

2.3 Relation to abstract definition of convex effect algebra

Effect algebras were originally introduced in [16] as a generalization of the projections in quantum theory and since then were heavily investigated [17–21], see [22] for a review. We are going to present the abstract definition as well as the results that put into context why we introduced the effect algebra the way we did.

Definition 9. An effect algebra is a system $(E, 0, 1, +)$ where E is a set, $0, 1 \in E$ and $+$ is a partially defined binary operation. Let $a, b \in E$ then we write $a \perp b$ if $a + b$ is defined. Let $a, b, c \in E$, then it must hold that

(E1) if $a \perp b$ then also $b \perp a$ and $a + b = b + a$,

(E2) if $a \perp b$ and $(a + b) \perp c$ then $b \perp c$, $a \perp (b + c)$ and $(a + b) + c = a + (b + c)$,

(E3) for every $a \in E$ there is $a' \in E$ such that $a \perp a'$ and $a + a' = 1$,

(E4) if $a \perp 1$ then $a = 0$.

Special class of effect algebras are convex effect algebras.

Definition 10. Effect algebra E is convex if for every $a \in E$ and $\lambda \in [0, 1] \subset \mathbb{R}$ there is an element $\lambda a \in E$ such that for all $\lambda, \mu \in [0, 1]$ and $a, b \in E$ we have

(C1) $\mu(\lambda a) = \lambda(\mu a)$,

(C2) if $\lambda + \mu \le 1$, then $\lambda a \perp \mu a$ and $\lambda a + \mu a = (\lambda + \mu)a$,

(C3) if $a \perp b$, then $\lambda a \perp \lambda b$ and $\lambda a + \lambda b = \lambda(a + b)$,

(C4) $1a = a$.

Special class of convex effect algebras are effect algebras which are intervals in real ordered vector spaces. As we will see, these effect algebras are closely related to our Def. 8.

Proposition 5. Let V be a real vector space with a pointed cone P, that is $P \cap -P = \{0\}$ and for $v, w \in V$ define the partial order $\ge$ given as $v \ge w$ if $v - w \ge 0$. Let $u \in P$ then the interval
$$[0, u] = \{v \in V : 0 \le v \le u\}$$
is a convex effect algebra with the partially defined binary operation of sum of vectors and $1 = u$. The convex structure is given by multiplication of vectors by scalars.

Proof. The proof is straightforward. Let $v, w, x \in [0, u]$, then $v + w = w + v$, and $v + w \le u$ if and only if $w + v \le u$. If $v + w \le u$ and $(v + w) + x \le u$ then from $(v + w) + x = v + w + x = v + (w + x)$ we have $w + x \le v + w + x \le u$. We define $v' = u - v$ as the unique element such that $v + v' = u$ and $v' \in [0, u]$ if and only if $v \in [0, u]$. At last if $v + u \le u$ then $v \le 0$ but also $v \ge 0$ so we must have $v = 0$. This shows that $[0, u]$ is an effect algebra.

Keep $v, w \in [0, u]$ and let $\lambda, \mu \in [0, 1]$. Clearly we have $\lambda(\mu v) = \lambda \mu v = \mu(\lambda v)$. If $\lambda + \mu \le 1$, then $\lambda v + \mu v = (\lambda + \mu)v \le v \le u$. If $v + w \le u$, then also $\lambda v + \lambda w = \lambda(v + w) \le v + w \le u$. At last, $1v = v$ is trivial. This shows that $[0, u]$ is convex effect algebra. $\qquad\square$

Definition 11. Let V be a real vector space with a pointed cone P and let $u \in P$, then we call $[0, u]$ linear effect algebra.

The following justifies why we used the term effect algebra in Def. 8.

Corollary 1. *Let K be a state space and let $E(K)$ be the effect algebra over K as given by Def. 8. Then $E(K)$ is a linear effect algebra with $u = 1$, $P = A(K)^+$ and $V = A(K)$.*

Proof. By definition we have $E(K) = [0, 1] \subset A(K)$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

We would like to mention one more result on the relation between linear effect algebras and convex effect algebras. As we have seen, every linear effect algebra is convex and one can actually prove that also every convex effect algebra is isomorphic to some linear effect algebra. To formulate the result we will first have to define the isomorphism.

Definition 12. Let E, F be effect algebras. A map $\Phi : E \to F$ is called additive if for $a, b \in E$, $a \perp b$ we have $\Phi(a) \perp \Phi(b)$ and $\Phi(a + b) = \Phi(a) + \Phi(b)$. An additive map Φ such that $\Phi(1) = 1$ is called morphism. A surjective morphism Φ such that for $a, b \in E$, $\Phi(a) \perp \Phi(b)$ implies $a \perp b$ is called isomorphism. If E, F are convex, then a morphism Φ such that for all $a \in E$ and $\lambda \in [0, 1]$ we have $\Phi(\lambda a) = \lambda \Phi(a)$ is called affine.

Theorem 3 (Gudder, Pulmannová). *Every convex effect algebra is affinely isomorphic to a linear effect algebra.*

Proof. See [23] for a proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

2.4 Duality between the state space and the effect algebra

There is an important duality in our construction that we will often use. To motivate what will follow consider this question: could two state spaces have the same effect algebra or is there a unique state space corresponding to a given effect algebra? What we will do to answer the question is to explore the set of all functionals that will have probabilities assigned by the effects; to do so we will have to look at some structure of the dual to $A(K)$. Since we intend to work with positivity of linear functionals the notion of dual cone will play an important role.

Definition 13. Let V be a real finite dimensional vector space and let $P \subset V$ be a cone. Let V^* denote the dual vector space of linear functionals then the cone dual to P is $P^* \subset V^*$ defined as

$$P^* = \{\psi \in V^* : \langle \psi, v \rangle \geq 0, \forall v \in P\}, \tag{1}$$

where for $v \in V$ and $\psi \in V^*$ we have denoted $\langle \psi, v \rangle$ the value that the functional ψ reaches on the vector v.

To motivate our exploration of the structure of the dual vector space of $A(K)$, note that for a fixed state $x \in K$ and some $f \in A(K)$ the evaluation functional $\phi_x : f \mapsto f(x)$ is a linear functional on $A(K)$, i.e. we can use any state to construct a linear functional on $A(K)$. Denote $A(K)^*$ the dual vector space to $A(K)$, it is uniquely defined as $A(K)$ is finite dimensional. For $\psi \in A(K)^*$ and $f \in A(K)$ we will denote $\langle \psi, f \rangle$ the value that ψ reaches on f. Dual cone to $A(K)^+$ is given as

$$A(K)^{*+} = \{\psi \in A(K)^* : \langle \psi, f \rangle \geq 0, \forall f \in A(K)^+\}.$$

Definition 14. Let $\mathfrak{S}(E(K)) \subset A(K)^*$ be given as

$$\mathfrak{S}(E(K)) = \{\psi \in A(K)^{*+} : \langle \psi, 1 \rangle = 1\}.$$

We will call $\mathfrak{S}(E(K))$ the state space over $E(K)$.

At this point one may be confused: K is the state space and $\mathfrak{S}(E(K))$ is a state space over $E(K)$, so it appears as a very poor naming convention. We will show that K is affinely isomorphic to $\mathfrak{S}(E(K))$ hence they can be considered the same object. Moreover this will be an important duality that we will use in the future. During the proof we will use the separation theorem which we will state for completeness.

Theorem 4 (Separation theorem)**.** *Let V be a finite dimensional real vector space and let $A, B \subset V$ be disjoint compact convex sets, i.e. $A \cap B = \emptyset$. Then there exists an affine function f on V such that*

$$\max_{x \in A} f(x) < 0 < \min_{y \in B} f(y).$$

Proof. See [11, Section 2.5] for a proof and note that as A, B are compact and disjoint, their Euclidean distance given by the standard Euclidean metric must be positive. □

The following is immediate and useful consequence.

Corollary 2. *Let $x, y \in K$, $x \neq y$ then there is $f \in E(K)$ such that $f(x) < f(y)$.*

Proof. Let $A = \{x\}$ and $B = \{y\}$ then as a result of Thm. 4 there is $f' \in A(K)$ such that $f'(x) < f'(y)$. If $f' \in A(K)^+$ then we can construct $f \in E(K)$ with said property as follows: let $M = \max_{z \in K} f'(z)$ then take $f = \frac{f'}{M}$, we have $f \in E(K)$.

If $f' \notin A(K)^+$ then let $m = \min_{z \in K} f'(z)$ and let $f'' = f' - m$; note that $f' \notin A(K)^+$ implies $m < 0$. $f'' \in A(K)^+$ and $f''(x) < f''(y)$. Now we can construct f as above. □

Note that in the same way we can use the separation theorem to prove that the effects are separated by states.

Theorem 5. $\mathfrak{S}(E(K))$ *is affinely isomorphic to K.*

Proof. We will proof the statement in two steps, first we will show that K is affinely isomorphic to a subset of $\mathfrak{S}(E(K))$ and then we will use the separation theorem to show that the subset is actually the whole set.

Let $x, y \in K$, $\lambda \in [0, 1]$ and $f \in A(K)$ and consider again the evaluation functional $\langle \phi_x, f \rangle = f(x)$. It is easy to see that $\phi_x \in A(K)^{*+}$ and $\langle \phi_x, 1 \rangle = 1$ for all $x \in K$, hence $\phi_x \in \mathfrak{S}(E(K))$. This gives rise to a map $\phi : K \to \mathfrak{S}(E(K))$, moreover this map is affine as

$$\langle \phi_{(\lambda x + (1-\lambda)y)}, f \rangle = f(\lambda x + (1 - \lambda)y) = \lambda f(x) + (1 - \lambda)f(y)$$
$$= \lambda \langle \phi_x, f \rangle + (1 - \lambda) \langle \phi_y, f \rangle$$

implies $\phi_{(\lambda x + (1-\lambda)y)} = \lambda \phi_x + (1 - \lambda)\phi_y$. Hence $\phi(K) \subset \mathfrak{S}(E(K))$, where $\phi(K) = \{\phi_x \in A(K)^* : x \in K\}$ and the map ϕ is affine. The fact that this is an isomorphism of K with a subset of $\mathfrak{S}(E(K))$ follows from the fact that for $x, y \in K$ if $x \neq y$ then $\phi_x \neq \phi_y$ follows from Coro. 2.

Now assume there is $\psi \in \mathfrak{S}(E(K))$ such that $\psi \notin \phi(K)$ then by Thm. 4 there is $f \in A(K)$ such that

$$\langle \psi, f \rangle < 0 < \min_{x \in K} f(x).$$

Observe that $0 < \min_{x \in K} f(x)$ implies that $f \in A(K)^+$ but then $\langle \psi, f \rangle < 0$ is a contradiction with $\psi \in A(K)^{*+}$. □

From now on we are going to drop the isomorphism ϕ and we will simply consider $K \subset A(K)^{*+}$. This will also allows us to define linear combinations of states but one must be careful as for $x, y \in K$, $\alpha, \beta \in \mathbb{R}$ and $f \in A(K)$ we will have

$$\langle \alpha x + \beta y, f \rangle = \alpha \langle x, f \rangle + \beta \langle y, f \rangle$$

and for $f = 1$

$$\langle \alpha x + \beta y, 1 \rangle = \alpha + \beta.$$

This may appear obvious now but it is very easy to make a blunder during calculations. Also for this reason we will append the coordinate expressions of states in Subsec. 2.6 with 1.

We already know that we could have started our construction by defining the effect algebra $E(K)$ instead of the state space K and we would obtain the same theory as $\mathfrak{S}(E(K)) = K$. Moreover Thm. 5 shows the same duality between the positive cones. We will proceed to explore the structure of $A(K)^{*+}$.

Definition 15. Let V be a vector space and let $P \subset V$ be a cone. The base of the cone P is a convex set $B \subset P$ such that for every $p \in P$, $p \neq 0$ there is a unique $x \in B$ and $\lambda \in \mathbb{R}^+$ such that

$$p = \lambda x.$$

Corollary 3. K *is the base of* $A(K)^{*+}$.

Proof. Let $\psi \in A(K)^{*+}$, $\psi \neq 0$, then we will show that $\langle \psi, 1 \rangle > 0$. If $\langle \psi, 1 \rangle = 0$ then also for every $f \in E(K)$ we have $\langle \psi, f \rangle \leq \langle \psi, 1 \rangle = 0$ as $f \leq 1$ so $\langle \psi, f \rangle = 0$. Then also for every $g \in A(K)^+$ we have $\langle \psi, g \rangle = 0$ as every $g \in A(K)^+$ is a multiple of some effect as we can just divide g by the maximum it takes on K and obtain an effect. In conclusion $\langle \psi, 1 \rangle = 0$ implies $\psi = 0$ for $\psi \in A(K)^{*+}$.

Take $\varphi = \frac{\psi}{\langle \psi, 1 \rangle}$, then $\varphi \in A(K)^{*+}$ and $\langle \varphi, 1 \rangle = 1$ so by Thm. 5 we have $\varphi \in K$. It is straightforward to see that the construction is unique. $\square$

Corollary 4. *The dual cone to* $A(K)^{*+}$ *is* $A(K)^+$, *i.e. we have*

$$A(K)^+ = \{f \in A(K) : \langle \psi, f \rangle \geq 0, \forall \psi \in A(K)^{*+}\}.$$

Proof. Note that $\langle \psi, f \rangle \geq 0$ for every $\psi \in A(K)^{*+}$ if and only if $\langle x, f \rangle \geq 0$ for every $x \in K$ as K is the base of $A(K)^{*+}$. The result follows. $\square$

For completeness we will show that any base of a convex, pointed and generating cone is a valid state space.

Proposition 6. *Let* V *be a finite dimensional real vector space, let* $P \subset V$ *be a convex, closed, pointed, generating cone and let* $K \subset P$ *be a base of* P, *then* K *is a state space.*

Proof. First of all we will show that we require P to be pointed in order to even have a base. Assume that P is not pointed and that it has a base K, then there is $0 \neq v \in P \cap (-P)$ such that for some $\lambda, \mu \in \mathbb{R}^+$ and $x, y \in K$ we have

$$\lambda x = v = -\mu y.$$

Then we have

$$\frac{\lambda}{\lambda + \mu} x + \frac{\mu}{\lambda + \mu} y = 0$$

so also $0 \in K$. This is a contradiction with K being a base of P, because now if $x \in K$ then also for all $\lambda \in [0,1]$ we have $\lambda x \in K$. It follows that if $p \in P$ such that for some $\mu \in \mathbb{R}^+$ and $y \in K$ we have $p = \mu y$ then the coefficient μ is not uniquely given as we also have $p = (2\mu)(\frac{1}{2}y)$ and we have already argued that $\frac{1}{2}y \in K$ as well.

Let $u : K \to \mathbb{R}$ be a constant function defined for $x \in K$ as $u(x) = 1$. We will extend u to a positive linear functional. Let $0 \neq p \in P$, then we have $p = \lambda x$ for some $\lambda \in \mathbb{R}^+$ and $x \in K$ and define $u(p) = \lambda$. This is well defined since λ is unique for the given p. Moreover let $u(0) = 0$ which is still consistent and continuous. Now we will show that u is affine on P: let $p, q \in P$ such that $p = \lambda x$, $q = \mu y$ for some $x, y \in K$ and $\lambda, \mu \in \mathbb{R}^+$ and let $\alpha \in [0,1]$, we have

$$\alpha p + (1 - \alpha)q = \alpha \lambda x + (1 - \alpha)\mu y$$
$$= (\alpha \lambda + (1 - \alpha)\mu) \left(\frac{\alpha \lambda}{\alpha \lambda + (1 - \alpha)\mu} x + \frac{(1 - \alpha)\mu}{\alpha \lambda + (1 - \alpha)\mu} y \right)$$

It follows that
$$u(\alpha p + (1 - \alpha)q) = \alpha \lambda + (1 - \alpha)\mu = \alpha u(p) + (1 - \alpha)u(q).$$

It is also quite straightforward that for $p \in P$ and $\lambda \in \mathbb{R}^+$ we have

$$u(\lambda p) = \lambda u(p)$$

so then if $p, q \in P$ and $\lambda, \mu \in \mathbb{R}^+$ we get

$$u(\lambda p + \mu q) = \lambda u(p) + \mu u(q)$$

by expressing $\lambda p + \mu q$ as a multiple of a convex combination and using the above properties of u. Now let $v \in V$, then there are $p, q \in P$ such that $v = p - q$ and we define

$$u(v) = u(p) - u(q).$$

To see that the definition is consistent, let $v = p - q = p' - q'$ for some $p, q, p', q' \in P$, then we have $p + q' = p' + q$ and

$$u(p) + u(q') = u(p') + u(q).$$

It follows that

$$u(p) - u(q) = u(p') - u(q').$$

Note that $u \in V^*$ is by construction the unique linear functional such that

$$K = \{p \in P : u(p) = 1\}.$$

Now we can show that K is closed. Let $\{x_i\}_{i=1}^\infty \subset K$ be a Cauchy sequence. Since P is closed there must be $p \in P$ that is the limit of the sequence, i.e. $p = \lim_{i \to \infty} x_i$. Since u is a linear functional, it must be continuous and we must have

$$\langle u, p \rangle = \lim_{i \to \infty} \langle u, x_i \rangle = 1.$$

It follows that $p \in K$ so K is closed.

The proof that K is bounded is rather technical. If K is not bounded then there must be a functional $\psi \in V^*$ such that for any $n \in \mathbb{N}$ there is $x_n \in K$ such that $\langle \psi, x_n \rangle > n$. Now let P^* denote the dual cone to P, we are going to show that $u \notin \text{int}(P^*)$, i.e. that u is not an interior point of P^*. Assume that $u \in \text{int}(P^*)$, then P^* contains an open ball $B(u) \subset P^*$ centered at u. It follows that for any $\varphi \in V^*$ we can always find a point $\xi \in B(u)$ such that u lies on the line segment connecting ξ and φ, i.e. that for some $\lambda \in (0, 1)$ we have

$$u = \lambda \xi + (1 - \lambda)\varphi.$$

Now let $\varphi = \psi$, we have

$$\lambda \xi = u - (1 - \lambda)\psi.$$

and for any $x \in K$ we must have

$$0 \leq \lambda \langle x, \xi \rangle = 1 - (1 - \lambda)\langle x, \psi \rangle.$$

Let $x = x_n$, where $x_n \in K$ is the point such that $\langle \psi, x_n \rangle \geq n$ introduced above, then

$$0 \leq 1 - (1 - \lambda)\langle x_n, \psi \rangle \leq 1 - (1 - \lambda)n$$

for all $n \in \mathbb{N}$ which is a contradiction as we can always find some $n \in \mathbb{N}$ such that $(1 - \lambda)n \geq 1$. So if K is not bounded, then $u \notin \text{int}(P^*)$. One can show that since P^* is a dual cone to P it is convex and closed, for every point $\varphi \in P^*$ such that $\varphi \notin \text{int}(P^*)$ there is some $p \in (P^*)^*$ such that

$$\langle \varphi, p \rangle = 0,$$

see [1, Theorem 11.5]. Now since we have already argued that if K is not bounded then $u \notin \text{int}(P)$ so let $p \in (P^*)^*$ be such that $\langle u, p \rangle = 0$. Moreover since P is convex and closed, one can show that $(P^*)^* = P$. It follows that $p \in P$ so there must be $y \in K$ and $\lambda \in \mathbb{R}^+$ such that $p = \lambda y$ and we get $\langle u, y \rangle = 0$ which is a contradiction. Hence K must be bounded. $\qquad \square$

We will proceed with examining the structure of the cone $A(K)^{*+}$. Note that the properties of $A(K)^{*+}$ will follow from the fact that it is a dual cone to $A(K)^+$.

Proposition 7. $A(K)^{*+}$ *is convex, closed, pointed and generating cone.*

Proof. It follows from the definition of $A(K)^{*+}$ as a cone of positive functionals that it is a convex cone.

To see that $A(K)^{*+}$ is closed we will use the fact that in the finite dimensional vector space all linear functionals are continuous. Hence if $\{\psi_n\}_{n=1}^\infty \subset A(K)^{*+}$ is a convergent sequence, such that $\psi = \lim_{n\to\infty} \psi_n$ then for every $f \in A(K)^+$ we have

$$\langle \psi, f \rangle = \lim_{n\to\infty} \langle \psi_n, f \rangle.$$

$\langle \psi, f \rangle \geq 0$ follows from $\langle \psi_n, f \rangle \geq 0$.

To see that $A(K)^{*+}$ is pointed assume there is $\psi \in A(K)^{*+} \cap -A(K)^{*+}$ then for every $f \in A(K)^+$ we must have that $\langle \psi, f \rangle = 0$ which implies $\psi = 0$ as $A(K)^+$ is generating.

To see that $A(K)^{*+}$ is generating remember that $A(K)^+$ is pointed, i.e. we have $A(K)^+ \cap -A(K)^+ = \{0\}$. Now assume that $A(K)^{*+}$ is not generating, i.e. that there is $\psi \in A(K)^*$ such that $\psi \notin A(K)^{*+} - A(K)^{*+}$. Then there is a function $f \in A(K)$ such that $\langle \psi, f \rangle \neq 0$ but for all $\varphi \in A(K)^{*+} - A(K)^{*+}$ we have $\langle \varphi, f \rangle = 0$. Note that we can construct such $f \in A(K)$ as $(A(K)^*)^* = A(K)$ and we can choose a suitable base in $A(K)^*$ that contains ϕ, find a dual base in $A(K)$ and use a suitable element of the dual base. It follows that $0 \neq f \in A(K)^+ \cap -A(K)^+$ which is a contradiction. $\square$

There is a natural norm on $A(K)^*$: the norm of linear functionals.

Definition 16. The norm of linear functionals on $A(K)^*$ is given for $\psi \in A(K)^*$ as

$$\|\psi\|_* = \sup\{\langle \psi, f \rangle : \|f\|_{\sup} \leq 1\}.$$

Proposition 8. $\|\cdot\|_*$ *is a norm.*

Proof. To prove that $\|\cdot\|_*$ is a norm we will show that it is positive-definite, homogeneous and subadditive. First of all note that if $\|\psi\|_* \geq 0$ as we always have $\langle \psi, 0 \rangle = 0$. Now assume that $\|\psi\|_* = 0$ but assume that there is $f \in A(K)$ such that $\langle \psi, f \rangle \neq 0$, i.e. that $\psi \neq 0$. We can assume $\langle \psi, f \rangle > 0$ as if $\langle \psi, f \rangle < 0$ then simply replace f by $-f$. Now let $f' = \frac{f}{\|f\|_{\sup}}$, then we have $\|f'\| = 1$ and we must have $\langle \psi, f' \rangle \leq 0$ which implies $\langle \psi, f \rangle = 0$ which is a contradiction. Hence $\|\psi\|_* = 0$ implies $\psi = 0$.

Let $\alpha \in \mathbb{R}$ then we have

$$\begin{aligned}
\|\alpha\psi\|_* &= \sup\{\langle \alpha\psi, f \rangle : \|f\|_{\sup} \leq 1\} \\
&= \sup\{|\alpha|\langle \psi, f \rangle : \|f\|_{\sup} \leq 1\} \\
&= |\alpha|\|\psi\|_*
\end{aligned}$$

as we can always replace f by $-f$ in case of $\alpha < 0$.

At last, let $\psi, \varphi \in A(K)^*$, then we have

$$\begin{aligned}
\|\psi + \varphi\|_* &= \sup\{\langle \psi + \varphi, f \rangle : \|f\|_{\sup} \leq 1\} \\
&= \sup\{\langle \psi, f \rangle + \langle \varphi, f \rangle : \|f\|_{\sup} \leq 1\} \\
&\leq \|\psi\|_* + \|\varphi\|_*.
\end{aligned}$$

$\square$

The supremum norm on $A(K)^*$ is useful in characterizing positive elements.

Proposition 9. *Let $\psi \in A(K)^*$ then we have $\psi \in A(K)^{*+}$ if and only if $\|\psi\|_* = \langle \psi, 1 \rangle$.*

Proof. Let $\psi \in A(K)^{*+}$ then there are $x \in K$ and $\lambda \in \mathbb{R}^+$ such that $\psi = \lambda x$ and we have $\langle \psi, 1 \rangle = \lambda$. Assume that $\|\psi\|_* > \lambda$, then there is $f \in A(K)$ such that $\|f\|_{\sup} \leq 1$ but $\langle x, f \rangle > 1$ which is a contradiction. So we have $\|\psi\|_* = \lambda$.

Let $\|\psi\|_* = \langle \psi, 1 \rangle$ and let $f \in A(K)$ such that $\|1 - f\|_{\sup} \leq 1$ then we

$$\langle \psi, f \rangle = \langle \psi, 1 \rangle - \langle \psi, 1 - f \rangle = \|\psi\|_* - \langle \psi, 1 - f \rangle \geq 0.$$

it follows that for every $f \in E(K)$ we have $\langle \psi, f \rangle \geq 0$ which implies $\psi \geq 0$. $\square$

Corollary 5. *Let $x \in K$ then $\|x\|_* = 1$.*

Proof. Since $K \subset A(K)^{*+}$ we have $\|x\|_* = \langle x, 1 \rangle = 1$. $\square$

2.5 No-Restriction hypothesis

So far we have been using the so-called No-Restriction hypothesis. The No-Restriction hypothesis states that every measurement that is mathematically well defined can be performed. Clearly this is a rather problematic assumption when constructing a physical theory as the physical theory does not have to be aware of our mathematics. On the other hand there are no known real world examples of theories that would violate the No-Restriction hypothesis. It may be that No-Restriction hypothesis follows from a more fundamental and better motivated axioms of a given theory as in [24]. The explanation we are going to provide is more pragmatic: if we assume No-Restriction hypothesis then we can prove results for all measurements and if one wishes to introduce restrictions later then it is only needed to check whether one has restricted to the point when the measurements needed for the proof are not in the theory anymore or whether we still have them.

There are two more important points to take into accounts when considering theories with restrictions. As first, let $m \in (0, \frac{1}{2}]$, then we can restrict the theory to include only effects that are affine functions $f : K \to [m, 1-m]$. One can see that this is the same as assuming that there is some minimal non-zero amount of white noise present in every measurement, quantified my m. But now one can see that the state space is not anymore uniquely specified by the restricted effect algebra so we should deal with a class of equivalence of state spaces instead of a fixed state space or assume that the state space is fixed by some other principle. Hence we either have to take into account the possible equivalence classes of state spaces or we have to introduce other higher principles to pick a unique state space.

The other possible problem with introducing restrictions is consistency. Given a set of measurements we can always construct new measurements by classical post-processing of results and randomization, together this is known as simulating measurements [25–28]. If one wishes to introduce restrictions then one has to be very careful not to exclude a measurement that can be simulated by measurements included in the theory as this would violate basic logical consistency.

For a more in-depth treatment of theories with restrictions see also [29]. We are going to keep assuming the No-Restriction hypothesis.

2.6 Examples

We are going to present several examples of theories. Many of these examples are important as they will show that the framework of GPTs is broad enough to describe many of the well-known state spaces. We also recommend the reader to familiarize themselves with the examples and how they are constructed as they can shed some light on the constructions we have presented in Sec. 2.

Example 1 (Classical theory). The first example is the most basic one. Let S_n denote a simplex with n vertexes, that is $S_n = \mathrm{conv}\left(\{s_1, \ldots, s_n\}\right)$ where $s_1, \ldots, s_n$ are affinely independent points, i.e. for $\alpha_i \in \mathbb{R}$, $i \in \{1, \ldots, n\}$, $\sum_{i=1}^{n} \alpha_i = 0$ we have $\sum_{i=1}^{n} \alpha_i s_i = 0$ if and only if $\alpha_i = 0$ for all $i \in \{1, \ldots, n\}$. The states $s_1, \ldots, s_n$ also form a basis of $A(S_n)^*$.

We will refer to systems with a state space that is a simplex as classical system or classical state space and we will refer to every other state space as non-classical.

The effect algebra $E(S_n)$ is generated by the functions $b_1, \ldots, b_n$ that are given as

$$\langle s_i, b_j \rangle = \delta_{ij}$$

for all $i, j \in \{1, \ldots, n\}$ and where δ_{ij} is the Kronecker delta. It follows that $b_i \geq 0$ for all $i \in \{1, \ldots, n\}$ as the points $s_1, \ldots, s_n$ are affinely independent. Moreover the functions $b_1, \ldots, b_n$ form a basis of $A(S_n)$ as for every function $f \in A(S_n)$ we have

$$f = \sum_{i=1}^{n} \langle s_i, f \rangle b_i.$$

Especially it follows that we have

$$1 = \sum_{i=1}^{n} b_i.$$

Now it is straightforward to see that for every $f \in A(S_n)^+$ there are $\mu_1, \ldots, \mu_n \in \mathbb{R}^+$ such that

$$f = \sum_{i=1}^{n} \mu_i b_i$$

and we have $f \in E(S_n)$ if and only if $0 \leq \mu_i \leq 1$. One can also show that these properties characterize the effect algebras of classical theories [21]. Moreover one can easily see that $f \in E(K)$ is an extreme point of the effect algebra if and only if $\mu_i \in \{0, 1\}$ for all $i \in \{1, \ldots, n\}$. It also follows that every point $s \in S_n$ has a unique decomposition to a convex combination of the extreme points $s_1, \ldots, s_n$, which is given as

$$s = \sum_{i=1}^{n} \langle s, b_i \rangle s_i.$$

It is easy to check the validity of the equation above; let $f \in E(S_n)$ be given as $f = \sum_{i=1}^{n} \mu_i b_i$, then

$$\langle s, f \rangle = \sum_{i=1}^{n} \mu_i \langle s, b_i \rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} \mu_i \langle s, b_j \rangle \langle s_j, b_i \rangle = \left\langle \sum_{j=1}^{n} \langle s, b_j \rangle s_j, f \right\rangle.$$

It follows that we can identify every point $s \in S_n$ with a probability distribution over the extreme points $s_1, \ldots, s_n$.

For $n = 2$ the state space S_2 can be represented by a line. We are going to denote the extreme points s_0, s_1 and we are going to use

$$s_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \qquad\qquad s_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

We are going to have only two functions that will generate $E(S_2)$ in the aforementioned way and we are going to denote them b and $1 - b$ as from Eq. (1) we will have that $b + (1 - b) = 1$ so the notation is quite natural. We have

$$b = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad\qquad 1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

where for $\psi \in A(S_2)^*$ and $f \in A(S)$,

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}, \qquad\qquad f = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}$$

the duality is given as

$$\langle \psi, f \rangle = \psi_1 f_1 + \psi_2 f_2.$$

We are going to use the same duality of real finite dimensional vector spaces in other examples. Now it is straightforward to check that we have

$$\begin{aligned} \langle s_0, b \rangle &= 0 & \langle s_1, b \rangle &= 1 \\ \langle s_0, 1 - b \rangle &= 1 & \langle s_1, 1 - b \rangle &= 0 \\ \langle s_0, 1 \rangle &= 1 & \langle s_1, 1 \rangle &= 1 \end{aligned}$$

as we should. This state space is well known as it represents the classical bit, s_0 and s_1 represent the two distinct states of the bit.

For $n = 3$ the state space S_3 is a triangle. We are going to denote the extreme points s_1, s_2, s_3 and we are going to use

$$s_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \qquad\qquad s_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \qquad\qquad s_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$
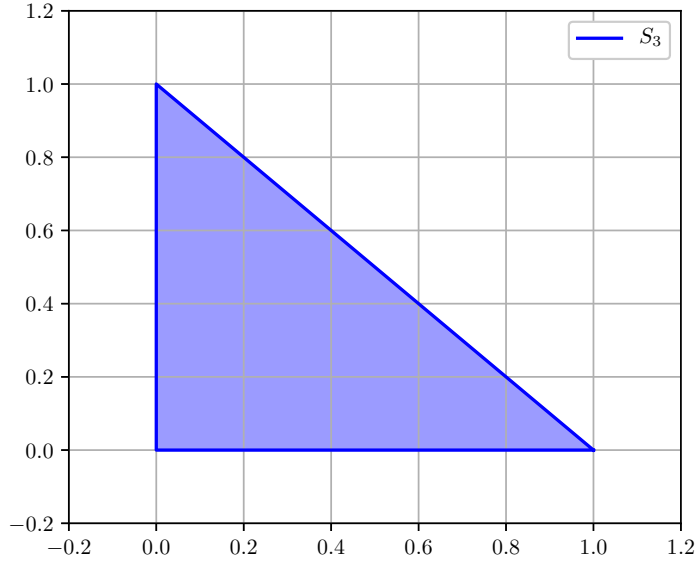
19

Figure 1: The simplex S_3 presented in Ex. 1.

as depicted in Fig. 1. The effects $b_1, b_2, b_3 \in E(S_3)$ generating the effect algebra are given as

$$b_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad b_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \qquad b_3 = \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}.$$

The wire-frame model of the effect algebra $E(S_3)$ is depicted in Fig. 2.

Example 2 (Quantum theory). Let $\mathcal{H}$ denote a finite dimensional complex Hilbert space. We will use the Dirac notation to denote the vectors $|\psi\rangle, |\varphi\rangle \in \mathcal{H}$, we will denote the inner product as $\langle\psi|\varphi\rangle$ and the vector norm as $\|\psi\| = \sqrt{\langle\psi|\psi\rangle}$. Let $B_h(\mathcal{H})$ denote the set of self-adjoint operators on $\mathcal{H}$. We say that an operator $A \in B_h(\mathcal{H})$ is positive semi-definite if and only if for all $|\psi\rangle \in \mathcal{H}$ we have $\langle\psi|A|\psi\rangle \geq 0$ and we write $A \geq 0$. For $A, B \in B_h(\mathcal{H})$ we write $A \geq B$ if and only if $A - B \geq 0$. $\mathrm{Tr}(A)$ will denote the trace of $A \in B_h(\mathcal{H})$. The set of states on $\mathcal{H}$ is

$$\mathfrak{D}_{\mathcal{H}} = \{\rho \in B_h(\mathcal{H}) : \rho \geq 0, \mathrm{Tr}(\rho) = 1\},$$

the effect algebra is

$$\mathcal{E}(\mathcal{H}) = \{E \in B_h(\mathcal{H}) : 0 \leq E \leq \mathbb{1}\},$$

where $\mathbb{1}$ denotes the identity matrix and the value of an effect E on a state ρ is given by the Hilbert-Schmidt inner product as $\mathrm{Tr}(\rho E)$. It is straightforward to see that pure states are the one dimensional projectors, i.e. operators of the form $|\psi\rangle\langle\psi|$ where $|\psi\rangle \in \mathcal{H}$, $\|\psi\| = 1$. Let $A \in B_h(\mathcal{H})$, then from $\mathrm{Tr}(|\psi\rangle\langle\psi|A) = \langle\psi|A|\psi\rangle$ it follows that $\mathcal{E}(\mathcal{H})$ really is the effect algebra corresponding to the state space $\mathfrak{D}_{\mathcal{H}}$. Moreover one can again see that the extreme points of $\mathcal{E}(\mathcal{H})$ are projectors, that is operators $P \in B_h(\mathcal{H})$ such that $P = P^2$.

This is a well known example of a finite-dimensional quantum theory referred to as qudit state space. If $\dim(\mathcal{H}) = 2$ then this corresponds to the qubit state space, the quantum equivalent of the classical bit. In this case $\mathfrak{D}_{\mathcal{H}}$ can be represented by a ball in $\mathbb{R}^3$, the so-called Bloch sphere.
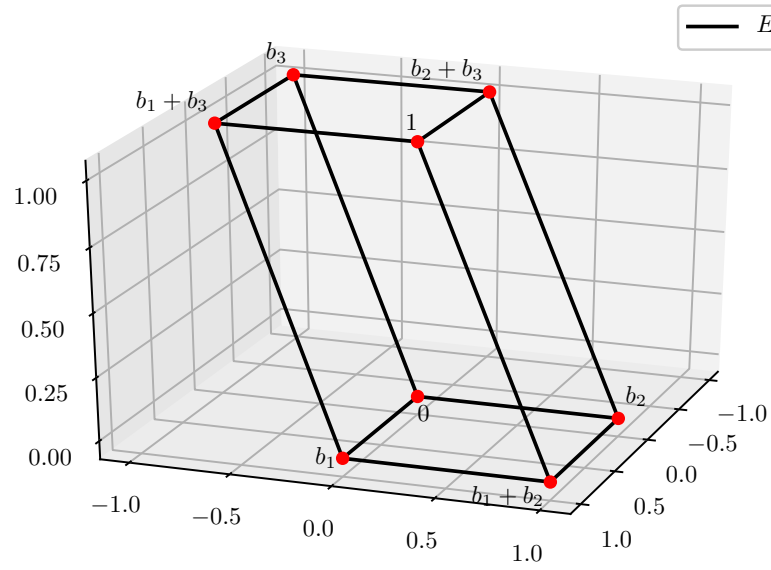
Figure 2: The effect algebra $E(S_3)$ corresponding to the simplex S_3 presented in Ex. 1.
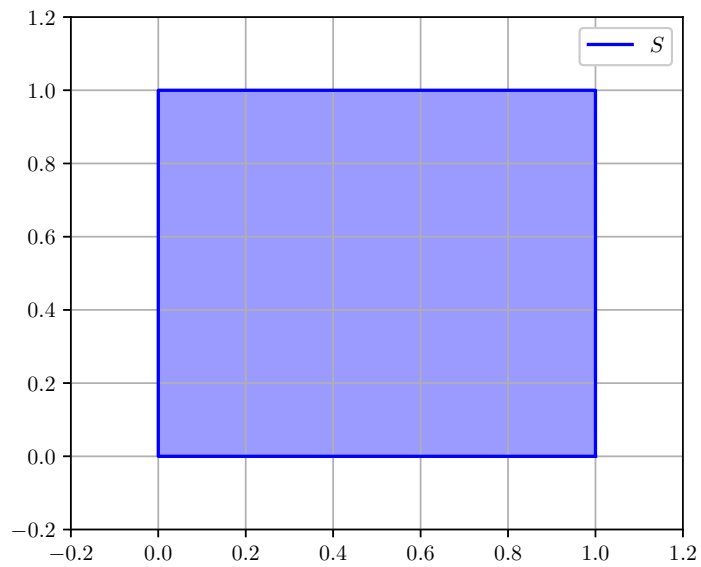
Figure 3: The square state space S presented in Ex. 3.
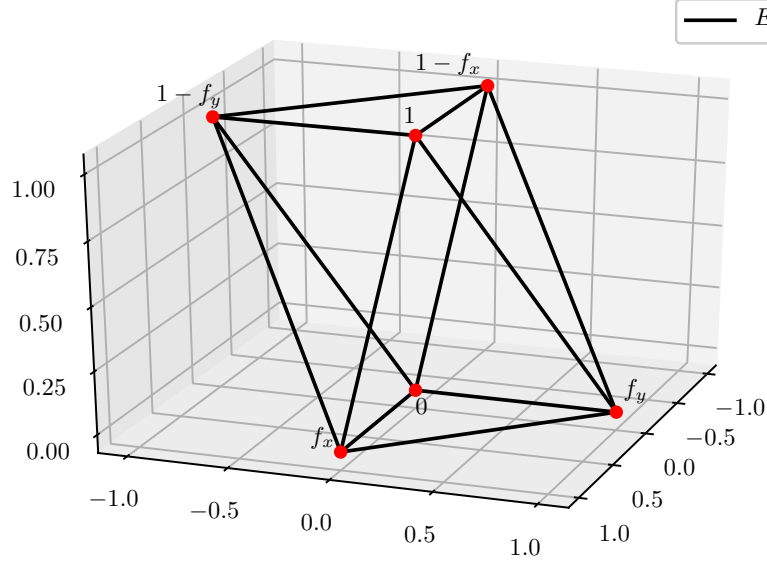
Figure 4: The effect algebra $E(S)$ corresponding to the square state space S presented in Ex. 3.

Example 3 (Square state space). Let $S = \mathrm{conv}(\{s_{00}, s_{10}, s_{01}, s_{11}\})$ where

$$s_{00} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \qquad\qquad s_{10} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

$$s_{01} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \qquad\qquad s_{11} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

are points such that $s_{00} + s_{11} = s_{10} + s_{01}$, i.e. S is a square as depicted in Fig 3.

The wire-frame model of $E(S)$ is depicted in Fig. 4. It is rather easy to find the extreme points of $E(S)$. Consider the general function $f \in A(S)$ to be given as

$$f = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

for some $a, b, c \in \mathbb{R}$. $f \in E(S)$ is equivalent to requiring $0 \le \langle s_{ij}, f \rangle \le 1$ for $i, j \in \{0, 1\}$ as s_{ij} are the extreme points of S, we get

$$0 \le c \le 1, \qquad\qquad 0 \le a + c \le 1,$$
$$0 \le b + c \le 1, \qquad\qquad 0 \le a + b + c \le 1.$$

To get extreme points we will assume that some of the inequalities become equalities and this will

yield the results. Let us denote

$$f_x = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad\qquad 1 - f_x = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

$$f_y = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \qquad\qquad 1 - f_y = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}$$

$$1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \qquad\qquad 0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Assume $c = 0$ and $a = 0$, then we can have both $b = 0$ and $b = 1$, this gives 0 and f_y. Assume $c = 0$ and $a = 1$, then we must have $b = 0$ and we get f_x. Let $c = 1$ and $a = 0$, then we can either have $b = 0$ or $b = -1$ which yields 1 and $1 - f_y$ respectively. If $c = 1$ and $a = -1$ then $b = 0$ follows and we get $1 - f_x$. This exhausts all of the possibilities. Hence the extreme points of the effect algebra $E(S)$ are f_x, f_y, $1 - f_x$, $1 - f_y$, 0 and 1.

3 Products of state spaces

In this section we are going to explore several standard constructions of products of state spaces. The most well known product is tensor product, that corresponds to creating a joint system of two systems. The motivation for using tensor products to describe the joint system of two systems is going to be similar to [30–33] and mathematically our explanation is going to be based on [3].

Assume that we have two systems K_A and K_B which are separated; our task is to find a joint description of both systems, i.e. to find the joint system K_{AB}. Let $f \in E(K_A)$ and $g \in E(K_B)$, then the joint description of K_A and K_B must be capable of describing the situation when we prepare $x \in K_A$ and $y \in K_B$ at the beginning of the experiment. Moreover we argue that for $x_1, x_2 \in K$ and $\lambda \in [0,1]$ it should be the same whether we prepare $\lambda x_1 + (1-\lambda)x_2 \in K_A$ and $y \in K_B$ or whether with probability λ we prepare $x_1 \in K_A$ and $y \in K_B$ and with probability $(1-\lambda)$ we prepare $x_2 \in K_A$ and $y \in K_B$. It follows that the preparation procedure of two systems should be affine in each state, and when extended as a map to $A(K_A)^*$ and $A(K_B)^*$ it should be bilinear. Since effects should map the bilinear preparation procedures to probabilities, they have to be elements of the dual vector space to bilinear preparations procedures and the dual is the tensor product of corresponding vector spaces. It follows then that the full set of states of the joint system K_{AB} must be described by the tensor product.

3.1 Tensor products

The most notable fact about tensor products of state spaces is that it is not uniquely defined but it has to be specified by the theory. Let V_A, V_B be finite-dimensional real vector spaces and let $K_A \subset V_A$, $K_B \subset V_B$ be two state spaces. We will denote $V_A \otimes V_B$ the tensor product of the vector spaces. Note that for $v \in V_A$, $w \in V_B$, $\psi \in V_A^*$ and $\varphi \in V_B^*$ we have

$$\langle \psi \otimes \varphi, v \otimes w \rangle = \langle \psi, v \rangle \langle \varphi, w \rangle.$$

Now assume that we want to construct a joint state space of the systems described by K_A and K_B and let $x \in K_A$, $y \in K_B$. We clearly want the joint state space to contain at least all possible states that describe the situation when the system K_A is prepared in the state x and the system K_B is prepared in the state y. This motivates the following definition

Definition 17. The minimal tensor product of the state spaces K_A and K_B is

$$K_A \dot\otimes K_B \subset V_A \otimes V_B$$

defined as

$$K_A \dot\otimes K_B = \text{conv}\{x \otimes y : x \in K_A, y \in K_B\}.$$

Proposition 10. $K_A \dot\otimes K_B$ *is a state space.*

Proof. It is straightforward to see that $K_A \dot\otimes K_B$ is convex by definition and from [3, Prop. 2.2] it follows that it is closed. It is also easy to see that it is bounded as $K_A \dot\otimes K_B \subset A(K_A)^* \otimes A(K_B)^*$ and $(A(K_A)^* \otimes A(K_B)^*)^* = A(K_A) \otimes A(K_B)$ and any functional of the form $f \otimes g \in A(K_A) \otimes A(K_B)$ is bounded on $K_A \dot\otimes K_B$. $\qquad\square$

We can apply the same logic to the effect algebras: let $f \in E(K_A)$ and $g \in E(K_B)$, then the minimal tensor product of the effect algebras, that will describe the measurements on the joint system, must at least contain the effect that describes that we apply f to K_A and g to K_B. This again motivates the following.

Definition 18. The minimal tensor product of the effect algebras $E(K_A)$ and $E(K_B)$ is

$$E(K_A) \dot\otimes E(K_B) \subset A(K_A) \otimes A(K_B)$$

defined as

$$E(K_A) \dot\otimes E(K_B) = \text{conv}\{f \otimes g : f \in E(K_A), g \in E(K_B)\}.$$

Proposition 11. $E(K_A)\dot{\otimes}E(K_B)$ *is a closed linear effect algebra, i.e. it is a linear effect algebra in the sense of Def. 8.*

Proof. Consider the convex cone P generated by $E(K_A)\dot{\otimes}E(K_B)$, i.e. $P = \text{conv}(\{f \otimes g : f \in A(K_A)^+, g \in A(K_B)^+\})$. It is again straightforward to see that P is closed as in the proof of Prop. 10. Now consider the interval $[0, 1 \otimes 1] \subset P$, we are going to show that $[0, 1 \otimes 1] = E(K_A)\dot{\otimes}E(K_B)$. Let $f \otimes g \in [0, 1 \otimes 1]$, then it follows that we must have $\|f\|_{\sup}\|g\|_{\sup} \leq 1$, $\frac{1}{\|f\|_{\sup}}f \in E(K_A)$. $\frac{1}{\|g\|_{\sup}} \in E(K_B)$ follows from $f \geq 0$ and $g \geq 0$ which we can assume without the loss of generality. Now assume that we have $\sum_{i=1}^n f_i \otimes g_i \in [0, 1 \otimes 1]$ and assume that $f_i \in A(K_A)^+$, $g_i \in A(K_B)^+$ for all $i \in \{1, \ldots, n\}$, one can see that every element from $[0, 1 \otimes 1]$ is of this form. It follows that we must have

$$\sum_{i=1} \|f_i\|_{\sup}\|g_i\|_{\sup} \leq 1$$

and

$$\sum_{i=1} f_i \otimes g_i = \sum_{i=1} \|f_i\|_{\sup}\|g_i\|_{\sup}\left(\frac{1}{\|f\|_{\sup}}f_i\right) \otimes \left(\frac{1}{\|g\|_{\sup}}g_i\right) + \left(1 - \sum_{i=1}\|f_i\|_{\sup}\|g_i\|_{\sup}\right)0 \otimes 0.$$

It follows that $\sum_{i=1} f_i \otimes g_i \in E(K_A)\dot{\otimes}E(K_B)$ as we have just expressed $\sum_{i=1} f_i \otimes g_i$ as a suitable convex combination. $\square$

Now we have defined a tensor product of state spaces and tensor product of effect algebras but clearly $E(K_A)\dot{\otimes}E(K_B)$ doesn't have to be the effect algebra corresponding to $K_A\dot{\otimes}K_B$, i.e. we can have $E(K_A)\dot{\otimes}E(K_B) \neq E(K_A\dot{\otimes}K_B)$.

Definition 19. The maximal tensor product of state spaces K_A and K_B is $K_A\hat{\otimes}K_B \subset V_A \otimes V_B$ defined as

$$K_A\hat{\otimes}K_B = \mathfrak{S}(E(K_A)\dot{\otimes}E(K_B))$$

and the maximal tensor product of the effect algebras $E(K_A)$ and $E(K_B)$ is $E(K_A)\hat{\otimes}E(K_B) \subset A(K_A) \otimes A(K_B)$ defined as

$$E(K_A)\hat{\otimes}E(K_B) = E(K_A\dot{\otimes}K_B).$$

We do not have to prove that $K_A\hat{\otimes}K_B$ is a well defined state space as this follows from the definition. The same holds for $E(K_A)\hat{\otimes}E(K_B)$.

Why we call these tensor products maximal is because if we would want to extend $K_A\hat{\otimes}K_B$ by adding a point $v \in V_A \otimes V_B$, then according to Thm. 4 there would be $f \in E(K_A)\dot{\otimes}E(K_B)$ such that $\langle v, f \rangle < 0$ and we have already argued that the effect algebra of the joint system should contain $E(K_A)\dot{\otimes}E(K_B)$. The same holds for $E(K_A)\hat{\otimes}E(K_B)$, extending the effect algebra by adding points would result in some states from $K_A\dot{\otimes}K_B$ not being a well-defined states. The following is an immediate result.

Proposition 12. *We have*

$$K_A\dot{\otimes}K_B \subset K_A\hat{\otimes}K_B,$$
$$E(K_A)\dot{\otimes}E(K_B) \subset E(K_A)\hat{\otimes}E(K_B).$$

Proof. The proof is straightforward. Let $x \in K_A$, $y \in K_B$ and $f \in E(K_A)$, $g \in E(K_B)$, then

$$\langle x \otimes y, f \otimes g \rangle = \langle x, f \rangle \langle y, g \rangle \geq 0$$

follows immediately. This shows that $x\otimes y \in K_A\hat{\otimes}K_B$ and $f\otimes g \in E(K_A)\hat{\otimes}E(K_B)$. As the extreme points of $K_A\dot{\otimes}K_B$ are by definition of the form $x \otimes y$ and the extreme points of $E(K_A)\dot{\otimes}E(K_B)$ are again by definition of the form $f \otimes g$ it follows that we must have $K_A\dot{\otimes}K_B \subset K_A\hat{\otimes}K_B$ and $E(K_A)\dot{\otimes}E(K_B) \subset E(K_A)\hat{\otimes}E(K_B)$. $\square$

The last tensor product to define is the so-called real tensor product. This tensor product is not uniquely defined, but has to be given by the underlying theory.

Definition 20. The real tensor product of state spaces K_A and K_B is denoted as $K_A \tilde{\otimes} K_B$ and it is a state space such that

$$K_A \dot{\otimes} K_B \subset K_A \tilde{\otimes} K_B \subset K_A \hat{\otimes} K_B.$$

Note that the name real tensor product was introduced by the author. In some publications $K_A \tilde{\otimes} K_B$ is referred to as composite system [32] and sometimes the definition coincides with the term parallel composition [34], but these terms could also include more general or different concepts.

Whenever needed in the future, we will assume that there is a well defined notion of real tensor product. We will also introduce the real tensor product of effect algebras for the sake of completeness as the effect algebra corresponding to the real tensor product of the state spaces.

Definition 21. We define the real tensor product of effect algebras $E(K_A)$ and $E(K_B)$ as

$$E(K_A) \tilde{\otimes} E(K_B) = E(K_A \tilde{\otimes} K_B).$$

We will also define the notions of separable and entangled states.

Definition 22. We say that a state $x \in K_A \tilde{\otimes} K_B$ is separable, if we have $x \in K_A \dot{\otimes} K_B$. We say that a state $x \in K_A \tilde{\otimes} K_B$ is entangled if it is not separable, i.e. if we have $x \in K_A \tilde{\otimes} K_B \setminus K_A \dot{\otimes} K_B$.

It may feel unsatisfactory that we have to define some extra composition rule for given theories but this is what we do in quantum theory as demonstrated by the following example.

Example 4. Let $\mathcal{H}$ be a Hilbert space and let $\mathfrak{D}_\mathcal{H}$ denote the corresponding state space. In quantum theory we define the real tensor product to be $\mathfrak{D}_\mathcal{H} \tilde{\otimes} \mathfrak{D}_\mathcal{H} = \mathfrak{D}_{\mathcal{H} \otimes \mathcal{H}}$. The set $\mathfrak{D}_\mathcal{H} \dot{\otimes} \mathfrak{D}_\mathcal{H}$ is the set of separable states and the set $\mathfrak{D}_\mathcal{H} \hat{\otimes} \mathfrak{D}_\mathcal{H}$ is the set of entanglement witnesses, that is for any entangled state $\rho \in \mathfrak{D}_{\mathcal{H} \otimes \mathcal{H}}$ there is some $A \in \mathfrak{D}_\mathcal{H} \hat{\otimes} \mathfrak{D}_\mathcal{H}$ such that $\mathrm{Tr}(\rho A) < 0$ [6, Theorem 6.39].

In a similar fashion to state spaces and effect algebras we can define tensor products of cones.

Definition 23. Let V, W be finite-dimensional real vector spaces and let $P \subset V$ and $Q \subset W$ be convex, generating and pointed cones. Let $P^* \subset V^*$ and $Q^* \subset W^*$ be the dual cones, i.e. the cones of positive linear functionals. The minimal tensor product of P and Q is denoted $P \dot{\otimes} Q$ and it is defined as

$$P \dot{\otimes} Q = \mathrm{conv}(\{p \otimes q : p \in P, q \in Q\})$$

and the maximal tensor product of the cones is denoted as $P \hat{\otimes} Q$ and it is defined as the dual cone to $P^* \dot{\otimes} Q^*$, i.e.

$$P \hat{\otimes} Q = (P^* \dot{\otimes} Q^*)^* = \{\psi \in V \otimes W : \langle p' \otimes q', \psi \rangle \geq 0, \forall p' \in P^*, \forall q' \in Q^*\}.$$

Proposition 13. *Let K_A and K_B be state spaces, then we have*

$$A(K_A)^{*+} \dot{\otimes} A(K_B)^{*+} = A(K_A \dot{\otimes} K_B)^{*+},$$
$$A(K_A)^{*+} \hat{\otimes} A(K_B)^{*+} = A(K_A \hat{\otimes} K_B)^{*+}.$$

Proof. Let $\varphi \in A(K_A)^{*+} \dot{\otimes} A(K_B)^{*+}$ such that $\varphi \neq 0$, then $\frac{1}{\langle \varphi, 1 \otimes 1 \rangle} \varphi$ is separable and normalized, so by Thm. 5 we have $\frac{1}{\langle \varphi, 1 \otimes 1 \rangle} \varphi \in K_A \dot{\otimes} K_B$. It follows that $A(K_A)^{*+} \dot{\otimes} A(K_B)^{*+} = A(K_A \dot{\otimes} K_B)^{*+}$. $A(K_A)^{*+} \hat{\otimes} A(K_B)^{*+} = A(K_A \hat{\otimes} K_B)^{*+}$ follows in a similar manner. $\square$

Proposition 14. *We have* $\mathrm{span}(A(K_A)^{*+} \dot{\otimes} A(K_B)^{*+}) = \mathrm{span}(A(K_A)^{*+} \hat{\otimes} A(K_B)^{*+})$, *i.e. both the minimal and maximal tensor products of the cones generate the same vector space.*

Proof. By definition we have $\mathrm{span}(E(K_A \hat{\otimes} K_B)) = \mathrm{span}(E(K_A) \dot{\otimes} E(K_B)) = A(K_A) \otimes A(K_B)$ and we must have

$$\mathrm{span}(A(K_A)^{*+} \hat{\otimes} A(K_B)^{*+}) = (A(K_A) \otimes A(K_B))^* = A(K_A)^* \otimes A(K_B)^*$$
$$= \mathrm{span}(A(K_A)^{*+} \dot{\otimes} A(K_B)^{*+}).$$

$\square$

Corollary 6. *We have* $\mathrm{span}(A(K_A)^+ \dot{\otimes} A(K_B)^+) = \mathrm{span}(A(K_A)^+ \hat{\otimes} A(K_B)^+)$.

Proof. The result follows from Prop. 14 by looking at the linear hulls of the dual cones. $\square$

3.2 Partial trace

We will proceed with defining the generalization of partial trace. It is clear that all of the effect algebras $E(K_A)\dot{\otimes}E(K_B)$, $E(K_A)\tilde{\otimes}E(K_B)$ and $E(K_A)\hat{\otimes}E(K_B)$ have the same unit effect $1 \otimes 1$. Let $\mathrm{id}_A : K_A \to K_A$ be the identity map and $f \in E(K_B)$, we are going to be interested in the map $\mathrm{id}_A \otimes f : K_A \tilde{\otimes} K_B \to K_A$.

Definition 24. Let $x \in K_A \tilde{\otimes} K_B$ and $f \in A(K_B)$, then we define $(\mathrm{id}_A \otimes f)(x) \in A(K_A)^*$ to be the unique functional such that for all $g \in A(K_A)$ we have

$$\langle (\mathrm{id}_A \otimes f)(x), g \rangle = \langle x, g \otimes f \rangle.$$

The following is immediate.

Proposition 15. Let $f \in A(K_B)$ then $\mathrm{id}_A \otimes f$ is an affine map.

Proof. Let $x, y \in K_A \tilde{\otimes} K_B$, $\lambda \in [0,1]$ and $g \in A(K_A)$, we have

$$\begin{aligned}
\langle (\mathrm{id}_A \otimes f)(\lambda x + (1-\lambda)y), g \rangle &= \langle (\lambda x + (1-\lambda)y, g \otimes f \rangle \\
&= \lambda \langle x, g \otimes f \rangle + (1-\lambda)\langle y, g \otimes f \rangle \\
&= \lambda \langle (\mathrm{id}_A \otimes f)(x), f \rangle + (1-\lambda)\langle (\mathrm{id}_A \otimes f)(y), g \rangle
\end{aligned}$$

so $(\mathrm{id}_A \otimes f)(\lambda x + (1-\lambda)y) = \lambda(\mathrm{id}_A \otimes f)(x) + (1-\lambda)(\mathrm{id}_A \otimes f)(y)$ follows. $\qquad \square$

One can also extend $\mathrm{id}_A \otimes f$ to a linear map $A(K_A)^* \otimes A(K_B)^* \to A(K_A)^*$. In a similar fashion we can define $g \otimes \mathrm{id}_B : K_A \tilde{\otimes} K_B \to K_B$ for any $g \in A(K_A)$. Besides that, the map is also linear in f as shown bellow.

Proposition 16. Let $f_1, f_2 \in A(K_B)$ and $\alpha \in \mathbb{R}$, then we have

$$\mathrm{id}_A \otimes (f_1 + \alpha f_2) = \mathrm{id}_A \otimes f_1 + \alpha\, \mathrm{id}_A \otimes f_2.$$

Proof. Let $x \in K_A \tilde{\otimes} K_B$ and $g \in A(K_A)$, we have

$$\begin{aligned}
\langle (\mathrm{id}_A \otimes (f_1 + \alpha f_2))(x), g \rangle &= \langle x, g \otimes (f_1 + \alpha f_2) \rangle \\
&= \langle x, g \otimes f_1 \rangle + \alpha \langle x, g \otimes f_2 \rangle \\
&= \langle (\mathrm{id}_A \otimes f_1)(x), g \rangle + \alpha \langle (\mathrm{id}_A \otimes f_2)(x), g \rangle.
\end{aligned}$$

$\square$

Proposition 17. Let $f \in A(K_B)^+$ then for every $x \in K_A \tilde{\otimes} K_B$ we have $(\mathrm{id}_A \otimes f)(x) \in A(K_A)^{*+}$. Moreover if $f \in E(K_B)$, then there are $y \in K_A$ and $\lambda \in [0,1]$ such that $(\mathrm{id}_A \otimes f)(x) = \lambda y$.

Proof. To see that $(\mathrm{id}_A \otimes f)(x) \in A(K_A)^{*+}$ just note that $\langle x, g \otimes f \rangle \geq 0$ for all $g \in A(K_B)^+$. It follows that we have $(\mathrm{id}_A \otimes f)(x) = \lambda y$ for some $y \in K_A$ and $\lambda \in \mathbb{R}^+$ as K_A is a base of the cone $A(K_A)^{*+}$.

Now assume that $f \in E(K)$, we have

$$\lambda = \langle \lambda y, 1 \rangle = \langle (\mathrm{id}_A \otimes f)(x), 1 \rangle = \langle x, 1 \otimes f \rangle \leq 1$$

which concludes the proof. $\qquad \square$

We are now going to consider the special case of the map $\mathrm{id}_A \otimes 1$. Let $x \in K_A \tilde{\otimes} K_B$, then we will call $(\mathrm{id}_A \otimes 1)(x)$ and $(1 \otimes \mathrm{id}_B)(x)$ the marginals of x and we are going to call the map $\mathrm{id}_A \otimes 1$ the partial trace as in quantum theory it corresponds to the partial trace [6, Definition 2.68].

Proposition 18. Let $x \in K_A \tilde{\otimes} K_B$ then $(\mathrm{id}_A \otimes 1)(x) \in K_A$.

Proof. As $1 \in E(K_B)$ by definition, we already know that $(\mathrm{id}_A \otimes 1)(x) \in A(K_A)^{*+}$ and the result follows from $\langle (\mathrm{id}_A \otimes 1)(x), 1 \rangle = \langle x, 1 \otimes 1 \rangle = 1$ by Thm. 5. $\qquad \square$

The following result is important and useful in many calculations.

Theorem 6 (Barnum, Barrett, Leifer, Wilcze). *Let $x \in K_A \tilde{\otimes} K_B$ and assume that $(\mathrm{id}_A \otimes 1)(x) = y$ is a pure state, then $x = y \otimes z$ for some $z \in K_B$.*

Proof. The proof can be found in [35, Lemma 3]. We will provide exactly the same proof, only formulated in terms that we have introduced so far.

Let $f \in E(K_B)$ then by Prop. 4 also $1 - f \in E(K_A)$. We have

$$y = (\mathrm{id}_A \otimes 1)(x) = (\mathrm{id}_A \otimes f)(x) + (\mathrm{id}_A \otimes (1 - f))(x)$$

and according to Prop. 17 we have $(\mathrm{id}_A \otimes f)(x) = \lambda w$ and $(\mathrm{id}_A \otimes (1 - f))(x) = \lambda' w'$ for some $w, w' \in K_A$ and $\lambda, \lambda' \in \mathbb{R}^+$. From $1 = \langle y, 1 \rangle = \lambda + \lambda'$ we get $\lambda' = 1 - \lambda$ so now we have $y = \lambda w + (1 - \lambda) w'$. Since y is pure we get $w = w' = y$ and we have $(\mathrm{id}_A \otimes f)(x) = \lambda y$. Moreover $\lambda = \langle x, 1 \otimes f \rangle = \langle (1 \otimes \mathrm{id}_B)(x), f \rangle$.

Let $g \in E(K_A)$, then we have

$$\begin{aligned}
\langle x, g \otimes f \rangle &= \langle (\mathrm{id}_A \otimes f)(x), g \rangle = \lambda \langle y, g \rangle \\
&= \langle (1 \otimes \mathrm{id}_B)(x), f \rangle \langle (\mathrm{id}_A \otimes 1)(x), g \rangle \\
&= \langle (\mathrm{id}_A \otimes 1)(x) \otimes (1 \otimes \mathrm{id}_B)(x), g \otimes f \rangle.
\end{aligned}$$

It follows that $x = (\mathrm{id}_A \otimes 1)(x) \otimes (1 \otimes \mathrm{id}_B)(x) = y \otimes z$ for $z = (1 \otimes \mathrm{id}_B)(x)$. $\qquad \square$

3.3 Existence of entanglement

A natural question may emerge: given two state spaces K_A and K_B, do we have $K_A \dot{\otimes} K_B \neq K_A \hat{\otimes} K_B$? This is an open problem and it is conjectured that $K_A \dot{\otimes} K_B \neq K_A \hat{\otimes} K_B$ if and only if both K_A and K_B are non-classical state spaces. We are going to present the current state of the research.

Proposition 19. *Let S_n be a simplex and K any state space, then $K \dot{\otimes} S_n = K \hat{\otimes} S_n$.*

Proof. Let $x \in K \hat{\otimes} S_n$ and let $s_1, \ldots, s_n$ be the extreme points of S_n, then $\{s_1, \ldots, s_n\}$ is a base of $A(S_n)^*$ which means that we have

$$x = \sum_{i=1}^{n} \varphi_i \otimes s_i$$

for some $\varphi_i \in A(K)^*$. Using the result of Prop. 17 we get

$$\varphi_i = (\mathrm{id} \otimes b_i)(x) \in A(K)^{*+}$$

which shows that $x \in K \dot{\otimes} S_n$. $\qquad \square$

Note that the result above also holds for the positive cones and is independent of the choice of the base.

Corollary 7. *Let V be a real finite-dimensional vector space, let $P \subset V$ be a pointed, generating, convex cone and let S_n be a simplex. We have*

$$P \dot{\otimes} A(S_n)^{*+} = P \hat{\otimes} A(S_n)^{*+}.$$

Proof. Let $K \subset P$ be a base of P, then $K \hat{\otimes} S_n$ is a base of $P \hat{\otimes} A(S_n)^{*+}$ and $K \dot{\otimes} S_n$ is a base of $P \dot{\otimes} A(S_n)^{*+}$. The result follows from Prop. 19. $\qquad \square$

Proposition 20. *Let K be a non-classical state space and let S be the square state space, then*

$$K \dot{\otimes} S \neq K \hat{\otimes} S.$$

Proof. See [36] for a proof. $\qquad \square$

Proposition 21. *Let K_A and K_B be non-classical polytopes, i.e. both K_A and K_B are not simplexes, then*

$$K_A \dot{\otimes} K_B \neq K_A \hat{\otimes} K_B.$$

Proof. The proof was communicated to the author by [37]. $\qquad \square$

There is also proof for centrally symmetric state spaces [38].

3.4 Multipartite entanglement

As we are using tensor products to describe bipartite systems it is also worth mentioning some of the specifics of describing multipartite systems, where number of parties is larger than 2. First of all there is the logical assumption of associativity of the real tensor product. Let K_A, K_B and K_C be state spaces, then we require

$$(K_A \tilde{\otimes} K_B) \tilde{\otimes} K_C = K_A \tilde{\otimes} (K_B \tilde{\otimes} K_C)$$

and in this case we denote the tripartite state space $K_A \tilde{\otimes} K_B \tilde{\otimes} K_C$. It is straightforward to prove that the minimal and maximal tensor products are associative as well as the real tensor product used in quantum theory.

Proposition 22. *Let K_A, K_B, K_C be state spaces, then*

$$(K_A \dot{\otimes} K_B) \dot{\otimes} K_C = K_A \dot{\otimes} (K_B \dot{\otimes} K_C).$$

and

$$(K_A \hat{\otimes} K_B) \hat{\otimes} K_C = K_A \hat{\otimes} (K_B \hat{\otimes} K_C).$$

Proof. We are going to begin with the minimal tensor products. By definition the set of extreme points of $K_A \dot{\otimes} K_B$ is

$$\text{ext}(K_A \dot{\otimes} K_B) = \{x \otimes y : x \in \text{ext}(K_A), y \in \text{ext}(K_B)\}.$$

We also have

$$\begin{aligned}
\text{ext}((K_A \dot{\otimes} K_B) \dot{\otimes} K_C) &= \{w \otimes z : w \in \text{ext}(K_A \dot{\otimes} K_B), z \in \text{ext}(K_C)\} \\
&= \{x \otimes y \otimes z : x \in \text{ext}(K_A), y \in \text{ext}(K_B), z \in \text{ext}(K_C)\}.
\end{aligned}$$

By the same argumentation we get

$$\text{ext}(K_A \dot{\otimes} (K_B \dot{\otimes} K_C)) = \{x \otimes y \otimes z : x \in \text{ext}(K_A), y \in \text{ext}(K_B), z \in \text{ext}(K_C)\}$$

and the associativity of the minimal tensor product follows by Thm. 2.

To show the associativity of the maximal tensor product, note that

$$(K_A \hat{\otimes} K_B) \hat{\otimes} K_C = \mathfrak{S}(E(\mathfrak{S}(E(K_A) \dot{\otimes} E(K_B))) \dot{\otimes} E(K_C)).$$

$E(\mathfrak{S}(E(K))) = E(K)$ for any state space K as a result of Thm. 5 and we get

$$\begin{aligned}
(K_A \hat{\otimes} K_B) \hat{\otimes} K_C &= \mathfrak{S}((E(K_A) \dot{\otimes} E(K_B)) \dot{\otimes} E(K_C)) \\
&= \mathfrak{S}(E(K_A) \dot{\otimes} E(K_B) \dot{\otimes} E(K_C))
\end{aligned}$$

where we argue that in the last step the associativity of minimal tensor products of effect algebras follows in the same way as associativity of the minimal tensor product of state spaces. By the same argumentation we get

$$K_A \hat{\otimes} (K_B \hat{\otimes} K_C) = \mathfrak{S}(E(K_A) \dot{\otimes} E(K_B) \dot{\otimes} E(K_C))$$

which concludes the proof. $\square$

Proposition 23. *The real tensor product used in quantum theory as introduced in Ex. 4 is associative.*

Proof. We have

$$(\mathfrak{D}_{\mathcal{H}} \tilde{\otimes} \mathfrak{D}_{\mathcal{H}}) \tilde{\otimes} \mathfrak{D}_{\mathcal{H}} = \mathfrak{D}_{\mathcal{H} \otimes \mathcal{H}} \tilde{\otimes} \mathfrak{D}_{\mathcal{H}} = \mathfrak{D}_{\mathcal{H} \otimes \mathcal{H} \otimes \mathcal{H}} = \mathfrak{D}_{\mathcal{H}} \tilde{\otimes} \mathfrak{D}_{\mathcal{H} \otimes \mathcal{H}} = \mathfrak{D}_{\mathcal{H}} \tilde{\otimes} (\mathfrak{D}_{\mathcal{H}} \tilde{\otimes} \mathfrak{D}_{\mathcal{H}}).$$

$\square$

It is tricky to define the separable and entangled states in the multipartite case. There might be states where two of the parties are entangled, but not all three and there might be states where all three parties are entangled but but none of the bipartite marginals of the state is entangled. These problems already manifest in quantum theory, see [6] for a short review.

3.5 Direct product and direct convex sum

At last, we will review other constructions with state spaces and effect algebras: the direct product and direct convex sum. Let V_1, V_2 be real, finite-dimensional vector spaces, then by $V_1 \times V_2$ we will denote their Cartesian product.

Definition 25. Let $K_A \subset V_A$ and $K_B \subset V_B$ be state spaces then the direct product of K_A and K_B is a state space $K_A \times K_B \subset V_A \times V_B$ given as

$$K_A \times K_B = \{(x, y) : x \in K_A, y \in K_B\}$$

with the convex combinations defined for $x_1, x_2 \in K_A$, $y_1, y_2 \in K_B$ and $\lambda \in [0, 1]$ as

$$\lambda(x_1, y_1) + (1 - \lambda)(x_2, y_2) = (\lambda x_1 + (1 - \lambda)x_2, \lambda y_1 + (1 - \lambda)y_2).$$

Proposition 24. $K_A \times K_B$ *is a state space.*

Proof. $K_A \times K_B$ is convex by definition. It is rather easy to see that it is also closed as K_A and K_B are closed [4, Proposition 8.1.4] and one can show in a very similar way that $K_A \times K_B$ is bounded as well. $\square$

The operational interpretation of the direct product of state spaces is as follows: assume that we want to keep track of two distinct possibilities of events, the first described by $x \in K_A$ and the second described by $y \in K_B$ so we can adopt the handy notation (x, y). This gives rise to the direct product of state spaces. Notice that we have already encountered the direct product of state spaces as $S = S_2 \times S_2$.

Definition 26. Let $K_A \subset V_A$ and $K_B \subset V_B$ be state spaces then the direct convex sum of K_A and K_B is a state space $K_A \oplus K_B$ given as

$$K_A \oplus K_B = \{(\lambda x, (1 - \lambda)y) : x \in K_A, y \in K_B, \lambda \in [0, 1]\}$$

with the convex combinations defined for $x_1, x_2 \in K_A$, $y_1, y_2 \in K_B$ and $\lambda_1, \lambda_2, \mu \in [0, 1]$ as

$$\mu(\lambda_1 x_1, (1 - \lambda_1)y_1) + (1 - \mu)(\lambda_2 x_2, (1 - \lambda_2)y_2) =$$
$$(\mu \lambda_1 x_1 + (1 - \mu)\lambda_2 x_2, \mu(1 - \lambda_1)y_1 + (1 - \mu)(1 - \lambda_2)y_2).$$

Proposition 25. $K_A \oplus K_B$ *is a state space.*

Proof. We are going to show that $K_A \oplus K_B$ is the base of the cone

$$A(K_A)^{*+} \times A(K_B)^{*+} = \{(\lambda x, \mu y) : x \in K_A, y \in K_B, \lambda, \mu \in \mathbb{R}^+\}.$$

It is immediate that $A(K_A)^{*+} \times A(K_B)^{*+}$ is closed convex cone. $A(K_A)^{*+} \times A(K_B)^{*+}$ generates the vector space $A(K_A)^* \times A(K_B)^*$ since $A(K_A)^{*+}$ and $A(K_B)^{*+}$ are generating. To show that $A(K_A)^{*+} \times A(K_B)^{*+}$ is pointed let $(v, w) \in A(K_A)^{*+} \times A(K_B)^{*+}$ and $-(v, w)A(K_A)^{*+} \times A(K_B)^{*+}$, we get $v \in A(K_A)^{*+} \cap (-A(K_A)^{*+})$ and $w \in A(K_B)^{*+} \cap (-A(K_B)^{*+})$. Since both $A(K_A)^{*+}$ and $A(K_B)^{*+}$ are pointed we get that $(v, w) = (0, 0)$.

Let $(1, 1) \in A(K_A)^+ \times A(K_B)^+$, then

$$K_A \oplus K_B = \{(\lambda x, \mu y) \in A(K_A)^{*+} \times A(K_B)^{*+} : \langle (\lambda x, \mu y), (1, 1) \rangle = 1\}$$

because $\langle (\lambda x, \mu y), (1, 1) \rangle = \lambda + \mu$. It follows that for $(\lambda x, \mu y) \in A(K_A)^{*+} \times A(K_B)^{*+}$ we have

$$(\lambda x, \mu y) = (\lambda + \mu)\left(\frac{\lambda}{\lambda + \mu}x, \frac{\mu}{\lambda + \mu}y\right) = \langle (\lambda x, \mu y), (1, 1) \rangle \left(\frac{\lambda}{\lambda + \mu}x, \frac{\mu}{\lambda + \mu}y\right)$$

so $K_A \oplus K_B$ is a base of the cone $A(K_A)^+ \times A(K_B)^+$. $\square$

The interpretation of direct convex sum is that we want to describe a preparation such that with a probability λ we get $x \in K_A$ and with the probability $1 - \lambda$ we get $y \in K_B$. This gives rise to a formal convex combination $(\lambda x, (1 - \lambda)y)$. This is knonw in quantum theory as superselection rules [39].

We will proceed with constructing the effect algebras for the state spaces $K_A \times K_B$ and $K_A \oplus K_B$. It is easier to characterize $E(K_A \oplus K_B)$ so we will begin there. Let $(\psi, \varphi) \in A(K_A)^* \times A(K_B)^*$ then from the trivial $(\psi, \varphi) = (\psi, 0) + (0, \varphi)$ and from the results of Prop. 7 we get that $A(K_A \oplus K_B)^* = A(K_A)^* \times A(K_B)^*$. The dual vector space is $A(K_A) \times A(K_B)$ with the positive cone

$$A(K_A)^+ \times A(K_B)^+ = \{(f, g) : f \in A(K_A)^+, g \in A(K_B)^+\}.$$

To find the constant function consider the following: let $x \in K_A$, $y \in K_B$ and $\lambda \in [0, 1]$, then for $(f, g) \in A(K_A) \times A(K_B)$ we have

$$\langle (\lambda x, (1 - \lambda)y), (f, g) \rangle = \lambda \langle x, f \rangle + (1 - \lambda)\langle y, g \rangle.$$

We require $\langle (\lambda x, (1 - \lambda)y), (f, g) \rangle = 1$ for all $x \in K_A$, $y \in K_B$. If we assume $\lambda = 0$, we get $f = 1$; if we assume $\lambda = 1$ we get $g = 1$. The unit effect is given as $(1, 1)$. Now let $(f, g) \in A(K_A) \times A(K_B)$, then we have $(0, 0) \leq (f, g) \leq (1, 1)$ if and only if $f \in E(K_A)$ and $g \in E(K_B)$. We conclude that $E(K_A \oplus K_B)$ is an effect algebra of ordered pairs of effects. This motivates the following definition.

Definition 27. The direct product of effect algebras $E(K_A)$ and $E(K_B)$ is denoted $E(K_A) \times E(K_B)$ and it is defined as

$$E(K_A) \times E(K_B) = E(K_A \oplus K_B).$$

Of course we could have first defined the direct product of effect algebras as the set of ordered pairs of effect as in [40], then proved that this is an effect algebra and then proved that $E(K_A) \times E(K_B) = E(K_A \oplus K_B)$. We have chosen to define $E(K_A) \times E(K_B)$ to be equal to $E(K_A \oplus K_B)$ as we find it more convenient.

Now let us construct $E(K_A \times K_B)$. Note that for every $(x, y) \in K_A \times K_B$ we have

$$\langle (x, y), (1, -1) \rangle = 0$$

so for $f_1, f_2 \in A(K_A)$ and $g_1, g_2 \in A(K_B)$ such that $(f_1, f_2) = (g_1, g_2) + \alpha(1, -1)$ for some $\alpha \in \mathbb{R}$ we have

$$\langle (x, y), (f_1, f_2) \rangle = \langle (x, y), (g_1, g_2) \rangle.$$

This introduces a relation of equivalence $\approx$ on $A(K_A) \times A(K_B)$: let $(f_1, f_2), (g_1, g_2) \in A(K_A) \times A(K_B)$, then $(f_1, f_2) \approx (g_1, g_2)$ if $(f_1, f_2) = (g_1, g_2) + \alpha(1, -1)$ for some $\alpha \in \mathbb{R}$. It is easy to see that $\approx$ is symmetric, reflexive and transitive. Let $(A(K_A) \times A(K_B))_\approx$ denote the vector space $A(K_A) \times A(K_B)$ factorized with respect to $\approx$, we will denote the equivalence class of the element $(f_1, f_2) \in A(K_A) \times A(K_B)$ as $(f_1, f_2)_\approx$. It is clear that $(A(K_A) \times A(K_B))_\approx$ must contain $E(K_A \times K_B)$ as we have just removed a redundant degree of freedom. Let $(A(K_A)^+ \times A(K_B)^+)_\approx$ denote the naturally induced positive cone, one can check that the cone is convex, closed, pointed and generating. The cone $(A(K_A)^+ \times A(K_B)^+)_\approx$ gives rise to an ordering $\leq_\approx$ such that we have $(0, 0) \leq_\approx (f_1, f_2)_\approx$ if and only if $(f_1, f_2)_\approx \in (A(K_A)^+ \times A(K_B)^+)_\approx$. The unit effect is the equivalence class of $(1, 0) \approx (0, 1)$ and we have

$$E(K_A \times K_B) = \{(f_1, f_2)_\approx \in (A(K_A) \times A(K_B))_\approx : (0, 0)_\approx \leq_\approx (f_1, f_2)_\approx \leq_\approx (1, 0)_\approx\}.$$

This construction is actually rather similar to the direct convex sum of state spaces, although it may not be obvious at first.

Definition 28. The direct convex sum of effect algebras $E(K_A)$ and $E(K_B)$ is denoted $E(K_A) \oplus E(K_B)$ and is defined as

$$E(K_A) \oplus E(K_B) = E(K_A \times K_B).$$

Again, we could have introduced the abstract definition of direct convex hull of effect algebras as in [40] and then proved that it is the same as the effect algebra on the direct product of state spaces. Again, we argue that the used approach was more convenient.

Note that we can extend the presented definitions of direct products and direct convex sums from two state spaces to any finite number of state spaces in a straightforward way. Moreover one can also show that the operations of direct product and direct convex sum are associative.

At last we will present few more result concerning direct products and direct convex sums of state spaces.

Proposition 26. *The set of extreme points of $K_A \times K_B$ is*

$$\text{ext}(K_A \times K_B) = \{(x, y) : x \in \text{ext}(K_A), y \in \text{ext}(K_B)\}.$$

Proof. Let $(x, y), (x_1, y_1), (x_2, y_2) \in K_A \times K_B$ such that for some $\lambda \in [0, 1]$ we have

$$(x, y) = \lambda(x_1, y_1) + (1 - \lambda)(x_2, y_2)$$

then also $x = \lambda x_1 + (1 - \lambda)x_2$ and $y = \lambda y_1 + (1 - \lambda)y_2$. If $x \in \text{ext}(K_A)$ and $y \in \text{ext}(K_B)$ we get $x_1 = x_2 = x$ and $y_1 = y_2 = y$ so $(x, y) = (x_1, y_1) = (x_2, y_2)$ and $(x, y) \in \text{ext}(K_A \times K_B)$. If $(x, y) \in \text{ext}(K_A \times K_B)$ then we must have $(x, y) = (x_1, y_1) = (x_2, y_2)$ so $x_1 = x_2 = x$, $y_1 = y_2 = y$ follows and we conclude that $x \in \text{ext}(K_A)$, $y \in \text{ext}(K_B)$. $\square$

Proposition 27. *The set of extreme points of $K_A \oplus K_B$ is given as*

$$\text{ext}(K_A \oplus K_B) = \{(x, 0) : x \in \text{ext}(K_A)\} \cup \{(0, y) : y \in \text{ext}(K_B)\}$$

Proof. Let $(\lambda x, (1 - \lambda)y) \in K_A \oplus K_B$ for some $x \in K_A$, $y \in K_B$ and $\lambda \in [0, 1]$ then clearly

$$(\lambda x, (1 - \lambda)y) = \lambda(x, 0) + (1 - \lambda)(0, y).$$

It follows that we have to require either $\lambda = 0$ or $\lambda = 1$. Moreover one can see that $(x, 0) \in \text{ext}(K_A \oplus K_B)$ if and only if $x \in \text{ext}(K_A)$ and similarly $(0, y) \in \text{ext}(K_A \oplus K_B)$ if and only if $y \in \text{ext}(K_B)$. $\square$

Proposition 28. *Let S_n be a simplex and let K be a state space, then*

$$K \dot\otimes S_n = K \oplus K \oplus \ldots \oplus K$$

where the sum on the right hand side contains exactly n copies of K.

Proof. Let $s_1, \ldots, s_n$ denote the extreme points of S_n, then every $x \in K \dot\otimes S_n$ can be written as

$$x = \sum_{i=1}^{n} \lambda_i y_i \otimes s_i$$

where $y_i \in K$ and $\lambda_i \in [0, 1]$ for all $i \in \{1, \ldots, n\}$ and $\sum_{i=1}^{n} \lambda_i = 1$. Since $s_1, \ldots, s_n$ is a basis of $A(S_n)^*$ the decomposition is unique. It follows that we can identify x with $(\lambda_1 y_1, \ldots, \lambda_n y_n) \in K \oplus \ldots \oplus K$. $\square$

We can also use similar ideas to characterize state spaces that are direct convex sums of other state spaces, see [41, Proposition 3.].

4 Transformations

In this section we are going to introduce transformations of state spaces, which will include channels, measurements and operations. The main message of this section is that we are going to show that all of the transformations can be understood as channels and that channels can naturally be identified with elements of some tensor product.

4.1 Channels

Let K_A and K_B be state spaces, then channel Φ is a transformation that takes a state from K_A as input and outputs a state from K_B. According to our interpretation of convex combinations we must require that channels are affine maps. As a result of affinity we can extend the channel to map $A(K_A)^*$ to $A(K_B)^*$ in a natural way, i.e. let $\psi \in A(K_A)^*$ be such that for $x, y \in K_A$ and $\lambda, \mu \in \mathbb{R}^+$ we have $\psi = \lambda x - \mu y$, then $\Phi(\psi) = \lambda \Phi(x) - \mu \Phi(y)$. To see that this is well defined, assume that

$$\psi = \lambda x - \mu y = \lambda' x' - \mu' y'$$

for some $x', y' \in K_A$ and $\lambda', \mu' \in \mathbb{R}^+$. If $\lambda + \mu' = 0$ then $\psi = 0$ hence we can assume $\lambda + \mu' > 0$. We have

$$\frac{\lambda}{\lambda + \mu'} x + \frac{\mu'}{\lambda + \mu'} y' = \frac{\lambda'}{\lambda + \mu'} x' + \frac{\mu}{\lambda + \mu'} y.$$

Note that $\lambda + \mu' = \lambda' + \mu$ as x, x', y, y' are all states, so we actually have the equality of two convex combinations. From the affinity of Φ we get

$$\Phi\left(\frac{\lambda}{\lambda + \mu'} x + \frac{\mu'}{\lambda + \mu'} y'\right) = \frac{\lambda}{\lambda + \mu'} \Phi(x) + \frac{\mu'}{\lambda + \mu'} \Phi(y')$$

and

$$\Phi\left(\frac{\lambda'}{\lambda + \mu'} x' + \frac{\mu}{\lambda + \mu'} y\right) = \frac{\lambda'}{\lambda + \mu'} \Phi(x') + \frac{\mu}{\lambda + \mu'} \Phi(y).$$

It follows that

$$\lambda \Phi(x) - \mu \Phi(y) = \lambda' \Phi(x') - \mu' \Phi(y')$$

so Φ is well defined. Moreover note that if $\psi \geq 0$ then also $\Phi(\psi) \geq 0$, i.e. Φ is a positive map. Also $1(\psi) = 1(\Phi(\psi))$, which can also be written as $1 = 1 \circ \Phi$, where $\circ$ denotes the concatenation of maps. We will call a map such that $1 \circ \Phi = 1$ trace-preserving as this is the name used in quantum theory.

Definition 29. A channel is an affine, positive and trace-preserving map. The set of all channels mapping a state space K_A to state space K_B will be denoted $\mathfrak{C}(K_A \to K_B)$.

We are going to revisit the assumption of positivity in Subsec. 4.2 where we will discuss complete positivity.

In what follows we will essentially use a construction similar to the ones that can be found in [3] and we will identify channels with elements of some tensor product. We strongly recommend the reader that may struggle to apprehend the following construction to look at the first chapter of [3].

Let $x \in K_A$ and $f \in E(K_B)$ and consider a channel $\Phi : K_A \to K_B$ and the expression $\langle \Phi(x), f \rangle$. On one hand we interpret this as that the effect f maps the state $\Phi(x)$ to probability, but on the other hand we may also see Φ as a bilinear form that assigns the value of $\langle \Phi(x), f \rangle$ to x and f. Hence we may write $\langle \Phi(x), f \rangle = \langle \Phi, x \otimes f \rangle$ where we have identified the pair x and f with the linear functional acting on the bilinear form Φ. Extending this expression by linearity we see that we can identify a channel $\Phi : K_A \to K_B$ with an element of $(A(K_A)^* \otimes A(K_B))^* = A(K_A) \otimes A(K_B)^*$. Then there are $g_i \in A(K_A)$ and $\psi_i \in A(K_B)^*$ for $i \in \{1, \dots, n\}$ such that we have

$$\Phi = \sum_{i=1}^n g_i \otimes \psi_i$$

where we omitted an isomophism between the representations of Φ as a map and as an element of the tensor product. For $x \in K_A$ we get

$$\Phi(x) = \sum_{i=1}^{n} \langle x, g_i \rangle \psi_i.$$

Note that the functions g_i nor the functionals ψ_i do not have to be positive, only the corresponding map Φ is positive and trace-preserving. Φ is positive if an only if $\Phi(x)$ is positive for any $x \in K_A$, which is the same as requiring that for every $f \in E(K_B)$ we have

$$0 \leq \langle \Phi(x), f \rangle = \langle \Phi, x \otimes f \rangle = \sum_{i=1}^{n} \langle x, g_i \rangle \langle \psi_i, f \rangle.$$

It follows that we must have

$$\Phi \in A(K_A)^{+} \hat{\otimes} A(K_B)^{*+}.$$

Φ is trace preserving if and only if for all $x \in K_A$ we have

$$\sum_{i=1}^{n} \langle x, g_i \rangle \langle \psi_i, 1 \rangle = 1$$

which is equivalent to

$$\sum_{i=1}^{n} g_i \langle \psi_i, 1 \rangle = 1 \circ \Phi = 1.$$

The above construction of assigning a vector in the tensor product to the channel is independent of positivity or trace-preserving. Matter of fact any linear map $L : A(K_A)^* \to A(K_B)^*$ can be identified with some vector from $A(K_A) \otimes A(K_B)^*$.

For $\Phi_1, \Phi_2 \in \mathfrak{C}(K_A \to K_B)$ we can define the convex combination of the channels with $\lambda \in [0, 1]$ as the unique channel such that for $x \in K_A$ we have

$$(\lambda \Phi_1 + (1 - \lambda)\Phi_2)(x) = \lambda \Phi_1(x) + (1 - \lambda)\Phi_2(x).$$

Note that $\mathfrak{C}(K_A \to K_B)$ is not a base of $A(K_A)^{+} \hat{\otimes} A(K_B)^{*+}$ as let $f \in E(K_A)$ be a non-constant effect and let $x \in K_B$, then $f \otimes x \in A(K_A)^{+} \hat{\otimes} A(K_B)^{*+}$ but it is not a multiple of any channel as $1 \circ (f \otimes x) = f$. But clearly the set of channels $K_A \to K_B$ is a base of some smaller cone.

Proposition 29. *Let $C = \text{span}(\mathfrak{C}(K_A \to K_B))$ then $\mathfrak{C}(K_A \to K_B)$ is a base of the cone $(A(K_A)^{+} \hat{\otimes} A(K_B)^{*+}) \cap C$.*

Proof. Let $\psi \in (A(K_A)^{+} \hat{\otimes} A(K_B)^{*+}) \cap C$. ψ is clearly positive and $1 \circ \psi = \lambda 1$ for some $\lambda \geq 0$. Assume $\lambda \neq 0$ then $\frac{1}{\lambda} \psi \in \mathfrak{C}(K_A \to K_B)$ as it is positive and trace-preserving. $\square$

One can again apply the same framework of GPTs to set of channels when we consider it as a base of cone, i.e. as some state space. In quantum theory this is referred to as the framework of PPOVMs [42] or quantum testers [43].

For channels mapping classical state spaces to classical state spaces we can obtain a more specific result.

Proposition 30. *Let S_{n_1} and S_{n_2} be classical state spaces and let $\nu \in \mathfrak{C}(S_{n_1} \to S_{n_2})$ then there are numbers $(\nu)_{ij} \in \mathbb{R}^+$, $i \in \{1, \ldots, n_1\}$, $j \in \{1, \ldots, n_2\}$ such that*

$$\nu = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (\nu)_{ij} b_i \otimes s_j.$$

Moreover it must hold that

$$\sum_{j=1}^{n_2} (\nu)_{ij} = 1.$$

for all $i \in \{1, \ldots, n_1\}$.

Proof. We already know that $\{b_1, \ldots, b_{n_1}\}$ is a base of $A(S_{n_1})$ and $\{s_1, \ldots, s_{n_2}\}$ is a base of $A(S_{n_2})^*$ and since $\nu \in A(S_{n_1}) \otimes A(S_{n_2})^*$ we must have

$$\nu = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (\nu)_{ij} b_i \otimes s_j.$$

for some $\nu_{ij} \in \mathbb{R}$, $i \in \{1, \ldots, n_1\}$, $j \in \{1, \ldots, n_2\}$. From

$$\langle \nu(s_i), b_j \rangle = (\nu)_{ij}$$

it follows that we must have $(\nu)_{ij} \in \mathbb{R}^+$ and $\sum_{j=1}^{n_2} (\nu)_{ij} = 1$. $\qquad \square$

Example 5. Consider $\mathfrak{C}(S_2 \rightarrow S_2)$. According to Prop. 30 every element of $\mathfrak{C}(S_2 \rightarrow S_2)$ is uniquely determined by the numbers $\nu_{11}, \nu_{12}, \nu_{21}, \nu_{22}$ such that $\nu_{12} = 1 - \nu_{11}$, $\nu_{22} = 1 - \nu_{21}$ and $\nu_{11}, \nu_{21} \in [0, 1]$. It follows that we can identify every point of $\mathfrak{C}(S_2 \rightarrow S_2)$ with a state of the square state space S by the map $(\nu_{11}, \nu_{12}, \nu_{21}, \nu_{22}) \mapsto (\nu_{11}, \nu_{21})$, moreover one can show that this map is an isomorphism. It follows that we have $\mathfrak{C}(S_2 \rightarrow S_2) = S$ up to the aforementioned isomorphism.

The last elementary construction for channels that we are going to introduce is the construction of adjoint channel, historically also called the Heisenberg picture. Let $\Phi \in \mathfrak{C}(K_A \rightarrow K_B)$, let $x \in K_A$ and $f \in E(K_B)$, and consider a function $h : K_A \rightarrow \mathbb{R}$ given as

$$h(x) = \langle \Phi(x), f \rangle.$$

It is straightforward to see that $h \in E(K_A)$.

Definition 30. Let $\Phi \in \mathfrak{C}(K_A \rightarrow K_B)$, then we define the linear map $\Phi^* : A(K_B) \rightarrow A(K_A)$ as the unique map such that for all $x \in K_A$ and $f \in E(K_B)$ we have

$$\langle \Phi(x), f \rangle = \langle x, \Phi^*(f) \rangle$$

and we call Φ^* the adjoint of Φ.

One can again see that the same construction can be used for any linear map $L : A(K_A)^* \rightarrow A(K_B)^*$. It is possible to see that both Φ and Φ^* correspond to the same vector in $A(K_A) \otimes A(K_B)^*$. Let $\Phi = \sum_{i=1}^n g_i \otimes \psi_i$, then for $f \in E(K_B)$ we have

$$\Phi^*(f) = \sum_{i=1}^n \langle \psi_i, f \rangle g_i.$$

Since we require a channel to be positive and trace-preserving by definition, we will look at what properties follow for the adjoint map.

Proposition 31. *Let $L : A(K_A)^* \rightarrow A(K_B)^*$ be a linear map, then L is an adjoint of a channel $\Phi \in \mathfrak{C}(K_A \rightarrow K_B)$, i.e. $L = \Phi^*$ if and only if L is positive and unital, i.e. $L(1) = 1$.*

Proof. If $L = \Phi^*$ then clearly $L(f) \geq 0$ for all $f \in A(K_B)^+$, i.e. L is positive. Moreover for any $x \in K_A$ we have

$$\langle x, \Phi^*(1) \rangle = \langle \Phi(x), 1 \rangle = 1$$

so $\Phi^*(1) = 1$, i.e. $L = \Phi^*$ is unital.

Now assume that L is unital and positive and let L^* be the adjoint of L, that is the unique linear map such that for all $x \in K_A$ and $f \in E(K_B)$ we have

$$\langle x, L(f) \rangle = \langle L^*(x), f \rangle.$$

It follows that L^* is positive as $\langle L^*(x), f \rangle \geq 0$ for all $f \in E(K_B)$. Moreover

$$\langle L^*(x), 1 \rangle = \langle x, L(1) \rangle = \langle x, 1 \rangle = 1$$

so L^* is also trace-preserving; it follows that $L^* \in \mathfrak{C}(K_A \rightarrow K_B)$. Finally it is rather easy to see that $(L^*)^* = L$ as for all $x \in K_A$ and $f \in E(K_B)$ we have

$$\langle x, L(f) \rangle = \langle L^*(x), f \rangle = \langle x, (L^*)^*(f) \rangle.$$

It follows that L is an adjoint of the channel L^*. $\qquad \square$

4.2 Complete positivity

We are going to revisit the requirement of positivity of channels; we are going to argue that in general circumstances requiring only positivity is not enough. Let K_A, K_B and K_C be state spaces such that $K_C \tilde{\otimes} K_A$ is defined and let $\Phi \in \mathfrak{C}(K_A \to K_B)$. Assuming that we can separate (in space or in time or in any other sense) the systems K_A and K_C, we can clearly apply the channel Φ to the K_A part of $K_C \tilde{\otimes} K_A$. In other words we assume that given $\Phi \in \mathfrak{C}(K_A \to K_B)$ and the tensor product $K_C \tilde{\otimes} K_A$ we can always construct the map $\mathrm{id}_C \otimes \Phi : A(K_C)^* \otimes A(K_A)^* \to A(K_C)^* \otimes A(K_B)^*$, where id_C is the identity map, i.e. for $\varphi \in A(K_C)^*$ and $\psi \in A(K_A)^*$ we have

$$(\mathrm{id}_C \otimes \Phi)(\varphi \otimes \psi) = \varphi \otimes \Phi(\psi).$$

We are going to proceed with the argument that $\mathrm{id}_C \otimes \Phi$ must be a well defined channel. It follows that we must also assume that $K_C \tilde{\otimes} K_B$ is defined. To see that the map $\mathrm{id}_C \otimes \Phi$ is trace-preserving, let $x \in K_C \tilde{\otimes} K_A$, then we have

$$\langle (\mathrm{id}_C \otimes \Phi)(x), 1 \otimes 1 \rangle = \langle x, 1 \otimes \Phi^*(1) \rangle = 1$$

where we have used that the adjoint of $\mathrm{id}_C \otimes \Phi$ is $(\mathrm{id}_C \otimes \Phi)^* = \mathrm{id}_C \otimes \Phi^*$ and the unitality of Φ^*.

The troublesome part is that $\mathrm{id}_C \otimes \Phi$ does not have to be positive even when Φ is positive! This problem is well known even in quantum theory and also can be used to our advantage to detect entanglement [6, Section 6.3.3]. In conclusion we have to add an additional requirement to Φ that also $\mathrm{id}_C \otimes \Phi$ is positive whenever $K_C \tilde{\otimes} K_A$ is defined. We call this requirement complete positivity.

Definition 31. We say that a channel $\Phi \in \mathfrak{C}(K_A \to K_B)$ is completely positive with respect to the tensor product $K_C \tilde{\otimes} K_A$ if $\mathrm{id}_C \otimes \Phi$ is positive.

Note that the notion of complete positivity does not depend on the state space, only on the cone $A(K_C)^{*+} \tilde{\otimes} A(K_A)^{*+}$. It follows that in the same way we can define the complete positivity of the adjoint channel mapping the effect algebras. A nice example of completely positive channel is the identity channel.

In most cases we will always assume that channels are completely positive when needed. Also in most of applications to quantum theory one always assumes complete positivity of all channels.

We will present two elementary results concerning the complete positivity.

Proposition 32. *A channel $\Phi \in \mathfrak{C}(K_A \to K_B)$ is completely positive with respect to $K_C \tilde{\otimes} K_A$ if and only if the adjoint channel Φ^* is completely positive with respect to $E(K_C) \tilde{\otimes} E(K_B) = E(K_C \tilde{\otimes} K_B)$.*

Proof. Let $x \in K_C \tilde{\otimes} K_A$ and $f \in E(K_C) \tilde{\otimes} E(K_B)$. Assume that $\Phi \in \mathfrak{C}(K_A \to K_B)$ is completely positive with respect to $K_C \tilde{\otimes} K_A$, then

$$0 \le \langle (\mathrm{id}_C \otimes \Phi)(x), f \rangle = \langle x, (\mathrm{id}_C \otimes \Phi^*)(f) \rangle$$

which shows that we must have $(\mathrm{id}_C \otimes \Phi^*)(f) \ge 0$.

Assume that Φ^* is completely positive with respect to $E(K_A) \tilde{\otimes} E(K_B)$, then the complete positivity of Φ with respect to $K_C \tilde{\otimes} K_A$ follows in a similar manner. $\square$

In case when the real tensor product coincides with either minimal or maximal tensor product we can show that the notions of positivity and complete positivity coincide.

Proposition 33. *Let K_A and K_C be state spaces and let $\Phi \in \mathfrak{C}(K_A \to K_B)$. Φ is completely positive with respect to $K_C \dot{\otimes} K_A$ and $K_C \hat{\otimes} K_A$.*

Proof. First consider the case of $K_C \dot{\otimes} K_A$ and let $x \in K_A$, $y \in K_C$. As we have $\Phi(x) \in K_B$ it follows that

$$(\mathrm{id}_C \otimes \Phi)(y \otimes x) = y \otimes \Phi(x) \in K_C \dot{\otimes} K_B.$$

Complete positivity of Φ with respect to $K_C \dot{\otimes} K_A$ follows as all of the extreme points of $K_C \dot{\otimes} K_A$ are of the form $y \otimes x$.

Now consider the case of $K_C \hat{\otimes} K_A$. We have $E(K_C \hat{\otimes} K_A) = E(K_A) \dot{\otimes} E(K_C)$ by definition. It is rather straighforward to check that Φ^* is always completely positive with respect to the minimal tensor product $E(K_A) \dot{\otimes} E(K_C)$ in the same way as in the first part of the proof. The complete positivity of Φ with respect to $K_C \hat{\otimes} K_A$ follows from Prop. 32. $\square$

4.3   Entanglement-breaking channels

We are going to consider a special class of channels such that whenever we apply them to a part of entangled state, we get a separable state.

Definition 32. Let $\Phi \in \mathfrak{C}(K_A \to K_B)$ be a channel, then we say that Φ is entanglement breaking with respect to $K_C \tilde{\otimes} K_A$ if for every $x \in K_C \tilde{\otimes} K_A$ we have

$$(\mathrm{id}_C \otimes \Phi)(x) \in K_C \dot{\otimes} K_B.$$

Note that as a result of the definition all entanglement-breaking channels are completely positive with respect to the same tensor product. To see that not all completely positive channels are entanglement-breaking, let K_A, K_B be state spaces such that $K_A \tilde{\otimes} K_B \neq K_A \dot{\otimes} K_B$, then one can see that the identity channel $\mathrm{id}_A \in \mathfrak{C}(K_A \to K_A)$ is completely positive but not entanglement breaking.

4.4   Measurements

In the beginning we were discussing measurements when introducing our postulates, now we will finally provide a proper definition. We have postponed providing the definition sooner as we believe than now it will be mathematically more straightforward and nicer.

The classical state space can be interpreted as the set of probability measures over a finite set of points (with respect to the sigma algebra of all subsets of given set). From the operational standpoint we require that the outcome of a measurement is a probability measure over the possible set of outcomes, hence a measurement is going to be an affine map that assigns a probability measure to a state. We are going to restrict to only finitely many outcomes.

Definition 33. n-outcome measurement m on the state space K is a channel $m : K \to S_n$.

The beauty in defining the measurements as special cases of channels is that we can already use all of the results we have derived for channel in Subsec. 4.1. Moreover the special structure of measurements allows us to obtain more specific results.

Let $m : K \to S_n$ be a measurement, then we already know that we have $m \in A(K)^+ \hat{\otimes} A(S_n)^{*+}$. According to Coro. 7 we must have $A(K)^+ \hat{\otimes} A(S_n)^{*+} = A(K)^+ \dot{\otimes} A(S_n)^{*+}$ and it follows that if we denote $\mathrm{ext}(S_n) = \{s_1, \ldots, s_n\}$ then there are unique functions $f_i \in A(K)^+$, $i \in \{1, \ldots, n\}$ such that

$$m = \sum_{i=1}^{n} f_i \otimes s_i.$$

m is trace-preserving if and only if we have $\sum_{i=1}^{n} f_i = 1$ which shows that we must have $f_i \in E(K)$ for all $i \in \{1, \ldots, n\}$. Here we would like to return to the question of interpreting effects as now we can provide a full answer. The effects are the building blocks of measurements and every measurement is isomorphic to a tuple of effects $(g_1, \ldots, g_n)$ such that $\sum_{i=1}^{n} g_i \leq 1$, the corresponding measurement is given as $m = \sum_{i=1}^{n} g_i \otimes s_i + (1 - \sum_{i=1}^{n} g_i) \otimes s_{n+1}$. We could have also started by defining measurements as tuples of effects.

Proposition 34. *Let K_A, K_B be state spaces and let $m : K_A \to S_n$ be a measurement and assume that $K_B \tilde{\otimes} K_A$ is defined, then m is entanglement-breaking with respect to $K_B \tilde{\otimes} K_A$.*

Proof. Let $x \in K_B \tilde{\otimes} K_A$ and let $m = \sum_{i=1}^{n} f_i \otimes s_i$ for some $f_i \in E(K_A)$, $i \in \{1, \ldots, n\}$. As a result of Prop. 17 we know that $(\mathrm{id}_B \otimes f_i)(x) \in A(K_B)^{*+}$. We have

$$(\mathrm{id}_B \otimes m)(x) = \sum_{i=1}^{n} (\mathrm{id}_B \otimes f_i)(x) \otimes s_i \in A(K_B)^{*+} \dot{\otimes} A(S_n)^{*+}.$$

$(\mathrm{id}_B \otimes m)(x) \in K_B \dot{\otimes} S_n$ follows from the trace-preserving of m. □

Corollary 8. *Every measurement is completely positive with respect to all tensor products.*

Proof. The result follows from Prop. 34. □

A special type of measurement is a two-outcome (also called dichotomic) measurement, that is a measurement $m : K \to S_2$. One can easily see that the set of two-outcome measurements is isomorphic to the effect algebra $E(K)$ using the aforementioned isomorphism between tuples of effects and measurements. Two-outcome measurements are important as they are the simplest non-trivial measurements that any theory provides.

4.5   Measure-and-prepare channels

Measure and prepare channels are conceptually similar to measurements; the operational idea is that we want to first measure and then prepare a state according to the measurement outcome we have obtained.

Definition 34. Let $\Phi \in \mathfrak{C}(K_A \to K_B)$, then Φ is a measure-and-prepare channel if there are $f_i \in E(K_A)$ and $x_i \in K_B$ for $i \in \{1, \dots, n\}$ such that $\sum_{i=1}^n f_i = 1$ and such that we have

$$\Phi = \sum_{i=1}^n f_i \otimes x_i.$$

Note that measurements are a special type of measure-and-prepare channels. The following result is immediate.

Proposition 35. *Let $\Phi \in \mathfrak{C}(K_A \to K_B)$ be measure-and-prepare, then Φ is entanglement-breaking with respect to any tensor product $K_C \tilde{\otimes} K_A$.*

Proof. The proof mimicks the steps of the proof of Prop. 34. Let $\Phi = \sum_{i=1}^n f_i \otimes x_i$ for some $f_i \in E(K_A)$ and $x_i \in K_B$, $i \in \{1, \dots, n\}$ such that $\sum_{i=1}^n f_i = 1$ and let $y \in K_C \tilde{\otimes} K_A$. We have

$$(\mathrm{id}_C \otimes \Phi)(y) = \sum_{i=1}^n (\mathrm{id}_C \otimes f_i)(y) \otimes x_i \in A(K_C)^{*+} \dot{\otimes} A(K_B)^{*+}.$$

$\square$

Corollary 9. *Let $\Phi \in \mathfrak{C}(K_A \to K_B)$ be measure-and-prepare, then Φ is completely positive with respect to any tensor product $K_C \tilde{\otimes} K_A$.*

Proof. The result again follows from Prop. 34. $\square$

It is known that in quantum theory all entanglement-breaking channels with respect to the real tensor product used in quantum theory are measure-and-prepare channels [44–46]. One can also obtain such result in a more general settings of weakly self-dual theories [47]. We will show that the same holds in any other theory if we consider a channel to be entanglement-breaking with respect to maximal tensor product with all possible compact convex sets.

Proposition 36. *Assume that a channel $\Phi \in \mathfrak{C}(K_A \to K_B)$ is entanglement-breaking with respect to all compact convex sets $K_C \subset V$, where V is a finite dimensional real vector space, and with respect to the maximal tensor products $K_C \hat{\otimes} K_A$, then Φ is measure-and-prepare.*

Proof. Let $x_i \in K_A$ for $i \in \{1, \dots, n\}$ be a set of states such that it is a basis of $A(K_A)^*$ and let $f_i \in A(K_A)$ for $i \in \{1, \dots, n\}$ be the dual basis, i.e. we have $\langle x_i, b_j \rangle = \delta_{ij}$ for all $i, j \in \{1, \dots, n\}$. Note that the functions b_i do not have to be positive and one can show that they are all positive if and only if K_A is a simplex. Now consider

$$\psi = \sum_{i=1}^n f_i \otimes x_i,$$

we are going to show that $\psi \in A(K_A)^+ \hat{\otimes} A(K_A)^{*+}$. Let $y \in K_A$ be a state such that $y = \sum_{i=1}^n \alpha_i x_i$ for some $\alpha_i \in \mathbb{R}$, $i \in \{1, \dots, n\}$ and let $g \in A(K_A)^+$ be such that $g = \sum_{i=1}^n \beta_i f_i$ for some $\beta_i \in \mathbb{R}$, $i \in \{1, \dots, n\}$, we have

$$\langle \psi, y \otimes g \rangle = \sum_{i=1}^n \langle f_i, y \rangle \langle x_i, g \rangle = \sum_{i=1}^n \alpha_i \beta_i = \langle y, g \rangle \geq 0.$$

Also we argue that for some suitable choice of base of the cone $A(K_A)^+$ the vector ψ is a state, e.g. let $u = \frac{1}{n} \sum_{i=1}^{n} x_i$ then $\langle \psi, u \otimes 1 \rangle = 1$. We have $\Phi = \sum_{i=1}^{n} h_i \otimes \varphi_i$ for some $h_i \in A(K_A)$ and $\varphi_i \in A(K_B)^*$, $i \in \{1, \ldots, n\}$ and

$$(\mathrm{id} \otimes \Phi)(\psi) = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} \langle h_j, x_i \rangle f_i \right) \otimes \varphi_j = \sum_{j=1}^{n} h_j \otimes \varphi_j.$$

It follows that if we require $(\mathrm{id} \otimes \Phi)(\psi) \in A(K_A)^+ \dot\otimes A(K_B)^{*+}$ then also $\Phi \in A(K_A)^+ \dot\otimes A(K_B)^{*+}$, i.e. Φ is measure-and-prepare. $\qquad \square$

4.6 Instruments and operations

When we were defining measurements, we were only concerned with what measurement outcome we obtain, but this is not entirely realistic model. In reality when we do a measurement we not only obtain a measurement outcome, but also some post-measurement state of the system. Therefore let K be a state space, then we should describe an n-outcome measurement by a channel $\Phi : K \to S_n \dot\otimes K$, where $S_n \dot\otimes K$ is the joint system of the measurement outcomes and the post-measurement state. We have assumed that the pre-measurement and post-measurement states belong to the same state space.

Definition 35. An instrument is a channel $\Phi : K \to S_n \dot\otimes K$.

It is straightforward that we can reconstruct the measurement from a given instrument simply by taking the partial trace over the post-measurement system K.

Definition 36. Let $m \in \mathfrak{C}(K \to S_n)$ be a measurement and let $\Phi \in \mathfrak{C}(K \to S_n \dot\otimes K)$ be an instrument, then we say that Φ is an instrument for the measurement m if

$$m = (\mathrm{id}_{S_n} \otimes 1) \circ \Phi$$

where id_{S_n} is the identity map on S_n, i.e. if m is the measurement that we obtain from Φ by taking the partial trace over the post-measurement system K.

We say that a map Φ is trace-non-increasing whenever we have $1 \circ \Phi \leq 1$. The following structural characterization is useful.

Proposition 37. *Let $\Phi : K \to K \dot\otimes S_n$ be an instrument, then there are positive, trace-non-increasing maps $\Phi_i : K \to A(K)^{*+}$, $i \in \{1, \ldots, n\}$ such that*

$$\Phi = \sum_{i=1}^{n} \Phi_i \otimes s_i,$$

i.e. for $x \in K$ we have $\Phi(x) = \sum_{i=1}^{n} \Phi_i(x) \otimes s_i$.

Proof. Since $s_1, \ldots, s_n$ is a basis of $A(S_n)^*$ and since $\Phi \in A(K) \otimes A(K)^* \otimes A(S_n)^*$ the we have

$$\Phi = \sum_{i=1}^{n} \Phi_i \otimes s_i$$

where $\Phi_i \in A(K) \otimes A(K)^*$, i.e. $\Phi_i : A(K)^* \to A(K)^*$ are linear maps for all $i \in \{1, \ldots, n\}$, it only remains to show that they are positive and trace-non-increasing. Let $b_i \in E(S_n)$ for $i \in \{1, \ldots, n\}$ be the functions such that $\langle s_i, b_j \rangle = \delta_{ij}$ for all $i, j \in \{1, \ldots, n\}$ and let $x \in K$, $f \in E(K)$, then we have

$$0 \leq \langle \Phi(x), f \otimes b_i \rangle = \langle \Phi_i(x), f \rangle$$

so the maps Φ_i are positive for all $i \in \{1, \ldots, n\}$. Since $b_i \in E(S_n)$ and $1 = \sum_{i=1}^{n} b_i$, we get that

$$\sum_{i=1}^{n} 1 \circ \Phi_i = 1$$

and $1 \circ \Phi_i \leq 1$ follows. $\qquad \square$

Clearly the set of positive, trace-non-increasing maps is of interest in a given theory.

Definition 37. Positive, trace-non-increasing map $\Phi : K \to A(K)^{*+}$ is called an operation.

Note that for simplicity we again require only positivity, but with regards to the same arguments as for channel, one should also require all instruments to be completely positive, as it is usually done in quantum theory.

We will shortly investigate the structure of instruments that give the same measurement.

Proposition 38. *Let* $\Phi_m \in \mathfrak{C}(K \to S_n \dot{\otimes} K)$ *be an instrument for a measurement* $m \in \mathfrak{C}(K \to S_n)$ *and let* $\Phi \in \mathfrak{C}(K \to K)$, $\mathrm{id}_{S_n} \in \mathfrak{C}(S_n \to S_n)$ *be the identity channel, then also* $(\mathrm{id}_{S_n} \otimes \Phi) \circ \Phi_m$ *is an instrument for the measurement* m.

Proof. We have
$$(\mathrm{id}_{S_n} \otimes 1) \circ (\mathrm{id}_{S_n} \otimes \Phi) \circ \Phi_m = (\mathrm{id}_{S_n} \otimes 1) \circ \Phi_m = m.$$

$\square$

One can ask whether there exist an instrument Φ_m such that every other instrument for the measurement $(\mathrm{id}_{S_n} \otimes 1) \circ \Phi_m$ is of the form $(\mathrm{id}_{S_n} \otimes \Phi) \circ \Phi_m$ for some $\Phi \in \mathfrak{C}(K \to K)$. This is true in some theories, specifically in quantum theory this is true for the Lüders instrument [6, Section 5.3].

5 Compatibility and similar notions

We have already explored entanglement as an aspect of non-classical theories; in this section we are going to look at incompatibility. Incompatibility is mostly associated with measurements and especially in quantum theory it is sometimes seen as generalization of non-commutativity of projection measures. Our approach is going to be more general. We will introduce compatibillity as property of channels, but in the special case of measurements in quantum theory, one can reconstruct the well known results [6].

Compatibility of measurements has been heavily investigated in quantum theory [48–53] but also in the framework of GPTs [47, 54, 55] and compatibility of channels in quantum theory has been considered before [56]. The approach that we are going to present is going to consider channels and we are only going to assume special properties of the channels only when needed.

5.1 Compatibility of channels

Let K_A, K_B and K_C be state spaces and let $\Phi_1 \in \mathfrak{C}(K_A \to K_B)$ and $\Phi_2 \in \mathfrak{C}(K_A \to K_C)$. Consider the following task: we are given an unknown $x \in K_A$ and we want to obtain the states $\Phi_1(x)$ and $\Phi_2(x)$, both at the same time and by using only single copy of the given $x \in K_A$. It would be tempting to say that we can just prepare the states $\Phi_1(x)$ and $\Phi_2(x)$ but that would assume that we would know x beforehand. One possible strategy is to look for a channel $\Phi \in \mathfrak{C}(K_A \to K_B \tilde{\otimes} K_C)$ such that

$$((\mathrm{id}_B \otimes 1) \circ \Phi)(x) = \Phi_1(x) \tag{2}$$

$$((1 \otimes \mathrm{id}_C) \circ \Phi)(x) = \Phi_2(x) \tag{3}$$

for all $x \in K_A$. If such channel Φ exists then we argue that we can implement Φ_1 and Φ_2 at the same time as we can obtain them as partial traces of Φ. Of course this can be generalized to more than two channels.

Definition 38. Let K_A, K_{B_i} where $i \in \{1, \ldots, n\}$ and let $\Phi_i \in \mathfrak{C}(K_A \to K_{B_i})$. We say that the channels $\Phi_1, \ldots, \Phi_n$ are compatible if there is a channel $\Phi \in \mathfrak{C}(K_A \to \tilde{\otimes}_{i=1}^n K_{B_i})$ such that

$$\left(\left(\mathrm{id}_{B_1} \otimes 1^{\otimes(n-1)} \right) \circ \Phi \right)(x) = \Phi_1(x)$$

$$\vdots$$

$$\left(\left(1^{\otimes(i-1)} \otimes \mathrm{id}_{B_i} \otimes 1^{\otimes(n-i)} \right) \circ \Phi \right)(x) = \Phi_i(x)$$

$$\vdots$$

$$\left(\left(1^{\otimes(n-1)} \otimes \mathrm{id}_{B_n} \right) \circ \Phi \right)(x) = \Phi_n(x)$$

where $1^{\otimes 2} = 1 \otimes 1$, $1^{\otimes 3} = 1 \otimes 1 \otimes 1$ and so on. Moreover we call Φ the joint channel of channels $\Phi_1, \ldots, \Phi_n$ and we call $\Phi_1, \ldots, \Phi_n$ the marginals of Φ.

Most of the time we are going to be interested in compatibility of two channels, but many of our results could be generalized to the case of n channels. Note that the joint channel does not have to be unique.

There is one special case of the described scenario of compatibility of two channels when we have $\Phi_1 = \Phi_2$, i.e. when we consider the compatibility of the channel with itself, also called the self-compatibility of the channel.

Definition 39. Let $\Phi \in \mathfrak{C}(K_A \to K_B)$ then we say that Φ is self-compatible if Φ is compatible with itself. We say that a channel is k-self-compatible if $k \in \mathbb{N}$ copies of the channel are compatible.

Note that one can find examples of channels that are e.g. 3-self-compatible but not 4-self-compatible. Self-compatibility of channels was investigated in quantum theory [46, 56–58] where one can find examples of channels that are not self-compatible. We are going to present such examples as well in Subsec. 5.2. Now we will investigate the general notion of compatibility of channels further. At first, we will consider the special case of measurements.

Proposition 39. *Let* $m_1 \in \mathfrak{C}(K \to S_{n_1})$ *and* $m_2 \in \mathfrak{C}(K \to S_{n_2})$ *be measurements given as*

$$m_1 = \sum_{i=1}^{n_1} f_i \otimes s_i$$

$$m_2 = \sum_{j=1}^{n_1} g_i \otimes s_j$$

for some $f_i, g_j \in E(K)$, $i \in \{1, \ldots, n_1\}$, $j \in \{1, \ldots, n_2\}$. *The measurements* m_1 *and* m_2 *are compatible if and only if there are effects* $h_{ij} \in E(K)$ *such that*

$$f_i = \sum_{j=1}^{n_2} h_{ij} \tag{4}$$

$$g_j = \sum_{i=1}^{n_1} h_{ij} \tag{5}$$

for all $i \in \{1, \ldots, n_1\}$ *and* $j \in \{1, \ldots, n_2\}$.

Proof. The proof is rather straightforward. Assume that m_1 and m_2 are compatible then there is $m \in \mathfrak{C}(K \to S_{n_1} \dot\otimes S_{n_2})$ such that

$$m = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} h_{ij} \otimes s_i \otimes s_j \tag{6}$$

and

$$m_1 = (\mathrm{id}_{S_{n_1}} \otimes 1) \circ m = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} h_{ij} \otimes s_i$$

$$m_2 = (1 \otimes \mathrm{id}_{S_{n_2}}) \circ m = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} h_{ij} \otimes s_j$$

from where the result follows.

If we assume that there are $h_{ij} \in E(K)$ such that Eq. (4) and (5) are satisfied then let m be given by Eq. (6). It is straightforward to verify that m is a measurement and that it is a joint measurement of m_1 and m_2. $\qquad\square$

A very useful tool when dealing with compatibility of channels is the notion of concatenation of channels and its relation to compatibility. In quantum theory, this was already investigated in [56].

Proposition 40. *Let* $\Phi_1 \in \mathfrak{C}(K_A \to K_{B_1})$, $\Phi_2 \in \mathfrak{C}(K_A \to K_{B_2})$ *and* $\Phi_3 \in \mathfrak{C}(K_{B_2} \to K_{B_3})$ *be channels, then* $\Phi_3 \circ \Phi_2 \in \mathfrak{C}(K_A \to K_{B_3})$. *If* Φ_1 *and* Φ_2 *are compatible, then also* Φ_1 *and* $\Phi_3 \circ \Phi_2$ *are compatible.*

Proof. The fact that $\Phi_3 \circ \Phi_2 \in \mathfrak{C}(K_A \to K_{B_3})$ is straightforward, it is a channel such that for $x \in K_A$ we have $(\Phi_3 \circ \Phi_2)(x) = \Phi_3(\Phi_2(x))$. Now assume that Φ_1 and Φ_2 are compatible with a joint channel $\Phi \in \mathfrak{C}(K_A \to K_{B_1} \tilde\otimes K_{B_2})$, i.e. we have $(\mathrm{id}_{B_1} \otimes 1) \circ \Phi = \Phi_1$ and $(1 \otimes \mathrm{id}_{B_2}) \circ \Phi = \Phi_2$ and consider the channel $(\mathrm{id}_{B_1} \otimes \Phi_3) \circ \Phi \in \mathfrak{C}(K_A \to K_{B_1} \tilde\otimes K_{B_3})$. We have

$$(\mathrm{id}_{B_1} \otimes 1) \circ (\mathrm{id}_{B_1} \otimes \Phi_3) \circ \Phi = (\mathrm{id}_{B_1} \otimes \Phi_3^*(1)) \circ \Phi = (\mathrm{id}_{B_1} \otimes 1) \circ \Phi = \Phi_1$$
$$(1 \otimes \mathrm{id}_{B_2}) \circ (\mathrm{id}_{B_1} \otimes \Phi_3) \circ \Phi = \Phi_3 \circ (1 \otimes \mathrm{id}_{B_2}) \circ \Phi = \Phi_3 \circ \Phi_2$$

which shows that Φ_1 and $\Phi_3 \circ \Phi_2$ are compatible. $\qquad\square$

Corollary 10. *Let $\Phi_1 \in \mathfrak{C}(K_A \to K_{B_1})$ and $\Phi_2 \in \mathfrak{C}(K_A \to K_{B_2})$ be measure-and-prepare channels given as*

$$\Phi_1 = \sum_{i=1}^{n_1} f_i \otimes x_i$$

$$\Phi_2 = \sum_{j=1}^{n_2} g_j \otimes y_j$$

for some $f_i, g_j \in E(K_A)$, $x_i \in K_{B_1}$ and $y_j \in K_{B_2}$ for $i \in \{1, \ldots, n_1\}$ and $j \in \{1, \ldots, n_2\}$. Let $m_1 \in \mathfrak{C}(K_A \to S_{n_1})$ and $m_2 \in \mathfrak{C}(K_A \to S_{n_2})$ be the corresponding measurements given as

$$m_1 = \sum_{i=1}^{n_1} f_i \otimes s_i$$

$$m_2 = \sum_{i=1}^{n_2} g_j \otimes s_j.$$

If m_1 and m_2 are compatible then also Φ_1 and Φ_2 are compatible and their joint channel can be chosen to be measure-and-prepare.

Proof. Consider the channels $\nu_1 \in \mathfrak{C}(S_{n_1} \to K_{B_1})$ and $\nu_2 \in \mathfrak{C}(S_{n_2} \to K_{B_2})$ given as

$$\nu_1(s_i) = x_i$$
$$\nu_2(s_j) = y_j$$

then $\Phi_1 = \nu_1 \circ m_1$ and $\Phi_2 = \nu_2 \circ m_2$, i.e. ν_1 and ν_2 are the preparation parts of the measure-and-prepare channels Φ_1 and Φ_2. The compatibility of Φ_1 and Φ_2 follows from the compatibility of m_1 and m_2 and using twice the result of Prop. 40.

To show that the joint channel can be selected to be measure-and-prepare let $m \in \mathfrak{C}(K \to S_{n_1} \dot{\otimes} S_{n_2})$ be the joint measurement of m_1 and m_2 and the construct $\Phi = (\nu_1 \otimes \nu_2) \circ m$. It is straightforward to check that Φ is measure and prepare and that it is a joint channel of Φ_1 and Φ_2. $\square$

Corollary 11. *Let $\Phi \in \mathfrak{C}(K_A \to K_B)$ be a measure-and-prepare channel, then Φ is self-compatible, i.e. it is compatible with itself.*

Proof. The result follows from Coro. 10. $\square$

It is straightforward to see that one can extend the results of Coro. 10 and Coro. 11 to more than 2 channels. The following result is inspired by [55] where similar result was proved for measurements but using a different approach.

Corollary 12. *Let $\Phi_1 \in \mathfrak{C}(K_A \to K_{B_1})$ and $\Phi_2 \in \mathfrak{C}(K_A \to K_{B_2})$ be channels, then they are compatible if there is a self-compatible channel $\Phi_3 \in \mathfrak{C}(K_A \to K_{B_3})$ and channels $\nu_1 \in \mathfrak{C}(K_{B_3} \to K_{B_1})$ and $\nu_2 \in \mathfrak{C}(K_{B_3} \to K_{B_2})$ such that*

$$\Phi_1 = \nu_1 \circ \Phi_3, \tag{7}$$
$$\Phi_2 = \nu_2 \circ \Phi_3. \tag{8}$$

Proof. It follows from Eq. (7) and (8), the self-compatibility of Φ_3 and Prop. 40 that the channels Φ_1 and Φ_2 are compatible. $\square$

In case of measurements one can prove a stronger result.

Corollary 13. *Let $m_1 \in \mathfrak{C}(K \to S_{n_1})$ and $m_2 \in \mathfrak{C}(K \to S_{n_2})$ be measurements, then they are compatible if and only if there is a measurement $m_3 \in \mathfrak{C}(K \to S_{n_3})$ and channels $\nu_1 \in \mathfrak{C}(S_{n_3} \to S_{n_1})$ and $\nu_2 \in \mathfrak{C}(S_{n_3} \to S_{n_2})$ such that*

$$m_1 = \nu_1 \circ m_3 \tag{9}$$
$$m_2 = \nu_2 \circ m_3. \tag{10}$$

Proof. Another version of the proof may be found in [55]. Assume that the measurements m_1 and m_2 are compatible and let $m \in \mathfrak{C}(K \to S_{n_1} \dot{\otimes} S_{n_2})$ be their joint measurement. Take $m_3 = m$, $\nu_1 = \mathrm{id} \otimes 1$ and $\nu_2 = 1 \otimes \mathrm{id}$, then the Eq. (9) and (10) become (2) and (3).

The converse statement follows from Coro. 12 and Coro. 11. □

5.2 Existence of incompatible channels

In this subsection we will argue that incompatibility exists in all non-classical theories and we will show how this implies a result about existence of entanglement as presented in Subsec. 3.3. First, we will show that if a state space is a simplex then all channels are compatible. Afterwards we will investigate the consequences of existence of incompatible measurements on a non-classical state space K.

Proposition 41. *Let S_n be a simplex, then k copies of the identity channel are compatible for any $k \in \mathbb{N}$, i.e. the* id *channel is self-k-compatible.*

Proof. Let $\mathrm{ext}(S_n) = \{s_1, \ldots, s_n\}$ and let $b_i \in E(S_n)$, for $i \in \{1, \ldots, n\}$ be the dual basis, i.e. we have $\langle s_i, b_j \rangle = \delta_{ij}$ for all $i, j \in \{1, \ldots, n\}$. Note that the fact that all of the functions $b_1, \ldots, b_n$ are positive is an exclusive property of the classical state spaces.

We have that

$$\mathrm{id}_{S_n} = \sum_{i=1}^{n} b_i \otimes s_i \in A(S_n)^+ \dot{\otimes} A(S_n)^{*+}$$

and it follows that

$$\Phi^{(k)} = \sum_{i=1}^{n} b_i \otimes (s_i^{\otimes k}) \in \mathfrak{C}(S_n \to S_n^{\otimes k})$$

is a well defined channel. One can see that $\Phi^{(k)}$ is the joint channel of k copies of the identity channel id_{S_n}. □

Corollary 14. *Let S_n be a simplex and $k \in \mathbb{N}$, then any set of k channels $\Phi_i \in \mathfrak{C}(S_n \to K_{B_i})$ for any state spaces K_{B_i} and $i \in \{1, \ldots, k\}$ is compatible.*

Proof. The result follows from the same idea as Prop. 40 and from the fact that we have $\Phi_i = \Phi_i \circ \mathrm{id}_{S_n}$. Let $\Phi^{(k)}$ be the joint of k copies of $\mathrm{id}_{S_n} \in \mathfrak{C}(S_n \to S_n)$ as constructed in the proof of Prop. 41 and let

$$\Phi = \left(\otimes_{i=1}^{k} \Phi_i \right) \circ \Phi^{(k)}$$

then it is straightforward to check that Φ is the joint channel of the channels $\Phi_1, \ldots, \Phi_k$. □

Proposition 42. *Let K be a state space then there exists a pair of incompatible two-outcome measurements $m_1, m_2 \in \mathfrak{C}(K \to S_2)$ whenever K is not a simplex.*

Proof. Constructive proof may be found in [59]. □

We are going to explore some of the consequences of existence of incompatible channels. The following result is known as no-broadcasting and it was originally proved in [35, 60].

Corollary 15 (No-broadcasting)**.** *Let K be a non-classical state space, i.e. K is not a simplex, then the identity channel* id *is not self-compatible.*

Proof. Assume that the identity channel id is self-compatible. Since for any channels $\Phi_1 \in \mathfrak{C}(K \to K_{B_1})$ and $\Phi_2 \in \mathfrak{C}(K \to K_{B_2})$ we have $\Phi_1 = \Phi_1 \circ \mathrm{id}$ and $\Phi_2 = \Phi_2 \circ \mathrm{id}$ it would follow from Coro. 12 that Φ_1 and Φ_2 would be compatible. This is in contradiction with the result of Prop.42. □

The following result will be about the existence of entanglement and it was also proved in [36] using a very different methods. In [36] the authors prove the statement of Prop. 20 which they use to conclude the following result. We will use the result of Coro. 15 to prove the following result.

Corollary 16. *Let K_A be a non-classical state space, then there exists a compact convex set $K_B \subset V$, where V is a finite dimensional real vector space, such that $K_A \dot{\otimes} K_B \neq K_A \hat{\otimes} K_B$.*

Proof. Let K_A be a non-classical state space and consider the identity channel id. We already know that id $\in A(K_A)^+ \hat{\otimes} A(K_A)^{*+}$; we must have id $\notin A(K_A)^+ \dot{\otimes} A(K_A)^{*+}$ as if id would be measure-and-prepare, then according to Coro. 11 it would be self-compatible which would be in contradiction with Coro. 15. It follows that we must have $A(K_A)^+ \dot{\otimes} A(K_A)^{*+} \neq A(K_A)^+ \hat{\otimes} A(K_A)^{*+}$.

To conclude the proof let K_B be a base of $A(K_A)^+$ then according to Prop. 6 we know that K_B is a compact convex set and we have $K_A \dot{\otimes} K_B \neq K_A \hat{\otimes} K_B$. $\qquad\square$

5.3 Degree of compatibility

Assume that we have two systems described by state spaces K_A and K_B and we want to to use a channel $\Phi \in \mathfrak{C}(K_A \to K_B)$ to communicate a message. That is we want to prepare a state $x \in K_A$, map it to $\Phi(x) \in K_B$ and then try to get as much information about the original state $x \in K_A$ from $\Phi(x) \in K_B$ as possible. Clearly our probability of being successful is going to depend on the channel Φ. For example if $K_A = K_B$ and $\Phi = $ id then the task is trivial as we are given the state itself. But now consider a different scenario, let $y \in K_B$ and let $\Phi(x) = y$ for all $x \in K_A$. Clearly there is nothing we can learn about x from $\Phi(x) = y$ and in this case we can see the channel $\Phi = 1 \otimes y$ as classical noise. We will introduce a name for the channels that represent noise.

Definition 40. Let $\Phi \in \mathfrak{C}(K_A \to K_B)$ be such that for all $x, y \in K_A$ we have $\Phi(x) = \Phi(y)$, then we call Φ a constant channel.

If a constant channel $m \in \mathfrak{C}(K \to S_n)$ is a measurement we may also refer to it as coin-toss measurement. The following is a trivial result

Lemma 1. *Let $\Phi_1 \in \mathfrak{C}(K_A \to K_{B_1})$ and let $\Phi_2 \in \mathfrak{C}(K_A \to K_{B_2})$. If Φ_2 is a constant channel then Φ_1 and Φ_2 are compatible.*

Proof. Let $y \in K_{B_2}$ be the state such that for all $x \in K_A$ we have $\Phi_2(x) = y$. Let $\Phi \in \mathfrak{C}(K_A \to K_{B_1} \tilde{\otimes} K_{B_2})$ be a channel defined for $x \in K_A$ as

$$\Phi(x) = \Phi_1(x) \otimes y.$$

It follows that Φ is a joint channel of Φ_1 and Φ_2. $\qquad\square$

Inspired by the simple result, consider the following question: let $\Phi_1, \Phi_{C_1} \in \mathfrak{C}(K_A \to K_{B_1})$ and $\Phi_2, \Phi_{C_2} \in \mathfrak{C}(K_A \to K_{B_2})$ such that Φ_{C_1} and Φ_{C_2} are constant channels, then for what value of $\lambda \in [0, 1]$ are the channels $\lambda \Phi_1 + (1 - \lambda)\Phi_{C_1}$ and $\lambda \Phi_2 + (1 - \lambda)\Phi_{C_2}$ compatible? Different aspects of the question were already investigated in [15, 54, 55, 61–64], we are only going to present the most basic definition and results.

Definition 41. Let $\Phi_1, \Phi_{C_1} \in \mathfrak{C}(K_A \to K_{B_1})$ and $\Phi_2, \Phi_{C_2} \in \mathfrak{C}(K_A \to K_{B_2})$ such that Φ_{C_1} and Φ_{C_2} are constant channels. The degree of compatibility of channels Φ_1 and Φ_2 is denoted $\mathrm{DegCom}(\Phi_1, \Phi_2)$ and it is defined as

$$\mathrm{DegCom}(\Phi_1, \Phi_2) = \sup_{\substack{\lambda \in [0,1], \\ \Phi_{C_1}, \Phi_{C_2}}} \{\lambda : \lambda \Phi_1 + (1 - \lambda)\Phi_{C_1}, \lambda \Phi_2 + (1 - \lambda)\Phi_{C_2} \text{ are comaptible}\}$$

Proposition 43. *Let $\Phi_1 \in \mathfrak{C}(K_A \to K_{B_1})$ and $\Phi_2 \in \mathfrak{C}(K_A \to K_{B_2})$ then*

$$\mathrm{DegCom}(\Phi_1, \Phi_2) \geq \frac{1}{2}.$$

Proof. The idea of the proof is well known and it can be also found in [54]. Let $\Phi_{C_1} \in \mathfrak{C}(K_A \to K_{B_1})$ and $\Phi_{C_2} \in \mathfrak{C}(K_A \to K_{B_2})$ be any constant channels such that for $x \in K_A$ we have $\Phi_{C_1}(x) = y_1$ and $\Phi_{C_2}(x) = y_2$. We will define a channel $\Phi \in \mathfrak{C}(K_A \to K_{B_1} \tilde{\otimes} K_{B_2})$ for $x \in K_A$ as

$$\Phi = \frac{1}{2}(\Phi_1(x) \otimes y_2 + y_1 \otimes \Phi_2(x)).$$

We have

$$(\mathrm{id}_{B_1} \otimes 1) \circ \Phi = \frac{1}{2}(\Phi_1 + \Phi_{C_1})$$

$$(1 \otimes \mathrm{id}_{B_2}) \circ \Phi = \frac{1}{2}(\Phi_2 + \Phi_{C_2})$$

hence $\frac{1}{2}(\Phi_1 + \Phi_{C_1})$ and $\frac{1}{2}(\Phi_2 + \Phi_{C_2})$ are compatible. $\mathrm{DegCom}(\Phi_1, \Phi_2) \geq \frac{1}{2}$ follows. $\qquad \square$

In some applications it may be of interest to look for channels $\Phi_1 \in \mathfrak{C}(K_A \to K_{B_1})$ and $\Phi_2 \in \mathfrak{C}(K_A \to K_{B_2})$ such that $\mathrm{DegCom}(\Phi_1, \Phi_2) = \dfrac{1}{2}$, for example one can show that the violation of Bell inequalities is bounded by the degree of compatibility [64]. Hence we introduce the following notion.

Definition 42. The channels $\Phi_1 \in \mathfrak{C}(K_A \to K_{B_1})$ and $\Phi_2 \in \mathfrak{C}(K_A \to K_{B_2})$ are called maximally incompatible if $\mathrm{DegCom}(\Phi_1, \Phi_2) = \dfrac{1}{2}$.

6 Correlations

We have already seen connections between incompatibility and entanglement in Coro. 16 where we have used existence of incompatible measurements to prove that entanglement exists. Now we are going to look at results that use both incompatibility and entanglement at the same time. We will go only as far as to formulate the basic definitions of steering and Bell non-locality for measurements only, as this is the standard formulation stemming from the EPR argument [65] and further works on the topic [66, 67] and formalized in [68]. The extensions of all of the definitions and results from measurements to channels as well as other results may be found in [69].

6.1 Steering

Steering as a phenomenon was first described in [66] and since then it has been heavily investigated in quantum theory [70–75] and in GPTs [30, 47, 76]. Also results on one-side device-independent protocols rely on steering [77].

Let K_A, K_B be a state spaces such that $K_A \tilde{\otimes} K_B$ is defined and let $m_i \in \mathfrak{C}(K_A \to S_{n_i})$ for $i \in \{1, \ldots, N\}$ be measurements given as

$$m_i = \sum_{j=1}^{n_i} f_j^{(i)} \otimes s_j$$

and let $x \in K_A \tilde{\otimes} K_B$. We are going to denote

$$y(j|i) = (f_j^{(i)} \otimes \mathrm{id})(x). \tag{11}$$

Using the arguments of local realism [65] one may argue that there must exist the so-called local hidden state model for $y(j|i)$, i.e. that we should have

$$y(j|i) = \sum_{\lambda=1}^{\Omega} p_\lambda p(j|i, \lambda) z_\lambda \tag{12}$$

for some $\Omega \in \mathbb{N}$, for all $i \in \{1, \ldots, N\}$ and for all $j \in \{1, \ldots, n_i\}$, where $p_\lambda \in \mathbb{R}^+$, $\sum_{\lambda=1}^{\Omega} p_\lambda = 1$ and $p(j|i, \lambda) \in \mathbb{R}^+$ such that $\sum_{j=1}^{n_i} p(j|i, \lambda) = 1$ and $z_\lambda \in K_B$. Of course Eq. (12) does not have to be satisfied for every choice of $x \in K_A \tilde{\otimes} K_B$ and measurements $m_1, \ldots, m_k$.

Definition 43. We say that a state $x \in K_A \tilde{\otimes} K_B$ is not steerable by measurements $m_1, \ldots, m_N$ if there is a local hidden state model for the ensemble $y(j|i)$ as given by Eq. (11), i.e. if there is a local hidden state model such that Eq. (12) is satisfied.

The following are standard results concerning steering.

Proposition 44. *A state $x \in K_A \tilde{\otimes} K_B$ is not steerable by measurements $m_1, \ldots, m_N$ if $x \in K_A \dot{\otimes} K_B$, i.e. if x is separable.*

Proof. The proof is rather simple and may be found in [69]. Let $x \in K_A \dot{\otimes} K_B$ and assume that $x = w \otimes z$ for some $w \in K_A$ and $z \in K_B$, it is sufficient to find local hidden state models for such states as they contain the extreme points of $K_A \dot{\otimes} K_B$. Let $m_i \in \mathfrak{C}(K_A \to S_{n_i})$ for $i \in \{1, \ldots, N\}$ be measurements given as

$$m_i = \sum_{j=1}^{n_i} f_j^{(i)} \otimes s_j$$

then we have

$$(f_j^{(i)} \otimes \mathrm{id})(x) = \langle w, f_j^{(i)} \rangle z$$

which is a local hidden variable model with $\langle w, f_j^{(i)} \rangle = p(j|i, \lambda)$, $z = z_\lambda$ and where the hidden parameter λ has just one value. We have

$$\sum_{j=1}^{n_i} \langle w, f_j^{(i)} \rangle = \langle w, 1 \rangle = 1$$

which concludes the proof. $\square$

Proposition 45. *A state $x \in K_A \tilde{\otimes} K_B$ is not steerable by measurements $m_1, \ldots, m_k$ if the measurements $m_1, \ldots, m_N$ are compatible.*

Proof. A proof may be found in [69] but it is rather straightforward. Let $m \in \mathfrak{C}(K \to S_n)$ such that there are channels $\nu_i \in \mathfrak{C}(S_n \to S_{n_i})$ such that $m_i = \nu_i \circ m$ and let m be given as

$$m = \sum_{k=1}^{n} g_k \otimes s_k.$$

We then have that

$$f_j^{(i)} = \sum_{k=1}^{n} (\nu_i)_{kj} g_k$$

where $(\nu_i)_{kj}$ are given as in Prop. 30 as

$$\nu_i \circ m = \sum_{k=1}^{n} g_k \otimes \nu_i(s_k) = \sum_{k=1}^{n} \sum_{j=1}^{n_i} (\nu_i)_{kj} g_k \otimes s_j.$$

We have

$$y(j|i) = (f_j^{(i)} \otimes \mathrm{id})(x) = \sum_{k=1}^{n} (\nu_i)_{kj} (g_k \otimes \mathrm{id})(x).$$

According to Prop. 17 we have $(g_k \otimes \mathrm{id})(x) = \mu_k z_k$ for some $z_k \in K_B$ and $\mu_k \in [0,1]$ for all $k \in \{1, \ldots, n\}$. It follows that

$$y(j|i) = \sum_{k=1}^{n} \mu_k (\nu_i)_{kj} z_k$$

is a local hidden state model with $k = \lambda$, $\mu_k = p_\lambda$ and $(\nu_i)_{kj} = p(j|i, \lambda)$. $\sum_{k=1}^{n} \mu_k = 1$ follows from

$$\sum_{k=1}^{n} \mu_k = \sum_{k=1}^{n} \langle (g_k \otimes \mathrm{id})(x), 1 \rangle = \langle x, 1 \otimes 1 \rangle = 1$$

and $\sum_{j=1}^{n_i} (\nu_i)_{kj} = 1$ follows from Prop. 30. $\qquad\square$

6.2 Bell non-locality

Bell non-locality as a phenomenon was established in [67] as a solution to the EPR paradox. The most well-known form of Bell non-locality today is given by the CHSH inequality [78]. Bell non-locality was also heavily investigated in quantum theory [51, 64, 79–86] and in GPTs as well [47, 87, 88], maximal violations of the CSHS inequality were investigated [89–92] and also in the context of device-independent protocols [93–96].

Bell non-locality is very similar to steering, with the twist that we apply measurements to both systems. So let again K_A and K_B be state spaces, let $m_i^A \in \mathfrak{C}(K_A \to S_{n_i^A})$ and $m_j^B \in \mathfrak{C}(K_B \to S_{n_j^B})$ for $i \in \{1, \ldots, N_A\}$ and $j \in \{1, \ldots, N_B\}$, where the measurements are given as

$$m_i^A = \sum_{k=1}^{n_i^A} f_k^{(i)} \otimes s_k,$$

$$m_j^B = \sum_{l=1}^{n_j^B} g_l^{(j)} \otimes s_l.$$

Let $x \in K_A \tilde{\otimes} K_B$, we are going to denote

$$p(k, l|i, j) = \langle x, f_k^{(i)} \otimes g_l^{(j)} \rangle \tag{13}$$

which corresponds to Eq. (11) in the case of steering. One can again use the arguments of local realism [65, 67] that we should have the so-called local hidden variable model given as

$$p(k, l|i, j) = \sum_{\lambda=1}^{\Omega} p_\lambda p(k|i, \lambda) p(l|j, \lambda) \tag{14}$$

for some $\Omega \in \mathbb{N}$, for all $i \in \{1, \ldots, N_A\}$, for all $j \in \{1, \ldots, N_B\}$, for all $k \in \{1, \ldots, n_i^A\}$, for all $l \in \{1, \ldots, n_j^B\}$, where $p_\lambda \in \mathbb{R}^+$, $\sum_{\lambda=1}^{\Omega} p_\lambda = 1$ and $p(k|i, \lambda), p(l|j, \lambda) \in \mathbb{R}^+$ such that $\sum_{k=1}^{n_i^A} p(k|i, \lambda) = 1$ and $\sum_{l=1}^{n_j^B} p(l|j, \lambda) = 1$. Again Eq. (14) does not have to be satisfied for any choice of $x \in K_A \tilde{\otimes} K_B$ and measurements $m_1^A, \ldots, m_{N_A}^A$ and $m_1^B, \ldots, m_{N_B}^B$.

Definition 44. The state $x \in K_A \tilde{\otimes} K_B$ is Bell local with respect to measurements $m_1^A, \ldots, m_{N_A}^A$, and $m_1^B, \ldots, m_{N_B}^B$ if there is a local hidden variable model for the ensemble $p(k, l|i, j)$ as given by Eq. (13). i.e. if there is a local hidden variable model such that Eq. (14) is satisfied.

We will again present a standard and basic result on the connection between Bell non-locality and steering.

Proposition 46. Let $x \in K_A \tilde{\otimes} K_B$ and let $m_i^A \in \mathfrak{C}(K_A \to S_{n_i^A})$ and $m_j^B \in \mathfrak{C}(K_B \to S_{n_j^B})$ for $i \in \{1, \ldots, N_A\}$ and $j \in \{1, \ldots, N_B\}$, where the measurements are given as

$$m_i^A = \sum_{k=1}^{n_i^A} f_k^{(i)} \otimes s_k,$$

$$m_j^B = \sum_{l=1}^{n_j^B} g_l^{(j)} \otimes s_l.$$

Let $y(k|i)$ be an ensemble given by Eq. (11), i.e.

$$y(k|i) = (f_k^{(i)} \otimes \mathrm{id})(x).$$

If x is not steerable by the measurements $m_1^A, \ldots, m_{N_A}^A$, then x is Bell local with respect to the measurements $m_1^A, \ldots, m_{N_A}^A$ and $m_1^B, \ldots, m_{N_B}^B$.

Proof. The proof is quite straightforward. Assume that we have

$$(f_k^{(i)} \otimes \mathrm{id})(x) = \sum_{\lambda=1}^{\Omega} p_\lambda p(k|i, \lambda) z_\lambda$$

as in Eq. (12), then

$$p(k, l|i, j) = \left\langle \sum_{\lambda=1}^{\Omega} p_\lambda p(k|i, \lambda) z_\lambda, g_l^{(j)} \right\rangle = \sum_{\lambda=1}^{\Omega} p_\lambda p(k|i, \lambda) \langle z_\lambda, g_l^{(j)} \rangle.$$

Denote $p(l|j, \lambda) = \langle z_\lambda, g_l^{(j)} \rangle$, then we have $\sum_{l=1}^{n_j^B} p(l|j, \lambda) = 1$ and we have obtained the local hidden variable model for $p(k, l|i, j)$. $\square$

We are going to use the result of Prop. 46 to get results equivalent to Prop. 44 and Prop. 45.

Corollary 17. Let $x \in K_A \dot{\otimes} K_B$, then there is a local hidden variable model for the ensemble $p(k, l|i, j)$ given by the state x and measurements $m_1^A, \ldots, m_{N_A}^A$ and $m_1^B, \ldots, m_{N_B}^B$, i.e. x is Bell local with respect to the measurements $m_1^A, \ldots, m_{N_A}^A$ and $m_1^B, \ldots, m_{N_B}^B$

Proof. The result follows from Prop. 44 and Prop. 46 as without entanglement we can not have steering and without steering we can't have Bell non-locality. $\square$

Corollary 18. *Assume that the measurements $m_1^A, \ldots, m_{N_A}^A$ are compatible then there is a local hidden variable model for the ensemble $p(k, l | i, j)$ given by the state $x \in K_A \tilde{\otimes} K_B$ and measurements $m_1^A, \ldots, m_{N_A}^A$ and $m_1^B, \ldots, m_{N_B}^B$, i.e. x is Bell local with respect to the measurements $m_1^A, \ldots, m_{N_A}^A$ and $m_1^B, \ldots, m_{N_B}^B$*

Proof. The result follows from Prop. 45 and Prop. 46 as if the measurements are compatible then they do not steer the state x and without steering we can not have Bell non-locality. $\qquad\square$

Conclusions

The aim of the review was to present a solid introduction to the framework of general probabilistic theories in a way that we believe to be self-contained and understandable to a theoretical physicist or mathematician. We hope that we have provided a complete review of the framework and we hope that the reader can now almost skip the first few pages of a research paper concerning general probabilistic theories.

This review is followed by papers written by the author and his co-authors where some of the recent results on compatibility, steering and Bell non-locality in general probabilistic theories are presented. The result of [59] on existence of incompatible two-outcome measurements were already mentioned in the review. In [15] the state spaces with maximally incompatible two-outcome measurements were characterized using the methods of linear programming. In [69] the problem of compatibility of channels was investigated and generalizations of steering and Bell non-locality for channels were introduced. The motivation behind these generalizations is different from the standard motivation using EPR paradox and local realism. It is shown in [69] that steering and Bell non-locality can be introduced as entanglement assisted incompatibility test. In [41] a comparison of different operational principles that use the notion of compatibility was presented. It was shown, that one can characterize state spaces where are measurements compatible with the identity channel and where are measurements compatible with every other measurement and one can also show that these two cases do not coincide.

The last paper [40] works with the more general framework of effect algebras in which a notion of spectrality is investigated. Spectrality in the form of spectral decompositions of self-adjoint operators is a well-known property of quantum theory. It is interesting to investigate spectrality in the framework of effect algebras (and also in the framework of GPTs). Spectrality frequently appears in derivations of quantum theory from operational axioms [24, 97, 98] and it was also studied in the framework of GPTs [99, 100].

The last two mentioned papers [40, 41] are unpublished as in the time of writing this review. [40] is in the second of peer-review and [41] already passed the first round of peer review and corrections to the paper are being made.

References

[1] R. T. Rockafellar, *Convex Analysis.* Princeton landmarks in mathematics and physics, Princeton University Press, 1997.

[2] L. Asimow and A. J. Ellis, *Convexity theory and its applications in functional analysis.* L.M.S. monographs, Academic Press, 1980.

[3] R. A. Ryan, *Introduction to Tensor Products of Banach Spaces.* Springer Monographs in Mathematics, Springer London, 2013.

[4] S. Morris, S. University of New England. Department of Mathematics, and C. Science, *Topology Without Tears.* University of New England, 1989.

[5] J. R. Munkres, *Topology.* Featured Titles for Topology Series, Prentice Hall, Incorporated, 2000.

[6] T. Heinosaari and M. Ziman, *The Mathematical Language of Quantum Theory. From Uncertainty to Entanglement.* Cambridge University Press, 2012.

[7] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory.* Publications of the Scuola Normale Superiore, Scuola Normale Superiore, 2011.

[8] J. Watrous, *The Theory of Quantum Information.* Cambridge University Press, 2018.

[9] G. Aubrun and S. Szarek, *Alice and Bob Meet Banach.* Mathematical Surveys and Monographs, American Mathematical Society, 2017.

[10] A. Barvinok, *A Course in Convexity.* Graduate studies in mathematics, American Mathematical Society, 2002.

[11] S. Boyd and L. Vandenberghe, *Convex Optimization.* Berichte über verteilte messysteme, Cambridge University Press, 2004.

[12] J. Sikora and J. Selby, "Simple proof of the impossibility of bit commitment in generalized probabilistic theories using cone programming," *Phys. Rev. A*, vol. 97, no. 4, p. 042302, 2018.

[13] J. H. Selby and J. Sikora, "How to make unforgeable money in generalised probabilistic theories," *Quantum*, vol. 2, p. 103, 2018.

[14] J. Sikora and J. H. Selby, "On the impossibility of coin-flipping in generalized probabilistic theories via discretizations of semi-infinite programs." 2019.

[15] A. Jenčová and M. Plávala, "Conditions on the existence of maximally incompatible two-outcome measurements in general probabilistic theory," *Phys. Rev. A*, vol. 96, no. 2, p. 022113, 2017.

[16] D. J. Foulis and M. K. Bennett, "Effect algebras and unsharp quantum logics," *Found. Phys.*, vol. 24, no. 10, pp. 1331–1352, 1994.

[17] G. Jenča, "Blocks of homogeneous effect algebras," *Bull. Aust. Math. Soc.*, vol. 64, no. 01, pp. 81–98, 2015.

[18] S. Gudder, S. Pulmannová, S. Bugajski, and E. Beltrametti, "Convex and linear effect algebras," *Reports Math. Phys.*, vol. 44, no. 3, pp. 359–379, 1999.

[19] S. Gudder, "Convex structures and effect algebras," *International Journal of Theoretical Physics*, vol. 38, no. 12, pp. 3179–3187, 1999.

[20] S. Gudder, "Sharply dominating effect algebras," *Tatra Mt. Math. Publ.*, vol. 15, pp. 23–30, 1998.

[21] S. Gudder, "Convex and Sequential Effect Algebras." 2018.

[22] A. Dvurecenskij and S. Pulmannová, *New Trends in Quantum Structures*. Mathematics and Its Applications, Springer Netherlands, 2013.

[23] S. Gudder and S. Pulmannová, "Representation theorem for convex effect algebras," *Comment. Math. Univ. Carolinae*, vol. 39, no. 4, pp. 645–660, 1998.

[24] G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Informational derivation of quantum theory," *Phys. Rev. A*, vol. 84, no. 1, p. 012311, 2011.

[25] L. Guerini, J. Bavaresco, M. T. Cunha, and A. Acín, "Operational framework for quantum measurement simulability," *J. Math. Phys.*, vol. 58, no. 9, p. 092102, 2017.

[26] M. Oszmaniec, L. Guerini, P. Wittek, and A. Acín, "Simulating Positive-Operator-Valued Measures with Projective Measurements," *Phys. Rev. Lett.*, vol. 119, no. 19, p. 190501, 2017.

[27] M. Oszmaniec, F. B. Maciejewski, and Z. Puchała, "All quantum measurements can be simulated using projective measurements and postselection." 2018.

[28] S. N. Filippov, T. Heinosaari, and L. Leppäjärvi, "Simulability of observables in general probabilistic theories," *Phys. Rev. A*, vol. 97, no. 6, p. 062102, 2018.

[29] P. Janotta and R. Lal, "Generalized probabilistic theories without the no-restriction hypothesis," *Phys. Rev. A*, vol. 87, no. 5, p. 052131, 2013.

[30] H. Barnum, C. P. Gaebler, and A. Wilce, "Ensemble Steering, Weak Self-Duality, and the Structure of Probabilistic Theories," *Found. Phys.*, vol. 43, no. 12, pp. 1411–1427, 2013.

[31] H. Barnum, J. Barrett, L. O. Clark, M. Leifer, R. Spekkens, N. Stepanik, A. Wilce, and R. Wilke, "Entropy and information causality in general probabilistic theories," *New J. Phys.*, vol. 14, no. 12, p. 129401, 2012.

[32] H. Barnum and A. Wilce, "Information processing in convex operational theories," *Electronic Notes in Theoretical Computer Science*, vol. 270, no. 1, pp. 3–15, 2011.

[33] J. Barrett, "Information processing in generalized probabilistic theories," *Phys. Rev. A*, vol. 75, no. 3, p. 032304, 2007.

[34] G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Probabilistic theories with purification," *Phys. Rev. A*, vol. 81, no. 6, p. 062348, 2010.

[35] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, "Cloning and Broadcasting in Generic Probabilistic Theories." 2006.

[36] I. Namioka and R. Phelps, "Tensor products of compact convex sets," *Pacific J. Math.*, vol. 31, no. 2, pp. 469–480, 1969.

[37] G. Aubrun, L. Lami, and C. Palazuelos. Private Communication, 2018.

[38] L. Lami, *Non-classical correlations in quantum mechanics and beyond*. PhD thesis, Universitat Autònoma de Barcelona, 2017.

[39] G. C. Wick, A. S. Wightman, and E. P. Wigner, "The intrinsic parity of elementary particles," *Physical Review*, vol. 88, no. 1, pp. 101–105, 1952.

[40] A. Jenčová and M. Plávala, "On the properties of spectral effect algebras." 2018.

[41] T. Heinosaari, L. Leppäjärvi, and M. Plávala, "No-free-information principle in general probabilistic theories." 2018.

[42] M. Ziman, "Process positive-operator-valued measure: A mathematical framework for the description of process tomography experiments," *Phys. Rev. A*, vol. 77, p. 062112, 2008.

[43] G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Memory Effects in Quantum Channel Discrimination," *Phys. Rev. Lett.*, vol. 101, no. 18, p. 180501, 2008.

[44] A. S. Holevo, "Coding theorems for quantum communication channels," *IEEE Int. Symp. Inf. Theory - Proc.*, p. 84, 1998.

[45] M. Horodecki, P. W. Shor, and M. B. Ruskai, "Entanglement Breaking Channels," *Rev. Math. Phys.*, vol. 15, pp. 629–641, aug 2003.

[46] Y. Kuramochi, "Entanglement-breaking channels with general outcome operator algebras," *J. Math. Phys.*, vol. 59, no. 10, p. 102206, 2018.

[47] A. Jenčová, "Incompatible measurements in a class of general probabilistic theories," *Phys. Rev. A*, vol. 98, no. 1, p. 012133, 2018.

[48] T. Heinosaari, T. Miyadera, and M. Ziman, "An Invitation to Quantum Incompatibility," *J. Phys. A Math. Theor.*, vol. 49, no. 12, p. 123001, 2015.

[49] M. Sedlák, D. Reitzner, G. Chiribella, and M. Ziman, "Incompatible measurements on quantum causal networks," *Phys. Rev. A*, vol. 93, no. 5, p. 052323, 2016.

[50] L.-Y. Hsu, "Exploring the joint measurability using an information-theoretic approach," *Quantum Inf. Process.*, vol. 15, p. 5167–5177, 2016.

[51] M. M. Wolf, D. Perez-Garcia, and C. Fernandez, "Measurements incompatible in quantum theory cannot be measured jointly in any other no-signaling theory," *Phys. Rev. Lett.*, vol. 103, no. 23, p. 230402, 2009.

[52] T. Heinosaari, "Simultaneous measurement of two quantum observables: Compatibility, broadcasting, and in-between," *Phys. Rev. A*, vol. 93, no. 4, p. 042118, 2016.

[53] C. Carmeli, T. Heinosaari, and A. Toigo, "Quantum Incompatibility Witnesses," *Phys. Rev. Lett.*, vol. 122, no. 13, p. 130402, 2019.

[54] P. Busch, T. Heinosaari, J. Schultz, and N. Stevens, "Comparing the degrees of incompatibility inherent in probabilistic physical theories," *EPL (Europhysics Lett.)*, vol. 103, no. 1, p. 10002, 2013.

[55] S. N. Filippov, T. Heinosaari, and L. Leppäjärvi, "Necessary condition for incompatibility of observables in general probabilistic theories," *Phys. Rev. A*, vol. 95, no. 3, p. 032127, 2017.

[56] T. Heinosaari and T. Miyadera, "Incompatibility of quantum channels," *J. Phys. A.*, vol. 50, p. 135302, 2017.

[57] R. F. Werner, "Optimal cloning of pure states," *Phys. Rev. A*, vol. 58, no. 3, pp. 1827–1832, 1998.

[58] M. Keyl and R. F. Werner, "Optimal cloning of pure states, testing single clones," *J. Math. Phys.*, vol. 40, no. 7, pp. 3283–3299, 1999.

[59] M. Plávala, "All measurements in a probabilistic theory are compatible if and only if the state space is a simplex," *Phys. Rev. A*, vol. 94, no. 4, p. 042108, 2016.

[60] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, "Generalized No-broadcasting theorem," *Phys. Rev. Lett.*, vol. 99, no. 24, p. 240501, 2007.

[61] T. Heinosaari, J. Schultz, A. Toigo, and M. Ziman, "Maximally incompatible quantum observables," *Phys. Lett. A*, vol. 378, no. 24-25, pp. 1695–1699, 2014.

[62] T. Heinosaari, J. Kiukas, and D. Reitzner, "Noise robustness of the incompatibility of quantum measurements," *Phys. Rev. A*, vol. 92, no. 2, p. 022115, 2015.

[63] E. Haapasalo, "Robustness of incompatibility for quantum devices," *J. Phys. A*, vol. 48, no. 25, p. 255303, 2015.

[64] M. Banik, M. R. Gazi, S. Ghosh, and G. Kar, "Degree of complementarity determines the nonlocality in quantum mechanics," *Phys. Rev. A*, vol. 87, no. 5, p. 052125, 2013.

[65] A. Einstein, B. Podolsky, and N. Rosen, "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?," *Physical Review*, vol. 47, pp. 777–780, 1935.

[66] E. Schrödinger, "Probability relations between separated systems," *Math. Proc. Cambridge Philos. Soc.*, vol. 32, no. 03, p. 446, 1936.

[67] J. S. Bell, "On the Einstein Podolski Rosen Paradox," *Physics*, vol. 1, no. 3, pp. 195–200, 1964.

[68] H. M. Wiseman, S. J. Jones, and A. C. Doherty, "Steering, entanglement, nonlocality, and the Einstein-Podolsky-Rosen paradox," *Phys. Rev. Lett.*, vol. 98, no. 14, p. 140402, 2007.

[69] M. Plávala, "Conditions for the compatibility of channels in general probabilistic theory and their connection to steering and Bell nonlocality," *Phys. Rev. A*, vol. 96, no. 5, p. 052127, 2017.

[70] M. Banik, S. Das, and A. S. Majumdar, "Measurement incompatibility and channel steering," *Phys. Rev. A*, vol. 91, no. 6, p. 062124, 2015.

[71] J. Bavaresco, M. T. Quintino, L. Guerini, T. O. Maciel, D. Cavalcanti, and M. T. Cunha, "Most incompatible measurements for robust steering tests," *Phys. Rev. A*, vol. 96, no. 2, p. 022110, 2017.

[72] D. Cavalcanti and P. Skrzypczyk, "Quantum steering: a review with focus on semidefinite programming," *Reports Prog. Phys.*, vol. 80, no. 2, p. 024001, 2017.

[73] M. T. Quintino, T. Vértesi, and N. Brunner, "Joint Measurability, Einstein-Podolsky-Rosen Steering, and Bell Nonlocality," *Phys. Rev. Lett.*, vol. 113, no. 16, p. 160402, 2014.

[74] R. Uola, C. Budroni, O. Gühne, and J. P. Pellonpää, "One-to-One Mapping between Steering and Joint Measurability Problems," *Phys. Rev. Lett.*, vol. 115, no. 23, p. 230402, 2015.

[75] R. Uola, T. Moroder, and O. Gühne, "Joint Measurability of Generalized Measurements Implies Classicality," *Phys. Rev. Lett.*, vol. 113, no. 16, p. 160403, 2014.

[76] M. Banik, "Measurement incompatibility and Schrödinger-Einstein-Podolsky-Rosen steering in a class of probabilistic theories," *J. Math. Phys.*, vol. 56, no. 5, p. 052101, 2015.

[77] C. Branciard, E. G. Cavalcanti, S. P. Walborn, V. Scarani, and H. M. Wiseman, "One-sided device-independent quantum key distribution: Security, feasibility, and the connection with steering," *Phys. Rev. A*, vol. 85, no. 1, p. 010301, 2012.

[78] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed Experiment to Test Local Hidden-Variable Theories," *Phys. Rev. Lett.*, vol. 23, no. 15, pp. 880–884, 1969.

[79] B. S. Cirel'son, "Quantum generalizations of Bell's inequality," *Lett. Math. Phys.*, vol. 4, no. 2, pp. 93–100, 1980.

[80] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts, "Nonlocal correlations as an information-theoretic resource," *Phys. Rev. A*, vol. 71, no. 2, p. 022101, 2005.

[81] Č. Brukner, M. Zukowski, J.-W. Pan, and A. Zeilinger, "Bell's Inequalities and Quantum Communication Complexity," *Phys. Rev. Lett.*, vol. 92, no. 12, p. 127901, 2004.

[82] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, "Bell nonlocality," *Rev. Mod. Phys.*, vol. 86, no. 2, pp. 419–478, 2014.

[83] D. Cavalcanti and P. Skrzypczyk, "Quantitative relations between measurement incompatibility, quantum steering, and nonlocality," *Phys. Rev. A*, vol. 93, no. 5, p. 052112, 2016.

[84] E. Bene and T. Vértesi, "Measurement incompatibility does not give rise to Bell violation in general," *New J. Phys.*, vol. 20, no. 1, p. 013021, 2018.

[85] P. Girdhar and E. G. Cavalcanti, "All two qubit states that are steerable via CHSH-type correlations are Bell-nonlocal," *Phys. Rev. A*, vol. 94, p. 032317, 2016.

[86] F. Hirsch, M. T. Quintino, and N. Brunner, "Quantum measurement incompatibility does not imply Bell nonlocality," *Phys. Rev. A*, vol. 97, no. 1, p. 012129, 2017.

[87] S. Popescu and D. Rohrlich, "Quantum nonlocality as an axiom," *Found. Phys.*, vol. 24, no. 3, pp. 379–385, 1994.

[88] P. Janotta, C. Gogolin, J. Barrett, and N. Brunner, "Limits on nonlocal correlations from the structure of the local state space," *New J. Phys.*, vol. 13, p. 063024, 2011.

[89] D. Beckman, D. Gottesman, M. A. Nielsen, and J. Preskill, "Causal and localizable quantum operations," *Phys. Rev. A*, vol. 64, no. 5, p. 052309, 2001.

[90] M. J. Hoban and A. B. Sainz, "A channel-based framework for steering, non-locality and beyond," *New J. Phys.*, vol. 20, no. 5, p. 053048, 2018.

[91] W. Van Dam, "Implausible consequences of superstrong nonlocality," *Nat. Comput.*, vol. 12, no. 1, pp. 9–12, 2013.

[92] M. Plávala and M. Ziman, "Popescu-Rohrlich box implementation in general probabilistic theory of processes." 2017.

[93] A. Acin, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, "Device-Independent Security of Quantum Cryptography against Collective Attacks," *Phys. Rev. Lett.*, vol. 98, p. 230501, 2007.

[94] F. Bischof, H. Kampermann, and D. Bruß, "Measurement-device-independent randomness generation with arbitrary quantum states," *Phys. Rev. A*, vol. 95, no. 6, p. 062305, 2017.

[95] R. Colbeck and A. Kent, "Private randomness expansion with untrusted devices," *J. Phys. A.*, vol. 44, no. 9, p. 095305, 2011.

[96] W.-Y. Hwang, H.-Y. Su, and J. Bae, "Improved measurement-device-independent quantum key distribution with uncharacterized qubits," *Phys. Rev. A*, vol. 95, no. 6, p. 062313, 2017.

[97] L. Hardy, "Quantum Theory From Five Reasonable Axioms." 2001.

[98] J. van de Wetering, "Reconstruction of quantum theory from universal filters." 2018.

[99] S. Gudder, "Spectral methods for a generalized probability theory," *Trans. Am. Math. Soc.*, vol. 119, no. 3, pp. 428–442, 1965.

[100] G. Chiribella and C. M. Scandolo, "Operational axioms for diagonalizing states," *Electron. Proc. Theor. Comput. Sci.*, vol. 195, no. Qpl, pp. 96–115, 2015.

List of attached papers

1. M. Plávala, "All measurements in a probabilistic theory are compatible if and only if the state space is a simplex," *Phys. Rev. A*, vol. 94, no. 4, p. 042108, 2016.

2. A. Jenčová and M. Plávala, "Conditions on the existence of maximally incompatible two-outcome measurements in general probabilistic theory," *Phys. Rev. A*, vol. 96, no. 2, p. 022113, 2017.

3. M. Plávala, "Conditions for the compatibility of channels in general probabilistic theory and their connection to steering and Bell nonlocality," *Phys. Rev. A*, vol. 96, no. 5, p. 052127, 2017.

4. T. Heinosaari, L. Leppäjärvi and M. Plávala, "No-free-information principle in general probabilistic theories." 2018.

5. A. Jenčová and M. Plávala, "On the properties of spectral effect algebras." 2018.

All measurements in a probabilistic theory are compatible if and only if the state space is a simplex

Martin Plávala[*]

Mathematical Institute, Slovak Academy of Sciences, Štefánikova 49, Bratislava, Slovakia
(Received 19 August 2016; published 12 October 2016)

We study the compatibility of measurements on finite-dimensional compact convex state space in the framework of general probabilistic theory. Our main emphasis is on formulation of necessary and sufficient conditions for two-outcome measurements to be compatible and we use these conditions to show that there exist incompatible measurements whenever the state space is not a simplex. We also formulate the linear programming problem for the compatibility of two-outcome measurements.

I. INTRODUCTION

Incompatibility lies deeply within quantum mechanics and many of the famous and key aspects of quantum theories have been traced to Heisenberg uncertainty principle, the no cloning theorem, violations of Bell inequalities, and other notions making use of compatibility (see [1] for recent review). In light of these discoveries compatibility in the framework of general probabilistic theories has been studied [2–5] in order to show the difference between classical and nonclassical probabilistic theories. Also the connection of compatibility and steering in general probabilistic theories has been studied [6–8].

Recently incompatibility of measurements on quantum channels and combs has been in question [9] as it potentially could be used as a resource in quantum theory in a similar way as an incompatibility of measurements on quantum states [1]. The degree of compatibility (also called robustness of incompatibility) has been studied for measurements on channels and combs [2,9,10].

In the present paper we study the notion of compatibility of measurements in the framework of probabilistic theories and we show that every two measurements are compatible if and only if the state space is a simplex. In one way this result has clear physical interpretation—classical state space is always a simplex and the existence of incompatible measurements is often seen as one of the main aspects of quantum theories.

The paper is organized as follows. Section II contains preliminary mathematical results and references. Note that Sec. II B contains the definition of maximal face that (to the best knowledge of the present author) was not defined elsewhere (even though it has close ties to the notion of tangent half space and tangent hyper-plane [11, p. 169]) and is later used in Sec. IV. In Sec. III the measurements are defined. In Sec. IV compatibility of measurements and degree of compatibility are defined and it is shown that all measurements are compatible if and only if the state space is a simplex. Also the linear program for compatibility of two two-outcome measurements is formulated.

II. PRELIMINARIES

We present preliminary mathematical knowledge used in the paper. In all of the paper E will denote a real,

finite-dimensional vector space equipped with the Euclidean topology and K will denote a nonempty compact convex subset of E. We will denote the convex hull of a set X as $\text{conv}(X)$, affine hull of a set X as $\text{aff}(X)$, and interior of a set X as $\text{int}(X)$, and by ∂K we will denote the boundary of K, i.e., $\partial K = K \setminus \text{int}(K)$ as K is closed.

A. Structure of $A(K)$

By $A(K)$ we will denote the set of real valued affine functions on K and by $A(K)^+$ we will denote the set of positive affine functions on K, i.e., $f \in A(K)^+$ if and only if $f(x) \geqslant 0$ for every $x \in K$. We will denote constant functions by the value they attain. Since K is compact and the functions $A(K)$ are continuous, every function reaches its maximum and minimum over K at some point of K and we can introduce the supremum norm for $f \in A(K)$ as

$$\|f\|_A = \sup_{x \in K} |f(x)|.$$

The set $A(K)^+$ is closed; convex, i.e., for $\lambda \in \mathbb{R}$, $0 \leqslant \lambda \leqslant 1$, $f_1, f_2 \in A(K)^+$ we have $\lambda f_1 + (1 - \lambda) f_2 \in A(K)^+$; a cone, i.e., for $\nu \in \mathbb{R}$, $\nu > 0$, $f \in A(K)^+$ we have $\nu f_1 \in A(K)^+$; pointed, i.e., $A(K)^+ \cap (-A(K)^+) = \{0\}$; and generating, i.e., for every $f \in A(K)$ we have $f_+, f_- \in A(K)^+$ such that $f = f_+ - f_-$.

The closed, pointed, convex cone $A(K)^+$ defines a partial order $\geqslant$ on $A(K)$ given for $f_1, f_2 \in A(K)$ as

$$f_1 \geqslant f_2 \Leftrightarrow f_1 - f_2 \in A(K)^+$$

or equivalently $f_1 \geqslant f_2 \Leftrightarrow (f_1 - f_2)(x) \geqslant 0, \forall x \in K$. The partial order $\geqslant$ will play a role in our formulation of linear program for incompatibility of two-outcome measurements.

Definition 1. We say that $e \in A(K)^+$ is an order unit if for every $f \in A(K)^+$ there is some $\nu \in \mathbb{R}$, $\nu > 0$ such that

$$\nu e \geqslant f.$$

In the current setting it is easy to see that every strictly positive function is an order unit. We will omit the simple proof of the following fact.

Proposition 1. $e \in A(K)^+$ is an order unit if and only if $e \in \text{int}(A(K)^+)$.

We will also use the notion of a base of a cone.

Definition 2. Let $Q \subset E$ be a cone, then a set $\mathcal{B} \subset Q$ is called base of Q if for every $0 \neq x \in Q$ there exist unique $y \in \mathcal{B}$ and $\lambda \in \mathbb{R}$ such that $x = \lambda y$.

[*]martin.plavala@mat.savba.sk

To formulate the linear programming problem we will also have to work with the dual space of $A(K)$; we will denote it $A(K)^*$. We will denote by $A(K)^{*+}$ the cone of positive functionals dual to $A(K)^+$, that is $\psi \in A(K)^{*+}$ if and only if for every $f \in A(K)^+$ we have $\psi(f) \geqslant 0$.

Proposition 2. $A(K)^{*+}$ is a closed pointed convex cone.

Proof. It is straightforward to see that $A(K)^{*+}$ is a closed convex cone. It is pointed because $A(K)^+$ is generating. ∎

We define the dual norm for $\psi \in A(K)^*$ as

$$\|\psi\|_* = \sup_{\|f\|_A \leqslant 1} |\psi(f)|.$$

For $x \in K$ let $\Phi_x \in A(K)^*$ be given for $f \in A(K)$ as

$$\Phi_x(f) = f(x).$$

The map $\Phi : K \to A(K)^*$ is called an evaluation map and it is affine. It is easy to see that $\Phi[K] = \{\Phi_x : x \in K\}$ contains only positive functionals with unit norm such that $\Phi_x(1) = 1$ for every $x \in K$. The converse is also true:

Proposition 3. $\Phi[K] = \{\psi \in A(K)^* : \|\psi\|_* = \psi(1) = 1\}$.

Proof. For proof see [12, Theorem 4.3]. Also note that $\|\psi\|_* = \psi(1) = 1$ implies $\psi \geqslant 0$. ∎

The set $\Phi[K]$ is sometimes referred to as the state space as in general applications it is often easier to work with $\Phi[K]$ rather than K.

B. Exposed faces and maximal faces of a convex set

In this subsection we will define faces, exposed faces, and maximal faces and prove Proposition 4.

Definition 3. Let $C \subset K$ be a convex set (that is, C is a convex set that is a subset of K). We say that C is a face of K if $x \in C, \lambda \in \mathbb{R}, 0 < \lambda < 1$ and $x = \lambda y + (1 - \lambda)z$ implies $y, z \in C$.

It is straightforward that K and the empty set are faces of K and they are called the trivial faces. Apart from the trivial faces it is known that all faces lie in ∂K [11, Corollary 18.1.3]. A face consisting of a single point is called an extreme point of K.

Definition 4. Let $C \subset K$ be a set where some affine function f reaches its maximum (or minimum) over K, i.e., if $\max_{x \in K} f(x) = M_f$, then $C = \{x \in K : f(x) = M_f\}$. Such C is called an exposed face of K.

Every exposed face is a face [11, p. 162]. An exposed face consisting of single point is called an exposed point. It will be important that the set of exposed points of K is dense in the set of extreme points of K [11, Theorem 18.6] and that every face of a closed convex set is closed [11, Corollary 18.1.1]. Also note that not every extreme point must be an exposed point; an example of this is presented in [11, p. 163].

We proceed by defining the notion of a maximal face. Maximal faces are generalizations of the $n - 1$-dimensional exposed faces of polytopes (that is, of convex sets that are convex hulls of a finite number of points).

Definition 5. Let $C \subset K$ be a nontrivial face, such that for every $x \in K \setminus C$ we have $\text{conv}(C \cup \{x\}) \cap \text{int}(K) \neq \emptyset$; then we say that C is a maximal face.

Note that we require maximal faces to be nontrivial, i.e., K itself is not a maximal face. One can show that every maximal face is exposed, because every maximal face is an intersection

of K and a hyperplane tangent to K. Also every intersection of K and a hyperplane tangent to K is a maximal face. We present a simple example of maximal faces of a triangle and circle.

Example 1. Assume that $K \subset \mathbb{R}^2$ is a triangle. The vertices of the triangle are extreme and exposed points of K, but they are not maximal faces. In this case maximal faces are the edges of the triangle. Now consider that $K \subset \mathbb{R}^2$ is the convex hull of the unit circle; then every extreme point of K is a maximal face.

Maximal faces will play a role in the notion of compatibility of measurements as the condition $\text{conv}(C \cup \{x\}) \cap \text{int}(K) \neq \emptyset$ will be of great importance.

Proposition 4. Let $K \subset \mathbb{R}^n$ be a nonempty convex compact set. Then for every point $x \in \partial K$ there are maximal faces C_1, C_2 such that $x \in C_1$ and $x \notin C_2$.

Proof. We will prove the statement in two steps. First we will prove that that every point of ∂K belongs to some maximal face. Then we prove that maximal faces that have a point in common cannot form ∂K.

Let $x \in \partial K$; then there exists a nonconstant affine function f that reaches its maximum over K in x [11, Colloraly 11.6.2]; let $f(x) = M_f$. The set $G_0 = \{x' \in K : f(x') = M_f\}$ is an exposed face. If G_0 is maximal face then the proof is finished; if G_0 is not a maximal face, then there must exist a point $y \in K \setminus G_0$ such that $\text{conv}(G_0 \cup \{y\}) \cap \text{int}(K) = \emptyset$. The set $\text{conv}(G_0 \cup \{y\})$ does not have to be a face itself, but since $\text{conv}(G_0 \cup \{y\}) \cap \text{int}(K) = \emptyset$ then there exists a nontrivial supporting hyperplane to K containing $\text{conv}(G_0 \cup \{y\})$ (see [11, Theorem 11.6] for a definition of a supporting hyperplane to K and proof of the statement). In other words there must exist a nonconstant affine function f_1 such that $\max_{y \in K} f_1(y) = M_{f_1}$ and $G_1 = \{x' \in K : f_1(x') = M_{f_1}\} \supset \text{conv}(G_0 \cup \{y\})$, i.e., G_1 is an exposed face of K and $x \in G_1$. Moreover for the dimensions of $\text{aff}(G_0)$ and $\text{aff}(G_1)$ we must have $\dim(\text{aff}(G_1)) > \dim(\text{aff}(G_0))$, because $y \in G_1$ and $y \notin G_0$. If G_1 is a maximal face then the proof is finished; if not then we can repeat the procedure to find exposed face $G_2 \supset G_1$.

Since the affine span of every maximal face can be at most $n - 1$ dimensional and the dimension of affine span of the exposed faces G_i is strictly growing with i it is clear that we can repeat this procedure at most $n - 1$ times to obtain a maximal face; hence in this way to every $x \in \partial K$ we can find a maximal face that contains it.

Now we will proceed with the second part of the proof. Take $x \in \partial K$, denote $\{C_i\}$ the set of all maximal faces of K, and assume $x \in \cap_i C_i$. Since every point of ∂K belongs to some maximal face we must have $\cup_i C_i = \partial K$. Let us define positive affine functions f_i, such that $C_i = \{y \in K : f_i(y) = 0\}$; then since a finite-dimensional convex compact set is an intersection of closed half spaces tangent to it [11, Theorem 18.8] we have $K = \{y \in \mathbb{R}^n : f_i(y) \geqslant 0, \forall i\}$. Since we have $f_i(x) = 0, \forall i$ then for any $\lambda \in \mathbb{R}, \lambda \geqslant 0$, and $z \in K$ we have

$$f_i(\lambda z + (1 - \lambda)x) = \lambda f_i(z) \geqslant 0$$

for every i. This implies that $\lambda z + (1 - \lambda)x \in K$, which is in contradiction with K being compact. ∎

III. MEASUREMENTS ON K

Let E be a finite-dimensional real vector space equipped with the Euclidean topology and let $K \subset E$ be a compact convex set. We will call K a state space as it represents a set of all possible states of some system and the convex combination is interpreted as probabilistic mixture. Let Ω be a nonempty compact Hausdorff space and let $\mathcal{P}(\Omega)$ denote the set of Borel probability measures on Ω.

Definition 6. Measurements (also called observables) on K with sample space Ω are affine mappings $m : K \to \mathcal{P}(\Omega)$.

The interpretation is that Ω represents all possible outcomes of a certain measurement and is usually referred to as sample space. For $x \in K$ the measure $m(x) \in \mathcal{P}(\Omega)$ is a generalized notion of assigning probabilities to the measurement outcomes. Our definition follows the usual definitions of measurements in probabilistic theories [4,6] but may be easily generalized to locally compact sample spaces Ω. Let $\sigma \subset \Omega$ be a measurable set; then by $m(x; \sigma)$ we will denote the measure of the set σ with respect to the measure $m(x)$.

Finite outcome measurements

Let the sample space $\Omega = \{\omega_1, \ldots, \omega_k\}$ be a finite set. Every Borel probability measure $\mu \in \mathcal{P}(\Omega)$ is of the form

$$\mu = \sum_{i=1}^{k} \lambda_i \delta_{\omega_i}$$

where δ_{ω_i} is the Dirac measure centered at ω_i and $\lambda_i \in \mathbb{R}, 0 \leqslant \lambda_i \leqslant 1, \sum_{i=1}^{k} \lambda_i = 1$. It follows that if m is a measurement on K with finite sample space Ω then there always are functions $f_j \in A(K)^+, \ 0 \leqslant f_j \leqslant 1$ for $j \in \{1, \ldots, k\}, \ \sum_{j=1}^{k} f_j = 1$ such that

$$m = \sum_{i=1}^{k} f_i \delta_{\omega_i}.$$

Remark 1. In the standard literature [13,14] usually it is instead of writing $m = \sum_{i=1}^{k} f_i \delta_{\omega_i}$ simply said that the function f_j represents the probability of the outcome ω_j. To simplify the notation we will use the formulation presented above.

IV. COMPATIBILITY OF MEASUREMENTS

Assume that we wish to perform two distinct measurements m_1, m_2 with two separate sample spaces Ω_1, Ω_2. We would like to know whether there exists a measurement that performs both m_1 and m_2 at the same time. To ask this question properly we will introduce the concept of marginal measurement. When working with the Cartesian product $\Omega_1 \times \Omega_2$ we will always consider the product topology on it given by the topologies of Ω_1, Ω_2.

Definition 7. Let $m : K \to \mathcal{P}(\Omega_1 \times \Omega_2)$ be a measurement on K with sample space $\Omega_1 \times \Omega_2$. We say that $m_1 : K \to \mathcal{P}(\Omega_1)$ is a marginal measurement of m if for every measurable set $\sigma \subset \Omega_1$ and $x \in K$ we have

$$m_1(x; \sigma) = m(x; \sigma \times \Omega_2).$$

This definition can be formally understood as

$$m_1(x; \sigma) = \int_{\Omega_2} m(x; \sigma \times d\omega_2)$$

for every measurable set $\sigma \subset \Omega_1$. For the finite outcome measurements the integral is replaced by a sum over the outcomes; i.e., for $m = \sum_{i,j=1}^{k} f_{ij} \delta_{(\omega_i, \omega_j)}$, where $f_{ij} \in A(K)^+$ and $\delta_{(\omega_i, \omega_j)}$ is the Dirac measure centered at (ω_i, ω_j), we have

$$m_1 = \sum_{i,j=1}^{k} f_{ij} \delta_{\omega_i}.$$

It is straightforward to see that m_1 is a measurement on K with sample space Ω_1 as the positivity and normalization to 1 follow from the properties of m. Now we are ready for the definition of compatibility.

Definition 8. We will say that measurements $m_1 : K \to \mathcal{P}(\Omega_1)$ and $m_2 : K \to \mathcal{P}(\Omega_2)$ are compatible if there exists a measurement $m : K \to \mathcal{P}(\Omega_1 \times \Omega_2)$ such that m_1, m_2 are marginal measurements of m.

This definition is the standard definition used for compatibility of measurements.

A natural question is, are there any incompatible measurements? It is of course long known that incompatible measurements in quantum mechanics exist, but mathematically it is interesting to ask what properties of K imply that all measurements are compatible.

Proposition 5. Let K be a simplex, that is, let $\{x_1, \ldots, x_n\}$ be the set of extreme points of K such that the points $x_1, \ldots, x_n$ are affinely independent. Then every measurement on K is compatible with every other measurement on K.

Proof. Let K be a simplex; then there exists affine functions $b_j : K \to \mathbb{R}, j \in \{1, \ldots, n\}$ defined by $b_j(x_i) = \delta_{ij}$. These functions are positive, because for every $y \in K$ we have $y = \sum_{i=1}^{n} \lambda_i x_i$ with $\sum_{i=1}^{n} \lambda_i = 1$ and $0 \leqslant \lambda_i \leqslant 1$ for every i.

Let m be a measurement on K with a sample space Ω; then for $y \in K, y = \sum_{i=1}^{n} \lambda_i x_i$ we have

$$m(y) = \sum_{i=1}^{n} \lambda_i m(x_i) = \sum_{i=1}^{n} b_i(y) m(x_i),$$

i.e., a measurement m on a simplex is uniquely described by the measures $m(x_i) \in \mathcal{P}(\Omega)$.

Now let m_1, m_2 be measurements on K with the sample spaces Ω_1, Ω_2, respectively; then for $y \in K$ we have as above

$$m_j(y) = \sum_{i=1}^{n} b_i(y) m_j(x_i),$$

for $j \in \{1, 2\}$. Let $(m_1 \times m_2)(x_i)$ denote the product measure obtained from the measures $m_1(x_i)$ and $m_2(x_i)$, that is, for measurable sets $\sigma_i \subset \Omega_i, i \in \{1, 2\}$ we have

$$(m_1 \times m_2)(x_i; \sigma_1 \times \sigma_2) = m_1(x_i; \sigma_1) m_2(x_i; \sigma_2).$$

Let the measurement $m : K \to \mathcal{P}(\Omega_1 \times \Omega_2)$ be given as

$$m(y) = \sum_{i=1}^{n} b_i(y)(m_1 \times m_2)(x_i);$$

then it is easy to verify that m_1 and m_2 are marginal measurements of m. ∎

Note that positivity of functions b_j plays a crucial role in the proof and these functions are positive only if K is a simplex. Next we introduce the concept of a coin-toss (also called trivial) measurement.

Definition 9. Let μ be some fixed Borel probability measure on sample space Ω; then by coin toss we will refer to the measurement given as

$$m(y) = \mu$$

for every $y \in K$.

Coin-toss measurements usually represent noise, that is, some random factor that affects the measurement outcomes. It can be also interpreted as the most simple measurement when we ignore any information about the state and simply "toss a coin" and return whatever value we obtain. It is straightforward that any coin-toss measurement is compatible with any other measurement.

In the following we state the usual definition of the degree of compatibility.

Definition 10. Let $i \in \{1,2\}$ and let $m_i : K \to \mathcal{P}(\Omega_i)$ be a measurement on K with sample space Ω_i. Let $\tau_i : K \to \mathcal{P}(\Omega_i)$ be some coin-toss measurements; then we define degree of compatibility of measurements m_1, m_2 as

$$\text{DegCom}(m_1, m_2) = \sup_{\substack{0 \leqslant \lambda \leqslant 1 \\ \tau_1, \tau_2}} \{\lambda : \lambda m_1 + (1-\lambda)\tau_1,$$

$$\lambda m_2 + (1-\lambda)\tau_2 \text{ are compatible}\}.$$

The reason for considering different trivial measurements τ_1, τ_2 is that the sample spaces may be different and even if they would be the same due to our definitions we cannot pick some preferred measure as, for example, a properly normed Lebesgue measure on a compact subset of $\mathbb{R}^k$. Note that the supremum is taken also over the coin-toss measurements τ_1, τ_2.

Based on the analysis of compatibility presented in [1] we obtain the following.

Proposition 6. For any two measurements $m_i : K \to \mathcal{P}(\Omega_i)$, $i \in \{1,2\}$, we have $\text{DegCom}(m_1, m_2) \geqslant \frac{1}{2}$.

Proof. The idea is that we can always toss a fair two sided coin, based on the result implementing one of the measurements and substituting the other by the respective coin-toss observable. In other words let μ_1, μ_2 be any Borel probability measures on Ω_1, Ω_2, respectively, that give rise to coin-toss measurements τ_i given as $\tau_i(y) = \mu_i$, $i \in \{1,2\}$. Consider the measurement $m : K \to \mathcal{P}(\Omega_1 \times \Omega_2)$ given for $y \in K$ as

$$m(y) = \tfrac{1}{2}(\mu_1 \times m_2(y) + m_1(y) \times \mu_2).$$

It is straightforward to verify that the measurements $\frac{1}{2}(m_1 + \tau_1)$ and $\frac{1}{2}(m_2 + \tau_2)$ are marginal measurements of m. ∎

A similar result has been observed even for compatibility of quantum channels [15].

A. Compatibility of two-outcome measurements

In general it may be hard to decide whether measurements m_1 and m_2 are compatible but in the case of two-outcome measurements, that is, in the case when Ω_1, Ω_2 contain only two points, we will formulate necessary and sufficient conditions for the measurements m_1, m_2 to be compatible. These conditions may be generalized in the same manner to general finite outcome measurements.

Let $\Omega_1 = \Omega_2 = \Omega = \{\omega_1, \omega_2\}$ be the sample space of the measurements m_1, m_2; then they are of the form

$$m_i = f_i \delta_{\omega_1} + (1 - f_i)\delta_{\omega_2}$$

for $i \in \{1,2\}$. Also every measurement m on K with sample space $\Omega \times \Omega$ is of the form

$$m = g_{11}\delta_{(\omega_1,\omega_1)} + g_{12}\delta_{(\omega_1,\omega_2)} + g_{21}\delta_{(\omega_2,\omega_1)} + g_{22}\delta_{(\omega_2,\omega_2)},$$

where $g_{11}, g_{12}, g_{21}, g_{22} \in A(K)^+$ and $\delta_{(\omega_j,\omega_k)}$ is a Dirac measure on $\Omega \times \Omega$ centered at $(\omega_j, \omega_k) \in \Omega \times \Omega$. Assume that m_1 and m_2 are marginal measurements of m; then we obtain

$$g_{11} + g_{12} = f_1, \tag{1}$$

$$g_{21} + g_{22} = 1 - f_1, \tag{2}$$

$$g_{11} + g_{21} = f_2, \tag{3}$$

$$g_{12} + g_{22} = 1 - f_2. \tag{4}$$

These equations imply $g_{11} + g_{12} + g_{21} + g_{22} = 1$, but not $g_{jk} \geqslant 0$, $j,k \in \{1,2\}$, and they in general do not have a unique solution. Let $g_{11} = p$, $0 \leqslant p \leqslant 1$; then a general solution to Eqs. (1)–(4) is

$$g_{12} = f_1 - p,$$
$$g_{21} = f_2 - p,$$
$$g_{22} = 1 - f_1 - f_2 + p,$$

which imply the inequalities

$$f_1 \geqslant p, \tag{5}$$

$$f_2 \geqslant p, \tag{6}$$

$$1 + p \geqslant f_1 + f_2, \tag{7}$$

that come from $g_{jk} \geqslant 0$ for all $j,k \in \{1,2\}$. In general there may not exist such p satisfying Eqs. (5)–(7); in that case the measurements are incompatible. But if m is a joint measurement of m_1, m_2 then Eqs. (5)–(7) must be satisfied and Eqs. (1)–(4) are satisfied simply because m_1 and m_2 are marginals of m. We have proved the following.

Proposition 7. Let m_1, m_2 be two-outcome measurements on K given as

$$m_i = f_i \delta_{\omega_1} + (1 - f_i)\delta_{\omega_2}$$

for $i \in \{1,2\}$; then they are compatible if and only if there is a function $p \in A(K)^+$, such that $0 \leqslant p \leqslant 1$ and Eqs. (5)–(7) are satisfied. Similar results in terms of operators in the case of measurements on states were obtained in [2,10].

Now we will proceed by deriving some conditions on the incompatibility of two-outcome measurements based on the results of Proposition 7 that will help us prove that there exist incompatible measurements if and only if K is not a simplex.

The main idea is that we will construct two functions $f_1, f_2 \in A(K)^+$ that reach both zero and one on K and for

the exposed faces

$$F_i = \{x \in K : f_i(x) = 0\}$$

$i \in \{1,2\}$ it holds that $\text{conv}(F_1 \cup F_2) \cap \text{int}(K) \neq \emptyset$. Then by the Eqs. (5) and (6) we have that $p(x) = 0$ for every $x \in \text{conv}(F_1 \cup F_2)$. Since $\text{conv}(F_1 \cup F_2) \cap \text{int}(K) \neq \emptyset$ and $p \geq 0$ we get $p = 0$. Then by Eq. (7) we must have $f_1 + f_2 \leq 1$ if the measurements are compatible so we will show that we can construct functions f_1, f_2 with the mentioned properties such that $f_1(y) + f_2(y) > 1$ for some $y \in K$ whenever K is not a simplex.

The ideas presented above were inspired by an example of incompatible measurements on a square presented in [4].

Proposition 8. Let $x \in K$ be an extreme point and let F be a maximal face disjoint from $\{x\}$; then there exist incompatible two-outcome measurements on K if F does not contain all other extreme points of K except for x.

Proof. For the definition of a maximal face see Definition 5 and remember that according to the definition K itself is not a maximal face. Note that closedness of K will play a role as it implies closedness of every face of K [11, Corollary 18.1.1].

Assume that there is one maximal face F disjoint from x, but F does not contain all extreme points of K except for x, i.e., there is an extreme point $y \in K$, such that $y \notin F$ and $y \neq x$. Since $F, \{x\}, \{y\}$ are closed sets and $\{y\}$ is disjoint from both F and $\{x\}$, then there exists some open neighborhood N_ε containing y, such that $x \notin N_\varepsilon$ and $F \cap N_\varepsilon = \emptyset$. There is an exposed point $z \in N_\varepsilon$ as the set of exposed points is dense in the set of extreme points of K [11, Theorem 18.6]. For the same reason we will consider x an exposed point as well. Now let us construct positive affine function f_1, f_x, f_z such that

$$F_1 = \{w \in K : f_1(w) = 0\},$$
$$\{x\} = \{w \in K : f_x(w) = 0\},$$
$$\{z\} = \{w \in K : f_z(w) = 0\},$$

and

$$\max_{w \in K} f_1(w) = \max_{w \in K} f_x(w) = \max_{w \in K} f_z(w) = 1.$$

The functions f_1, f_x, f_z give rise to two-outcome measurements m_1, m_x, m_z given as

$$m_1 = f_1 \delta_{\omega_1} + (1 - f_1) \delta_{\omega_2},$$
$$m_x = f_x \delta_{\omega_1} + (1 - f_x) \delta_{\omega_2},$$
$$m_z = f_z \delta_{\omega_1} + (1 - f_z) \delta_{\omega_2}.$$

Since we have

$$\text{conv}(F_1 \cup \{x\}) \cap \text{int}(K) \neq \emptyset,$$
$$\text{conv}(F_1 \cup \{z\}) \cap \text{int}(K) \neq \emptyset,$$

we must have by Proposition 7

$$f_1 + f_x \leq 1, \quad f_1 + f_z \leq 1,$$

for the measurements m_1, m_x and m_1, m_z to be compatible. From $f_1 + f_x \leq 1$ we get $\{w \in K : f_1(w) = 1\} = \{x\}$ and from $f_1 + f_z \leq 1$ we get $\{w \in K : f_1(w) = 1\} = \{z\}$, which is a contradiction with $x \neq z$ implied by $x \notin N_\varepsilon$ and $z \in N_\varepsilon$. ∎

Proposition 9. Let $K \subset \mathbb{R}^n$ be a compact convex set; then there exist incompatible measurements on K whenever K is not a simplex.

Proof. We will rely on the results of Proposition 8. Assume that $x \in K$ is an extreme point that is affinely dependent on other extreme points, i.e., there are extreme points $\{y_1, \ldots, y_n\} \subset K$ such that $x = \sum_{i=1}^n \alpha_i y_i$ with $\sum_{i=1}^n \alpha_i = 1$ and let F denote the maximal face disjoint from $\{x\}$. Now let us construct a nonconstant positive affine function $f \in A(K)^+$ such that

$$F = \{z \in K : f(z) = 0\}.$$

Again the function f exists as F is an exposed face. Since $x = \sum_{i=1}^n \alpha_i y_i$, $\{y_1, \ldots, y_n\} \in F$ and f is affine, we have

$$f(x) = \sum_{i=1}^n \alpha_i f(y_i) = 0$$

and we must have $x \in F$, which is a contradiction. Hence the set of exposed points must be affinely independent and finite and K must be a simplex. ∎

It is an open question whether it can be in an easier fashion showed that the compactness and convexity of K together with compatibility of every two-outcome measurement imply the Riesz decomposition property [16, p. 84] as it is known that it is equivalent to K being a simplex [16, Corollary II.3.11]. It is also known that in more general settings of effect algebras the result does not hold, i.e., there are effect algebras that are compatible but that do not satisfy the Riesz decomposition property (see [17, Example 3.6] for an example).

B. Linear programming problem for compatibility of two-outcome measurements

We will formulate the problem of compatibility of two two-outcome measurements as a problem of linear programming [18] similar to the one obtained in [2]. We will start with the results of Proposition 7 and we will construct the linear programming problem from there.

Let m_1, m_2 be two-outcome measurements with sample space $\Omega = \{\omega_1, \omega_2\}$ given as

$$m_i = f_i \delta_{\omega_1} + (1 - f_i) \delta_{\omega_2}$$

for $i \in \{1, 2\}$ and let τ represent a coin-toss measurement given as

$$\tau = \tfrac{1}{2}(\delta_{\omega_1} + \delta_{\omega_2}).$$

In the following calculations we will restrict ourselves only to this special coin-toss observable as it is sufficient to determine whether the measurements m_1, m_2 are compatible.

We want to know what is the highest possible $\lambda \in [\frac{1}{2}, 1]$, such that the measurements $\lambda m_1 + (1 - \lambda)\tau, \lambda m_2 + (1 - \lambda)\tau$ are compatible. In terms of Proposition 7 we want to know what is the highest value of λ such that there exists $\tilde{p} \in A(K)^+$ such that the conditions

$$\lambda f_1 + \frac{1 - \lambda}{2} \geq \tilde{p}, \quad \lambda f_2 + \frac{1 - \lambda}{2} \geq \tilde{p},$$
$$1 + \tilde{p} \geq \lambda(f_1 + f_2) + (1 - \lambda)$$

are satisfied. Denoting $p = \frac{\tilde{p}}{\lambda}$ and $\mu = \frac{1-\lambda}{\lambda}$ we obtain

$$\frac{\mu}{2} - p \geqslant -f_1, \tag{8}$$

$$\frac{\mu}{2} - p \geqslant -f_2, \tag{9}$$

$$p \geqslant f_1 + f_2 - 1. \tag{10}$$

Now it is important to realize that maximizing λ is equivalent to minimizing μ. In the following we will introduce new partially ordered vector spaces and a linear map as the problem of linear programming will be formulated in their terms.

Let $x \in \mathbb{R} \times A(K)$; then $x = (\alpha, g)$ for $\alpha \in \mathbb{R}$ and $g \in A(K)$. We introduce partial ordering on $\mathbb{R} \times A(K)$ by the relation

$$(\alpha, g) = x \geqslant 0 \Leftrightarrow \alpha \geqslant 0, g \in A(K)^+.$$

The topological dual to $\mathbb{R} \times A(K)$ is $\mathbb{R} \times A(K)^*$; for $x = (\alpha, g)$, $\tilde{c} \in \mathbb{R} \times A(K)^*$, $\tilde{c} = (\beta, \psi)$, $\beta \in \mathbb{R}$, $\psi \in A(K)^*$ we have

$$\langle \tilde{c}, x \rangle = \alpha\beta + \psi(g).$$

We will also use $A(K) \times A(K) \times A(K)$ equipped with the following partial order: let $(g_1, g_2, g_3) \in A(K) \times A(K) \times A(K)$; then $(g_1, g_2, g_3) \geqslant 0$ if and only if $g_i \geqslant 0$ for every $i \in \{1, 2, 3\}$.

Let $T : \mathbb{R} \times A(K) \to A(K) \times A(K) \times A(K)$ be a linear map given as

$$T(\alpha, g) = \left(-g + \frac{\alpha}{2}, -g + \frac{\alpha}{2}, g \right),$$

where $\frac{\alpha}{2}$ stands for the constant function attaining the value $\frac{\alpha}{2}$. It is straightforward to see that T is linear.

Proposition 10. Let $c \in \mathbb{R} \times A(K)^*$, $c = (1, 0)$, $F \in A(K) \times A(K) \times A(K)$, $F = (-f_1, -f_2, f_1 + f_2 - 1)$ and $x \in \mathbb{R} \times A(K)$, $x = (\mu, p)$; then

$$\inf \langle c, x \rangle$$
$$x \geqslant 0$$
$$Tx \geqslant F$$

is a primal linear programming problem. When the reached minimum is zero then the measurements m_1, m_2 are compatible. Moreover, there always exists a primal feasible plan.

Proof. The proof is straightforward. We have $\langle c, x \rangle = \mu$ for the given c, $x \geqslant 0$ translates to $\mu \geqslant 0$, and $p \geqslant 0$. Note that $\mu \geqslant 0$ corresponds to $\lambda \leqslant 1$. $Tx \geqslant F$ is the same as

$$\left(-p + \frac{\mu}{2}, -p + \frac{\mu}{2}, p \right) \geqslant (-f_1, -f_2, f_1 + f_2 - 1)$$

which is in turn equivalent to conditions (8)–(10).

Since $\mu = \frac{1-\lambda}{\lambda}$ then $\mu = 0$ implies $\lambda = 1$. There always exists a primal feasible plan as we know that for $\lambda = \frac{1}{2}$ the measurements are always compatible (see Proposition 6). ∎

Now that we have the primal problem we will find the dual problem to obtain another condition on the compatibility of measurements m_1, m_2.

Proposition 11. The dual problem to the problem introduced in Proposition 10 is given as

$$\sup \langle F, l \rangle$$
$$T^* l \leqslant c$$
$$l \geqslant 0$$

where $l \in A(K)^* \times A(K)^* \times A(K)^*$ and T^* is given by the relation $\langle \tilde{l}, T\tilde{x} \rangle = \langle T^*\tilde{l}, \tilde{x} \rangle$ for every $\tilde{l} \in A(K)^* \times A(K)^* \times A(K)^*$ and $\tilde{x} \in \mathbb{R} \times A(K)$, i.e., $T^* : A(K)^* \times A(K)^* \times A(K)^* \to \mathbb{R} \times A(K)^*$, such that for $(\psi_1, \psi_2, \psi_3) \in A(K)^* \times A(K)^* \times A(K)^*$ we have

$$T^*(\psi_1, \psi_2, \psi_3) = \left(\tfrac{1}{2}(\psi_1 + \psi_2)(1), -\psi_1 - \psi_2 + \psi_3 \right)$$

where 1 stands for the constant function on K and $\psi_i(1)$ is the value of functional ψ_1 on this function, that is, for some $z_{11}, z_{12} \in K$ and $a_1, a_2 \in \mathbb{R}$, $a_1 \geqslant 0$, $a_2 \geqslant 0$ we have $\psi_1 = a_1\phi_{z_1} - a_2\phi_{z_2}$ and $\psi_1(1) = a_1 - a_2$.

Proof. The only thing we need to do is to find T^*; the rest follows from the relation between primal and dual problems [18, p. 163].

From the relation $\langle \tilde{l}, T\tilde{x} \rangle = \langle T^*\tilde{l}, \tilde{x} \rangle$ for $\tilde{l} = (\psi_1, \psi_2, \psi_3) \in A(K)^* \times A(K)^* \times A(K)^*$ and $\tilde{x} = (\alpha, g) \in \mathbb{R} \times A(K)$ we get

$$\begin{aligned}
\langle \tilde{l}, T\tilde{x} \rangle &= \left\langle (\psi_1, \psi_2, \psi_3), \left(-g + \frac{\alpha}{2}, -g + \frac{\alpha}{2}, g \right) \right\rangle \\
&= \frac{\alpha}{2}(\psi_1 + \psi_2)(1) + (-\psi_1 - \psi_2 + \psi_3)(g) \\
&= \left\langle \left(\frac{1}{2}(\psi_1 + \psi_2)(1), -\psi_1 - \psi_2 + \psi_3 \right), (\alpha, g) \right\rangle \\
&= \langle T^*\tilde{l}, \tilde{x} \rangle.
\end{aligned}$$
∎

Proposition 12. The duality gap between the primal problem given by Proposition 10 and the dual problem given by Proposition 11 is zero.

Proof. The duality gap is zero if there is a primal feasible plan and the cone

$$Q = \{ (T\tilde{x}, \langle c, \tilde{x} \rangle) : \tilde{x} \in \mathbb{R} \times A(K), \tilde{x} \geqslant 0 \},$$
$$Q \subset A(K) \times A(K) \times A(K) \times \mathbb{R},$$

where $c = (1, 0)$ as in Proposition 10 is closed [18, Theorem 7.2]. To show that Q is closed we will use the fact that if V, W are topological vector spaces then $Q_V \subset V$ is a cone with compact convex base and $T_V : V \to W$ is a continuous linear transformation, such that $\ker(T_V) \cap Q_V = \{0\}$; then the cone $T_V(Q_V)$ is closed [18, Lemma 7.3].

Because the cone $A(K)^+$ is generating there exists a base of positive functions $h_1, \ldots h_n$ such that for every $\tilde{h} \in A(K)^+$ we have $\tilde{h} = \sum_{i=1}^{n} \lambda_i h_i$ for $\lambda_i \geqslant 0$. We introduce the L^1 norm on $A(K)$: for $h' \in A(K)$, $h' = \sum_{i=1}^{n} v_i h_i$ we have $\|h'\|_{L1} = \sum_{i=1}^{n} |v_i|$. Note that this norm is an affine function on $A(K)^+$.

We can introduce a norm on $\mathbb{R} \times A(K)$ as follows: let $\tilde{x} = (\alpha, g) \in \mathbb{R} \times A(K)$; then

$$\|\tilde{x}\|_{\mathbb{R} \times A(K)} = |\alpha| + \|g\|_{L1}.$$

The base of the positive cone in $\mathbb{R} \times A(K)$ is the set

$$\mathcal{K} = \{ \tilde{x} \in \mathbb{R} \times A(K) : \|\tilde{x}\|_{\mathbb{R} \times A(K)} = 1 \}.$$

$\mathcal{K}$ is compact and convex, because the norm $\| \cdot \|_{\mathbb{R} \times A(K)}$ is continuous and for $\alpha \geqslant 0$ and $g \in A(K)^+$ it is affine.

The map $T' : \mathbb{R} \times A(K) \to A(K) \times A(K) \times A(K) \times \mathbb{R}$ given as

$$T'\tilde{x} = (T\tilde{x}, \langle c, \tilde{x} \rangle)$$

is linear and continuous. If for $(\alpha, g) = \tilde{x} \in \mathbb{R} \times A(K)$ it holds that $T'\tilde{x} = 0$ then we have to have $\tilde{x} = (0,0)$ as $\langle c, \tilde{x} \rangle = 0$ implies $\alpha = 0$ and $T\tilde{x} = (0,0,0)$ implies $g = 0$. In conclusion we have $\ker(T') = \{(0,0)\}$.

This shows that the cone Q is closed and since we have already showed in Proposition 10 that a primal feasible plan exists the duality gap is zero. ∎

We will proceed with rewriting the dual problem from Proposition 11 into a more usable form to obtain a necessary and sufficient condition for two two-outcome measurements to be incompatible. We will start from the dual problem stated in Proposition 11. Since $l \in A(K)^* \times A(K)^* \times A(K)^*$ and $l \geqslant 0$ we must have some $z_1, z_2, z_3 \in K$ and $a_1, a_2, a_3 \in \mathbb{R}$, $a_i \geqslant 0$, $i \in \{1,2,3\}$, such that $l = (a_1\phi_{z_1}, a_2\phi_{z_2}, a_3\phi_{z_3})$ in the formalism of Sec. II A. From $T^*l \leqslant c$ we obtain the conditions

$$\tfrac{1}{2}(a_1 + a_2) \leqslant 1, \quad a_3\phi_{z_3} \leqslant a_1\phi_{z_1} + a_2\phi_{z_2}.$$

Moreover we have

$$\langle F, l \rangle = -a_1 f_1(z_1) - a_2 f_2(z_2) + a_3(f_1(z_3) + f_2(z_3) - 1).$$

Thus we have proved the following.

Proposition 13. The two-outcome measurements m_1, m_2 corresponding to the functions f_1, f_2 are incompatible if and only if there exist positive numbers $a_1, a_2, a_3 \in \mathbb{R}$ and $z_1, z_2, z_3 \in K$ such that Eqs. (11) are satisfied and $\langle F, l \rangle > 0$.

To make $\langle F, l \rangle > 0$ we could first consider $f_1(z_1) = f_2(z_2) = 0$ as then only $f_1(z_3) + f_2(z_3) > 1$ would be required. In this case it would be easy to satisfy Eqs. (11) by suitable choice of a_1, a_2, a_3 whenever $\text{conv}(\{z_1, z_2\}) \cap \text{int}(K) \neq \emptyset$ as then for some $\nu \in [0,1]$ we would have $\nu z_1 + (1 - \nu)z_2 \in \text{int}(K)$ and $\phi_{\nu z_1 + (1-\nu)z_2}$ would be an order unit in $A(K)^*$. As a matter of fact, this is exactly the idea we used to prove Proposition 9. By similar methods of semidefinite programming it was shown that in the case of measurements on states the value of $\langle F, l \rangle$ corresponds to maximal violation of the Clauser-Horne-Shimony-Holt Bell inequality [2].

V. CONCLUSIONS

Incompatibility of measurements is one of the key aspects of quantum theories and, as our results have shown, in finite-dimensional cases it only differentiates classical probabilistic theories from general probabilistic theories. The quest for finding some essentially quantum restriction on probabilistic theories also considered in [3,5] is not over as such restriction would probably help us to understand quantum theories better and more deeply.

It is an open question whether our results hold also in the infinite-dimensional case. A possible approach to generalize our results would be to prove it using the Riesz decomposition property and to observe whether the proof may be generalized for infinite-dimensional state space. It was already proved in the framework of quantum logic [19] that if the state space is a Bauer simplex then the set of yes-no observables is compatible in quantum logic sense. There are hints that the compatibility in quantum logic sense and compatibility as defined in Definition 8 coincide as it can be showed that the conditions for compatibility of two-outcome measurements given by Proposition 7 may be rewritten in a similar way as in [19, Theorem 2].

ACKNOWLEDGMENTS

[1] T. Heinosaari, T. Miyadera, and M. M. Ziman, J. Phys. A **49**, 123001 (2016).

[2] M. M. Wolf, D. Perez-Garcia, and C. Fernandez, Phys. Rev. Lett. **103**, 230402 (2009).

[3] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, Phys. Rev. Lett. **99**, 240501 (2007).

[4] P. Busch, T. Heinosaari, J. Schultz, and N. Stevens, Europhys. Lett. **103**, 10002 (2013).

[5] S. Aravinda, R. Srikanth, and A. Pathak, arXiv:1607.01768.

[6] M. Banik, J. Math. Phys. **56**, 052101 (2015).

[7] R. Uola, T. Moroder, and O. Gühne, Phys. Rev. Lett. **113**, 160403 (2014).

[8] M. T. Quintino, T. Vértesi, and N. Brunner, Phys. Rev. Lett. **113**, 160402 (2014).

[9] M. Sedlák, D. Reitzner, G. Chiribella, and M. Ziman, Phys. Rev. A **93**, 052323 (2016).

[10] T. Heinosaari, J. Kiukas, and D. Reitzner, Phys. Rev. A **92**, 022115 (2015).

[11] R. T. Rockafellar, *Convex Analysis*, Princeton Landmarks in Mathematics and Physics (Princeton University, Princeton, NJ, 1997).

[12] L. Asimow and A. J. Ellis, *Convexity Theory and its Applications in Functional Analysis*, L.M.S. Monographs (Academic, New York, 1980).

[13] T. Heinosaari and M. Ziman, *The Mathematical Language of Quantum Theory. From Uncertainty to Entanglement* (Cambridge University, Cambridge, England, 2012).

[14] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory*, Publications of the Scuola Normale Superiore (Scuola Normale Superiore, Pisa, 2011).

[15] T. Heinosaari and T. Miyadera, arXiv:1608.01794.

[16] E. M. Alfsen, *Compact Convex Sets and Boundary Integrals*, Ergebnisse der Mathematik und ihrer Grenzgebiete (Springer-Verlag, Berlin, Heidelberg, New York, 1971).

[17] G. Jenča, Bull. Aust. Math. Soc. **64**, 81 (2001).

[18] A. Barvinok, *A Course in Convexity*, Graduate Studies in Mathematics (American Mathematical Society, Providence, RI, 2002).

[19] J. Pykacz, Int. J. Theor. Phys. **22**, 97 (1983).

Conditions on the existence of maximally incompatible two-outcome measurements in general probabilistic theory

Anna Jenčová[*] and Martin Plávala[†]

Mathematical Institute, Slovak Academy of Sciences, Štefánikova 49, Bratislava, Slovakia

We formulate the necessary and sufficient conditions for the existence of a pair of maximally incompatible two-outcome measurements in a finite-dimensional general probabilistic theory. The conditions are on the geometry of the state space; they require the existence of two pairs of parallel exposed faces with an additional condition on their intersections. We introduce the notion of discrimination measurement and show that the conditions for a pair of two-outcome measurements to be maximally incompatible are equivalent to requiring that a (potential, yet nonexisting) joint measurement of the maximally incompatible measurements would have to discriminate affinely dependent points. We present several examples to demonstrate our results.

I. INTRODUCTION

General probabilistic theories (GPTs) form a general framework that provides a unified description of all physical systems known today. In such theories, the central object is the state space represented by a convex set and the measurements (or more general devices) are seen as certain maps on the state space (see, e.g., [1]). The study of such theories began some time ago, related to mathematical foundations of quantum mechanics, but has gained a great deal of attention recently in connection with information theory. It was identified that several nonclassical effects that we know from quantum mechanics, such as steering and Bell nonlocality [2], can be found in this broader framework.

It has been known since the beginning of quantum theory that some quantum mechanical measurements cannot be implemented simultaneously; this phenomenon is referred to as incompatibility of measurements. It was shown that such measurements appear in any nonclassical GPT [3]. Moreover, one can violate even the bounds that hold in quantum mechanics: In finite dimensions, the minimal degree of compatibility of quantum measurements is bounded from below by a dimension-dependent constant [4], while a GPT may admit pairs of maximally incompatible two-outcome measurements [5], i.e., two-outcome measurements such that their degree of compatibility attains the minimal value of $\frac{1}{2}$.

A long-standing question is what the properties that single out quantum mechanics in the framework of GPTs are. To answer this question, it is important to know the relation of the manifestations of nonclassical effects to the geometry of state spaces. One step in this direction was made in [3], where it was proved that the nonexistence of incompatible pairs of two-outcome measurements characterizes classical theories, that is, such GPTs where the state space is a simplex. In the present work we aim at the other extreme, namely, we find necessary and sufficient conditions for a pair of maximally incompatible measurements to exist in a given GPT.

This question is of interest also because of the relation of incompatibility to other nonclassical features: It was shown

that one would need maximally incompatible measurements for maximal violation of Bell inequalities [6,7]. We obtain conditions that restrain the geometry of the state spaces for which this is possible. The essence of what is required was already captured in [5] in the example of the square state space. On the other hand, we prove that maximally incompatible measurements exist if the state space is the set of quantum channels; this is related to the results recently found in [8], where a somewhat different notion of the compatibility of measurements on quantum channels and combs was studied. We also introduce the notion of discrimination two-outcome measurement and show how the concept of discrimination measurements is connected to maximally incompatible measurements. Our results are demonstrated on some examples.

The paper is organized as follows. In Sec. II we provide a quick review of GPTs and of the notation we will use. Section III deals with measurements and their compatibility, especially the two-outcome measurements. In Sec. IV we introduce the degree of compatibility and show its relation to a linear program for the compatibility of two-outcome measurements. In Sec. V we formulate and prove the necessary and sufficient conditions for the existence of maximally incompatible two-outcome measurements. In Sec. VI we introduce the concept of discrimination measurements and study their (in)compatibility. We summarize in Sec. VII.

II. STRUCTURE OF A GENERAL PROBABILISTIC THEORY

We present the standard definition of a finite-dimensional GPT in a quick review just to settle the notation. See, e.g., [9] for more information.

In a GPT, a state represents a mathematical description of a procedure for preparation of a physical system. To express the possibility of forming probabilistic mixtures of such procedures, it is assumed that the state space is a convex subset of a vector space $\mathcal{V}$. The convex combinations are interpreted operationally (see, e.g., [10], Pt. 2). We will assume below that $\mathcal{V}$ is finite dimensional and the state space is a compact convex subset $K \subset \mathcal{V}$.

Measurements on the system are represented by maps assigning to each state the corresponding outcome probabilities.

---

[*]jenca@mat.savba.sk
[†]martin.plavala@mat.savba.sk

These maps are assumed affine, that is, mapping a convex mixture of states to the corresponding convex mixture of probability distributions on the set of outcomes. The two-outcome measurements (yes-no experiments) are represented by affine maps $f : K \to [0,1]$, mapping each state $x \in K$ to the corresponding probability of success (the "yes" outcome). Such maps are called effects.

We next list some basic definitions and briefly present the framework of ordered vector spaces of affine functions, which will be useful in the following. A good handbook for some of the standard notions from convex analysis is [11].

Let $\mathcal{V}$ be a finite-dimensional real vector space. For any $X \subset \mathcal{V}$, $\text{conv}(X)$ will denote the convex hull of X and $\text{aff}(X)$ the affine hull of X. If X is convex, then a face of X is a convex subset $F \subseteq X$ such that $\lambda x + (1 - \lambda) \in F$ for some $x,y \in X$ and $\lambda \in (0,1)$ implies that $x,y \in F$.

Let $K \subset \mathcal{V}$ be a compact convex subset and let $A(K)$ denote the set of affine functions $f : K \to \mathbb{R}$. Then $A(K)$ is a finite-dimensional real vector space. The partial order on $A(K)$ is introduced in a natural way: Let $f,g \in A(K)$; then $f \geqslant g$ if and only if $f(x) \geqslant g(x)$ for all $x \in K$. The corresponding positive cone is the convex cone of positive affine functions, which will be denoted by $A(K)^+$. The constant functions are denoted by the value they attain, i.e., $1(x) = 1$ and $0(x) = 0$ for all $x \in K$.

The set of effects on K will be denoted by $E(K)$, that is, $E(K) = \{f \in A(K) : 1 \geqslant f \geqslant 0\}$. For any $f \in E(K)$, the set

$$f^{-1}(0) = \{x \in K, \ f(x) = 0\}$$

is a face of K. A face of this form is called an exposed face. Two exposed faces F_0 and F_1 of $E(K)$ are parallel if $F_0 = f^{-1}(0)$ and $F_1 = f^{-1}(1) = (1 - f)^{-1}(0)$.

Let $A(K)^*$ be the vector space dual to $A(K)$ and let $\langle \psi, f \rangle$ denote the value of the functional $\psi \in A(K)^*$ on $f \in A(K)$. The positive cone $A(K)^+$ defines the dual order on $A(K)^*$ as follows: For $\psi_1, \psi_2 \in A(K)^*$, $\psi_1 \geqslant \psi_2$ if and only if $\langle \psi_1, f \rangle \geqslant \langle \psi_2, f \rangle$ for every $f \in A(K)^+$. Here the positive cone is the dual cone $A(K)^{*+} = \{\psi \in A(K)^* : \psi \geqslant 0\}$ of positive functionals.

For any $x \in K$ let ϕ_x denote the functional in $A(K)^*$, given by the evaluation $\langle \phi_x, f \rangle = f(x)$. Then clearly ϕ_x is positive and normalized: $\langle \phi_x, 1 \rangle = 1$. On the other hand, it can be seen that every positive normalized functional $\psi \in A(K)^*$ is of the form $\psi = \phi_x$ for some $x \in K$ (see, e.g., [12]). This implies that the set $\mathfrak{S}_K = \{\phi_x : x \in K\}$ is a base of the cone $A(K)^{*+}$, i.e., for every $\psi \in A(K)^{*+}$, $\psi \neq 0$, there is a unique $x \in K$ and unique $\alpha \in \mathbb{R}$, $\alpha > 0$, such that $\psi = \alpha \phi_x$.

The simplest example of a state space is an $(n - 1)$-dimensional simplex Δ_n; this describes the state space of a classical system. We have $A(\Delta_n) \simeq \mathbb{R}^n$; $A(\Delta_n)^+$ is identified with the set of vectors with non-negative entries and the effects are given by vectors with entries in $[0,1]$.

A quantum state space is the set of density operators $\mathfrak{S}(\mathcal{H})$ on a finite-dimensional Hilbert space $\mathcal{H}$. The space $A(\mathfrak{S}(\mathcal{H}))$ is identified with the space $B_h(\mathcal{H})$ of self-adjoint operators on $\mathcal{H}$ by $A(\rho) \equiv \text{Tr} \, A\rho$. Here the positive cone is the cone of positive-semidefinite operators $B(\mathcal{H})^+$, the constant 1 is the identity operator $\mathbb{1}$, and the effects satisfy $0 \leqslant E \leqslant \mathbb{1}$.

The ensuing two further examples will be important in the following.

Example 1. The square state space is defined as the convex hull $S := \text{conv}\{x_{00}, x_{01}, x_{10}, x_{11}\}$ of points $x_{ij} \in \mathcal{V}$, satisfying $x_{00} + x_{11} = x_{01} + x_{10}$. The space $A(S)$ can be identified with $\mathbb{R}^3$, with a positive cone $A(S)^+$ whose base is again a square. The vertices of this square correspond to the extremal effects, determined by

$$f_{k,l}(x_{n_1,n_2}) = \begin{cases} 1 & \text{for } n_k = l \\ 0 & \text{otherwise,} \end{cases}$$

where $n_1, n_2, k, l \in \{0,1\}$. The set of effects $E(S)$ is a double pyramid, with base formed by the above square and the two apexes given by the 0 and 1 functionals.

Example 2. Let $\mathcal{C}(\mathcal{H})$ denote the set of completely positive trace-preserving maps $B(\mathcal{H}) \to B(\mathcal{H})$; such maps are often called quantum channels. We will use the standard and well-known definitions that may be found in [10].

Let $\mathcal{H} \otimes \mathcal{H}$ denote the tensor product of $\mathcal{H}$ with itself and let Tr_1 denote the partial trace. Using the Choi representation of quantum channels, we have the identification

$$\mathcal{C}(\mathcal{H}) \equiv \{C \in B_h(\mathcal{H} \otimes \mathcal{H}) : \text{Tr}_1(C) = \mathbb{1}, C \geqslant 0\}.$$

This is clearly a finite-dimensional state space. Affine functions on $\mathcal{C}(\mathcal{H})$ have the form $C \mapsto \text{Tr} \, CA$ for some $A \in B_h(\mathcal{H} \otimes \mathcal{H})$, but note that A and $A + \mathbb{1} \otimes X$ define the same function if $X \in B_h(\mathcal{H})$ and $\text{Tr} \, X = 0$. Moreover, all elements of $A(\mathcal{C}(\mathcal{H}))^+$ are given by some positive operator A and all effects have the form $C \mapsto \text{Tr} \, CM$, where $M \in B_h(\mathcal{H} \otimes \mathcal{H})$ is such that

$$0 \leqslant M \leqslant \mathbb{1} \otimes \sigma$$

for some $\sigma \in \mathfrak{S}(\mathcal{H})$ [1,13], so that effects are given by two-outcome process positive-operator-valued measures defined in [14].

III. COMPATIBLE MEASUREMENTS IN GPT

Let K be a state space and let Ω be the set of all possible outcomes of some measurement. In the most general case, Ω carries the structure of a measurable space, but since we are mostly interested in two-outcome measurements, we will assume that Ω is a finite set. Let $\mathcal{P}(\Omega)$ be the set of all probability measures on Ω.

Definition 1. A measurement m on K with sample space Ω is an affine map

$$m : K \to \mathcal{P}(\Omega).$$

Let $\Omega = \{\omega_1, \ldots, \omega_k\}$ and let m be a measurement on K with sample space Ω. Then $x \mapsto m(x; \omega_i) := m(x)(\omega_i)$ is clearly an effect on K with $\sum_i m(\cdot; \omega_i) = 1$ and any such k-tuple of effects determines a measurement on K. In particular, the general form of a two-outcome measurement is

$$m_f(x) := f(x)\delta_1 + (1 - f)(x)\delta_2,$$

where $f \in E(K)$ and $\delta_1 = \delta_{\omega_1}$ and $\delta_2 = \delta_{\omega_2}$ are the two Dirac measures on $\Omega = \{\omega_1, \omega_2\}$. In other words, this means that $f(x)$ is the probability of getting the outcome ω_1 by the measurement m_f if the system is prepared in the state $x \in K$.

We will present the standard definition of compatibility of measurements [15].

Definition 2. Let m_1 and m_2 be measurements on K with sample spaces Ω_1 and Ω_2, respectively. We say that the measurements m_1 and m_2 are compatible if there exists a measurement m on K with sample space $\Omega_1 \times \Omega_2$ such that m_1 and m_2 are marginals of m: For all $x \in K$ and $A_1 \subset \Omega_1$ and $A_2 \subset \Omega_2$ we have

$$m_1(x; A_1) = m(x; A_1 \times \Omega_2),$$
$$m_2(x; A_2) = m(x; \Omega_1 \times A_2).$$

In this case, m is called a joint measurement of m_1 and m_2.

In the following, we deal with the compatibility of two-outcome measurements, given by two effects $f,g \in E(K)$.

Proposition 1. The two-outcome measurements m_f and m_g are compatible if and only if there exists a function $p \in E(K)$ such that

$$f \geqslant p, \tag{1}$$

$$g \geqslant p, \tag{2}$$

$$1 + p \geqslant f + g. \tag{3}$$

Moreover, any joint measurement of m_f and m_g is of the form

$$\begin{aligned} m = {}& p\delta_{(1,1)} + (f - p)\delta_{(1,2)} + (g - p)\delta_{(2,1)} \\ & + (1 + p - f - g)\delta_{(2,2)}, \end{aligned} \tag{4}$$

where $\delta_{(i,j)} = \delta_{(\omega_i,\omega_j)}$ and $p \in E(K)$ satisfies (1)–(3).

Proof. Let $\Omega = \{\omega_1, \omega_2\}$ and let $m : K \to \mathcal{P}(\Omega \times \Omega)$ be a joint measurement of m_f and m_g. Let us define $h_{ij} := m(\cdot, (\omega_i, \omega_j)) \in E(K)$, $i,j = 1,2$. Then we must have

$$f = h_{11} + h_{12}, \quad 1 - f = h_{21} + h_{22},$$
$$g = h_{11} + h_{21}, \quad 1 - g = h_{12} + h_{22},$$

which follows from Definition 2. Denoting $h_{11} = p$, one can show that this is equivalent to the conditions (1)–(3) and m is given by (4). See [3] for a more throughout derivation of these conditions. ∎

Proposition 2. m_f and m_g are compatible if and only if $m_{(1-f)}$ and m_g are compatible.

Proof. Assume that m_f and m_g are compatible and let $p \in E(K)$ satisfy (1)–(3). If we let $p' = g - p$, then Eq. (2) implies $p' \geqslant 0$, Eq. (3) implies $1 - f \geqslant p'$, $p \geqslant 0$ implies $g \geqslant p'$, and Eq. (1) implies $1 + p' \geqslant (1 - f) + g$. Since $1 - (1 - f) = f$ it is clear that the compatibility of $m_{(1-f)}$ and m_g implies the compatibility of m_f and m_g in the same manner. ∎

IV. DEGREE OF COMPATIBILITY

A degree of compatibility gives a way to quantify the (in)compatibility of a pair of measurements. One possibility to introduce such a degree in any GPT is to use the least amount of noise needed to make the measurements compatible [see [16,17] for some different (but related) definitions]. We first introduce the coin-toss measurements that will represent the noise.

Definition 3. A coin-toss measurement on K is a constant map

$$\tau(x) = \mu \in \mathcal{P}(\Omega), \quad x \in K.$$

Such a measurement ignores the input state and just returns the outcomes according to some fixed probability distribution. It is straightforward that a coin-toss measurement is compatible with any other measurement. Observe also that any pair of measurements m_1 and m_2 can be made compatible by mixing with a coin toss. Indeed, let τ be a coin toss and let

$$m'_i = \tfrac{1}{2}m_i + \tfrac{1}{2}\tau.$$

The measurements m'_1 and m'_2 are compatible: The joint measurement consists of choosing one of the measurements at random (by flipping a coin) and replacing the other by τ. This observation leads to the following definition of the degree of compatibility, introduced in [4].

Definition 4. Let m_1 and m_2 be two measurements on K with sample space Ω. The degree of compatibility of m_1 and m_2 is defined as

$$\begin{aligned} \mathrm{DegCom}(m_1, m_2) = \sup_{\substack{0 \leqslant \lambda \leqslant 1 \\ \tau_1, \tau_2}} \{\lambda : \lambda m_1 + (1 - \lambda)\tau_1, \lambda m_2 \\ + (1 - \lambda)\tau_2 \text{ are compatible}\}, \end{aligned}$$

where the supremum is taken over all coin-toss measurements τ_1 and τ_2.

The interpretation of this measure of compatibility is clear: The convex combination is a mathematical representation of making the measurements m_1 and m_2 less sharp by adding noise in the form of the coin tosses. As the value of λ decreases, the measurements get less and less sharp, until at some point they become compatible. If for a pair of measurements this happens at a larger value of λ than for another pair, we may say that the first pair is more compatible.

It can be seen from the remarks after Definition 3 that the lowest possible value of degree of compatibility is $\frac{1}{2}$. If this happens for a pair of measurements, it means that the only way to make them compatible is to discard one of them completely and replace it by a coin toss. It is known that such pairs of measurements exist for some state spaces [5], but not in finite-dimensional quantum mechanics [4].

Definition 5. We will say that two measurements are maximally incompatible if $\mathrm{DegCom}(m_f, m_g) = \frac{1}{2}$.

We will now turn to the study of the degree of compatibility of two-outcome measurements m_f and m_g. The following statement follows immediately from Proposition 2.

Corollary 1. Let m_f and m_g be two-outcome measurements. Then

$$\begin{aligned} \mathrm{DegCom}(m_f, m_g) &= \mathrm{DegCom}(m_{(1-f)}, m_g) \\ &= \mathrm{DegCom}(m_f, m_{(1-g)}) \\ &= \mathrm{DegCom}(m_{(1-f)}, m_{(1-g)}). \end{aligned}$$

It was shown in [17] that compatibility of m_f and m_g can be formulated as a problem of linear programming. A similar linear program was introduced in [3] and it was shown that the

dual program is of the form

$$\sup \left(a_3 [f(z_3) + g(z_3) - 1] - a_1 f(z_1) - a_2 g(z_2) \right),$$
$$a_1 + a_2 \leqslant 2, \quad a_3 \phi_{z_3} \leqslant a_1 \phi_{z_1} + a_2 \phi_{z_2},$$

where $z_1, z_2, z_3 \in K$ and a_1, a_2, a_3 are non-negative numbers. Let β denote the supremum. Then we have

$$\beta = \frac{1 - \mathrm{DegCom}_{1/2}(m_f, m_g)}{\mathrm{DegCom}_{1/2}(m_f, m_g)},$$

where

$$\mathrm{DegCom}_{1/2}(m_f, m_g) := \sup_{0 \leqslant \lambda \leqslant 1} \{\lambda : \lambda m_f + (1-\lambda)\tau, \lambda m_g$$
$$+ (1-\lambda)\tau \text{ are compatible}\}$$

is the degree of compatibility provided by mixing the measurements m_f and m_g with the fixed coin-toss measurement $\tau = \frac{1}{2}(\delta_1 + \delta_2)$. The measurements m_f and m_g are compatible if and only if $\beta = 0$. We clearly have

$$\mathrm{DegCom}_{1/2}(m_1, m_2) \leqslant \mathrm{DegCom}(m_f, m_g),$$

so $\mathrm{DegCom}_{1/2}(m_f, m_g) = 1$ implies $\mathrm{DegCom}(m_f, m_g) = 1$ and $\mathrm{DegCom}(m_f, m_g) = \frac{1}{2}$ implies $\mathrm{DegCom}_{1/2}(m_f, m_g) = \frac{1}{2}$.

We next show that if m_f and m_g are incompatible, the supremum in the above program is reached with $a_1 + a_2 = 2$, which allows us to rewrite the program in a more convenient way. So assume that the measurements m_f and m_g are incompatible. Then we have $\beta > 0$, which implies $a_1 + a_2 > 0$. Assume that the supremum is reached for some a_1, a_2, a_3 and z_1, z_2, z_3 such that $a_1 + a_2 < 2$. Define

$$a_1' = \frac{2}{a_1 + a_2} a_1,$$
$$a_2' = \frac{2}{a_1 + a_2} a_2,$$
$$a_3' = \frac{2}{a_1 + a_2} a_3.$$

It is straightforward to see that $a_3' \phi_{z_3} \leqslant a_1' \phi_{z_1} + a_2' \phi_{z_2}$. Moreover,

$$\beta < a_3'[f(z_3) + g(z_3) - 1] - a_1' f(z_1) - a_2' g(z_2),$$

which is a contradiction. It follows that in the case when the measurements m_f and m_g are incompatible we can write the linear program as

$$\sup 2(\eta[f(z_3) + g(z_3) - 1] - \nu f(z_1) - (1-\nu) g(z_2)),$$
$$\eta \phi_{z_3} \leqslant \nu \phi_{z_1} + (1-\nu) \phi_{z_2}, \tag{5}$$

where we have set $a_1 + a_2 = 2$ and used the substitutions $2\nu = a_1$ and $2\eta = a_3$. Also note that $\eta \phi_{z_3} \leqslant \nu \phi_{z_1} + (1-\nu)\phi_{z_2}$ implies that there exists $z_4 \in K$ such that

$$\nu z_1 + (1-\nu) z_2 = \eta z_3 + (1-\eta) z_4.$$

V. MAXIMALLY INCOMPATIBLE TWO-OUTCOME MEASUREMENTS

In this section we find necessary and sufficient conditions for the existence of maximally incompatible measurements m_f and m_g on a given state space K. A sufficient condition was

proved in [5]: A pair of maximally incompatible two-outcome measurements exists if K is the square state space of Example 1 or, more generally, there are two pairs of parallel hyperplanes tangent to K such that the corresponding exposed faces contain the edges of a square. Besides the square, such state spaces include the cube, pyramid, double pyramid, cylinder, etc. We will show that this condition is also necessary so that it characterizes state spaces admitting a pair of maximally incompatible two-outcome measurements.

The following notation will be used throughout:

$$F_0 = \{z \in K : f(z) = 0\},$$
$$F_1 = \{z \in K : f(z) = 1\},$$
$$G_0 = \{z \in K : g(z) = 0\},$$
$$G_1 = \{z \in K : g(z) = 1\}.$$

We begin by rephrasing the above sufficient condition. For completeness, we add a proof along the lines of [5].

Proposition 3. Assume there are some points $x_{00} \in F_0 \cap G_0$, $x_{10} \in F_1 \cap G_0$, $x_{01} \in F_0 \cap G_1$, and $x_{11} \in F_1 \cap G_1$ such that

$$\tfrac{1}{2}(x_{00} + x_{11}) = \tfrac{1}{2}(x_{10} + x_{01}).$$

Then $\mathrm{DegCom}(m_f, m_g) = \frac{1}{2}$.

Proof. Let p be any positive affine function on K. Then we have

$$p(x_{11}) + p(x_{00}) = p(x_{10}) + p(x_{01})$$

and

$$p(x_{11}) \leqslant p(x_{10}) + p(x_{01})$$

follows. Let $\tau_1 = \mu_1 \delta_{\omega_1} + (1-\mu_1)\delta_{\omega_2}$ and $\tau_2 = \mu_2 \delta_{\omega_1} + (1-\mu_2)\delta_{\omega_2}$ be coin-toss measurements. Then the conditions (1)–(3) for $\lambda m_f + (1-\lambda)\tau_1$ and $\lambda m_g + (1-\lambda)\tau_2$ take the form

$$\lambda f + (1-\lambda)\mu_1 \geqslant p,$$
$$\lambda g + (1-\lambda)\mu_2 \geqslant p,$$
$$1 + p \geqslant \lambda(f + g) + (1-\lambda)(\mu_1 + \mu_2).$$

Expressing some of these conditions at the points x_{10}, x_{01}, x_{11}, we get

$$1 + p(x_{11}) \geqslant 2\lambda + (1-\lambda)(\mu_1 + \mu_2), \tag{6}$$
$$(1-\lambda)\mu_1 \geqslant p(x_{01}), \tag{7}$$
$$(1-\lambda)\mu_2 \geqslant p(x_{10}). \tag{8}$$

From (6) we obtain

$$2\lambda \leqslant 1 + p(x_{11}) - (1-\lambda)(\mu_1 + \mu_2)$$

and since from (7) and (8) we have

$$p(x_{11}) \leqslant p(x_{10}) + p(x_{01}) \leqslant (1-\lambda)(\mu_1 + \mu_2),$$

it follows that $\lambda \leqslant \frac{1}{2}$. ∎

At this point we can demonstrate that maximally incompatible two-outcome measurements exist for the set $\mathcal{C}(\mathcal{H})$ of quantum channels (see Example 2).

Example 3. Let $K = \mathcal{C}(\mathcal{H})$, with $\dim(\mathcal{H}) = 2$. Let $|0\rangle, |1\rangle$ be an orthonormal basis of $\mathcal{H}$ and let $M, N \in B_h(\mathcal{H} \otimes \mathcal{H})$ be

given as

$$M = |0\rangle\langle 0| \otimes |0\rangle\langle 0|,$$
$$N = |0\rangle\langle 0| \otimes |1\rangle\langle 1|.$$

Then $0 \leqslant M \leqslant \mathbb{1} \otimes |0\rangle\langle 0|$ and $0 \leqslant N \leqslant \mathbb{1} \otimes |1\rangle\langle 1|$, so $f(C) = \operatorname{Tr} CM$ and $g(C) = \operatorname{Tr} CN$ define effects on $\mathcal{C}(\mathcal{H})$. Let

$$C_{00} = |1\rangle\langle 1| \otimes \mathbb{1},$$
$$C_{10} = |0\rangle\langle 0| \otimes |0\rangle\langle 0| + |1\rangle\langle 1| \otimes |1\rangle\langle 1|,$$
$$C_{01} = |0\rangle\langle 0| \otimes |1\rangle\langle 1| + |1\rangle\langle 1| \otimes |0\rangle\langle 0|,$$
$$C_{11} = |0\rangle\langle 0| \otimes \mathbb{1}.$$

It is easy to check that $C_{00}, C_{10}, C_{01}, C_{11} \in \mathcal{C}(\mathcal{H})$. Moreover,

$$C_{00} + C_{11} = \mathbb{1} \otimes \mathbb{1} = C_{10} + C_{01}$$

and

$$\operatorname{Tr}(C_{00}M) = \operatorname{Tr}(C_{00}N) = 0,$$
$$\operatorname{Tr}(C_{10}M) = 1, \quad \operatorname{Tr}(C_{10}N) = 0,$$
$$\operatorname{Tr}(C_{01}M) = 0, \quad \operatorname{Tr}(C_{01}N) = 1,$$
$$\operatorname{Tr}(C_{11}M) = \operatorname{Tr}(C_{11}N) = 1.$$

In conclusion, C_{00}, C_{10}, C_{01}, and C_{11} satisfy the properties in Proposition 3, so the two-outcome measurements m_f and m_g are maximally incompatible. An analogical fact was also observed in [1,8] in different circumstances.

We proceed to prove some necessary conditions.

Proposition 4. $\operatorname{DegCom}(m_f, m_g) = \frac{1}{2}$ only if

$$F_0 \cap G_0 \neq \emptyset,$$
$$F_0 \cap G_1 \neq \emptyset,$$
$$F_1 \cap G_0 \neq \emptyset,$$
$$F_1 \cap G_1 \neq \emptyset.$$

Proof. Let $F_1 \cap G_1 = \emptyset$. Then $f + g < 2$. Let $\tau = \delta_{\omega_2}$ and consider the measurements $\lambda m_f + (1 - \lambda)\tau = m_{\lambda f}$ and $\lambda m_g + (1 - \lambda)\tau = m_{\lambda g}$, $\lambda \in [0,1]$. Since $f + g < 2$, we can choose $\lambda > \frac{1}{2}$ such that $1 \geqslant \lambda(f + g)$, so $p = 0$ satisfies Eqs. (1)–(3) for $m_{\lambda f}$ and $m_{\lambda g}$.

The result for the other sets follows by using Corollary 1.∎

The conditions given by Proposition 4 are not sufficient, as we will demonstrate in the following example.

Example 4. Let K be a simplex with vertices x_1, x_2, x_3, x_4 and let b_1, b_2, b_3, b_4 denote positive affine functions such that

$$b_i(x_j) = \delta_{ij}.$$

Such functions exist because K is a simplex. Let

$$f = b_1 + b_2, \quad g = b_1 + b_3.$$

Then we have

$$F_1 \cap G_1 = \{x_1\},$$
$$F_1 \cap G_0 = \{x_2\},$$
$$F_0 \cap G_1 = \{x_3\},$$
$$F_0 \cap G_0 = \{x_4\},$$

but clearly the measurements m_f and m_g must be compatible as K is a simplex. As a matter of fact, Eqs. (1)–(3) are satisfied with $p = b_1$.

Proposition 5. $\operatorname{DegCom}(m_f, m_g) = \frac{1}{2}$ if and only if there exist points $x_{00}, x_{01}, x_{10}, x_{11}$ such that $x_{00} \in F_0 \cap G_0$, $x_{10} \in F_1 \cap G_0$, $x_{01} \in F_0 \cap G_1$, $x_{11} \in F_1 \cap G_1$, and

$$\tfrac{1}{2}(x_{00} + x_{11}) = \tfrac{1}{2}(x_{10} + x_{01}).$$

Proof. The "if" part is proved in Proposition 3. Conversely, if $\operatorname{DegCom}(m_f, m_g) = \frac{1}{2}$ then according to the results of Sec. IV, the supremum in (5) must be equal to 1, so we must have

$$\eta[f(z_3) + g(z_3) - 1] - \nu f(z_1) - (1 - \nu)g(z_2) = \tfrac{1}{2} \quad (9)$$

for some $\eta, \nu \in [0,1]$ and $z_1, z_2, z_3 \in K$ such that

$$\nu \phi_{z_1} + (1 - \nu)\phi_{z_2} \geqslant \eta \phi_{z_3}.$$

It follows that

$$\nu \phi_{z_1} \geqslant \eta \phi_{z_3} - (1 - \nu)\phi_{z_2}, \quad (10)$$
$$(1 - \nu)\phi_{z_2} \geqslant \eta \phi_{z_3} - \nu \phi_{z_1}. \quad (11)$$

Rewriting Eq. (9) we get

$$\langle \eta \phi_{z_3} - \nu \phi_{z_1}, f \rangle + \langle \eta \phi_{z_3} - (1 - \nu)\phi_{z_2}, g \rangle - \eta = \tfrac{1}{2}.$$

We clearly have $\langle \eta \phi_{z_3} - \nu \phi_{z_1}, f \rangle \leqslant \eta$, but Eq. (11) implies $\langle \eta \phi_{z_3} - \nu \phi_{z_1}, f \rangle \leqslant 1 - \nu$ and thus we must have $\langle \eta \phi_{z_3} - \nu \phi_{z_1}, f \rangle \leqslant \min(\eta, 1 - \nu)$. Similarly, we get $\langle \eta \phi_{z_3} - (1 - \nu)\phi_{z_2}, g \rangle \leqslant \min(\eta, \nu)$ and

$$\tfrac{1}{2} \leqslant \min(\eta, \nu) + \min(\eta, 1 - \nu) - \eta = \min(\nu, 1 - \nu, \eta, 1 - \eta),$$

which implies $\nu = \eta = \frac{1}{2}$. Moreover, there must be some $z_4 \in K$ such that

$$\tfrac{1}{2}(z_1 + z_2) = \tfrac{1}{2}(z_3 + z_4). \quad (12)$$

Equation (9) now becomes

$$f(z_3) + g(z_3) - f(z_1) - g(z_2) = 2,$$

which implies $f(z_3) = g(z_3) = 1$ and $f(z_1) = g(z_2) = 0$ as $f, g \in E(K)$. From Eq. (12) we get

$$f(z_2) = 1 + f(z_4),$$

which implies $f(z_2) = 1$, $f(z_4) = 0$, and

$$g(z_1) = 1 + g(z_4),$$

which implies $g(z_1) = 1$ and $g(z_4) = 0$. Together we get

$$z_3 \in F_1 \cap G_1,$$
$$z_2 \in F_1 \cap G_0,$$
$$z_1 \in F_0 \cap G_1,$$
$$z_4 \in F_0 \cap G_0.$$

∎

We now give a characterization of state spaces admitting a pair of maximally incompatible two-outcome measurements.

Theorem 1. Let K be a state space. A pair of maximally incompatible two-outcome measurements on K exists if and only if K contains a square (as defined in Example 1) whose opposite edges lie in parallel exposed faces of K. The effects
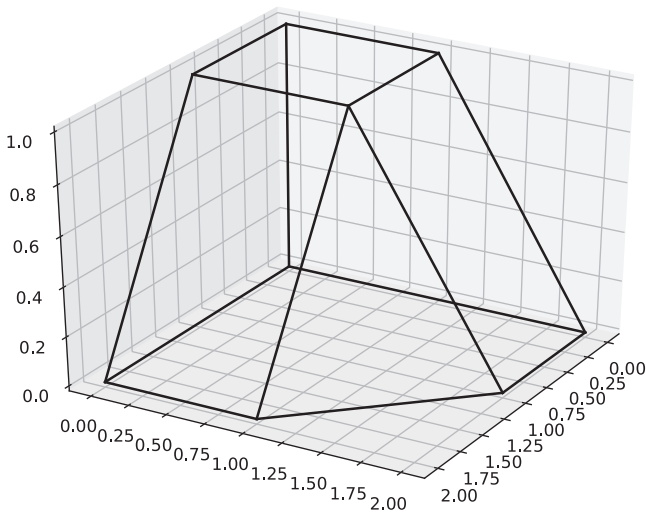
FIG. 1. State space K used in Example 5.

determining these exposed faces then constitute the pair of maximally incompatible measurements.

Proof. Let m_f and m_g be maximally incompatible. Then F_0, F_1 and G_0, G_1 are parallel exposed faces that contain the edges of a square by Proposition 5. Conversely, assume that the condition holds. Let $x_{00}, x_{01}, x_{10}, x_{11} \in K$ be the vertices of the square and let F_0, F_1 and G_0, G_1 be pairs of parallel exposed faces containing the opposite edges, so that $x_{ij} \in F_i \cap G_j$, $i,j \in \{1,2\}$. Then there are some $f,g \in E(K)$ such that $F_0 = f^{-1}(0), F_1 = f^{-1}(1)$ and $G_0 = g^{-1}(0), G_1 = g^{-1}(1)$. By Proposition 3, m_f and m_g are maximally incompatible. ∎

In the remainder of this section, we aim to give some geometric insight into the above condition.

Corollary 2. Let K be a state space. If $\dim(\mathrm{aff}(K)) = 2$, then maximally incompatible two-outcome measurements exist on K if and only if K is a square. In general, such measurements exist only if there is an affine subspace $V \subset \mathrm{aff}(K)$, $\dim(V) = 2$, such that $S = V \cap K$ is a square.

Proof. The first statement is immediate from Theorem 1. The principal idea for the second statement is that $V = \mathrm{aff}(S)$, where S is the square in question. Let us assume that there exist maximally incompatible measurements m_f and m_g on K and let S be the square as in Theorem 1, with vertices $x_{00}, x_{01}, x_{10}, x_{11}$. Let $V = \mathrm{aff}(x_{00}, x_{10}, x_{01})$ and let F_0, F_1 and G_0, G_1 be the parallel exposed faces of K, containing the edges of S. It is easy to see that $F_i \cap V$ and $G_i \cap V$ are faces of $K \cap V$ and they coincide with the edges of S. It is now obvious that $S = V \cap K$. ∎

We will present an example to show that the condition in Corollary 2 is not sufficient, even if the square $V \cap K$ is an exposed face of K.

Example 5. Let $K \subset \mathbb{R}^3$ be defined as

$$K = \mathrm{conv}\{(0,0,0),(2,0,0),(0,2,0),(2,1,0),(1,2,0),(1,1,1),$$
$$\times (1,0,1),(0,1,1),(0,0,1)\}$$

(see Fig. 1). Let

$$V = \{(x_1,x_2,x_3) \in \mathbb{R}^3 : x_3 = 1\}.$$

Then $K \cap V = S$, where

$$S = \mathrm{conv}\{(1,1,1),(1,0,1),(0,1,1),(0,0,1)\}$$

is an exposed face and a square.

To see that there is not a pair of maximally incompatible measurements m_f and m_g, corresponding to S, it is enough to realize that the effects f and g would have to reach the values 0 and 1 on maximal faces that are not parallel, i.e., we would have to have $\mathrm{aff}(F_0) \cap \mathrm{aff}(F_1) \neq \emptyset$ and $\mathrm{aff}(G_0) \cap \mathrm{aff}(G_1) \neq \emptyset$, which is impossible. On the other hand, the examples of a double pyramid or a cylinder show that maximally incompatible two-outcome measurements may exist on K even if the square $V \cap K$ described in Corollary 2 is not a face of K.

VI. DISCRIMINATION MEASUREMENTS

We now introduce a type of measurement that will allow us to formulate the conditions for existence of maximally incompatible two-outcome measurements in a clearer way.

Definition 6. We say that a two-outcome measurement m_f discriminates the sets $E_0, E_1 \subset K$ if

$$E_0 \subset f^{-1}(0), \quad E_1 \subset f^{-1}(1).$$

We call such measurement a discrimination measurement.

The idea of the definition is simple: Assume that a system is in an unknown state, but we know that it belongs to either E_0 or E_1. By performing the discrimination measurement m_f we can tell with 100% accuracy whether the state of the system belongs to E_0 or E_1. Clearly, m_f is a discrimination measurement if and only if both $f^{-1}(0)$ and $f^{-1}(1)$ are nonempty.

The definition can be generalized in an obvious way to general measurements that can discriminate more than two sets. The most well-known discrimination measurements used in quantum mechanics are projective measurements consisting of rank-1 projections that discriminate the states corresponding to the projections.

We are ready to reformulate the necessary and sufficient condition for maximal incompatibility of two-outcome measurements.

Proposition 6. The measurements m_f and m_g are maximally incompatible if and only if there is a square $S \subseteq K$ such that m_f and m_g discriminate the opposite edges of S.

Proof. First assume that there is a square S whose opposite edges can be discriminated by measurements m_f and m_g. Denote the vertices of S by x_{00}, x_{10}, x_{01}, and x_{11}. Then it is clear that the requirements of Proposition 3 are satisfied and thus we must have $\mathrm{DegCom}(m_f, m_g) = \frac{1}{2}$.

Conversely, assume that $\mathrm{DegCom}(m_f, m_g) = \frac{1}{2}$. Then by Proposition 5 there is a square S with vertices x_{00}, x_{10}, x_{01}, and x_{11} such that $x_{i0}, x_{i1} \in F_i$ and $x_{0j}, x_{1j} \in G_j$, $i,j \in \{0,1\}$. By convexity, m_f discriminates the edges between x_{00}, x_{01} and x_{10}, x_{11} and similarly m_g discriminates the other parallel pair. ∎

We finish with a necessary condition for compatibility of two-outcome discrimination measurements. Assume that m_f and m_g are such that $F_i \cap G_j \neq \emptyset$ (obviously, m_f and m_g are discrimination measurements in this case) and let $x_{ij} \in F_i \cap G_j$, $i,j \in \{0,1\}$. Assume also that m_f and m_g are compatible, so that (1)–(3) hold. Inserting $x_{11} \in F_1 \cap G_1$ into Eq. (3),

we get

$$p(x_{11}) \geqslant 1,$$

which together with $p \in E(K)$ implies $p(x_{11}) = 1$. Equations (1) and (2) and the positivity of p imply

$$p(x_{00}) = p(x_{10}) = p(x_{01}) = 0.$$

Expressing also the functions $(f - p)$, $(g - p)$, and $(1 + p - f - g)$ on the points x_{00}, x_{10}, x_{01}, and x_{11}, we get

$$(f - p)(x_{00}) = (f - p)(x_{01}) = (f - p)(x_{11}) = 0,$$
$$(f - p)(x_{10}) = 1,$$
$$(g - p)(x_{00}) = (g - p)(x_{10}) = (g - p)(x_{11}) = 0,$$
$$(g - p)(x_{01}) = 1,$$
$$(1 + p - f - g)(x_{00}) = (1 + p - f - g)(x_{10}) = 0,$$
$$(1 + p - f - g)(x_{01}) = 0,$$
$$(1 + p - f - g)(x_{11}) = 1.$$

This shows that the joint measurement m given by (4) discriminates the points x_{00}, x_{10}, x_{01}, and x_{11}. In particular, this implies that these points must be affinely independent. We have proved the following.

Proposition 7. Let m_f and m_g be compatible discrimination measurements such that $F_i \cap G_j \neq \emptyset$ for all $i, j \in \{0, 1\}$. Then any joint measurement of m_f and m_g must discriminate $F_i \cap G_j$.

VII. CONCLUSION

We have shown that the existence of maximally incompatible two-outcome measurements in GPT is equivalent to a geometric condition on the state space K. The essence of this condition is covered by the example of square state space in [5]. The importance of this result lies in its connection to maximal violation of Bell's inequalities and therefore to possible realizations of Popescu-Rorlich boxes [18], which are studied as potentially powerful resources in information theory. The example of the state space $\mathcal{C}(\mathcal{H})$ (Example 3) is particularly interesting in this respect, since it shows that maximal incompatibility can be achieved by devices existing in quantum theory.

The geometric interpretation of the minimal degree of compatibility that can be attained on a state space K is an interesting question for future research. It would be also of interest whether the connection between discriminating certain sets and compatibility of measurements could be fruitful from an information-theoretic viewpoint. This area of research might also yield some insight into why there exist maximally incompatible measurements on quantum channels although they do not exist on quantum states.

[1] T. Heinosaari, T. Miyadera, and M. Ziman, J. Phys. A: Math. Theor. **49**, 123001 (2015).

[2] H. M. Wiseman, S. J. Jones, and A. C. Doherty, Phys. Rev. Lett. **98**, 140402 (2007).

[3] M. Plávala, Phys. Rev. A **94**, 042108 (2016).

[4] T. Heinosaari, J. Schultz, A. Toigo, and M. Ziman, Phys. Lett. A **378**, 1695 (2014).

[5] P. Busch, T. Heinosaari, J. Schultz, and N. Stevens, Europhys. Lett. **103**, 10002 (2013).

[6] N. Stevens and P. Busch, Phys. Rev. A **89**, 022123 (2014).

[7] M. Banik, M. R. Gazi, S. Ghosh, and G. Kar, Phys. Rev. A **87**, 052125 (2013).

[8] M. Sedlák, D. Reitzner, G. Chiribella, and M. Ziman, Phys. Rev. A **93**, 052323 (2016).

[9] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. A **81**, 062348 (2010).

[10] T. Heinosaari and M. Ziman, *The Mathematical Language of Quantum Theory: From Uncertainty to Entanglement* (Cambridge University Press, Cambridge, 2012).

[11] R. T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, 1997).

[12] L. Asimow and A. J. Ellis, *Convexity Theory and its Applications in Functional Analysis* (Academic, New York, 1980), Theorem 4.3.

[13] A. Jenčová, J. Math. Phys. **53**, 012201 (2012).

[14] M. Ziman, Phys. Rev. A **77**, 062112 (2008).

[15] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (Scuola Normale Superiore, Pisa, 2011), Chap. 2.

[16] E. Haapasalo, J. Phys. A: Math. Theor. **48**, 255303 (2015).

[17] M. M. Wolf, D. Perez-Garcia, and C. Fernandez, Phys. Rev. Lett. **103**, 230402 (2009).

[18] S. Popescu and D. Rohrlich, Found. Phys. **24**, 379 (1994).

Conditions for the compatibility of channels in general probabilistic theory and their connection to steering and Bell nonlocality

Martin Plávala[*]

Mathematical Institute, Slovak Academy of Sciences, Štefánikova 49, Bratislava, Slovakia

We derive general conditions for the compatibility of channels in general probabilistic theory. We introduce formalism that allows us to easily formulate steering by channels and Bell nonlocality of channels as generalizations of the well-known concepts of steering by measurements and Bell nonlocality of measurements. The generalization does not follow the standard line of thinking stemming from the Einstein-Podolsky-Rosen paradox, but introduces steering and Bell nonlocality as entanglement-assisted incompatibility tests. We show that all of the proposed definitions are, in the special case of measurements, the same as the standard definitions, but not all of the known results for measurements generalize to channels. For example, we show that for quantum channels, steering is not a necessary condition for Bell nonlocality. We further investigate the introduced conditions and concepts in the special case of quantum theory and we provide many examples to demonstrate these concepts and their implications.

I. INTRODUCTION

Incompatibility of measurements is the well-known quantum phenomenon that gives rise to steering and Bell nonlocality. Historically, the idea of measurement incompatibility dates back to Bohr's principle of complementarity. Steering was first described by Schrödinger [1] and Bell nonlocality was first introduced by Bell [2], both as a reply to the paradox of Einstein, Podolsky, and Rosen (EPR) [3]. It is known that incompatibility of measurements is necessary and in some cases sufficient for both steering and Bell nonlocality, but the operational connection between incompatibility, steering, and Bell nonlocality was so far not described in general terms that would also fit channels, not only measurements.

There was extensive research into properties of quantum incompatibility of measurements [4,5], quantum incompatibility of measurements and its noise robustness, or degree of compatibility [6,7], connection of quantum incompatibility of measurements and steering [8–12], connection of quantum incompatibility of measurements and Bell nonlocality [13–16], and connection between steering and Bell nonlocality [17,18] (for a recent review, see [19]). In recent years, the problems of incompatibility of measurements on channels [20], compatibility of channels [21], the connection of channel steering to measurement incompatibility [22], and incompatibility in general probabilistic theory [23–26] were all studied.

The aim of this paper is to heavily generalize the recent results of [26], where compatibility, steering, and Bell nonlocality of measurements were formulated using convex analysis and the geometry of tensor products. In this paper, we will generalize the ideas and results of [26] for the case of two channels in general probabilistic theory. The generalizations are not straightforward and we will have to introduce several operational ideas and definitions, e.g., we introduce the operational interpretation of direct products of state spaces and we define steering and Bell nonlocality as very simple entanglement-assisted incompatibility test, that

boil down to the problem whether there exists a multipartite state with given marginal states.

During all of our calculations we will restrict ourselves to finite-dimensional general probabilistic theory and to only the case of two channels. We will restrict to only two channels just for simplicity, as one may easily formulate many of our results for more than two channels using the same operational ideas as we will present.

The paper is organized as follows: In Sec. II we describe our motivation for using general probabilistic theory. We provide several references to known applications and their connections to each other. In Sec. III we introduce general probabilistic theory. Note that in Sec. III D we introduce the operational interpretations of direct products in general probabilistic theory. In Sec. IV we define compatibility of channels and we derive a condition for compatibility of channels. In Sec. V we show that our condition for compatibility of channels yields the condition for compatibility of measurements that was presented in Ref. [26]. In Sec. VI we derive specific conditions for the compatibility of quantum channels. In Sec. VII we propose an idea for a test of incompatibility of channels, that will not work at first, but will eventually lead to both steering and Bell nonlocality. In Sec. VIII we define steering by channels as one-sided entanglement assisted incompatibility test and we derive some basic results. In Sec. IX we show that for the special case of measurements our definition of steering leads to the standard definition of steering [27] in the formalism of [26]. In Sec. X we derive the specific conditions for steering by quantum channels, we show that every pair of incompatible channels may be used for steering of maximally entangled state, and that there are entangled states that are not steerable by any pair of channels, among other results. In Sec. XI we define Bell nonlocality of channels as a two-sided entanglement assisted incompatibility test and we derive some basic results. Then, in Sec. XII we show that, when applied to measurement, the general definition of Bell nonlocality yields the standard definition of Bell nonlocality [27] in the formalism of [26], and we also show that for measurements steering is a necessary condition for Bell nonlocality. In Sec. XIII we derive conditions for the Bell nonlocality of quantum channels, we

———————
[*]martin.plavala@mat.savba.sk

formulate a generalized version of the Clauser-Horne-Shimony-Holt (CHSH) inequality, we show that for such inequality Tsirelson bound [28] both holds and is reached, we show an example of violation of the generalized version of CHSH inequality, and we build on the example from Sec. X of an entangled state not steerable by any pair of channels to show that, even though the state is not steerable by any pair of channels, it leads to Bell nonlocality, which shows that steering is not a necessary condition for Bell nonlocality for quantum channels. In Sec. XIV we conclude the paper by presenting the many open questions and possible areas of research opened by our paper.

II. MOTIVATIONS FOR USING GENERAL PROBABILISTIC THEORY

There are few motivations to using general probabilistic theory. The first motivation is mathematical as general probabilistic theory is a unified framework capable of describing both classical and quantum theory, as well as other theories. In this paper, the mathematical motivation is (according to the personal opinion of the author) even stronger as some of the formulations of the presented ideas and some of the proofs of the presented theorems turn out to be clearer in the framework of general probabilistic theory. The second motivation comes from foundations of quantum theory as general probabilistic theory provides insight into the structure of entanglement and incompatibility. The third and most promising motivation comes from information theory. There were developed several models [29–31] that have very interesting information-theoretic properties and that can be described by general probabilistic theory, albeit sometimes it needs to be extended even more [32]. Apart from the well-known results on the properties of Popescu-Rohrlich boxes [33,34], it was showed that there are theories in which one can search an N-item database in $O(\sqrt[3]{N})$ queries [35] and that there is a general probabilistic theory that can be simulated by a probabilistic classical computer that can perform Deutsch-Jozsa and Simon's algorithm [36].

The aforementioned results show that studying general probabilistic theory is interesting even from a practical viewpoint and that it could have potential applications in information processing.

III. INTRODUCTION TO GENERAL PROBABILISTIC THEORY

General probabilistic theory is a unified framework to describe the kinematics of different systems in a mathematically unified fashion. The idea of general probabilistic theory is an operational approach to setting the axioms and then carrying forward using convex analysis. Useful books on convex analysis are [37,38]. The beautiful aspect of general probabilistic theory is that it is only little bit more general than dealing with the different systems on their own, but we do not have to basically rewrite the same calculations over and over again for different theories.

During our calculations, we will use two recurring examples: one will be finite-dimensional classical theory and the other will be finite-dimensional quantum theory. The finite-dimensional classical theory is closely tied to the known results about incompatibility, steering, and Bell nonlocality of measurements and we will mainly use it to verify that the definitions we will propose are, in the special case of measurements, the same as the known definitions. The quantum theory is our main concern as this is the theory we are mostly interested in. Some results, that we will only prove for quantum theory, may be generalized for general probabilistic theory, but we will limit the generality of our calculations to make them more understandable to readers that are not so far familiar with general probabilistic theory.

Given that we will work with many different spaces, their duals, their tensor products, and many isomorphic sets, all isomorphisms will be omitted unless explicitly stated otherwise.

A. State space and the effect algebra of general probabilistic theory

There are two central notions in general probabilistic theory: the state space that describes all possible states of the system and the effect algebra that describes the measurements on the system. We will begin our construction from the state space and then define the effect algebra, but we will show how one can go the other way and start from an effect algebra and obtain state space afterwards. We will restrict ourselves to finite-dimensional spaces and always use the Euclidean topology.

Let V denote a real, finite-dimensional vector space and let $X \subset V$, then by $\mathrm{conv}(X)$ we will denote the convex hull of X, by $\mathrm{aff}(X)$ we will denote the affine hull of X. We will proceed with the definition of relative interior of a set $X \subset V$.

Definition 1. Let $X \subset V$, then the relative interior of X, denoted $\mathrm{ri}(X)$ is the interior of X when it is considered as a subset of $\mathrm{aff}(X)$.

For a more thorough discussion of relative interior, see [37, p. 44].

Let K be a compact convex subset of V, then K is a state space. The points $x \in K$ represent the states of some system and their convex combination is interpreted operationally, that is, for $x,y \in K, \lambda \in [0,1] \subset \mathbb{R}$ the state $\lambda x + (1 - \lambda)y$ corresponds to having prepared x with probability λ and y with probability $1 - \lambda$.

To define measurements, we have to be able to assign probabilities to states, that is, we have to have a map $f : K \to [0,1]$ such that, to follow the operational interpretation of convex combination, we have assigned the convex combination of probabilities to the convex combination of respective states. In other words, for $x,y \in K, \lambda \in [0,1]$ we have

$$f(\lambda x + (1 - \lambda)y) = \lambda f(x) + (1 - \lambda)f(y),$$

which means that f is an affine function. Such functions are called effects because they correspond to assigning probabilities of measurement outcomes to states. We will proceed with a more formal definition of effects and of effect algebra.

Let $A(K)$ denote the set of affine functions $K \to \mathbb{R}$. $A(K)$ is itself a real linear space, moreover it is ordered as follows: let $f,g \in A(K)$, then $f \geqslant g$ if $f(x) \geqslant g(x)$ for every $x \in K$. There are two special functions 0 and 1 in $A(K)$, such that $0(x) = 0$ and $1(x) = 1$ for all $x \in K$.

The set $A(K)^+ = \{f \in A(K) : f \geqslant 0\}$ is the convex, closed cone of positive functions. The cone $A(K)^+$ is

generating, that is, for every $f \in A(K)$ we have $f_+, f_- \in A(K)^+$ such that $f = f_+ - f_-$, and it is pointed, that is, if $f \geqslant 0$ and $-f \geqslant 0$, then $f = 0$.

Although we will provide a proper definition of measurement in Sec. III E, we will now introduce the concept of yes or no measurement, or two-outcome measurement, that will motivate the definition of the effect algebra. Our notion of measurement might seem different to the standard understanding and one may argue that what we will refer to as measurements should be called entanglement-breaking maps, but this way of defining measurement is standard in general probabilistic theory, hence, we will use it. A measurement is a procedure that assigns probabilities to possible outcomes based on the state that is measured. If we have only two outcomes and we know that the probability of the first outcome is $p \in [0,1]$, then, by normalization, the probability of the second outcome must be $1 - p$. This shows that a two-outcome measurement needs to assign only probability to one outcome and the other probability follows.

Since assigning probabilities to states is a function $f : K \to [0,1]$ and due to our operational interpretation of convex combination we want such function to be affine. Traditionally, the functions that assign probabilities to states are called effects and the set of all effects is called effect algebra.

Definition 2. The set $E(K) = \{f \in A(K) : 0 \leqslant f \leqslant 1\}$ is called the effect algebra.

In general, one may define effect algebra in more general fashion, using the partially defined operation of addition and a unary operation $\perp$, that would in our case correspond to $f^\perp = 1 - f$ (see [39] for a more thorough treatment).

Let $f \in E(K)$, then the two-outcome measurement m_f corresponding to the effect f is the procedure that for $x \in K$ assigns the probability $f(x)$ to the first outcome and the probability $1 - f(x)$ to the second outcome. Note that we did not mention any labels of the outcomes. Usually, the outcomes are labeled yes and no, or 0 and 1, or -1 and 1, but from an operational perspective this does not matter.

We provide two standard examples of special cases of our definitions.

Example 1 (Classical theory). In classical theory, the state space K is a simplex, that is the convex hull of a set of affinely independent points $x_1, \ldots, x_n$. The special property of the simplex is that every point $x \in K$ can be uniquely expressed as convex combination of the points $x_1, \ldots, x_n$, due to their affine independence.

Example 2 (Quantum theory). Let $\mathcal{H}$ denote a finite-dimensional complex Hilbert space, let $B_h(\mathcal{H})$ denote the real linear space of self-adjoint operators on $\mathcal{H}$, for $A \in B_h(\mathcal{H})$ let $\mathrm{Tr}(A)$ denote the trace of the operator A, and let $A \geqslant 0$ denote that A is positive semidefinite. We say that $A \leqslant B$ if $0 \leqslant B - A$. Let $B_h(\mathcal{H})^+ = \{A \in B_h(\mathcal{H}) : A \geqslant 0\}$ denote the cone of positive-semidefinite operators.

In quantum theory the state space is given as

$$\mathfrak{D}_{\mathcal{H}} = \{\rho \in B_h(\mathcal{H}) : \rho \geqslant 0, \mathrm{Tr}(\rho) = 1\}$$

which is the set of density operators on $\mathcal{H}$. The effect algebra $E(\mathfrak{D}_{\mathcal{H}})$ is given as

$$E(\mathfrak{D}_{\mathcal{H}}) = \{M \in B_h(\mathcal{H}) : 0 \leqslant M \leqslant \mathbb{1}\}.$$

The value of the effect $M \in E(\mathfrak{D}_{\mathcal{H}})$ on the state $\rho \in \mathfrak{D}_{\mathcal{H}}$ is given as

$$M(\rho) = \mathrm{Tr}(\rho M).$$

B. Structure of general probabilistic theory

This section will be rather technical, but we will introduce several mathematical results that we will use later on.

Let $x \in K$ and consider the map $\overline{x} : A(K) \to \mathbb{R}$, that to $f \in A(K)$ assigns the value $f(x)$. This is clearly a linear functional on $A(K)$. Moreover for $x, y \in K, \lambda \in [0,1]$ we have

$$\overline{\lambda x + (1 - \lambda)y} = \lambda \overline{x} + (1 - \lambda)\overline{y}$$

as the functions in $A(K)$ are affine by definition. We conclude that the state space K must be affinely isomorphic to some subset of the dual of $A(K)$. Since the aforementioned isomorphism is going to be extremely useful in later calculations, we will describe it in more detail. Let $A(K)^*$ denote the dual of $A(K)$, that is, the space of all linear functionals on $A(K)$. For $\psi \in A(K)^*$ and $f \in A(K)$ we will denote the value the functional ψ reaches on f as $\langle \psi, f \rangle$. The dual cone to $A(K)^+$ is the cone $A(K)^{*+} = \{\psi \in A(K)^* : \langle \psi, f \rangle \geqslant 0, \forall f \in A(K)^+\}$ that gives rise to the ordering on $A(K)^*$ given as follows: let $\psi, \varphi \in A(K)^*$, then $\psi \geqslant \varphi$ if and only if $(\psi - \varphi) \in A(K)^{*+}$, i.e., if $\psi - \varphi \geqslant 0$.

It is straightforward that the state space K is isomorphic to a subset of the cone $A(K)^{*+}$, moreover, it is straightforward to realize that the functionals isomorphic to K must map the function $1 \in A(K)$ to the value 1.

Definition 3. Let $\mathfrak{S}_K = \{\psi \in A(K)^{*+} : \langle \psi, 1 \rangle = 1\}$. We call $\mathfrak{S}_K$ the state space of the effect algebra $E(K)$.

It might be confusing at this point why we call $\mathfrak{S}_K$ a state space, but this will be cleared by the following.

Proposition 1. $\mathfrak{S}_K$ is affinely isomorphic to K.

Proof. It is clear that the map $x \to \overline{x}$ maps K to a convex subset of $\mathfrak{S}_K$. It is easy to show the inclusion of $\mathfrak{S}_K$ in the image of K using Hahn-Banach separation theorem (see [40, Chap. 1, Theorem 4.3] for a proof). ∎

We will omit the isomorphism between K and $\mathfrak{S}_K$, so for any $x, y \in K, \alpha \in \mathbb{R}$ we will write $\alpha x + y$ instead of $\alpha \overline{x} + \overline{y} \in A(K)^*$. Still, one must be careful when omitting this isomorphism because if $0 \in V$ denotes the zero vector and $0 \in K$, then $\overline{0} \in A(K)^*$ is not the zero functional as by construction we have $\langle \overline{0}, 1 \rangle = 1$. We will do our best to avoid such possible problems by choosing appropriate notation.

There are two more results we will heavily rely on:

Proposition 2. $\mathfrak{S}_K$ is a base of $A(K)^{*+}$, that is, for every $\psi \in A(K)^{*+}, \psi \neq 0$, there is a unique $x \in K$ and $\lambda \in \mathbb{R}, \lambda \geqslant 0$, such that $\psi = \lambda x$.

Proof. Let $\psi \in A(K)^{*+}, \psi \neq 0$, then $\langle \psi, 1 \rangle \neq 0$ as if $\langle \psi, 1 \rangle = 0$ and $\psi \geqslant 0$, then $\psi = 0$, because $1 \in ri(A(K)^+)$. Let $\psi' = \frac{1}{\langle \psi, 1 \rangle} \psi$. It is straightforward that $\psi' \in \mathfrak{S}_K$. ∎

Proposition 3. $A(K)^{*+}$ is a generating cone in $A(K)^*$, that is, for every $\psi \in A(K)$ there are $\varphi_+, \varphi_- \in A(K)^{*+}$ such that $\psi = \varphi_+ - \varphi_-$.

Proof. The result follows from the fact that $A(K)^+$ is a pointed cone (see [38, Sec. 2.6.1]). ∎

C. Tensor products of state spaces and effect algebras

Tensor products are a way to describe joint systems of several other systems. There are several approaches to introducing a tensor product in general probabilistic theory. There is a category theory based approach [41] that is a viable way to introduce the tensor products, but we will use a simpler, operational approach. Note that the state space of the joint system will be a compact convex subset of a real, finite-dimensional vector space as it itself must be a state space of some general probabilistic theory. Also, keep in mind that describing a tensor product of state spaces K_A, K_B is equivalent to describing the tensor product of the cones $A(K_A)^{*+}$, $A(K_B)^{*+}$. This is going to be useful as some things are easier to express in terms of the positive cones.

Let V, W be real finite-dimensional vector spaces and let $v \in V$, $w \in W$. $v \otimes w$ will refer to the element of the algebraic tensor product $V \otimes W$ (see, e.g., [42]). We will first describe the minimal and maximal tensor products of state spaces that set bounds on the real state space of the joint system. Note that when describing the joint state space of two state spaces or states of two systems, we will refer to them as bipartite state space or bipartite states.

Let K_A, K_B denote two state spaces of Alice and Bob, respectively. For every $x_A \in K_A$, $x_B \in K_B$ there must be a state of the joint system describing the situation that Alice's system is in the state x_A and Bob's system is in the state x_B. We will denote such state $x_A \otimes x_B$ and we will call it a product state. Since the state space must be convex, the state space of the joint system must contain at least the convex hull of the product states. This leads to the definition of minimal tensor product.

Definition 4. The minimal tensor product of state spaces K_A and K_B, denoted $K_A \dot{\otimes} K_B$, is the compact convex set

$$K_A \dot{\otimes} K_B = \mathrm{conv}(\{x_A \otimes x_B : x_A \in K_A, x_B \in K_B\}).$$

The bipartite states $y \in K_A \dot{\otimes} K_B$ are also called separable states. For the positive cones we get

$$A(K_A \dot{\otimes} K_B)^{*+}$$
$$= \mathrm{conv}(\{\psi_A \otimes \psi_B : \psi_A \in A(K_A)^{*+}, \psi_B \in A(K_B)^{*+}\}).$$

Example 3. In quantum theory, the minimal tensor product $\mathfrak{D}_{\mathcal{H}} \dot{\otimes} \mathfrak{D}_{\mathcal{H}}$ is the set of all separable states, that is, of all states of the form $\sum_{i=1}^{n} \lambda_i \rho_i \otimes \sigma_i$ for $n \in \mathbb{N}$ and $\rho_i \in \mathfrak{D}_{\mathcal{H}}$, $\sigma_i \in \mathfrak{D}_{\mathcal{H}}$, $0 \leqslant \lambda_i$ for $i \in \{1, \ldots, n\}$, $\sum_{i=1}^{n} \lambda_i = 1$.

In a similar fashion, let $f_A \in E(K_A)$, $f_B \in E(K_B)$, then we can define a function $f_A \otimes f_B$ as the unique affine function such that for $x_A \in K_A$, $x_B \in K_B$ we have

$$(f_A \otimes f_B)(x_A \otimes x_B) = f_A(x_A) f_B(x_B).$$

This function is used in the most simple measurement on the joint system, such that Alice applies the two-outcome measurement m_{f_A} and Bob applies the two-outcome measurement m_{f_B}, so $f_A \otimes f_B$ must be an effect on the joint state space. This leads to the definition of the maximal tensor product.

Definition 5. The maximal tensor product of the state spaces K_A and K_B, denoted $K_A \hat{\otimes} K_B$, is defined as

$$K_A \hat{\otimes} K_B = \{\psi \in A(K_A)^* \otimes A(K_B)^* : \forall f_A \in A(K_A)^+,$$
$$\forall f_B \in A(K_B)^+, \langle \psi, f_A \otimes f_B \rangle \geqslant 0\}.$$

States in $K_A \hat{\otimes} K_B \setminus K_A \dot{\otimes} K_B$ are called entangled states. Equivalent definition, in terms of the positive cones, would be

$$A(K_A \hat{\otimes} K_B)^{*+} = (A(K_A)^+ \dot{\otimes} A(K_B)^+)^{*+},$$

where

$$A(K_A)^+ \dot{\otimes} A(K_B)^+$$
$$= \mathrm{conv}(\{f_A \otimes f_B : f_A \in A(K_A)^+, f_B \in A(K_B)^+\}).$$

As we see, the definition of tensor product of cones of positive functionals goes hand in hand with the definition of tensor product of cones of positive functions.

Example 4. In quantum theory, the maximal tensor product of the cones $B_h(\mathcal{H})^+ \hat{\otimes} B_h(\mathcal{H})^+$ is the cone of entanglement witnesses [43, Sec. 6.3.1], i.e., $W \in B_h(\mathcal{H})^+ \hat{\otimes} B_h(\mathcal{H})^+$ if for every $\rho \in \mathfrak{D}_{\mathcal{H}}$, $\sigma \in \mathfrak{D}_{\mathcal{H}}$ we have $\mathrm{Tr}(W\rho \otimes \sigma) \geqslant 0$. Note that this does not imply the positivity of W.

From the constructions it is clear that the state space of the joint system has to be a subset of the maximal tensor product and it has to contain the minimal tensor product. But, there is no other specification of the state space of the joint system in general; it has to be provided by the theory we are working with.

Definition 6. We will call the joint state space of the systems described by the state spaces K_A and K_B the real tensor product of K_A and K_B and we will denote it $K_A \tilde{\otimes} K_B$. We always have

$$K_A \dot{\otimes} K_B \subseteq K_A \tilde{\otimes} K_B \subseteq K_A \hat{\otimes} K_B.$$

Example 5. In quantum theory, the real tensor product of the state spaces is defined as the set of density matrices on the tensor product of the Hilbert spaces, that is,

$$\mathfrak{D}_{\mathcal{H}} \tilde{\otimes} \mathfrak{D}_{\mathcal{H}} = \mathfrak{D}_{\mathcal{H} \otimes \mathcal{H}}.$$

It is tricky to work with the tensor products in general probabilistic theory as the real tensor product is not always specified, or it may not be clear what it should be. We will always assume that every tensor product we need to be defined is defined. Moreover, when working with a tensor product of more than two state spaces, say K_A, K_B, K_C, we will always assume that

$$(K_A \tilde{\otimes} K_B) \tilde{\otimes} K_C = K_A \tilde{\otimes} (K_B \tilde{\otimes} K_C)$$

and we will simply write $K_A \tilde{\otimes} K_B \tilde{\otimes} K_C$. In the applications of general probabilistic theory to quantum and classical theory it will always be clear how to construct the needed tensor products and we consider this sufficient for us since we are mainly interested in the applications of our results.

We will state and prove a result about classical state spaces that we will use several times later on.

Proposition 4. Let S be a simplex with the extremal points $x_1, \ldots, x_n$, i.e., $S = \mathrm{conv}(\{x_1, \ldots, x_n\})$ and let K be any state space, then we have

$$S \dot{\otimes} K = S \hat{\otimes} K.$$

Proof. Let S be a simplex and let $x_i \in A(S)^{*+}$, $i \in \{1, \ldots, n\}$, be the extreme points of S. The points $x_1, \ldots, x_n$ form a basis of $A(S)^*$. Let $\psi \in S \hat{\otimes} K$, then we have

$$\psi = \sum_{i=1}^{n} x_i \otimes \varphi_i$$

for some $\varphi_i \in A(K)^*$. Our aim is to prove that $\varphi_i \in A(K)^{*+}$, then $\psi \in S\dot{\otimes}K$ follows by definition.

Let $b_1, \ldots, b_n$ denote the basis of $A(S)$ dual to the basis $x_1, \ldots, x_n$ of $A(S)^*$, i.e., we have $b_i(x_j) = \delta_{ij}$, where $i, j \in \{1, \ldots, n\}$ and δ_{ij} is the Kronecker delta. We have $b_i \in E(S)$ because S is a simplex. For any $f \in E(K)$ we have

$$0 \leqslant (\psi, b_i \otimes f) = (\varphi_i, f)$$

for all $i \in \{1, \ldots, n\}$, which implies $\varphi_i \in A(K)^{*+}$. ∎

Note that tensor product of the simplexes S_1, S_2 is also a simplex, so we have

$$K\hat{\otimes}S_1\hat{\otimes}S_2 = K\dot{\otimes}S_1\dot{\otimes}S_2.$$

D. Direct product of state spaces and effect algebras

For certain reasons, we will need to use direct products together with tensor products. The idea of why they will be used is going to be clear in the end, but now we will present several of their properties that will be required later. As in Sec. III C, we will work mostly with the cones of the positive functionals.

Let K_A, K_B be two state spaces. Given $A(K_A)^{*+}$ and $A(K_B)^{*+}$, there are two ways to define the direct product of these cones. The first is to use the cone $A(K_A)^{*+} \times A(K_B)^{*+}$. The second is to realize that we can construct $K_A \times K_B$ that will be a compact and convex set, i.e., a state space that gives rise to the cone $A(K_{B_1} \times K_{B_2})^{*+}$.

It may seem that these cones are fairly similar, but they are not and they have different physical interpretations. Let $\psi \in A(K_A \times K_B)^{*+}$, then there are unique $\lambda \in \mathbb{R}$, $x_A \in K_A$, $x_B \in K_B$ such that $\psi = \lambda(x_A, x_B)$. Now, let $\varphi \in A(K_A)^{*+} \times A(K_B)^{*+}$, then there are $y_A \in K_A$, $y_B \in K_B$, $\alpha_A, \alpha_B \in \mathbb{R}$, $\alpha_A, \alpha_B \geqslant 0$ such that $\varphi = (\alpha_A y_A, \alpha_B y_B)$. In other words, the normalization may be different in every component of the product. This can be rewritten as

$$\varphi = (\alpha_A y_A, \alpha_B y_B)$$
$$= (\alpha_A + \alpha_B)\left(\frac{\alpha_A}{\alpha_A + \alpha_B}y_A, \frac{\alpha_B}{\alpha_A + \alpha_B}y_B\right)$$
$$= (\alpha_A + \alpha_B)\left(\frac{\alpha_A}{\alpha_A + \alpha_B}(y_A, 0) + \frac{\alpha_B}{\alpha_A + \alpha_B}(0, y_B)\right)$$

that shows that every element of $A(K_A)^{*+} \times A(K_B)^{*+}$ can be uniquely expressed as a multiple of a convex combination of elements of the form $(y_A, 0)$ and $(0, y_B)$. The operational interpretation of such states is that we do not even know which system we are working with, but we know that with some probability p we have the first system and with probability $1 - p$ we have the second system.

The operational interpretation of $A(K_A \times K_B)^{*+}$ is a bit harder to grasp. We may understand $\psi \in A(K_A \times K_B)^{*+}$ as a (multiple of) conditional state. That is, we will interpret the object (x_A, x_B) as a state that corresponds to making a choice in the past between the systems K_A and K_B and keeping track of both of the outcomes at once. The cone $A(K_A \times K_B)^{*+}$ will play a central role in our results on incompatibility, steering, and Bell nonlocality because in the problem of incompatibility, we wish to implement two channels at the same time, and in

steering and Bell nonlocality we are choosing between two incompatible channels.

At last, we will need to describe the set $A(K_A \times K_B)$ and its structure with respect to the sets $A(K_A)$ and $A(K_B)$. We will show that $A(K_A \times K_B)$ corresponds to a certain subset of $A(K_A) \times A(K_B)$ by using the following two ideas: since all of the vector spaces are finite dimensional, we have that $A(K_A) \times A(K_B)$ is the dual to $A(K_A)^* \times A(K_B)^*$ and $A(K_A \times K_B)^*$ can be identified with a subset of $A(K_A)^* \times A(K_B)^*$. Note that this identification holds only between the vector spaces and not between the corresponding state spaces.

Proposition 5. We have

$$A\left(K_{B_1} \times K_{B_2}\right)^{*+} \subset A\left(K_{B_1}\right)^{*+} \times A\left(K_{B_2}\right)^{*+}.$$

Proof. The idea of the proof is that if we have $\varphi \in A(K_A)^{*+} \times A(K_B)^{*+}$ such that $\varphi = (\alpha_A y_A, \alpha_B y_B)$, then $\varphi \in A(K_A \times K_B)^{*+}$ if and only if $\alpha_A = \alpha_B$. Therefore, we can identify $A(K_A \times K_B)^{*+}$ with the set $\{\psi \in A(K_A)^{*+} \times A(K_B)^{*+} : \langle \psi, (1, -1)\rangle = 0\}$. It is easy to verify this constraint on the positive cones and since it is linear it must hold everywhere else. ∎

The above proof shows that the function $(1, -1) \in A(K_A) \times A(K_B)$ is equal to zero when restricted to $A(K_A \times K_B)^*$ or, in other words, $(1, 0) = (0, 1)$ when restricted to $A(K_A \times K_B)^*$. We introduce a relation of equivalence on $A(K_A) \times A(K_B)$ as follows: for $f, g \in A(K_A) \times A(K_B)$ we say that f and g are equivalent and we write $f \sim g$ if $f - g = k(1, -1)$ for some $k \in \mathbb{R}$. Equivalently, $f \sim g$ if for every $\psi \in A(K_A \times K_B)^*$ we have $\langle \psi, f\rangle = \langle \psi, g\rangle$. $A(K_A \times K_B)$ corresponds to the set of equivalence classes of $A(K_A) \times A(K_B)$ with respect to the relation of equivalence $\sim$.

To demonstrate this, consider the constant function $1 \in E(K_A \times K_B)$ and let $x \in K_A$, $y \in K_B$, then we have

$$\langle (x, y), (1, 0)\rangle = \langle x, 1\rangle = 1 = \langle (x, y), 1\rangle,$$
$$\langle (x, y), (0, 1)\rangle = \langle y, 1\rangle = 1 = \langle (x, y), 1\rangle.$$

This is not a coincidence because $(1, 0) - (0, 1) = (1, -1)$, so we have $(1, 0) \sim (0, 1)$.

E. Channels and measurements in general probabilistic theory

It is not easy to define channels in general probabilistic theory as we would like all of the channels to be completely positive. We will use the following definition:

Definition 7. Let K_A, K_B be state spaces, then channel Φ is a linear map

$$\Phi : A(K_A)^* \to A(K_B)^*$$

that is positive, i.e., for every $\psi \in A(K_A)^{*+}$ we have $\Phi(\psi) \in A(K_B)^{*+}$ and that for $\psi \in K_A$ we have $\Phi(\psi) \in K_B$.

One may also require a channel to be completely positive, that is, if K_C is some state space such that we can define $K_C\bar{\otimes}K_A$, then we can consider the map $\mathrm{id} \otimes \Phi : K_C\bar{\otimes}K_A \to K_C\hat{\otimes}K_B$ and require it to be positive. In the applications of general probabilistic theory to classical and quantum theories, we always know how to create joint systems of given two systems, so in the examples we will always require complete positivity of channels, but one still has to bear in mind that

in the general case, complete positivity is not a well-defined concept.

One can identify the channel $\Phi : A(K_A)^* \to A(K_B)^*$ with an element of $A(K_A) \otimes A(K_B)^*$ as follows: let $x \in K_A$ and $f \in A(K_B)$, then the expression $\langle \Phi(x), f \rangle$ gives rise to a linear functional on $A(K_A)^* \otimes A(K_B)$. This means that we have $\Phi \in A(K_A) \otimes A(K_B)^*$, where we omit the isomorphism between the channel and the functional. If we also consider the positivity of the channel on the elements of the form $x \otimes f \in K_A \dot{\otimes} E(K_B)$, we get

$$\Phi \in A(K_A)^+ \hat{\otimes} A(K_B)^{*+}.$$

This is a well-known construction that may be also used to define the tensor product of linear spaces [42, Chap. 1.3].

There is one more construction with channels that will be important in our formulation of compatibility of channels: compositions with effect. Let $\Phi : K_A \to K_B$ be a channel and let $f \in E(K_B)$, then they give rise to an effect $(f \circ \Phi) \in E(K_A)$ defined for $x_A \in K_A$ as

$$\langle x_A, (f \circ \Phi) \rangle = \langle \Phi(x_A), f \rangle.$$

By the same idea, we can define a map $f \otimes \mathrm{id} : A(K_B)^* \otimes A(K_C)^* \to A(K_C)^*$ such that for $x_B \in K_B$ and $x_C \in K_C$ we have $(f \otimes \mathrm{id})(x_B \otimes x_C) = f(x_B)x_C$ and we extend the map by linearity. Also, given a channel $\Phi : K_A \to K_B \tilde{\otimes} K_C$ we can compose the map $f \otimes \mathrm{id}$ with the channel Φ to obtain $(f \otimes \mathrm{id}) \circ \Phi' : A(K_A)^* \to A(K_C)^*$ such that the corresponding functional on $A(K_A) \otimes A(K_C)^*$ is for $x_A \in K_A$ and $g \in A(K_C)$ given as

$$\langle (f \otimes \mathrm{id}) \circ \Phi, x_A \otimes g \rangle = \langle \Phi(x_A), f \otimes g \rangle.$$

Specifically, we will be interested in the expressions $(1 \otimes \mathrm{id}) \circ \Phi$ and $(\mathrm{id} \otimes 1) \circ \Phi$. If Φ is a channel, then $(1 \otimes \mathrm{id}) \circ \Phi$ and $(\mathrm{id} \otimes 1) \circ \Phi$ are channels as well and they are called marginal channels of Φ.

A special type of channel is a measurement.

Definition 8. A channel $m : K_A \to K_B$ is called a measurement if K_B is a simplex.

The interpretation is simple: the vertices of the simplex correspond to the possible measurement outcomes and the resulting state is a probability distribution over the measurement outcomes, i.e., an assignment of probabilities to the possible outcomes. Since we require all state spaces to be finite dimensional, this implies that we consider only finite-outcome measurements. Let K_B be a simplex with vertices $\omega_1, \ldots, \omega_n$, then we can identify a measurement m with an element of $A(K_A)^+ \hat{\otimes} A(K_B)^{*+}$ of the form

$$m = \sum_{i=1}^{n} f_i \otimes \delta_{\omega_i},$$

where for $i \in \{1, \ldots, n\}$ we have $f_i \in E(K_A)$, $\sum_{i=1}^{n} f_i = 1$ and $\delta_{\omega_i} \in \mathfrak{S}(K_B)$ are the functionals corresponding to the extreme points of K_B (where we have not omitted the isomorphism this time). This expression has an operational interpretation that for $x \in K_A$ the measurement m assigns the probability $f_i(x)$ to the outcome ω_i.

Example 6. Quantum channels are completely positive, trace-preserving maps $\Phi : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$. The complete positiv-

ity means that for any $\rho \geqslant 0$ we have $(\mathrm{id} \otimes \Phi)(\rho) \geqslant 0$. We denote the set of channels $\Phi : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$ as $\mathfrak{C}_{\mathcal{H} \to \mathcal{H}}$.

Let $|1\rangle, \ldots, |n\rangle$ be an orthonormal base of $\mathcal{H}$. To every quantum channel we may assign its unique Choi matrix $C(\Phi)$ defined as

$$C(\Phi) = (\Phi \otimes \mathrm{id}) \left(\sum_{i,j=1}^{n} |ii\rangle\langle jj| \right),$$

where we use the shorthand $|ii\rangle = |i\rangle \otimes |i\rangle$. Note that $C(\Phi) \geqslant 0$ and $\mathrm{Tr}_1(C(\Phi)) = \mathbb{1}$, where Tr_1 denotes the partial trace. Also, every matrix $C \in B_h(\mathcal{H} \otimes \mathcal{H})$ such that $C \geqslant 0$ and $\mathrm{Tr}_1(C) = \mathbb{1}$ is a Choi matrix of some channel (see [43, Sec. 4.4.3]).

The Choi matrix $C(\Phi)$ is isomorphic to a state $\frac{1}{\dim(\mathcal{H})} C(\Phi)$, which corresponds to the channel $\Phi \otimes \mathrm{id}$ acting on the maximally entangled state $|\psi^+\rangle\langle\psi^+|$, where

$$|\psi^+\rangle = \frac{1}{\sqrt{\dim(\mathcal{H})}} \sum_{i=1}^{n} |ii\rangle.$$

IV. COMPATIBILITY OF CHANNELS

Definition 9. Let K_A, K_{B_1}, K_{B_2} be state spaces and let Φ_1, Φ_2 be channels

$$\Phi_1 : K_A \to K_{B_1},$$
$$\Phi_2 : K_A \to K_{B_2}.$$

We say that Φ_1, Φ_2 are compatible if and only if there exists a channel

$$\Phi : K_A \to K_{B_1} \tilde{\otimes} K_{B_2}$$

such that Φ_1 and Φ_2 are the marginal channels of Φ, i.e., we have

$$\Phi_1 = (\mathrm{id} \otimes 1) \circ \Phi, \tag{1}$$
$$\Phi_2 = (1 \otimes \mathrm{id}) \circ \Phi. \tag{2}$$

The channel Φ is also called the joint channel of the channels Φ_1, Φ_2.

The operational meaning of compatibility of channels is that if the channels Φ_1, Φ_2 are compatible, then we can apply them both to the input state at once and select which one we actually want the output from latter. If the channels are incompatible, we have to choose from which one we want the output before applying anything. For a more in-depth explanation, see [19]. The important thing is that there is a choice from which channel we want to get the output so we can expect to see $A(K_{B_1} \times K_{B_2})^{*+}$ come up in the calculations.

Consider the channel $\Phi : K_A \to K_{B_1} \tilde{\otimes} K_{B_2}$. One can realize that the maps $(\mathrm{id} \otimes 1) : \Phi \mapsto (\mathrm{id} \otimes 1) \circ \Phi$ and $(1 \otimes \mathrm{id}) : \Phi \mapsto (1 \otimes \mathrm{id}) \circ \Phi$ are linear maps of channels. Moreover, Eqs. (1) and (2) both have Φ on the right-hand side in the same position. We are going to exploit this to obtain a simpler condition for compatibility of the channels Φ_1, Φ_2. To do so, we have to introduce a new map J.

Let us define a map $J : A(K_A) \otimes A(K_{B_1})^* \otimes A(K_{B_2})^* \to A(K_A) \otimes A(K_{B_1} \times K_{B_2})^*$ given for $\Xi \in A(K_A) \otimes A(K_{B_1})^* \otimes A(K_{B_2})^*$ as

$$J(\Xi) = ((\mathrm{id} \otimes 1) \circ \Xi, (1 \otimes \mathrm{id}) \circ \Xi).$$

For $\Xi = f \otimes \psi \otimes \varphi$, we have

$$J(\Xi) = f \otimes (\langle \varphi, 1 \rangle \psi, \langle \psi, 1 \rangle \varphi).$$

Proposition 6. J is a linear mapping.

Proof. Let $\Xi_1, \Xi_2 \in A(K_A) \otimes A(K_{B_1} \otimes K_{B_2})^*$ and $\lambda \in \mathbb{R}$, then we have

$$\begin{aligned}
J(\lambda \Xi_1 + \Xi_2) &= (\lambda(\mathrm{id} \otimes 1) \circ \Xi_1 + (\mathrm{id} \otimes 1) \circ \Xi_2, 0) \\
&\quad + (0, \lambda(1 \otimes \mathrm{id}) \circ \Xi_1 + (1 \otimes \mathrm{id}) \circ \Xi_2) \\
&= \lambda((\mathrm{id} \otimes 1) \circ \Xi_1, (1 \otimes \mathrm{id}) \circ \Xi_1) \\
&\quad + ((\mathrm{id} \otimes 1) \circ \Xi_2, (1 \otimes \mathrm{id}) \circ \Xi_2) \\
&= \lambda J(\Xi_1) + J(\Xi_2).
\end{aligned}$$
∎

Assume that the channels Φ_1, Φ_2 are compatible and that Φ is their joint channel, then we must have

$$J(\Phi) = (\Phi_1, \Phi_2)$$

which is just a more compact form of Eqs. (1) and (2).

Proposition 7. The channels Φ_1, Φ_2 are compatible if and only if there is $\Phi \in A(K_A)^+ \hat{\otimes} A(K_{B_1} \tilde{\otimes} K_{B_2})^{*+}$ such that

$$J(\Phi) = (\Phi_1, \Phi_2). \tag{3}$$

Proof. If the channels Φ_1, Φ_2 are compatible, then Eq. (3) must hold for their joint channel Φ. If Eq. (3) holds for some $\Phi \in A(K_A)^+ \hat{\otimes} A(K_{B_1} \tilde{\otimes} K_{B_2})^{*+}$, then the channels Φ_1, Φ_2 are compatible and Φ is their joint channel. ∎

The operational interpretation is that (Φ_1, Φ_2) represents a conditional channel in the same way as the states from $A(K_{B_1} \times K_{B_2})^{*+}$ represent conditional states that keep track of some choice made in the past. If the channels are compatible, then we actually do not have to make the choice of either using Φ_1 or Φ_2, but we can use their joint channel, that has the property that its marginals reproduce the outcomes of the two channels Φ_1, Φ_2. We will investigate several of the properties of the map J.

Proposition 8. For every $(\xi_1, \xi_2) \in A(K_A) \otimes A(K_{B_1} \times K_{B_2})^*$ there is a $\Xi \in A(K_A) \otimes A(K_{B_1})^* \otimes A(K_{B_2})^*$ such that

$$J(\Xi) = (\xi_1, \xi_2).$$

Moreover, if we have

$$(1, 1) \circ (\xi_1, \xi_2) = 1,$$

then

$$(1 \otimes 1) \circ \Xi = 1.$$

Proof. Let $f_1, \ldots, f_n$ be a basis of $A(K_A)$, then we have

$$\xi_1 = \sum_{i=1}^{n} f_i \otimes \psi_i, \quad \xi_2 = \sum_{i=1}^{n} f_i \otimes \varphi_i$$

for some $\psi_i \in A(K_{B_1})^*$ and $\varphi_i \in A(K_{B_2})^*$. Since we must have

$$(1, 0) \circ (\xi_1, \xi_2) = (0, 1) \circ (\xi_1, \xi_2),$$

we obtain

$$\sum_{i=1}^{n} \langle \psi_i, 1 \rangle f_i = \sum_{i=1}^{n} \langle \varphi_i, 1 \rangle f_i,$$

which implies

$$\langle \psi_i, 1 \rangle = \langle \varphi_i, 1 \rangle = k_i$$

for all $i \in \{1, \ldots, n\}$ as $f_1, \ldots, f_n$ is linearly independent. Let

$$\Xi = \sum_{i=1}^{n} k_i^{-1} f_i \otimes \psi_i \otimes \varphi_i,$$

then we have

$$\begin{aligned}
J(\Xi) &= \sum_{i=1}^{n} k_i^{-1} f_i \otimes \left(\langle \varphi_i, 1_{B_2} \rangle \psi_i, \langle \psi_i, 1_{B_1} \rangle \varphi_i \right) \\
&= \sum_{i=1}^{n} f_i \otimes (\psi_i, \varphi_i).
\end{aligned}$$

If we have $1 \circ (\xi_1, \xi_2) = 1$, then

$$\sum_{i=1}^{n} k_i f_i = 1$$

and we get

$$\begin{aligned}
(1 \otimes 1) \circ \Xi &= (1 \otimes 1) \circ \left(\sum_{i=1}^{n} k_i^{-1} f_i \otimes \psi_i \otimes \varphi_i \right) \\
&= \sum_{i=1}^{n} k_i^{-1} \langle \psi_i, 1 \rangle \langle \varphi_i, 1 \rangle f_i = 1.
\end{aligned}$$
∎

Proposition 9. We have

$$\begin{aligned}
J \left(A(K_A)^+ \dot{\otimes} A\left(K_{B_1} \right)^{*+} \dot{\otimes} A\left(K_{B_2} \right)^{*+} \right) \\
= A(K_A)^+ \dot{\otimes} A\left(K_1 \times K_{B_2} \right)^{*+}.
\end{aligned}$$

Proof. Let $(\xi_1, \xi_2) \in A(K_A)^+ \dot{\otimes} A(K_{B_1} \times K_{B_2})^{*+}$, then as in the proof of Proposition 8 we have

$$\xi_1 = \sum_{i=1}^{n} f_i \otimes \psi_i, \quad \xi_2 = \sum_{i=1}^{n} f_i \otimes \varphi_i,$$

but now we have $f_i \geqslant 0$, $\psi_i \geqslant 0$, and $\varphi_i \geqslant 0$ for $i \in \{1, \ldots, n\}$. It follows by the same construction as in the proof of Proposition 8 that we can construct $\Xi = \sum_{i=1}^{n} k_i^{-1} f_i \otimes \psi_i \otimes \varphi_i$ and we get $\Xi \in A(K_A)^+ \dot{\otimes} A(K_{B_1})^{*+} \dot{\otimes} A(K_{B_2})^{*+}$.

Let $\Xi \in A(K_A)^+ \dot{\otimes} A(K_{B_1})^{*+} \dot{\otimes} A(K_{B_2})^{*+}$, then we have $\Xi = \sum_{i=1}^{n} f_i \otimes \psi_i \otimes \varphi_i$ such that $f_i \geqslant 0$, $\psi_i \geqslant 0$, $\varphi_i \geqslant 0$ for all $i \in \{1, \ldots, n\}$, moreover, without lack of generality we can assume $\langle \psi_i, 1_{B_1} \rangle = \langle \varphi_i, 1_{B_2} \rangle = 1$. We have

$$J(\Xi) = \sum_{i=1}^{n} f_i \otimes (\psi_i, \varphi_i) \in A(K_A)^+ \dot{\otimes} A\left(K_{B_1} \times K_{B_2} \right)^{*+},$$

which concludes the proof. ∎

It would be very useful to know what is the image of the cone $A(K_A)^+ \hat{\otimes} A(K_{B_1} \tilde{\otimes} K_{B_2})^{*+}$ when mapped by J. We will denote the resulting cone $Q = J(A(K_A)^+ \hat{\otimes} A(K_{B_1} \tilde{\otimes} K_{B_2})^{*+})$. The cone is important due to the following:

Corollary 1. The channels Φ_1, Φ_2 are compatible if and only if

$$(\Phi_1, \Phi_2) \in Q = J\left(A(K_A)^+ \hat{\otimes} A\left(K_{B_1} \tilde{\otimes} K_{B_2} \right)^{*+} \right).$$

Proof. Follows from Proposition 7.   ∎

Proposition 10. $A(K_A)^+ \dot\otimes A(K_{B_1} \times K_{B_2})^{*+} \subset Q$.

Proof. Since

$$A(K_A)^+ \hat\otimes A\big(K_{B_1} \dot\otimes K_{B_2}\big)^{*+} \subset A(K_A)^+ \hat\otimes A\big(K_{B_1} \tilde\otimes K_{B_2}\big)^{*+},$$

we must have

$$J\big(A(K_A)^+ \hat\otimes A\big(K_{B_1} \dot\otimes K_{B_2}\big)^{*+}\big) \subset Q.$$

The result follows from Proposition 9.   ∎

Proposition 11. $Q \subset A(K_A)^+ \hat\otimes A(K_{B_1} \times K_{B_2})^{*+}$.

Proof. Since we have

$$A(K_A)^+ \hat\otimes A\big(K_{B_1} \tilde\otimes K_{B_2}\big)^{*+} \subset A(K_A)^+ \hat\otimes A\big(K_{B_1} \hat\otimes K_{B_2}\big)^{*+},$$

we must have

$$Q \subset J\big(A(K_A)^+ \hat\otimes A\big(K_{B_1} \hat\otimes K_{B_2}\big)^{*+}\big).$$

Let $\Xi \in A(K_A)^+ \hat\otimes A(K_{B_1} \hat\otimes K_{B_2})^{*+}$, then for $\psi \in A(K_A)^{*+}$ and $(f_1, f_2) \in A(K_{B_1} \times K_{B_2})^+$ we get

$$\langle J(\Xi), x \otimes f \rangle = \langle ((\text{id} \otimes 1) \circ \Xi, (1 \otimes \text{id}) \circ \Xi), x \otimes f \rangle$$
$$= \langle \Xi(x), f_1 \otimes 1 \rangle + \langle \Xi(x), 1 \otimes f_2 \rangle \geqslant 0,$$

that shows we have $J(A(K_A)^+ \hat\otimes A(K_{B_1} \hat\otimes K_{B_2})^{*+}) \subset A(K_A)^+ \hat\otimes A(K_{B_1} \times K_{B_2})^{*+}$, which concludes the proof. ∎

We can also construct Q as the cone we get when we factorize the cone $A(K_A)^+ \hat\otimes A(K_{B_1} \tilde\otimes K_{B_2})^{*+}$ with respect to the relation of equivalence given as follows: $\Xi_1 \approx \Xi_2$ if and only if $J(\Xi_1) = J(\Xi_2)$ or, equivalently, if and only if $\Xi_1 = \Xi_2 + \Xi$, such that $J(\Xi) = 0$.

Note that since J is a linear map, as we showed in Proposition 6, it is clear that Q is a convex cone. For two given channels $\Phi_1 : K_A \to K_{B_1}$, $\Phi_2 : K_A \to K_{B_1}$ one may write a primal linear program that would check the condition for compatibility given by Corollary 1. We will write such linear program for quantum channels later.

V. COMPATIBILITY OF MEASUREMENTS

We will apply the results of Sec. IV to the problem of compatibility of measurements. We will obtain the same results that were recently presented in Ref. [26], that are a generalization of [44].

Let K_A be a state space and let S_1, S_2 be simplexes and let $m_1 : K_A \to S_1$, $m_2 : K_A \to S_2$ be measurements. According to Proposition 7, the measurements m_1, m_2 are compatible if and only if

$$(m_1, m_2) \in J(A(K_A)^+ \hat\otimes A(S_1 \tilde\otimes S_2)^{*+}).$$

Since both S_1 and S_2 are simplexes, then we have $S_1 \tilde\otimes S_2 = S_1 \dot\otimes S_2$ and the condition for compatibility reduces according to Proposition 9 to

$$(m_1, m_2) \in A(K_A)^+ \dot\otimes A(S_1 \times S_2)^{*+}.$$

Due to the simpler structure of simplexes, one may get even more specific results about measurements (see [26]).

For demonstration of the derived conditions we will reconstruct the result of [44] about compatibility of two-outcome measurements. According to our definition, a measurement is two-outcome if the simplex it has as a target space has two vertexes, i.e., it is a line segment. Let K be a state space,

$f, g \in E(K)$ and $m_f : K \to S$, $m_g : K \to S$ be two-outcome measurements given as

$$m_f = f \otimes \delta_{\omega_1} + (1 - f) \otimes \delta_{\omega_2},$$
$$m_g = g \otimes \delta_{\omega_1} + (1 - g) \otimes \delta_{\omega_2}.$$

The state space given by $A(S \times S)^{*+}$ is a square given as $\text{conv}((\delta_{\omega_1}, \delta_{\omega_1}), (\delta_{\omega_1}, \delta_{\omega_2}), (\delta_{\omega_2}, \delta_{\omega_1}), (\delta_{\omega_2}, \delta_{\omega_2}))$, that is just affinely isomorphic to $S \times S$. We have

$$(m_1, m_2) = f \otimes \big(\delta_{\omega_1}, 0\big) + (1 - f) \otimes \big(\delta_{\omega_2}, 0\big)$$
$$+ g \otimes \big(0, \delta_{\omega_1}\big) + (1 - g) \otimes \big(0, \delta_{\omega_2}\big)$$
$$= f \otimes \big(\delta_{\omega_1}, \delta_{\omega_2}\big) + (1 - f) \otimes \big(\delta_{\omega_2}, \delta_{\omega_2}\big)$$
$$+ g \otimes \big(0, \delta_{\omega_1} - \delta_{\omega_2}\big),$$

where in the second step we have used the basis $(\delta_{\omega_1}, \delta_{\omega_2})$, $(\delta_{\omega_2}, \delta_{\omega_2})$, $(0, \delta_{\omega_1} - \delta_{\omega_2})$ of $A(S \times S)^*$ to express (m_1, m_2) in a more reasonable form. To have $(m_1, m_2) \in A(K)^+ \hat\otimes A(S \times S)^{*+}$ we must have

$$(m_1, m_2) = h_{11} \otimes \big(\delta_{\omega_1}, \delta_{\omega_1}\big) + h_{12} \otimes \big(\delta_{\omega_1}, \delta_{\omega_2}\big)$$
$$+ h_{21} \otimes \big(\delta_{\omega_2}, \delta_{\omega_1}\big) + h_{22} \otimes \big(\delta_{\omega_2}, \delta_{\omega_2}\big)$$
$$= (h_{11} + h_{12}) \otimes \big(\delta_{\omega_1}, \delta_{\omega_2}\big)$$
$$+ (h_{21} + h_{22}) \otimes \big(\delta_{\omega_2}, \delta_{\omega_2}\big)$$
$$+ (h_{11} + h_{21}) \otimes \big(0, \delta_{\omega_1} - \delta_{\omega_2}\big)$$

for some $h_{11}, h_{12}, h_{21}, h_{22} \in E(K)$. This implies the standard conditions for the compatibility of two-outcome measurements m_f, m_g:

$$f = h_{11} + h_{12},$$
$$1 - f = h_{21} + h_{22},$$
$$g = h_{11} + h_{21}$$

(see, e.g., [45]).

VI. COMPATIBILITY OF QUANTUM CHANNELS

In this section, we will derive results more specific to the compatibility of quantum channels. Let $\Phi_1 : \mathfrak{D}_\mathcal{H} \to \mathfrak{D}_\mathcal{H}$, $\Phi_2 : \mathfrak{D}_\mathcal{H} \to \mathfrak{D}_\mathcal{H}$ be quantum channels, then, according to Proposition 7 they are compatible if and only if there is a channel $\Phi : \mathfrak{D}_\mathcal{H} \to \mathfrak{D}_{\mathcal{H} \otimes \mathcal{H}}$ such that for all $\rho \in \mathfrak{D}_\mathcal{H}$ we have

$$(\Phi_1(\rho), \Phi_2(\rho)) = (\text{Tr}_2(\Phi(\rho)), \text{Tr}_1(\Phi(\rho))). \quad (4)$$

This is equivalent to the definition of compatibility of quantum channels already stated in Ref. [21]. It is straightforward that Eq. (4) implies that

$$(C(\Phi_1), C(\Phi_2)) = (\text{Tr}_2(C(\Phi)), \text{Tr}_1(C(\Phi))),$$

we will show that they are equivalent. This will help us to get rid of the state ρ in Eq. (4).

Proposition 12. The channels $\Phi_1 : \mathfrak{D}_\mathcal{H} \to \mathfrak{D}_\mathcal{H}$, $\Phi_2 : \mathfrak{D}_\mathcal{H} \to \mathfrak{D}_\mathcal{H}$ are compatible if and only if there exists a channel $\Phi : \mathfrak{D}_\mathcal{H} \to \mathfrak{D}_{\mathcal{H} \otimes \mathcal{H}}$ such that

$$(C(\Phi_1), C(\Phi_2)) = (\text{Tr}_2(C(\Phi)), \text{Tr}_1(C(\Phi))).$$

Proof. Let $\rho \in \mathfrak{D}_{\mathcal{H}}$, then we have

$$\mathrm{Tr}_2(\Phi(\rho)) = \mathrm{Tr}_{2,E}(C(\Phi)\mathbb{1} \otimes \mathbb{1} \otimes \rho^T)$$
$$= \mathrm{Tr}_E(\mathrm{Tr}_2(C(\Phi))\mathbb{1} \otimes \rho^T)$$
$$= \mathrm{Tr}_E(C(\Phi_1)\mathbb{1} \otimes \rho^T) = \Phi_1(\rho).$$

The same follows for Φ_2. ∎

As we already showed in Sec. IV, the cone $Q = J(A(\mathfrak{D}_{\mathcal{H}})^+ \hat{\otimes} A(\mathfrak{D}_{\mathcal{H}\otimes\mathcal{H}})^{*+})$ is of interest for the compatibility of channels. In the case of quantum channels we will use Proposition 12 to formulate similar cone in terms of Choi matrices of the channels and we will write a semidefinite program for the compatibility of quantum channels based on this approach.

Denote $P = \{(\mathrm{Tr}_2(C), \mathrm{Tr}_1(C)) : C \in \mathfrak{C}_{\mathcal{H}\to\mathcal{H}\otimes\mathcal{H}}\}$, then according to Proposition 12 the channels $\Phi_1 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$, $\Phi_2 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$ are compatible if and only if

$$(C(\Phi_1), C(\Phi_2)) \in P.$$

Note that, by our definition, P is not a cone, but it generates some cone just by adding all of the operators of the form λC, where $C \in P$ and $\lambda \in \mathbb{R}, \lambda \geqslant 0$.

It would be very interesting to obtain more specific results on the structure of P, but the task is not trivial. To make it simpler, we will investigate the structure of the dual cone P^* given as

$$P^* = \{(A,B) \in B_h(\mathcal{H}) \times B_h(\mathcal{H}) :$$
$$\langle C, (A,B) \rangle \geqslant 0, \ \forall \ C \in P\}.$$

Notice that $(A,B) \in B_h(\mathcal{H}) \times B_h(\mathcal{H})$ is simply a block-diagonal matrix having blocks A and B. Also, every $C \in P$ is a block-diagonal matrix, let $C = (C_1, C_2)$, then

$$\langle (C_1,C_2), (A,B) \rangle = \mathrm{Tr}(C_1 A) + \mathrm{Tr}(C_2 B).$$

Let $C \in P$, then by definition there exists a channel $\Phi : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}\otimes\mathcal{H}}$ such that

$$C = (\mathrm{Tr}_2(C(\Phi)), \mathrm{Tr}_1(C(\Phi))).$$

Let $(A,B) \in P^*$, then we have

$$\langle C, (A,B) \rangle = \mathrm{Tr}(\mathrm{Tr}_2(C(\Phi))A + \mathrm{Tr}_1(C(\Phi))B)$$
$$= \mathrm{Tr}(C(\Phi)(\tilde{A} + \mathbb{1} \otimes B)) \geqslant 0,$$

where $\tilde{A}$ is the operator such that $\mathrm{Tr}(\mathrm{Tr}_2(C(\Phi))A) = \mathrm{Tr}(C(\Phi)\tilde{A})$. If $A = A_1 \otimes A_2$, then $\tilde{A} = A_1 \otimes \mathbb{1} \otimes A_2$. In general, one can write A as a sum of factorized operators and express $\tilde{A}$ in such way because the map $A \mapsto \tilde{A}$ is linear.

The result is that $\tilde{A} + \mathbb{1} \otimes B$ must correspond to a positive function on quantum channels, hence, we must have $\tilde{A} + \mathbb{1} \otimes B \geqslant 0$ (see [46,47]). We have proved the following:

Proposition 13. The channels $\Phi_1 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$, $\Phi_2 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$ are compatible if and only if

$$\mathrm{Tr}(C(\Phi_1)A) + \mathrm{Tr}(C(\Phi_2)B) \geqslant 0$$

for all $A, B \in B_h(\mathcal{H} \otimes \mathcal{H})$ such that

$$\tilde{A} + \mathbb{1} \otimes B \geqslant 0.$$

This allows us to formulate the semidefinite program [38] for the compatibility of quantum channels as follows:

Proposition 14. Given channels $\Phi_1 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$, $\Phi_2 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$, the semidefinite program for the compatibility of quantum channels is

$$\inf_{A,B} \mathrm{Tr}(C(\Phi_1)A) + \mathrm{Tr}(C(\Phi_2)B), \quad \tilde{A} + \mathbb{1} \otimes B \geqslant 0,$$

where $\tilde{A}$ is given as above.

If the reached infimum is negative, then the channels are incompatible; if the reached infimum is 0, then the channels are compatible.

Proof. The result follows from Proposition 13. One may see that the infimum is at most 0 because one may always chose $A = B = 0$. ∎

VII. PRELUDE TO STEERING AND BELL NONLOCALITY

We will propose a possible test for the compatibility of channels that will not work, but it will motivate our definitions of steering and Bell nonlocality.

Let K_A, K_{B_1}, K_{B_2} be state spaces and let $\Phi_1 : K_A \to K_{B_1}$, $\Phi_2 : K_A \to K_{B_2}$ be channels. The channels Φ_1, Φ_2 are compatible if Eq. (3) is satisfied for some channel $\Phi : K_A \to K_{B_1} \tilde{\otimes} K_{B_2}$. This is the same as saying the channels Φ_1, Φ_2 are compatible if for all $x \in K_A$ we have

$$(\Phi_1(x), \Phi_2(x)) = (((\mathrm{id} \otimes 1) \circ \Phi)(x), ((1 \otimes \mathrm{id}) \circ \Phi)(x)). \quad (5)$$

If the channels Φ_1 and Φ_2 are compatible, then for every $x \in K_A$ there must exist a state $y \in K_{B_1} \tilde{\otimes} K_{B_2}$ such that

$$\Phi_1(x) = (\mathrm{id} \otimes 1)(y), \quad (6)$$
$$\Phi_2(x) = (1 \otimes \mathrm{id})(y). \quad (7)$$

Would it be a reasonable test for the compatibility of the channels Φ_1 and Φ_2 if we considered the state $x \in K_A$ fixed and we would test whether, for the fixed state x, there exists $y \in K_{B_1} \tilde{\otimes} K_{B_2}$ such that Eqs. (6) and (7) are satisfied? It would not, because for a fixed $x \in K_A$ one always has $\Phi_1(x) \otimes \Phi_2(x) \in K_{B_1} \tilde{\otimes} K_{B_2}$ that satisfies Eqs. (6) and (7).

Still, throwing away this line of thinking would not be a good choice because going further, one may ask the following: If there would be another system K_C, such that $K_C \tilde{\otimes} K_A$ is defined, then what if we would use the entanglement between the systems K_A and K_C to obtain a better condition for the compatibility of the channels Φ_1, Φ_2 using the very same line of thinking? As we will see, this approach leads to the notions of steering and Bell nonlocality.

VIII. STEERING

Steering is one of the puzzling phenomena we find in quantum theory but not in classical theory. It is usually described as a two-party protocol, that allows one side to alter the state of the other in a way that would not be possible in classical theory by performing a measurement and announcing the outcome. Although originally discovered by Schrödinger [1], steering was formalized in Ref. [27]. Recently, there was introduced a new formalism for steering in Ref. [26].

So far, it was always only considered that during steering one party performs a measurement. Since a measurement is a special case of a channel, one may ask whether it is possible

to define steering by channels. We will use our formalism for compatibility of channels to introduce steering by channels by continuing the line of thoughts presented in Sec. VII. We will have to formulate steering in a little different way than it usually is formulated for measurements, but we will show that for measurements we will obtain the known results.

Let $K_A, K_{B_1}, K_{B_2}, K_C$ be finite-dimensional state spaces, such that $K_C \tilde{\otimes} K_A$ is defined and let

$$\Phi_1 : K_A \to K_{B_1},$$

$$\Phi_2 : K_A \to K_{B_2}$$

be channels. We can construct channels

$$\text{id} \otimes \Phi_1 : A(K_C)^{*+} \tilde{\otimes} A(K_A)^{*+} \to A(K_C)^{*+} \hat{\otimes} A(K_{B_1})^{*+},$$

$$\text{id} \otimes \Phi_2 : A(K_C)^{*+} \tilde{\otimes} A(K_A)^{*+} \to A(K_C)^{*+} \hat{\otimes} A(K_{B_2})^{*+}.$$

Moreover, we can construct the conditional channel

$$\text{id} \otimes (\Phi_1, \Phi_2) : A(K_C)^{*+} \tilde{\otimes} A(K_A)^{*+}$$
$$\to A(K_C)^{*+} \hat{\otimes} A(K_{B_1} \times K_{B_2})^{*+}.$$

These channels play a central role in steering and we will keep this notation throughout this section. First, we will introduce a handy name for the output state of $\text{id} \otimes (\Phi_1, \Phi_2)$.

Definition 10. Let $\psi \in K_C \tilde{\otimes} K_A$ be a bipartite state, then we call $(\text{id} \otimes (\Phi_1, \Phi_2))(\psi)$ a bipartite conditional state.

Steering may be seen as a three-party protocol that tests the compatibility of channels. The parties in question will be named Alice, Bob, and Charlie. Alice and Charlie share a bipartite state $\psi \in K_C \tilde{\otimes} K_A$ and Alice has the channels Φ_1 and Φ_2 at her disposal, that would send her part of the state ψ to Bob. Since Alice can choose between the channels Φ_1 and Φ_2, she will be, in our formalism, applying the conditional channel (Φ_1, Φ_2) and the resulting state will be a bipartite state from $A(K_C)^{*+} \tilde{\otimes} A(K_{B_1} \times K_{B_2})^{*+}$. The structure of the resulting bipartite conditional state $(\text{id} \otimes (\Phi_1, \Phi_2))(\psi)$ will not only depend on the input state ψ, but also on the compatibility of the channels Φ_1 and Φ_2. Let us assume that the channels Φ_1 and Φ_2 are compatible, then there is a channel $\Phi : K_A \to K_{B_1} \tilde{\otimes} K_{B_2}$ such that $(\Phi_1, \Phi_2) = J(\Phi)$ and we have

$$(id \otimes (\Phi_1, \Phi_2))(\psi) = (\text{id} \otimes J(\Phi))(\psi)$$
$$= (\text{id} \otimes J')((\text{id} \otimes \Phi)(\psi)),$$

where $J' : A(K_{B_1} \tilde{\otimes} K_{B_2})^* \to A(K_{B_1} \times K_{B_2})^*$, $J'(\psi) = ((id \otimes 1)(\psi), (1 \otimes id)(\psi))$. The calculation shows that if the channels Φ_1, Φ_2 are compatible, then we must have

$$(\text{id} \otimes (\Phi_1, \Phi_2))(\psi) \in (\text{id} \otimes J')(K_C \tilde{\otimes} K_{B_1} \tilde{\otimes} K_{B_2})$$

which does not have to hold in general if the channels are not compatible. This shows that we can define steering of a state by channels as an entanglement-assisted incompatibility test.

Definition 11. The bipartite state $\psi \in A(K_C)^{*+} \tilde{\otimes} A(K_A)^{*+}$ is steerable by channels $\Phi_1 : A(K_A)^{*+} \to A(K_{B_1})^{*+}$, $\Phi_2 : A(K_A)^{*+} \to A(K_{B_2})^{*+}$ if

$$(\text{id} \otimes (\Phi_1, \Phi_2))(\psi) \notin (\text{id} \otimes J')(K_C \tilde{\otimes} K_{B_1} \tilde{\otimes} K_{B_2}).$$

Now, we present the standard result about the connection between compatibility of the channels and steering. The result follows from our definition immediately.

Corollary 2. The bipartite state $\psi \in A(K_C)^{*+} \tilde{\otimes} A(K_A)^{*+}$ is not steerable by channels $\Phi_1 : A(K_A)^{*+} \to A(K_B)^{*+}$, $\Phi_2 : A(K_A)^{*+} \to A(K_B)^{*+}$ if the channels Φ_1 and Φ_2 are compatible.

Proof. If the channels Φ_1, Φ_2 are compatible, then we have $(\Phi_1, \Phi_2) = J(\Phi)$ for some $\Phi : K_A \to K_{B_1} \tilde{\otimes} K_{B_2}$ and for every $\psi \in K_C \tilde{\otimes} K_A$ we have

$$(\text{id} \otimes (\Phi_1, \Phi_2))(\psi) \in (\text{id} \otimes J')(K_C \tilde{\otimes} (K_{B_1} \tilde{\otimes} K_{B_2})).$$

∎

Proposition 15. The bipartite state $\psi \in A(K_C)^{*+} \tilde{\otimes} A(K_A)^{*+}$ is not steerable by channels $\Phi_1 : A(K_A)^{*+} \to A(K_B)^{*+}$, $\Phi_2 : A(K_A)^{*+} \to A(K_B)^{*+}$ if $\psi \in A(K_C)^{*+} \dot{\otimes} A(K_A)^{*+}$, i.e., if ψ is separable.

Proof. Every separable state is by definition a convex combination of product states, i.e., of states of the form $x_C \otimes x_A$, where $x_A \in K_A$, $x_C \in K_C$. Since the maps $\text{id} \otimes (\Phi_1, \Phi_2)$ and $\text{id} \otimes J'$ are linear, it is sufficient to prove that for every product state $x_C \otimes x_A \in K_C \dot{\otimes} K_A$ we have $(\text{id} \otimes (\Phi_1, \Phi_2))(x_C \otimes x_A) \in (\text{id} \otimes J')(K_C \tilde{\otimes} K_{B_1} \tilde{\otimes} K_{B_2})$. It follows by our construction in Sec. VII that product states are not steerable by any channels as one can always take $x_C \otimes \Phi_1(x_A) \otimes \Phi_2(x_A)$. Remember that during steering, we fix not only the channels, but also the bipartite state, so the presented construction is valid. ∎

IX. STEERING BY MEASUREMENTS

We will show that the definition of steering given by Definition 11 follows the standard definition of steering [27] in the formalism introduced in Ref. [26], when we replace measurements by channels.

Proposition 16. Let S_1, S_2 be simplexes and let $m_1 : K_A \to S_1$, $m_2 : K_A \to S_2$ be measurements, then, a state $\psi \in K_C \tilde{\otimes} K_A$ is steerable by m_1, m_2 if and only if

$$(\text{id} \otimes (m_1, m_2))(\psi) \notin K_C \dot{\otimes} (S_1 \times S_2).$$

Proof. The result follows from the fact that $K_C \tilde{\otimes} S_1 \tilde{\otimes} S_2 = K_C \dot{\otimes} S_1 \dot{\otimes} S_2$. ∎

To obtain the standard definition of steering, one only needs to note that if $\xi \in K_C \dot{\otimes}(S_1 \times S_2)$, then there are $x_i \in K_C$, $s_i \in S_1 \times S_2$ and $0 \leqslant \lambda_i \leqslant 1$ for $i \in \{1, \ldots, n\}$ such that

$$\xi = \sum_{i=1}^n \lambda_i x_i \otimes s_i, \tag{8}$$

where the interpretation of s_i is that it is a conditional probability, conditioned by the choice of the measurement. At this point, it is straightforward to see that Eq. (8) corresponds to [27, Eq. (5)].

X. STEERING BY QUANTUM CHANNELS

Steering plays an important role in quantum theory. It has found so far applications in quantum cryptography [48] as an intermediate step between quantum key distribution and device-independent quantum key distribution.

We will prove several results and present a simple example of steering by quantum channels. Given the standard, operational, interpretation of steering by measurements the example may seem strange, but rather expected.

Let $\Phi_1 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}, \Phi_2 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$ be channels and let $|\psi^+\rangle = [\dim(\mathcal{H})]^{-\frac{1}{2}} \sum_{i=1}^{\dim(\mathcal{H})} |ii\rangle$ be the maximally entangled vector. We will show that the maximally entangled state $|\psi^+\rangle\langle\psi^+|$ is steerable by the channels Φ_1, Φ_2 whenever they are incompatible.

The proof is rather simple as the bipartite conditional state we obtain is $(id \otimes (\Phi_1, \Phi_2))(|\psi^+\rangle\langle\psi^+|)$. If the channels Φ_1, Φ_2 are compatible, then the state $|\psi^+\rangle\langle\psi^+|$ is not steerable by compatible channels. Now, let us assume that there is a state in $\rho \in \mathfrak{D}_{\mathcal{H}\otimes\mathcal{H}\otimes\mathcal{H}}$ such that we have

$$(id \otimes (\Phi_1, \Phi_2))(|\psi^+\rangle\langle\psi^+|) = (id \otimes J')(\rho), \qquad (9)$$

i.e., that the state $|\psi^+\rangle\langle\psi^+|$ is not steerable by the channels Φ_1, Φ_2. Equation (9) implies that we must have

$$(id \otimes \Phi_1)(|\psi^+\rangle\langle\psi^+|) = \mathrm{Tr}_3(\rho),$$

that, after taking trace over the second Hilbert space, gives

$$\frac{1}{\dim(\mathcal{H})}\mathbb{1} = \mathrm{Tr}_{23}(\rho). \qquad (10)$$

Now, the picture becomes clear: $(id \otimes \Phi_1)(|\psi^+\rangle\langle\psi^+|)$ is isomorphic to the Choi matrix $C(\Phi_1)$ and Eq. (10) implies that the state ρ must be isomorphic to a Choi matrix of some channel Φ. This together with Proposition 12 means that Eq. (9) holds if and only if the channels are compatible. Thus, we have proved the following:

Proposition 17. The maximally entangled state $|\psi^+\rangle\langle\psi^+|$ is steerable by channels $\Phi_1 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}, \Phi_2 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$ if and only if they are incompatible.

We will investigate steering by unitary channels. We will see a phenomenon that is impossible to happen for steering by measurements: it is possible to steer a state when the two channels we are testing for incompatibility are two copies of the same channel. Let U, V be unitary matrices, i.e., $UU^* = VV^* = \mathbb{1}$, where U^* denotes the conjugate transpose matrix to U and let Φ_U, Φ_V be the corresponding unitary channels, that is, for $\rho \in \mathfrak{D}_{\mathcal{H}}$ we have

$$\Phi_U(\rho) = U\rho U^*, \quad \Phi_V(\rho) = V\rho V^*.$$

Note that we have $\Phi_{\mathbb{1}} = id$, i.e., the unitary channel given by an identity matrix is the identity channel.

Proposition 18. The bipartite state $\rho \in \mathfrak{D}_{\mathcal{H}\otimes\mathcal{H}}$ is steerable by the unitary channels Φ_U, Φ_V if and only if it is steerable by two copies of the identity channel id.

Proof. The state $\rho \in \mathfrak{D}_{\mathcal{H}\otimes\mathcal{H}}$ is steerable by the channels Φ_U, Φ_V if and only if there is a state $\sigma \in \mathfrak{D}_{\mathcal{H}\otimes\mathcal{H}\otimes\mathcal{H}}$ such that

$$\mathrm{Tr}_3(\sigma) = (id \otimes \Phi_U)(\rho), \quad \mathrm{Tr}_2(\sigma) = (id \otimes \Phi_V)(\rho).$$

If such state σ exists, then for $\tilde{\sigma} = (id \otimes \Phi_{U^*} \otimes \Phi_{V^*})(\sigma)$ we have

$$\mathrm{Tr}_3(\tilde{\sigma}) = \rho, \quad \mathrm{Tr}_2(\tilde{\sigma}) = \rho,$$

i.e., the state ρ is not steerable by two copies of id. The same holds the other way around by almost the same construction; if the state ρ is not steerable by two copies of id, then it is not steerable by any unitary channels Φ_U, Φ_V. $\blacksquare$

Note that a similar result would hold if only one of the channels would be unitary, but then only that one unitary channel would be replaced by the identity map id. Clearly, if

the state ρ would be separable, then it would not be steerable by any channel. The converse does not hold, even if the state ρ is entangled it still may not be steerable by any channels. We will provide a useful condition for the steerability of a given state $\rho \in \mathfrak{D}_{\mathcal{H}\otimes\mathcal{H}}$ that will help us to show that even if the state ρ is entangled, it does not have to be steerable by any pair of channels $\Phi_1 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}, \Phi_2 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$.

Proposition 19. The state $\rho \in \mathfrak{D}_{\mathcal{H}\otimes\mathcal{H}}$ is steerable by the channels $\Phi_1 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}, \Phi_2 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$ only if it is steerable by two copies of the identity channel id : $\mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$.

Proof. Assume that the state $\rho \in \mathfrak{D}_{\mathcal{H}\otimes\mathcal{H}}$ is not steerable by two copies of the identity channel id : $\mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$, then there exists a state $\sigma \in \mathfrak{D}_{\mathcal{H}\otimes\mathcal{H}\otimes\mathcal{H}}$ such that

$$\mathrm{Tr}_3(\sigma) = \rho, \quad \mathrm{Tr}_2(\sigma) = \rho.$$

Let $\Phi_1 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}, \Phi_2 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$ be any two channels and denote

$$\tilde{\sigma} = (id \otimes \Phi_1 \otimes \Phi_2)(\sigma),$$

then we have

$$\mathrm{Tr}_3(\tilde{\sigma}) = (id \otimes \Phi_1)(\rho), \quad \mathrm{Tr}_2(\tilde{\sigma}) = (id \otimes \Phi_2)(\rho),$$

so the state ρ is not steerable by the channels Φ_1, Φ_2. $\blacksquare$

Note that one may get other conditions for steering by replacing only one of the channels by the identity map id. One may generalize this result to the general probabilistic theory, but it may be rather restrictive and not as general as one would wish. One may also use the idea of the proof of Proposition 19 together with the result of Proposition 17 to obtain the results on compatibility of channels that are concatenations of other channels, similar to the results obtained in Ref. [21].

We will present an example of an entangled state that is not steerable by any pair of channels.

Example 7. Let $\dim(\mathcal{H}) = 2$ with the standard basis $|0\rangle, |1\rangle$ and let $|W\rangle \in \mathcal{H} \otimes \mathcal{H} \otimes \mathcal{H}$ be given as

$$|W\rangle = \frac{1}{\sqrt{3}}(|001\rangle + |010\rangle + |100\rangle).$$

The projector $|W\rangle\langle W| \in \mathfrak{D}_{\mathcal{H}\otimes\mathcal{H}\otimes\mathcal{H}}$ is known as W state. We have

$$\rho_W = \mathrm{Tr}_2(|W\rangle\langle W|) = \mathrm{Tr}_3(|W\rangle\langle W|) \in \mathfrak{D}_{\mathcal{H}\otimes\mathcal{H}},$$

that shows that the state ρ_W is not steerable by a pair of the identity channels id : $\mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$, which as a result of Proposition 19 means that it is not steerable by any channels $\Phi_1 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}, \Phi_2 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$. Moreover it is known that the state ρ_W is entangled [43, Example 6.70].

Since it will be useful in later calculations, we will show that the state $|W\rangle\langle W|$ is the only state from $\mathfrak{D}_{\mathcal{H}\otimes\mathcal{H}\otimes\mathcal{H}}$ such that $\rho_W = \mathrm{Tr}_2(|W\rangle\langle W|) = \mathrm{Tr}_3(|W\rangle\langle W|)$. Let $|\varphi\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$, then we have

$$\rho_W = \frac{1}{3}|00\rangle\langle 00| + \frac{2}{3}|\varphi\rangle\langle\varphi|.$$

Let $\sigma \in \mathfrak{D}_{\mathcal{H}\otimes\mathcal{H}\otimes\mathcal{H}}$ denote the state such that $\rho_W = \mathrm{Tr}_2(\sigma) = \mathrm{Tr}_3(\sigma)$. We have $\rho_W|11\rangle = 0$ that implies $\mathrm{Tr}(\sigma|11\rangle\langle 11| \otimes \mathbb{1}) = \mathrm{Tr}(\sigma|1\rangle\langle 1| \otimes \mathbb{1} \otimes |1\rangle\langle 1|) = 0$ that implies $\langle 111|\sigma|111\rangle = \langle 110|\sigma|110\rangle = \langle 101|\sigma|101\rangle = 0$ as $\sigma \geqslant 0$. We will show that this implies $\sigma|111\rangle = \sigma|110\rangle = \sigma|101\rangle = 0$.

Let $A \in B_h(\mathcal{H})$, $A \geqslant 0$, and let $|\psi\rangle \in \mathcal{H}$. Let $\|\psi\| = \sqrt{\langle \psi | \psi \rangle}$ denote the norm given by inner product. Assume that we have $\langle \psi | A | \psi \rangle = 0$, then

$$\|\sqrt{A}\psi\|^2 = \langle \sqrt{A}\psi | \sqrt{A}\psi \rangle = \langle \psi | A | \psi \rangle = 0$$

and in conclusion we have $\sqrt{A}|\psi\rangle = 0$ and

$$A|\psi\rangle = \sqrt{A}(\sqrt{A}|\psi\rangle) = 0.$$

Finally, let us denote $|\varphi^\perp\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$. We have $\rho_W |\varphi^\perp\rangle = 0$ that implies $\mathrm{Tr}(\sigma |\varphi^\perp\rangle\langle\varphi^\perp| \otimes \mathbb{1}) = 0$ which yields $\sigma |\varphi^\perp 0\rangle = \sigma |\varphi^\perp 1\rangle = 0$. We still use the shorthand $|\varphi^\perp 0\rangle = |\varphi^\perp\rangle \otimes |0\rangle$.

The eight vectors $|000\rangle$, $|001\rangle$, $|\varphi 0\rangle$, $|\varphi 1\rangle$, $|\varphi^\perp 0\rangle$, $|\varphi^\perp 1\rangle$, $|110\rangle$, $|111\rangle$ form an orthonormal basis of $\mathcal{H} \otimes \mathcal{H} \otimes \mathcal{H}$. We have already showed that we must have

$$\sigma |\varphi^\perp 0\rangle = \sigma |\varphi^\perp 1\rangle = \sigma |110\rangle = \sigma |111\rangle = 0,$$

so in general we must have

$$\begin{aligned}
\sigma = \; & a_{00}|000\rangle\langle 000| + a_{01}|001\rangle\langle 001| + a_{\varphi 0}|\varphi 0\rangle\langle\varphi 0| \\
& + a_{\varphi 1}|\varphi 1\rangle\langle\varphi 1| + b_1|000\rangle\langle 001| + \bar{b}_1|001\rangle\langle 000| \\
& + b_2|000\rangle\langle\varphi 0| + \bar{b}_2|\varphi 0\rangle\langle 000| + b_3|000\rangle\langle\varphi 1| \\
& + \bar{b}_3|\varphi 1\rangle\langle 000| + b_4|001\rangle\langle\varphi 0| + \bar{b}_4|\varphi 0\rangle\langle 001| \\
& + b_5|001\rangle\langle\varphi 1| + \bar{b}_5|\varphi 1\rangle\langle 001| + b_6|\varphi 0\rangle\langle\varphi 1| \\
& + \bar{b}_6|\varphi 1\rangle\langle\varphi 0|.
\end{aligned}$$

Using the above expression for σ we get

$$\begin{aligned}
\mathrm{Tr}_2(\sigma) = \; & a_{00}|00\rangle\langle 00| + a_{01}|01\rangle\langle 01| + \frac{a_{\varphi 0}}{2}\mathbb{1} \otimes |0\rangle\langle 0| \\
& + \frac{a_{\varphi 1}}{2}\mathbb{1} \otimes |1\rangle\langle 1| + b_1|00\rangle\langle 01| + \bar{b}_1|01\rangle\langle 00| \\
& + \frac{b_2}{\sqrt{2}}|00\rangle\langle 10| + \frac{\bar{b}2}{\sqrt{2}}|10\rangle\langle 00| + \frac{b_3}{\sqrt{2}}|00\rangle\langle 11| \\
& + \frac{\bar{b}_3}{\sqrt{2}}|11\rangle\langle 00| + \frac{b_4}{\sqrt{2}}|01\rangle\langle 10| + \frac{\bar{b}_4}{\sqrt{2}}|10\rangle\langle 01| \\
& + \frac{b_5}{\sqrt{2}}|01\rangle\langle 11| + \frac{\bar{b}_5}{\sqrt{2}}|11\rangle\langle 01| + \frac{b_6}{2}\mathbb{1} \otimes |0\rangle\langle 1| \\
& + \frac{\bar{b}_6}{2}\mathbb{1} \otimes |1\rangle\langle 0|,
\end{aligned}$$

that implies $a_{00} = a_{\varphi 1} = 0$, $a_{\varphi 0} = \frac{2}{3}$, $a_{01} = \frac{1}{3}$, $b_1 = b_2 = b_3 = b_5 = b_6 = 0$, and $b_4 = \frac{\sqrt{2}}{3}$. In conclusion, we have

$$\begin{aligned}
\sigma = \; & \tfrac{1}{3}(|001\rangle\langle 001| + 2|\varphi 0\rangle\langle\varphi 0| \\
& + \sqrt{2}|001\rangle\langle\varphi 0| + \sqrt{2}|\varphi 0\rangle\langle 001|) \\
= \; & |W\rangle\langle W|.
\end{aligned}$$

XI. BELL NONLOCALITY

Bell nonlocality is, similarly to steering, a phenomenom that we do not find in classical theory, but is often used in quantum theory. Bell nonlocality [2] was formulated as a response to the well-known EPR paradox [3]. Although in the original formulation the operational idea was different than

the one we will present, we will see that Bell nonlocality may be understood as an incompatibility test, in the same way as steering.

Let us assume that we have four parties: Alice, Bob, Charlie, and Dan. Alice has two channels $\Phi_1^A : K_A \to K_{B_1}$ and $\Phi_2^A : K_A \to K_{B_2}$ that she can use to send a state to Bob, and Charlie has two channels $\Phi_1^C : K_C \to K_{D_1}$ and $\Phi_2^C : K_C \to K_{D_2}$ that he can use to send a state to Dan. Assume that $K_C \tilde{\otimes} K_A$ is defined and let $\psi \in K_C \tilde{\otimes} K_A$ be a bipartite state shared by Alice and Charlie. The idea that we use to define Bell nonlocality is very simple: if we were able to use $(\mathrm{id} \otimes (\Phi_1^A, \Phi_2^A))(\psi)$ and $((\Phi_1^C, \Phi_2^C) \otimes \mathrm{id})(\psi)$ as nontrivial incompatibility test, we may as well investigate whether $((\Phi_1^C, \Phi_2^C) \otimes (\Phi_1^A, \Phi_2^A))(\psi)$ provides an incompatibility test in the same manner.

Definition 12. Let $\psi \in K_C \tilde{\otimes} K_A$ and let

$$\begin{aligned}
\Phi_1^A : K_A &\to K_{B_1}, \\
\Phi_2^A : K_A &\to K_{B_2}, \\
\Phi_1^C : K_C &\to K_{D_1}, \\
\Phi_2^C : K_C &\to K_{D_2}
\end{aligned}$$

be channels. We call the state $((\Phi_1^C, \Phi_2^C) \otimes (\Phi_1^A, \Phi_2^A))(\psi)$ bipartite biconditional state.

Assume that the channels Φ_1^A and Φ_2^A are compatible, so that we have $(\Phi_1^A, \Phi_2^A) = J(\Phi^A)$ for some channel $\Phi^A : K_A \to K_{B_1} \tilde{\otimes} K_{B_2}$ and also that the channels Φ_1^C and Φ_2^C are compatible, so there is a channel $\Phi^C : K_C \to K_{D_1} \tilde{\otimes} K_{D_2}$ such that $(\Phi_1^A, \Phi_2^A) = J(\Phi^A)$. Let $\psi \in K_C \tilde{\otimes} K_A$, then we have

$$\big((\Phi_1^C, \Phi_2^C) \otimes (\Phi_1^A, \Phi_2^A)\big)(\psi) = (J' \otimes J')((\Phi_C \otimes \Phi_A)(\psi)),$$

where the maps J' are defined as before, with the exception that we denote them the same even though they map different spaces.

We present a definition of Bell nonlocality using the same line of thinking as we used in Definition 11. For simplicity we will denote

$$Q_{DC} = (J' \otimes J')\big(K_{D_1} \tilde{\otimes} K_{D_2} \tilde{\otimes} K_{C_1} \tilde{\otimes} K_{C_2}\big).$$

Definition 13. Let $\psi \in K_C \tilde{\otimes} K_A$ be a bipartite state and let $\Phi_1^A : K_A \to K_{B_1}$, $\Phi_2^A : K_A \to K_{B_2}$, $\Phi_1^C : K_A \to K_{C_1}$, and $\Phi_2^C : K_A \to K_{D_2}$ be channels. We say that the bipartite biconditional state $((\Phi_1^C, \Phi_2^C) \otimes (\Phi_1^A, \Phi_2^A))(\psi)$ is Bell nonlocal if

$$\big((\Phi_1^C, \Phi_2^C) \otimes (\Phi_1^A, \Phi_2^A)\big)(\psi) \notin Q_{DC}.$$

Otherwise, we call the bipartite biconditional state Bell local.

The following result follows immediately from Definition 13.

Corollary 3. Let $\psi \in K_C \tilde{\otimes} K_A$ be a bipartite state and let $\Phi_1^A : K_A \to K_{B_1}$, $\Phi_2^A : K_A \to K_{B_2}$, $\Phi_1^C : K_A \to K_{C_1}$, and $\Phi_2^C : K_A \to K_{D_2}$ be channels. The bipartite biconditional state $((\Phi_1^C, \Phi_2^C) \otimes (\Phi_1^A, \Phi_2^A))(\psi)$ is Bell nonlocal only if the channels Φ_1^A, Φ_2^A and Φ_1^C, Φ_2^C are incompatible.

We will show that entanglement plays a key role in Bell nonlocality.

Proposition 20. Let $\psi \in K_C \dot{\otimes} K_A$ be a separable bipartite state and let $\Phi_1^A : K_A \to K_{B_1}$, $\Phi_2^A : K_A \to K_{B_2}$, $\Phi_1^C : K_A \to$

K_{C_1}, and $\Phi_2^C : K_A \to K_{D_2}$ be channels. The bipartite biconditional state $((\Phi_1^C, \Phi_2^C) \otimes (\Phi_1^A, \Phi_2^A))(\psi)$ is Bell local.

Proof. It is again sufficient to consider $\psi = x_C \otimes x_A$ for $x_A \in K_A$, $x_C \in K_C$ due to the linearity of the maps (Φ_1^A, Φ_2^A) and (Φ_1^C, Φ_2^C). Consider the state $\varphi \in K_{D_1} \tilde{\otimes} K_{D_2} \tilde{\otimes} K_{C_1} \tilde{\otimes} K_{C_2}$ given as

$$\varphi = \Phi_1^C(x_C) \otimes \Phi_2^C(x_C) \otimes \Phi_1^A(x_A) \otimes \Phi_2^A(x_A),$$

then we have

$$\big((\Phi_1^C, \Phi_2^C) \otimes (\Phi_1^A, \Phi_2^A)\big)(\psi) = (J' \otimes J')(\varphi).$$

∎

XII. BELL NONLOCALITY OF MEASUREMENTS

We will again show that Definition 13 follows the standard definition of Bell nonlocality [27] in the formalism of [26].

Proposition 21. Let S_1^A, S_2^A, S_1^C, and S_2^C be simplexes and let $m_1^A : K_A \to S_1^A$, $m_2^A : K_A \to S_2^A$, $m_1^C : K_C \to S_1^C$, $m_2^C : K_C \to S_2^C$ be measurements. Let $\psi \in K_C \tilde{\otimes} K_A$, then the bipartite biconditional state $((m_1^C, m_2^C) \otimes (m_1^A, m_2^A))(\psi)$ is Bell nonlocal if

$$\big((m_1^C, m_2^C) \otimes (m_1^A, m_2^A)\big)(\psi) \notin \big(S_1^C \times S_2^C\big) \dot{\otimes} \big(S_1^A \times S_2^A\big).$$

Proof. By direct calculation we have

$$\begin{aligned} Q_{CD} &= (J' \otimes J')\big(S_1^C \dot{\otimes} S_2^C \dot{\otimes} S_1^A \dot{\otimes} S_2^A\big) \\ &= \big(S_1^C \times S_2^C\big) \dot{\otimes} \big(S_1^A \times S_2^A\big). \end{aligned}$$

∎

One may again use the interpretation that both $S_1^C \times S_2^C$ and $S_1^A \times S_2^A$ are spaces of conditional measurement probabilities, so if we have $\psi \in (S_1^C \times S_2^C) \dot{\otimes} (S_1^A \times S_2^A)$, then we must have $0 \leqslant \lambda_i \leqslant 1$, for $i \in \{1, \dots, n\}$, $\sum_{i=1}^n \lambda_i = 1$, such that

$$\psi = \sum_{i=1}^n \lambda_i s_i^C \otimes s_i^A,$$

where in standard formulations both $s_i^C \in S_1^C \times S_2^C$ and $s_i^A \in_1^A \times S_2^A$ are represented by probabilities, i.e., by numbers, so the tensor product between them is omitted.

We will provide proof of the standard and well-known result about connection of steering and Bell nonlocality of measurements.

Proposition 22. Let S_1^A, S_2^A, S_1^C, and S_2^C be simplexes and let $m_1^A : K_A \to S_1^A$, $m_2^A : K_A \to S_2^A$, $m_1^C : K_C \to S_1^C$, $m_2^C : K_C \to S_2^C$ be measurements. Let $\psi \in K_C \tilde{\otimes} K_A$. If

$$\big(\mathrm{id} \otimes (m_1^A, m_2^A)\big)(\psi) \in K_C \dot{\otimes} \big(S_1^A \times S_2^A\big),$$

i.e., if the bipartite state is not steerable by measurements m_1^A, m_2^A, then

$$\big((m_1^C, m_2^C) \otimes (m_1^A, m_2^A)\big)(\psi) \in \big(S_1^C \times S_2^C\big) \dot{\otimes} \big(S_1^A \times S_2^A\big).$$

Proof. Let

$$\big(\mathrm{id} \otimes (m_1^A, m_2^A)\big)(\psi) \in K_C \dot{\otimes} \big(S_1^A \times S_2^A\big),$$

then for $n \in \mathbb{N}$, $i \in \{1, \dots, n\}$, there are $0 \leqslant \lambda_i \leqslant 1$, $x_i \in K_C$ and $s_i \in S_1^A \times S_2^A$, $\sum_{i=1}^n \lambda_i = 1$, such that we have

$$\big(\mathrm{id} \otimes (m_1^A, m_2^A)\big)(\psi) = \sum_{i=1}^n \lambda_i x_i \otimes s_i.$$

We get

$$\big((m_1^C, m_2^C) \otimes (m_1^A, m_2^A)\big)(\psi) = \sum_{i=1}^n \lambda_i \big(m_1^C, m_2^C\big)(x_i) \otimes s_i,$$

and since we have $(m_1^C, m_2^C)(x_i) = (m_1^C(x_i), m_2^C(x_i)) \in S_1^C \times S_2^C$ we have

$$\big((m_1^C, m_2^C) \otimes (m_1^A, m_2^A)\big)(\psi) \in \big(S_1^C \times S_2^C\big) \dot{\otimes} \big(S_1^A \times S_2^A\big).$$

∎

Note that the same result would also hold for steering by the measurements m_1^C, m_2^C.

One may think that steering is somehow half of Bell nonlocality, or that it is some middle step towards Bell nonlocality as even our constructions in Secs. VIII and XI would point to such a result. We will show that this is not true in general, as we will provide a counterexample using quantum channels in Example 9.

XIII. BELL NONLOCALITY OF QUANTUM CHANNELS

Bell nonlocality of quantum measurements is a deeply studied topic in quantum theory, with several applications in various device-independent protocols [49–52], randomness generation and randomness expansion [53,54], and others (for a recent review on Bell nonlocality see [55]).

Bell nonlocality of quantum channels follows very similar rules to steering by quantum channels. We will derive results specific for quantum theory in the same manner as in Sec. X.

Proposition 23. Let $\rho \in \mathfrak{D}_{\mathcal{H} \otimes \mathcal{H}}$ and let $\Phi_1^1 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$, $\Phi_2^1 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$, $\Phi_1^2 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$, $\Phi_2^2 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$ be channels. The bipartite biconditional state $((\Phi_1^1, \Phi_2^1) \otimes (\Phi_1^2, \Phi_2^2))(\rho)$ is Bell nonlocal only if the bipartite biconditional state $((\mathrm{id}, \mathrm{id}) \otimes (\mathrm{id}, \mathrm{id}))(\rho)$ is Bell nonlocal.

Proof. If the bipartite biconditional state $((\mathrm{id}, \mathrm{id}) \otimes (\mathrm{id}, \mathrm{id}))(\rho)$ is Bell local, then there exist $\sigma \in \mathfrak{D}_{\mathcal{H} \otimes \mathcal{H} \otimes \mathcal{H} \otimes \mathcal{H}}$ such that

$$\begin{aligned} \mathrm{Tr}_{24}(\sigma) &= \rho, \\ \mathrm{Tr}_{23}(\sigma) &= \rho, \\ \mathrm{Tr}_{14}(\sigma) &= \rho, \\ \mathrm{Tr}_{13}(\sigma) &= \rho. \end{aligned}$$

Let

$$\tilde{\sigma} = \big(\Phi_1^1 \otimes \Phi_2^1 \otimes \Phi_1^2 \otimes \Phi_2^2\big)(\sigma),$$

then

$$\begin{aligned} \mathrm{Tr}_{24}(\tilde{\sigma}) &= \big(\Phi_1^1 \otimes \Phi_1^2\big)(\rho), \\ \mathrm{Tr}_{23}(\tilde{\sigma}) &= \big(\Phi_1^1 \otimes \Phi_2^2\big)(\rho), \\ \mathrm{Tr}_{14}(\tilde{\sigma}) &= \big(\Phi_2^1 \otimes \Phi_1^2\big)(\rho), \\ \mathrm{Tr}_{13}(\tilde{\sigma}) &= \big(\Phi_2^1 \otimes \Phi_2^2\big)(\rho). \end{aligned}$$

∎

Note that again we do not have to replace all of the channels by the identity channels id, but we may replace only some.

Proposition 24. Let $\rho \in \mathfrak{D}_{\mathcal{H} \otimes \mathcal{H}}$ and let $\Phi_1^1 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$, $\Phi_2^1 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$, $\Phi_1^2 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$, $\Phi_2^2 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$ be channels, moreover, let $\Phi_1^1 = \Phi_U$ be a unitary channel given by the unitary matrix U, then the bipartite biconditional state $((\Phi_U, \Phi_2^1) \otimes (\Phi_1^2, \Phi_2^2))(\rho)$ is Bell nonlocal if and only if the bipartite biconditional state $((id, \Phi_2^1) \otimes (\Phi_1^2, \Phi_2^2))(\rho)$ is Bell nonlocal.

Proof. Using the very same idea as before, if the bipartite biconditional state $((\Phi_U, \Phi_2^1) \otimes (\Phi_1^2, \Phi_2^2))(\rho)$ is Bell local, then there is $\sigma \in \mathfrak{D}_{\mathcal{H} \otimes \mathcal{H} \otimes \mathcal{H} \otimes \mathcal{H}}$ such that

$$\mathrm{Tr}_{24}(\sigma) = \left(\Phi_U \otimes \Phi_1^2\right)(\rho),$$
$$\mathrm{Tr}_{23}(\sigma) = \left(\Phi_U \otimes \Phi_2^2\right)(\rho),$$
$$\mathrm{Tr}_{14}(\sigma) = \left(\Phi_2^1 \otimes \Phi_1^2\right)(\rho),$$
$$\mathrm{Tr}_{13}(\sigma) = \left(\Phi_2^1 \otimes \Phi_2^2\right)(\rho).$$

Let

$$\tilde{\sigma} = (\Phi_{U^*} \otimes \mathrm{id} \otimes \mathrm{id} \otimes \mathrm{id})(\sigma),$$

then we get

$$\mathrm{Tr}_{24}(\tilde{\sigma}) = \left(\mathrm{id} \otimes \Phi_1^2\right)(\rho),$$
$$\mathrm{Tr}_{23}(\tilde{\sigma}) = \left(\mathrm{id} \otimes \Phi_2^2\right)(\rho),$$
$$\mathrm{Tr}_{14}(\tilde{\sigma}) = \left(\Phi_2^1 \otimes \Phi_1^2\right)(\rho),$$
$$\mathrm{Tr}_{13}(\tilde{\sigma}) = \left(\Phi_2^1 \otimes \Phi_2^2\right)(\rho).$$

$\blacksquare$

One may obtain similar results if some other of the channels $\Phi_1^1, \Phi_2^1, \Phi_1^2, \Phi_2^2$ is unitary as well as if more or even all of them are unitary.

The most iconic and most studied aspects of Bell nonlocality are the Bell inequalities. We are going to present a version of CHSH inequality for quantum channels. Assume that $\dim(\mathcal{H}) = 2$ and let $|0\rangle, |1\rangle$ denote any orthonormal basis of $\mathcal{H}$. We will use the shorthand $|00\rangle = |0\rangle \otimes |0\rangle$. Let $i, j \in \{1, 2\}$ and let

$$\begin{aligned}
E\left(\Phi_i^1, \Phi_j^2\right) &= \langle 00|\left(\Phi_i^1 \otimes \Phi_j^2\right)(\rho)|00\rangle \\
&\quad - \langle 01|\left(\Phi_i^1 \otimes \Phi_j^2\right)(\rho)|01\rangle \\
&\quad - \langle 10|\left(\Phi_i^1 \otimes \Phi_j^2\right)(\rho)|10\rangle \\
&\quad + \langle 11|\left(\Phi_i^1 \otimes \Phi_j^2\right)(\rho)|11\rangle \\
&= \mathrm{Tr}\left(\left(\Phi_i^1 \otimes \Phi_j^2\right)(\rho)A\right),
\end{aligned}$$

where

$$A = |00\rangle\langle 00| - |01\rangle\langle 01| - |10\rangle\langle 10| + |11\rangle\langle 11|.$$

The quantity $E(\Phi_i^1, \Phi_j^2)$ is to be interpreted as the correlation between the marginals $\mathrm{Tr}_1((\Phi_i^1 \otimes \Phi_j^2)(\rho))$ and $\mathrm{Tr}_2((\Phi_i^1 \otimes \Phi_j^2)(\rho))$. Since we have $-\mathbb{1} \leqslant A \leqslant \mathbb{1}$ it is straightforward that we have $-1 \leqslant E(\Phi_i^1, \Phi_j^2) \leqslant 1$. Define a quantity

$$X_\rho = E\left(\Phi_1^1, \Phi_1^2\right) + E\left(\Phi_1^1, \Phi_2^2\right) + E\left(\Phi_2^1, \Phi_1^2\right) - E\left(\Phi_2^1, \Phi_2^2\right),$$

we will show that X_ρ corresponds to the quantity used in CHSH inequality. It is straightforward to see that $-4 \leqslant X_\rho \leqslant 4$ is the algebraic bound on X_ρ.

Proposition 25. If the biconditional bipartite state $((\Phi_1^1, \Phi_2^1) \otimes (\Phi_1^2, \Phi_2^2))(\rho)$ is Bell local, then we have $-2 \leqslant X_\rho \leqslant 2$.

Proof. If the biconditional bipartite state $((\Phi_1^1, \Phi_2^1) \otimes (\Phi_1^2, \Phi_2^2))(\rho)$ is Bell local then there is $\sigma \in \mathfrak{D}_{\mathcal{H} \otimes \mathcal{H} \otimes \mathcal{H} \otimes \mathcal{H}}$ such that

$$\mathrm{Tr}_{24}(\sigma) = \left(\Phi_1^1 \otimes \Phi_1^2\right)(\rho),$$
$$\mathrm{Tr}_{23}(\sigma) = \left(\Phi_1^1 \otimes \Phi_2^2\right)(\rho),$$
$$\mathrm{Tr}_{14}(\sigma) = \left(\Phi_2^1 \otimes \Phi_1^2\right)(\rho),$$
$$\mathrm{Tr}_{13}(\sigma) = \left(\Phi_2^1 \otimes \Phi_2^2\right)(\rho).$$

This yields

$$\begin{aligned}
E\left(\Phi_1^1, \Phi_1^2\right) &= \mathrm{Tr}\left(\left(\Phi_1^1 \otimes \Phi_1^2\right)(\rho)A\right) = \mathrm{Tr}(\mathrm{Tr}_{24}(\sigma)A) \\
&= \mathrm{Tr}(\sigma(|0\rangle\langle 0| \otimes \mathbb{1} \otimes |0\rangle\langle 0| \otimes \mathbb{1} \\
&\quad - |0\rangle\langle 0| \otimes \mathbb{1} \otimes |1\rangle\langle 1| \otimes \mathbb{1} \\
&\quad - |1\rangle\langle 1| \otimes \mathbb{1} \otimes |0\rangle\langle 0| \otimes \mathbb{1} \\
&\quad + |1\rangle\langle 1| \otimes \mathbb{1} \otimes |1\rangle\langle 1| \otimes \mathbb{1})).
\end{aligned}$$

In the same manner, we get

$$\begin{aligned}
E\left(\Phi_1^1, \Phi_2^2\right) &= \mathrm{Tr}(\sigma(|0\rangle\langle 0| \otimes \mathbb{1} \otimes \mathbb{1} \otimes |0\rangle\langle 0| \\
&\quad - |0\rangle\langle 0| \otimes \mathbb{1} \otimes \mathbb{1} \otimes |1\rangle\langle 1| \\
&\quad - |1\rangle\langle 1| \otimes \mathbb{1} \otimes \mathbb{1} \otimes |0\rangle\langle 0| \\
&\quad + |1\rangle\langle 1| \otimes \mathbb{1} \otimes \mathbb{1} \otimes |1\rangle\langle 1|)),
\end{aligned}$$

$$\begin{aligned}
E\left(\Phi_2^1, \Phi_1^2\right) &= \mathrm{Tr}(\sigma(\mathbb{1} \otimes |0\rangle\langle 0| \otimes |0\rangle\langle 0| \otimes \mathbb{1} \\
&\quad - \mathbb{1} \otimes |0\rangle\langle 0| \otimes |1\rangle\langle 1| \otimes \mathbb{1} \\
&\quad - \mathbb{1} \otimes |1\rangle\langle 1| \otimes |0\rangle\langle 0| \otimes \mathbb{1} \\
&\quad + \mathbb{1} \otimes |1\rangle\langle 1| \otimes |1\rangle\langle 1| \otimes \mathbb{1})),
\end{aligned}$$

and

$$\begin{aligned}
E\left(\Phi_2^1, \Phi_2^2\right) &= \mathrm{Tr}(\sigma(\mathbb{1} \otimes |0\rangle\langle 0| \otimes \mathbb{1} \otimes |0\rangle\langle 0| \\
&\quad - \mathbb{1} \otimes |0\rangle\langle 0| \otimes \mathbb{1} \otimes |1\rangle\langle 1| \\
&\quad - \mathbb{1} \otimes |1\rangle\langle 1| \otimes \mathbb{1} \otimes |0\rangle\langle 0| \\
&\quad + \mathbb{1} \otimes |1\rangle\langle 1| \otimes \mathbb{1} \otimes |1\rangle\langle 1|)).
\end{aligned}$$

Together we get

$$\begin{aligned}
X_\rho = 2\,\mathrm{Tr}(\sigma(&|0000\rangle\langle 0000| + |0001\rangle\langle 0001| - |0010\rangle\langle 0010| \\
&- |0011\rangle\langle 0011| + |0100\rangle\langle 0100| - |0101\rangle\langle 0101| \\
&+ |0110\rangle\langle 0110| - |0111\rangle\langle 0111| - |1000\rangle\langle 1000| \\
&+ |1001\rangle\langle 1001| - |1010\rangle\langle 1010| + |1011\rangle\langle 1011| \\
&- |1100\rangle\langle 1100| - |1101\rangle\langle 1101| + |1110\rangle\langle 1110| \\
&+ |1111\rangle\langle 1111|))
\end{aligned}$$

that implies $-2 \leqslant X_\rho \leqslant 2$. $\blacksquare$

At this point, one may ask whether there exists an equivalent of Tsirelson bound [28] for the inequality given by Proposition 25, or what is the maximum violation of the

aforementioned inequality. We will show that the Tsirelson bound $2\sqrt{2}$ is both reachable and maximum violation by quantum channels.

Proposition 26. For any state $\rho \in \mathfrak{D}_{\mathcal{H} \otimes \mathcal{H}}$ and any four channels $\Phi_1^1 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$, $\Phi_2^1 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$, $\Phi_1^2 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$, $\Phi_2^2 : \mathfrak{D}_{\mathcal{H}} \to \mathfrak{D}_{\mathcal{H}}$ we have

$$X_\rho \leqslant 2\sqrt{2}.$$

Proof. We define the adjoint channel Φ_1^{1*} to channel Φ_1^1 as the linear map $\Phi_1^{1*} : B_h(\mathcal{H}) \to B_h(\mathcal{H})$ such that for all $\sigma \in \mathfrak{D}_{\mathcal{H}}$ and $E \in B_h(\mathcal{H}), 0 \leqslant E \leqslant \mathbb{1}$, we have

$$\mathrm{Tr}\big(\Phi_1^1(\sigma)E\big) = \mathrm{Tr}\big(\sigma \Phi_1^{1*}(E)\big).$$

Since Φ_1^1 is a channel, we have $0 \leqslant \Phi_1^{1*}(E) \leqslant \mathbb{1}$ and $\Phi_1^{1*}(\mathbb{1}) = \mathbb{1}$. This approach of mapping effects instead of states is called the Heisenberg picture.

Let $i,j \in \{1,2\}$, then we have

$$\mathrm{Tr}\big((\Phi_i^1 \otimes \Phi_j^2)(\rho)|00\rangle\langle 00|\big) = \mathrm{Tr}\big(\rho \Phi_i^{1*}(|0\rangle\langle 0|) \otimes \Phi_j^{2*}(|0\rangle\langle 0|)\big).$$

Denoting

$$M_i^1 = \Phi_i^{1*}(|0\rangle\langle 0|),$$
$$M_j^2 = \Phi_j^{1*}(|0\rangle\langle 0|),$$

we see that we have

$$\begin{aligned}
E\big(\Phi_i^1, \Phi_j^2\big) &= \mathrm{Tr}\big(\rho M_i^1 \otimes M_j^2\big) - \mathrm{Tr}\big(\rho(\mathbb{1} - M_i^1) \otimes M_j^2\big) \\
&\quad - \mathrm{Tr}\big(\rho M_i^1 \otimes (\mathbb{1} - M_j^2)\big) \\
&\quad + \mathrm{Tr}\big(\rho(\mathbb{1} - M_i^1) \otimes (\mathbb{1} - M_j^2)\big) \\
&= E\big(M_i^1, M_j^2\big),
\end{aligned}$$

where $E(M_i^1, M_j^2)$ is a correlation for the two-outcome measurements given by the effects M_i^1 and M_j^2. It is a well-known result [28] that we always have

$$\begin{aligned}
&E\big(M_1^1, M_1^2\big) + E\big(M_1^1, M_2^2\big) + E\big(M_2^1, M_1^2\big) - E\big(M_2^1, M_2^2\big) \\
&\leqslant 2\sqrt{2}.
\end{aligned}$$

∎

It is very intuitive that the Tsirelson bound, reachable by measurements, will be also reachable by channels. To prove this, let $M, N \in B_h(\mathcal{H}), 0 \leqslant M \leqslant \mathbb{1}, 0 \leqslant N \leqslant \mathbb{1}$, and define channels $\Phi_M : B_h(\mathcal{H}) \to B_h(\mathcal{H})$, $\Phi_N : B_h(\mathcal{H}) \to B_h(\mathcal{H})$ such that for $\sigma \in \mathfrak{D}_{\mathcal{H}}$ we have

$$\Phi_M(\sigma) = \mathrm{Tr}(\sigma M)|0\rangle\langle 0| + \mathrm{Tr}(\sigma(\mathbb{1} - M))|1\rangle\langle 1|,$$
$$\Phi_N(\sigma) = \mathrm{Tr}(\sigma N)|0\rangle\langle 0| + \mathrm{Tr}(\sigma(\mathbb{1} - N))|1\rangle\langle 1|.$$

It is easy to verify that the maps Φ_M, Φ_N are quantum channels and that they are also measurements as they map the state space $\mathfrak{D}_{\mathcal{H}}$ to the simplex $\mathrm{conv}\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$. Let $\rho \in \mathfrak{D}_{\mathcal{H} \otimes \mathcal{H}}$, then we have

$$\begin{aligned}
\mathrm{Tr}((\Phi_M \otimes \Phi_N)(\rho)A) &= \mathrm{Tr}(\rho(M \otimes N - (\mathbb{1} - M) \otimes N) \\
&\quad - \mathrm{Tr}(\rho(M \otimes (\mathbb{1} - N))) \\
&\quad + \mathrm{Tr}(\rho((\mathbb{1} - M) \otimes (\mathbb{1} - N))) \\
&= E(M, N).
\end{aligned}$$

This proves that any set of correlations and any violation of CHSH inequality reachable by measurements are also reachable by quantum channels as a violation of the bound given by Proposition 25.

To generalize the proposed inequality, one may replace the projectors $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$ by any pair of effects $M, N \in B_h(\mathcal{H}), 0 \leqslant M \leqslant \mathbb{1}, 0 \leqslant N \leqslant \mathbb{1}$, and have

$$\begin{aligned}
A &= M \otimes N - (\mathbb{1} - M) \otimes N - M \otimes (\mathbb{1} - N) + (\mathbb{1} - M) \\
&\quad \otimes (\mathbb{1} - N).
\end{aligned}$$

From now on, we will consider a special case. Keep $\dim(\mathcal{H}) = 2$ and let

$$|\psi^+\rangle\langle\psi^+| = \tfrac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 11|)$$

be the maximally entangled state, let U_1, U_2, V_1, V_2 be unitary matrices and let $\Phi_1^1 = \Phi_{U_1}$, $\Phi_2^1 = \Phi_{U_2}$, $\Phi_1^2 = \Phi_{V_1}$, $\Phi_2^2 = \Phi_{V_2}$ be unitary channels given by the respective unitary matrices. We will consider the bipartite biconditional state $((\Phi_{U_1}, \Phi_{U_2}) \otimes (\Phi_{V_1}, \Phi_{V_2}))(|\psi^+\rangle\langle\psi^+|)$ and we will show that the correlations for the given bipartite biconditional state are of a particular nice form. We have

$$\big(\Phi_{U_i} \otimes \Phi_{V_j}\big)(|\psi^+\rangle\langle\psi^+|) = \big(\mathrm{id} \otimes \Phi_{V_j U_i^T}\big)(|\psi^+\rangle\langle\psi^+|),$$

where $i,j \in \{1,2\}$ and for U^T denotes the transpose of the matrix U. For the correlation we have

$$\begin{aligned}
E\big(\Phi_{U_i}, \Phi_{V_j}\big) &= \mathrm{Tr}\big((\mathrm{id} \otimes \Phi_{V_j U_i^T})(|\psi^+\rangle\langle\psi^+|)A\big) \\
&= \tfrac{1}{2}\big(|\langle 0|V_j U_i^T|0\rangle|^2 + |\langle 1|V_j U_i^T|1\rangle|^2 \\
&\quad - |\langle 0|V_j U_i^T|1\rangle|^2 - |\langle 1|V_j U_i^T|0\rangle|^2\big). \quad (11)
\end{aligned}$$

We will provide an example of a violation of the bound given by Proposition 25 by incompatible unitary channels.

Example 8. Let $\dim(\mathcal{H}) = 2$ and let $\vartheta \in \mathbb{R}$ be a parameter. Let U_1, U_2, V_1, V_2 be unitary matrices given as

$$U_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

$$U_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$V_1 = \frac{1}{\sqrt{1+\vartheta}}\begin{pmatrix} \sqrt{\vartheta} & 1 \\ 1 & -\sqrt{\vartheta} \end{pmatrix},$$

$$V_2 = \frac{1}{\sqrt{1+\vartheta}}\begin{pmatrix} 1 & \sqrt{\vartheta} \\ \sqrt{\vartheta} & -1 \end{pmatrix}.$$

Consider the bipartite biconditional state $(\Phi_{U_1}, \Phi_{U_2}) \otimes (\Phi_{V_1}, \Phi_{V_2})(|\psi^+\rangle\langle\psi^+|)$. Using Eq. (11), we can obtain $X_{|\psi^+\rangle\langle\psi^+|}$ as a function of ϑ. The function is plotted in Fig. 1, where it is shown that for certain values of ϑ the bipartite biconditional state violates the bound given by Proposition 25.

It is also easy to see that the bipartite biconditional state $((\mathrm{id}, \mathrm{id}) \otimes (\mathrm{id}, \mathrm{id}))(|\psi^+\rangle\langle\psi^+|)$ does not violate the bound given by Proposition 25 because all of the correlations are the same, yet according to Proposition 24 we know that it must be a Bell nonlocal bipartite biconditional state. This shows that not all Bell nonlocal bipartite biconditional states violate the inequality given by Proposition 25.
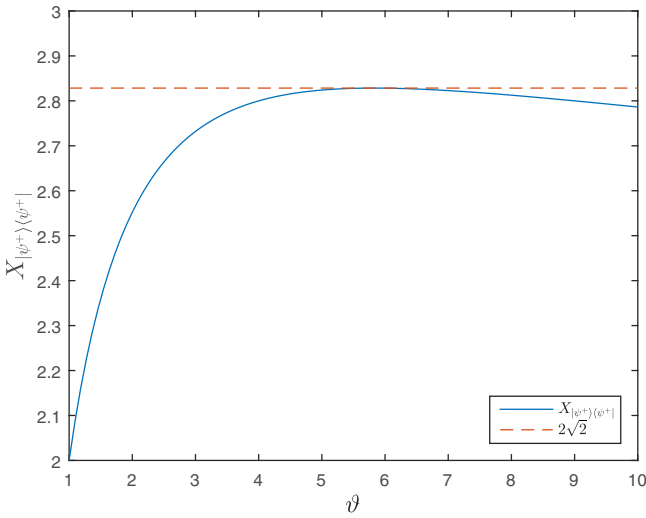
FIG. 1. The blue solid line is $X_{|\psi^+\rangle\langle\psi^+|}$ as a function of the parameter $\vartheta \in [1, 10]$ when we consider the bipartite biconditional state $(\Phi_{U_1}, \Phi_{U_2}) \otimes (\Phi_{V_1}, \Phi_{V_2})(|\psi^+\rangle\langle\psi^+|)$ from Example 8. The red dashed line corresponds to the Tsirelson bound $2\sqrt{2}$.

One may wonder whether there is or is not a connection between steering and Bell nonlocality. As we have already showed in Proposition 22, for measurements Bell nonlocality implies steering. We will show that for channels the same does not hold.

Example 9. Let $\dim(\mathcal{H}) = 2$. Let $\rho_W \in \mathfrak{D}_{\mathcal{H} \otimes \mathcal{H}}$ be given as in Example 7 as a partial trace over the state $|W\rangle\langle W|$. We already know that the state ρ_W is not steerable by any pair of channels. Consider the bipartite biconditional state $((\mathrm{id}, \mathrm{id}) \otimes (\mathrm{id}, \mathrm{id}))(\rho_W)$, if it is Bell local, then there must be a state $\sigma \in \mathfrak{D}_{\mathcal{H} \otimes \mathcal{H} \otimes \mathcal{H} \otimes \mathcal{H}}$ such that

$$\mathrm{Tr}_{13}(\sigma) = \mathrm{Tr}_{14}(\sigma) = \mathrm{Tr}_{23}(\sigma) = \mathrm{Tr}_{24}(\sigma) = \rho_W.$$

Observe that $\mathrm{Tr}_1(\sigma) \in \mathfrak{D}_{\mathcal{H} \otimes \mathcal{H} \otimes \mathcal{H}}$ is such that $\mathrm{Tr}_3(\mathrm{Tr}_1(\sigma)) = \mathrm{Tr}_4(\mathrm{Tr}_1(\sigma)) = \rho_W$ which implies that, according to our calculations in Example 7, we must have

$$\mathrm{Tr}_1(\sigma) = |W\rangle\langle W|.$$

According to [56, Lemma 3] this implies that there is a state $\rho \in \mathfrak{D}_{\mathcal{H}}$ such that $\sigma = \rho \otimes |W\rangle\langle W|$. This implies that we have $\mathrm{Tr}_{23}(\sigma) = \rho \otimes \frac{1}{3}(2|0\rangle\langle 0| + |1\rangle\langle 1|)$ which is clearly a separable state. This is a contradiction as we should have had $\mathrm{Tr}_{23}(\sigma) = \rho_W$, which is an entangled state.

XIV. CONCLUSIONS

We have introduced the general definition of compatibility of channels in general probabilistic theory through the idea of conditional channels. We have also shown that a naive idea for a compatibility test leads to a simple and straightforward formulation of steering and Bell nonlocality. These formulations of steering and Bell nonlocality are different even when we consider only measurements instead of channels. Throughout the paper, we have shown that all of our definitions and results are in correspondence with the known result for measurements

and we have also provided several examples and results about the introduced concepts in quantum theory.

The paper has opened several questions and areas of research. For example, a possible area of research would be to look at the structure of conditional states and conditional channels and to try to connect them to Bayesian theory.

Concerning the compatibility of channels, one may formulate different notions of degree of (in)compatibility or of robustness of compatibility in general probabilistic theory and look at their properties, in a similar way as it was already done in quantum theory [57]. For quantum channels, one may wonder which types of channels are compatible. This would generalize the no broadcasting theorem [56,58] which states that two unitary channels can not be compatible.

One may also consider our formulations of steering and Bell nonlocality as a case of the problem of finding a multipartite state with given marginals. Such problems were studied in recent years [59,60], but not in the form that would be applicable to the problems of steering and Bell nonlocality as incompatibility tests. This opens questions as to whether one may characterize the structure of the cone Q_{CD} and of other cones of interest in quantum theory. From a geometrical viewpoint, this question is closely tied to the question of existence of other Bell inequalities for channels than the one we presented. Existence and exact form of the generalized Bell inequalities are also a very interesting possible area of research.

We may also consider the use of steering and Bell nonlocality of channels in the context of quantum information theory and quantum communication. Both steering and Bell nonlocality of measurement were used to formulate quantum protocols and it is of great interest whether exploiting the steering and Bell nonlocality of channels may lead to even better or more useful applications.

One may also try and clarify the lack of connection between steering and Bell nonlocality of channels. As we have showed in Example 9, even if two channels can not steer a state, when applied to both parts of the state the resulting biconditional bipartite state may be Bell nonlocal. This may even have interesting applications in quantum theory of information as so far steering has been considered to lead to one-side device-independent protocols that were seen as a middle step between the original protocol and device-independent protocol.

It may also be interesting to consider the resource theories of channel incompatibility, of steering by channels, and of Bell nonlocality of channels. Several similar resource theories were already constructed (see [61] for a review).

[1] E. Schrödinger, Math. Proc. Cambridge Philos. Soc. **32**, 446 (1936).

[2] J. S. Bell, Physics (College. Park. Md) **1**, 195 (1964).

[3] A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. **47**, 777 (1935).

[4] T. Heinosaari, D. Reitzner, and P. Stano, Found. Phys. **38**, 1133 (2008).

[5] L.-Y. Hsu, Quantum Inf. Proc. **15**, 5167 (2016).

[6] T. Heinosaari, J. Kiukas, and D. Reitzner, Phys. Rev. A **92**, 022115 (2015).

[7] T. Heinosaari, J. Schultz, A. Toigo, and M. Ziman, Phys. Lett. A **378**, 1695 (2014).

[8] J. Bavaresco, M. T. Quintino, L. Guerini, T. O. Maciel, D. Cavalcanti, and M. T. Cunha, Phys. Rev. A **96**, 022110 (2017).

[9] D. Cavalcanti and P. Skrzypczyk, Rep. Prog. Phys. **80**, 024001 (2017).

[10] R. Uola, T. Moroder, and O. Gühne, Phys. Rev. Lett. **113**, 160403 (2014).

[11] R. Uola, C. Budroni, O. Gühne, and J. P. Pellonpää, Phys. Rev. Lett. **115**, 230402 (2015).

[12] M. T. Quintino, T. Vértesi, and N. Brunner, Phys. Rev. Lett. **113**, 160402 (2014).

[13] M. M. Wolf, D. Perez-Garcia, and C. Fernandez, Phys. Rev. Lett. **103**, 230402 (2009).

[14] M. Banik, Md. R. Gazi, S. Ghosh, and G. Kar, Phys. Rev. A **87**, 052125 (2013).

[15] F. Hirsch, M. T. Quintino, and N. Brunner, arXiv:1707.06960.

[16] B. Erika and T. Vértesi, arXiv:1705.10069.

[17] P. Girdhar and E. G. Cavalcanti, Phys. Rev. A **94**, 032317 (2016).

[18] D. Cavalcanti and P. Skrzypczyk, Phys. Rev. A **93**, 052112 (2016).

[19] T. Heinosaari, T. Miyadera, and M. Ziman, J. Phys. A: Math. Theor. **49**, 123001 (2016).

[20] M. Sedlák, D. Reitzner, G. Chiribella, and M. Ziman, Phys. Rev. A **93**, 052323 (2016).

[21] T. Heinosaari and T. Miyadera, J. Phys. A: Math. Theor. **50**, 135302 (2017).

[22] M. Banik, S. Das, and A. S. Majumdar, Phys. Rev. A **91**, 062124 (2015).

[23] S. N. Filippov, T. Heinosaari, and L. Leppäjärvi, Phys. Rev. A **95**, 032127 (2017).

[24] M. Banik, J. Math. Phys. **56**, 052101 (2015).

[25] N. Stevens and P. Busch, Phys. Rev. A **89**, 022123 (2014).

[26] A. Jenčová, arXiv:1705.08008.

[27] H. M. Wiseman, S. J. Jones, and A. C. Doherty, Phys. Rev. Lett. **98**, 140402 (2007).

[28] B. S. Cirel'son, Lett. Math. Phys. **4**, 93 (1980).

[29] J. Barrett, Phys. Rev. A **75**, 032304 (2007).

[30] R. W. Spekkens, Phys. Rev. A **75**, 032110 (2007).

[31] S. Popescu and D. Rohrlich, Found. Phys. **24**, 379 (1994).

[32] P. Janotta and R. Lal, Phys. Rev. A **87**, 052131 (2013).

[33] W. Van Dam, Nat. Comput. **12**, 9 (2013).

[34] Y. Quek and P. W. Shor, Phys. Rev. A **95**, 052329 (2017).

[35] S. Aaronson, Phys. Rev. A **71**, 032325 (2005).

[36] N. Johansson and J.-Å. Larsson, Quantum Inf. Proc. **16**, 233 (2017).

[37] R. T. Rockafellar, *Convex Analysis*, Princeton Landmarks in Mathematics and Physics (Princeton University Press, Princeton, NJ, 1997).

[38] S. Boyd and L. Vandenberghe, *Convex Optimization*, Berichte über verteilte messyseme (Cambridge University Press, Cambridge, 2004).

[39] D. J. Foulis and M. K. Bennett, Found. Phys. **24**, 1331 (1994).

[40] L. Asimow and A. J. Ellis, *Convexity Theory and its Applications in Functional Analysis*, L.M.S. Monographs (Academic Press, New York, 1980).

[41] A. Kissinger and S. Uijlen, arXiv:1701.04732.

[42] R. A. Ryan, *Introduction to Tensor Products of Banach Spaces*, Springer Monographs in Mathematics (Springer, London, 2013).

[43] T. Heinosaari and M. Ziman, *The Mathematical Language of Quantum Theory. From Uncertainty to Entanglement* (Cambridge University Press, Cambridge, 2012).

[44] I. Namioka and R. Phelps, Pacific J. Math. **31**, 469 (1969).

[45] M. Plávala, Phys. Rev. A **94**, 042108 (2016).

[46] A. Jenčová, arXiv:1207.5420.

[47] M. Ziman, Phys. Rev. A **77**, 062112 (2008).

[48] C. Branciard, E. G. Cavalcanti, S. P. Walborn, V. Scarani, and H. M. Wiseman, Phys. Rev. A **85**, 010301(R) (2012).

[49] A. K. Ekert, Phys. Rev. Lett. **67**, 661 (1991).

[50] A. Acin, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Phys. Rev. Lett. **98**, 230501 (2007).

[51] Č. Brukner, M. Zukowski, J.-W. Pan, and A. Zeilinger, Phys. Rev. Lett. **92**, 127901 (2004).

[52] W.-Y. Hwang, H.-Y. Su, and J. Bae, Phys. Rev. A **95**, 062313 (2017).

[53] R. Colbeck and A. Kent, J. Phys. A: Math. Theor. **44**, 095305 (2011).

[54] F. Bischof, H. Kampermann, and D. Bruß, Phys. Rev. A **95**, 062305 (2017).

[55] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Rev. Mod. Phys. **86**, 419 (2014).

[56] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, arXiv:quant-ph/0611295.

[57] E. Haapasalo, J. Phys. A: Math. Theor. **48**, 255303 (2015).

[58] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, Phys. Rev. Lett. **99**, 240501 (2007).

[59] A. Klyachko, arXiv:quant-ph/0409113.

[60] N. Wyderka, F. Huber, and O. Gühne, Phys. Rev. A **96**, 010102 (2017).

[61] M. Horodecki and J. Oppenheim, Int. J. Mod. Phys. B **27**, 1345019 (2013).

No-free-information principle in general probabilistic theories

Teiko Heinosaari[*] and Leevi Leppäjärvi[†]

QTF Centre of Excellence, Department of Physics and Astronomy, University of Turku, Turku 20014, Finland

Martin Plávala[‡]

Mathematical Institute, Slovak Academy of Sciences, Štefánikova 49, Bratislava, Slovakia

In quantum theory, the no-information-without-disturbance and no-free-information theorems express that those observables that do not disturb the measurement of another observable and those that can be measured jointly with any other observable must be trivial, i.e., coin tossing observables. We show that in the framework of general probabilistic theories these statements do not hold in general and continue to completely specify these two classes of observables. In this way, we obtain characterizations of the probabilistic theories where these statements hold. As a particular class of state spaces we consider the polygon state spaces, in which we demonstrate our results and show that while the no-information-without-disturbance principle always holds, the validity of the no-free-information principle depends on the parity of the number of vertices of the polygons.

I. INTRODUCTION

Quantum theory implies three simple, yet significant and powerful theorems: the no-broadcasting theorem, the no-information-without-disturbance theorem, and the no-free-information theorem. The no-broadcasting theorem says that quantum states cannot be copied; the no-information-without-disturbance theorem states that a quantum observable that can be measured without any disturbance must be trivial, meaning that it does not give any information on the input state; and the no-free-information theorem states that a quantum observable that can be measured jointly with any other observable must be a trivial observable. In other words, there is no free information, in the sense that a measurement of any non-trivial observable precludes the measurement of some other observable.

Each of the previous three statements can be formulated in the framework of general probabilistic theories (GPTs for short). In this context we find it better to call them *principles* instead of theorems as they are not valid in all probabilistic theories. In particular, the no-broadcasting principle is known to be valid in any non-classical general probabilistic theory [1, 2]. In this work, we concentrate on the latter two principles and investigate their validity in the realm of GPTs. The no-information-without-disturbance principle has been shown to hold within GPTs with some additional assumptions, such as purification [3]; however, the validity of this principle has only been mentioned
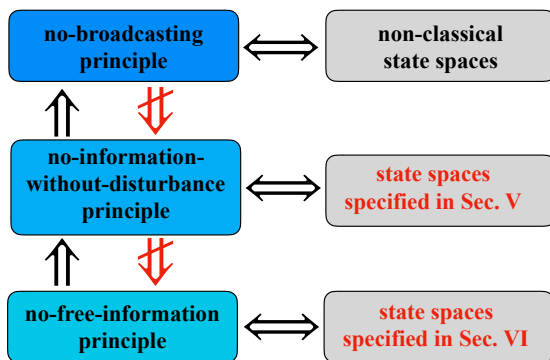

FIG. 1. Summary of the main results.

[*] teiko.heinosaari@utu.fi

[†] leille@utu.fi

[‡] martin.plavala@mat.savba.sk

in [4] but never fully investigated in all probabilistic theories. The no-free-information principle seems to have not been investigated at all in any other theory than quantum theory.

Amongst these principles, no-free-information principle is conceptually the weakest, with no-broadcasting the strongest: If the no-free-information principle is valid in some GPT—that is, for every non-trivial observable there exists another incompatible with it—then the no-information-without-disturbance principle must also be valid, as a non-disturbing observable would be compatible with every other observable. Furthermore, if the no-information-without-disturbance principle is valid and hence no non-trivial observable is non-disturbing, then the no-broadcasting principle has to hold, otherwise we would be capable of using the broadcasting map to create non-trivial non-disturbing observables.

We will define three classes of observables, the first one consisting of those observables that always yield a constant outcome independent of the measured state, the second one consisting of those observables that can be measured without any disturbance and the third one consisting of those observables that are compatible with any other observable. We will then characterize these classes, enabling us to show that the properties are different in some GPTs. We will also derive a necessary and sufficient criterion for a GPT to have both the no-information-without-disturbance principle and no-free-information principle valid be valid. Finally, we demonstrate the difference between the three principles by analyzing them in polygon state spaces. The main results of our investigation are summarized in Fig. 1.

II. MOTIVATING EXAMPLE

In this section we will present a simple example to motivate our current investigation. A proper mathematical formulation of the general framework will follow in later sections; in the following example we are going to work with the set $B_h(\mathcal{H})$ of square self-adjoint matrices over a finite dimensional Hilbert space $\mathcal{H}$. We denote by $\mathbb{1}$ the identity matrix and 0 the zero matrix. For $A \in B_h(\mathcal{H})$, we write $A \geq 0$ if A is positive-semidefinite. Let $A, B \in B_h(\mathcal{H})$, then if $A \geq 0$ and $\text{Tr}(A) = 1$, then A is a state and if $0 \leq B \leq \mathbb{1}$, then B is an effect. We refer the reader to [5] for a more throughout treatment of states and effects and their operational meanings in quantum theory.

Imagine that we have an imperfect state preparation device that is meant to prepare qubits in a state ρ, but may malfunction and prepare a qutrit in a state σ. Moreover we assume that the machine malfunctions with a probability p_e, thereby the final state should be a mixture of ρ and σ with probabilities $1 - p_e$ and p_e, respectively. This means that the machine is going to output a state Ψ that should formally be given as $\Psi = (1 - p_e)\rho + p_e\sigma$. But how does one understand the mixture of the 2×2 matrix ρ and the 3×3 matrix σ? And how does one describe the output state-space of such a machine?

Qubits are effectively a spin-$\frac{1}{2}$ systems and qutrits a spin-1 systems, hence the joint Hilbert space $\mathcal{H}$ containing both representations of the group $\text{SU}(2)$ is going to be 5 dimensional and divided into two superselection sectors [6] of dimensions 2 and 3, corresponding to the qubit and qutrit respectively. The output state Ψ is going to be a block-diagonal 5×5 matrix given as

$$\Psi = \begin{pmatrix} (1 - p_e)\rho & 0 \\ 0 & p_e\sigma \end{pmatrix}.$$

Let M be an effect on $\mathcal{H}$, then M is of the form

$$M = \begin{pmatrix} M_1 & M_3 \\ M_3^* & M_2 \end{pmatrix},$$

where M_1, M_2, M_3 are matrices of corresponding sizes. We have

$$\text{Tr}(\Psi M) = \text{Tr}\begin{pmatrix} (1 - p_e)\rho M_1 & (1 - p_e)\rho M_3 \\ p_e\sigma M_3^* & p_e\sigma M_2 \end{pmatrix} = (1 - p_e)\text{Tr}(\rho M_1) + p_e\text{Tr}(\sigma M_2),$$

hence from the operational viewpoint we may set $M_3 = 0$ without loss of generality.

Let N be an effect given as

$$N = \begin{pmatrix} \mathbb{1} & 0 \\ 0 & 0 \end{pmatrix}$$

then N and $\mathbb{1} - N$ form a projective POVM. Moreover, both N and $\mathbb{1} - N$ commute with all other block-diagonal effects, hence we conclude that the observable corresponding to the POVM N, $\mathbb{1} - N$ is compatible with every other measurement.

This is hardly a surprise, rather a known property of the superselection sectors. Yet this opens the questions of whether this is the only case when an observable is compatible with every other observable; whether no-information-without-disturbance still holds; and whether an observable does not disturb any other observables if it is compatible with them all.

As we saw in this example, we need to at least describe the set of states containing only block-diagonal matrices. For this reason we will work in the GPT formalism as it will provide a unified, cleaner and better suited apparatus for our calculations.

III. PRELIMINARIES

In the GPT framework we assume that a state space is convex as we want to interpret convex combinations as mixtures of states. To describe observables, we will introduce effects as functions that assign probabilities to states.

A. Structure of general probabilistic theories

A state space $\mathcal{S}$ is a compact convex subset of an ordered real finite-dimensional vector space $\mathcal{V}$ such that $\mathcal{S}$ is a compact base for a generating positive cone $\mathcal{V}_+ = \{x \in \mathcal{V} \,|\, x \geq 0\}$. Let $\mathcal{V}^*$ denote the dual vector space to $\mathcal{V}$, then the effect algebra $\mathcal{E}(\mathcal{S}) \subset \mathcal{V}^*$ is the set of linear functionals $e : \mathcal{V} \to \mathbb{R}$ such that $0 \leq e(x) \leq 1$ for every $x \in \mathcal{S}$. The zero and the unit effects $o \in \mathcal{E}(\mathcal{S})$ and $u \in \mathcal{E}(\mathcal{S})$ are the unique effects satisfying $o(x) = 0$ and $u(x) = 1$ for all $x \in \mathcal{S}$.

The state space can be expressed as

$$\mathcal{S} = \{x \in \mathcal{V} \,|\, x \geq 0, \ u(x) = 1\},$$

i.e. as an intersection of the positive cone $\mathcal{V}_+$ and an affine hyperplane determined by the unit effect u on $\mathcal{V}$. Similarly we can define subnormalised states as

$$\mathcal{S}^{\leq 1} = \{x \in \mathcal{V} \,|\, x \geq 0, \ u(x) \leq 1\}.$$

If $\dim(\mathrm{aff}(\mathcal{S})) = d$, we say that the state space $\mathcal{S}$ is d-dimensional, and then we can choose $\mathcal{V}$ such that $\dim(\mathcal{V}) = \dim(\mathcal{V}^*) = d + 1$. It follows that the effects can be expressed as linear functionals on $\mathcal{V}$ such that

$$\mathcal{E}(\mathcal{S}) = \{e \in \mathcal{V}^* \,|\, o \leq e \leq u\},$$

where the partial order in the dual space is the dual order defined by the positive dual cone $\mathcal{V}_+^* = \{f \in \mathcal{V}^* \,|\, f(x) \geq 0 \text{ for all } x \in \mathcal{V}_+\}$ of $\mathcal{V}_+$. In fact $\mathcal{E}(\mathcal{S})$ is then just the intersection of the positive dual cone $\mathcal{V}_+^*$ and the set $u - \mathcal{V}_+^*$.

We say that a non-zero effect $e \in \mathcal{E}(\mathcal{S})$ is indecomposable if a decomposition $e = e_1 + e_2$ for some effects $e_1, e_2 \in \mathcal{E}(\mathcal{S})$ is possibly only if e_1 and e_2 are positive scalar multiples of e [7]. The indecomposable effects are exactly the ones that lie on the extreme rays of the positive dual cone $\mathcal{V}_+^*$.

When dealing with systems composed of several systems we have to prescribe a procedure for how to construct a joint state space of the composed system. Mathematically, this amounts of specifying a tensor product. We are going to use a tensor product only in cases where the other state space is classical. Therefore, there is a unique choice known as the minimal tensor product [8].

Definition 1. Let $\mathcal{S}_1$, $\mathcal{S}_2$ be state spaces, then their minimal tensor product, denoted as $\mathcal{S}_1 \dot{\otimes} \mathcal{S}_2$, is given as

$$\mathcal{S}_1 \dot{\otimes} \mathcal{S}_2 = \mathrm{conv}\left(\{x_1 \otimes x_2 \,|\, x_1 \in \mathcal{S}_1, x_2 \in \mathcal{S}_2\}\right).$$

B. Observables and channels

In this section we will introduce the main objects of interest to us - observables, channels and compatibility. We will begin with observables and their compatibility, and build our way towards channels.

Definition 2. An observable A with a finite outcome set Ω_A on a state space $\mathcal{S}$ is a mapping $\mathsf{A} : x \to \mathsf{A}_x$ from the outcome set Ω_A to the set of effects $\mathcal{E}(\mathcal{S})$ such that $\sum_{x \in \Omega_\mathsf{A}} \mathsf{A}_x = u$. The set of observables on $\mathcal{S}$ is denoted by $\mathcal{O}(\mathcal{S})$. For each $\mathsf{A} \in \mathcal{O}(\mathcal{S})$ we refer to Ω_A as the outcome set of A.

Let $A, B \in \mathcal{O}(\mathcal{S})$ with respective outcome sets Ω_A, Ω_B. We say that B is a post-processing of A, denoted by $A \to B$, if there is a right-stochastic matrix ν with elements ν_{xy}, $x \in \Omega_A$, $y \in \Omega_B$, $0 \le \nu_{xy} \le 1$, $\sum_{y \in \Omega_B} \nu_{xy} = 1$ such that

$$B_y = \sum_{x \in \Omega_A} \nu_{xy} A_x \,,$$

in which case we also write $B = \nu \circ A$. The operational interpretation is straightforward: we have $A \to B$ only if we can obtain the probabilities given by B from the probabilities given by A. The condition $\sum_{y \in \Omega_B} \nu_{xy} = 1$ follows from $\sum_{y \in \Omega_B} B_y = u$.

Definition 3. A collection of m observables $A^{(1)}, \ldots, A^{(m)} \in \mathcal{O}(\mathcal{S})$ is compatible if there exists an observable $J_{A^{(1)}, \ldots, A^{(m)}} \in \mathcal{O}(\mathcal{S})$ such that $J_{A^{(1)}, \ldots, A^{(m)}} \to A^{(i)}$ for all $i = 1, \ldots, m$. If two observables A and B are compatible we denote it $A \circledcirc B$.

Compatibility of observables and of observables and channels will play a central role in our calculations.

Definition 4. Let $\mathcal{S}_1$, $\mathcal{S}_2$ be a state spaces. An operation is an affine map $\Psi : \mathcal{S}_1 \to \mathcal{S}_2^{\le 1}$. A channel is an affine map $\Phi : \mathcal{S}_1 \to \mathcal{S}_2$. The set of channels from $\mathcal{S}_1$ to $\mathcal{S}_2$ is denoted by $\mathfrak{C}(\mathcal{S}_1, \mathcal{S}_2)$ and in the special case where $\mathcal{S}_1 = \mathcal{S}_2 \equiv \mathcal{S}$ we denote it by $\mathfrak{C}(\mathcal{S})$.

In quantum theory we also require channels to be completely positive, but we omit this within GPTs as in general it is problematic to specify what complete positivity means.

Let $\mathcal{S}$ be a state space and let $A \in \mathcal{O}(\mathcal{S})$ with an outcome set Ω_A of n elements. We can identify the points of Ω_A with the extreme points of a simplex, which allows us to form convex combinations of the points of Ω_A. Moreover we will denote this simplex $\mathcal{P}(\Omega_A)$ and its extreme points $\delta_1, \ldots, \delta_n$ as they correspond to classical measures on Ω_A supported on a single point. Now we can see the observable A as a channel $A : \mathcal{S} \to \mathcal{P}(\Omega_A)$. Furthermore, a post-processing ν can be seen as a channel mapping the classical state spaces corresponding to outcome sets of observables.

As mentioned above, similarly to compatibility of measurements, we can introduce the compatibility of a measurement and a channel. The central role is going to be played by a generalization of partial trace, which is as follows: let $\mathcal{S}_1$, $\mathcal{S}_2$ be state spaces and let $x \in \mathcal{S}_1 \dot{\otimes} \mathcal{S}_2$, then by definition we have $x = \sum_{i=1}^n \lambda_i x_i^1 \otimes x_i^2$ for some $x_i^1 \in \mathcal{S}_1$, $x_i^2 \in \mathcal{S}_2$, $\lambda_i \ge 0$ for $i \in \{1, \ldots, n\}$ and $\sum_{i=1}^n \lambda_i = 1$. We then define the maps $u_1 : \mathcal{S}_1 \dot{\otimes} \mathcal{S}_2 \to \mathcal{S}_2$ and $u_2 : \mathcal{S}_1 \dot{\otimes} \mathcal{S}_2 \to \mathcal{S}_1$ as

$$u_1(x) = \sum_{i=1}^n \lambda_i u(x_i^1) x_i^2 = \sum_{i=1}^n \lambda_i x_i^2 \,,$$
$$u_2(x) = \sum_{i=1}^n \lambda_i u(x_i^2) x_i^1 = \sum_{i=1}^n \lambda_i x_i^1 \,.$$

The maps u_1, u_2 are direct generalizations of partial traces. The definitions may be easily generalized also for entangled states but this is out of the scope of what we will need in future calculations.

Definition 5. A channel $\Phi : \mathcal{S} \to \mathcal{S}$ is compatible with an observable $A \in \mathcal{O}(\mathcal{S})$ with outcome set Ω_A if and only if there is a channel $\tilde{\Phi} : \mathcal{S} \to \mathcal{S} \dot{\otimes} \mathcal{P}(\Omega_A)$ such that for all $x \in \mathcal{S}$ we have

$$\Phi(x) = (u_2 \circ \tilde{\Phi})(x),$$
$$A(x) = (u_1 \circ \tilde{\Phi})(x),$$

where $\circ$ denotes the composition of maps.

If the channel Φ were an observable, we would obtain a definition of compatibility of observables which can be shown to be equivalent to Def. 3; see [9]. In a similar fashion one may also formulate the definition of compatibility of channels [10].

We will start with a simple lemma for the compatibility of an observable and a channel.

Lemma 1. *A channel $\Phi \in \mathfrak{C}(\mathcal{S})$ and an n-outcome observable $A \in \mathcal{O}(\mathcal{S})$ are compatible if and only if for $i \in \{1, \ldots, n\}$ there are operations $\Phi_i : \mathcal{S} \to \mathcal{V}+$ such that*

$$\Phi = \sum_{i=1}^n \Phi_i, \tag{1}$$

$$A_i = u \circ \Phi_i. \tag{2}$$

Proof. Let $\Omega_A = \{\delta_1, \ldots, \delta_n\}$ denote the outcome space with n points. Moreover let $b_1, \ldots, b_n$ denote the dual base of affine functions $\mathcal{P}(\Omega_A) \to \mathbb{R}$, such that $b_i(\delta_j) = 1$ if and only if $i = j$. If Φ and A are compatible, then there exists a channel $\tilde{\Phi} : \mathcal{S} \to \mathcal{S} \dot{\otimes} \mathcal{P}(\Omega_A)$ such that $\Phi = u_2 \circ \tilde{\Phi}$ and $A = u_1 \circ \tilde{\Phi}$.

In general, we have $\tilde{\Phi} \in \mathcal{V}^* \otimes \mathcal{V} \otimes \mathcal{P}(\Omega_A)$, i.e.

$$\tilde{\Phi} = \sum_{i=1}^n \sum_{j \in J} f_{ij} \otimes \psi_j \otimes \delta_i$$

for some $f_{ij} \in \mathcal{V}^*$ and $\psi_j \in \mathcal{V}$ and for some index j from a finite index set J. Denote $\Phi_i = \sum_{j \in J} f_{ij} \otimes \psi_j$ and notice that Φ_i are linear maps $\mathcal{V} \to \mathcal{V}$.

Since $\tilde{\Phi}$ must be a channel then $b_i \circ \tilde{\Phi} : \mathcal{S} \to \mathcal{S}$ must also be a positive map and since $b_i \circ \tilde{\Phi} = \Phi_i$, we see that Φ_i are positive maps. Since $\tilde{\Phi}$ is a joint channel of Φ and A we must have

$$\Phi = u_2 \circ \tilde{\Phi} = \sum_{i=1}^n \Phi_i,$$

$$A = u_1 \circ \tilde{\Phi} = \sum_{i=1}^n (u \circ \Phi_i) \otimes \delta_i.$$

$\sum_{i=1}^n (u \circ \Phi_i)(x) = 1$ for all $x \in \mathcal{S}$ implies that Φ_i are operations.

If there exist operations Φ_i satisfying (1) and (2), then define $\tilde{\Phi} = \sum_{i=1}^n \Phi_i \otimes \delta_i$. Positivity and normalisation of $\tilde{\Phi}$ follows from the positivity of Φ_i and (2). The fact that $\tilde{\Phi}$ is a joint channel of Φ and A follows from (1) and (2). $\qquad\square$

IV. FORMULATION OF THE TWO PRINCIPLES

The purpose of measuring an observable is to learn something about the input state via the obtained measurement outcome probability distribution. An observable is called *trivial* if it cannot provide any information on input states. More precisely, this means that a trivial observable T assigns the same measurement outcome probability distribution to all states, i.e., $T = pu$ for some probability distribution p on Ω_T. Physically speaking, a measurement of a trivial observable can be implemented simply by rolling a dice and producing a probability distribution independently of the input state. We denote by $\mathcal{T}_1$ the set of all trivial observables, i.e.,

$$\mathcal{T}_1 = \{T \in \mathcal{O}(\mathcal{S}) \,|\, T_x(s) = T_x(s') \; \forall x \in \Omega_T, \; \forall s, s' \in \mathcal{S}\}$$
$$= \{T \in \mathcal{O}(\mathcal{S}) \,|\, \exists p \in \mathcal{P}(\Omega_T) : \; T_x = p(x)u \; \forall x \in \Omega_T\}.$$

From the banal structure of trivial observables it follows that any such observable is compatible with every other observable. Formally, if $T = pu$ is a trivial observable and A is some other observable, then we can define an observable $J_{T,A}$ with effects $J_{T,A}(x, y) = p(x)A_y$, and we have $\sum_x J_{T,A}(x, y) = A_y$ and $\sum_y J_{T,A}(x, y) = T_x$.

Furthermore, a trivial observable is compatible with every channel. Namely, if $T = pu$ is a trivial observable and Φ is a channel, then we can define operations $\Phi_i : \mathcal{S} \to \mathcal{V}_+$ as $\Phi_i = p(i)\Phi$ for all $i \in \Omega_T$. Clearly, then $\sum_{i \in \Omega_T} \Phi_i = \Phi$ and $(u \circ \Phi_i)(x) = p(i) = T_i(x)$ for all $i \in \Omega_T$ so that by Lemma 1 we conclude that T and Φ are compatible.

These two features of trivial observables raise natural questions: are there observables other than trivial ones that have these features? If so, what is the structure of such observables? As we have seen in Sec. II, the answer to the first question is affirmative, hence the second question urges an investigation.

To properly analyze the two mentioned features, we consider them as independent properties that determine a subclass of observables. Hence, for a state space $\mathcal{S}$, we define the following subsets of observables:

$$\mathcal{T}_2 = \{T \in \mathcal{O}(\mathcal{S}) \,|\, T \infty \Phi \; \forall \Phi \in \mathfrak{C}(\mathcal{S})\},$$
$$\mathcal{T}_3 = \{T \in \mathcal{O}(\mathcal{S}) \,|\, T \infty A \; \forall A \in \mathcal{O}(\mathcal{S})\}.$$

If an observable T is compatible with the identity channel id, then T is compatible with any channel $\Phi \in \mathfrak{C}(\mathcal{S})$. Namely, suppose that T is compatible with id, so there exist operations $\Psi_i : \mathcal{S} \to \mathcal{V}_+$ such that $\sum_{i \in \Omega_T} \Psi_i = id$ and $u \circ \Psi_i = T_i$. Then we can define a new set of operations as $\Phi \circ \Psi_i$, and these operations give $\sum_{i \in \Omega_T} \Phi \circ \Psi_i = \Phi \circ id = \Phi$ and $u \circ (\Phi \circ \Psi_i) = (u \circ \Phi) \circ \Phi_i = u \circ \Phi_i = T_i$. Therefore, we can concisely write

$$\mathcal{T}_2 = \{T \in \mathcal{O}(\mathcal{S}) \,|\, T \infty id\}.$$

We conclude that $\mathcal{T}_2$ *is the set of observables that can be measured without causing any disturbance.*

Now, suppose that $\mathsf{T} \in \mathcal{T}_2$, so there exist operations $\Phi_i : \mathcal{S} \to \mathcal{V}_+$ such that $\sum_{i \in \Omega_\mathsf{T}} \Phi_i = id$ and $u \circ \Phi_i = \mathsf{T}_i$ for all $i \in \Omega_\mathsf{T}$. If $\mathsf{A} \in \mathcal{O}(\mathcal{S})$, we define a joint observable G of A and T by $\mathsf{G}_{ij} = \mathsf{A}_j \circ \Phi_i$ for all $i \in \Omega_\mathsf{T}$ and $j \in \Omega_\mathsf{A}$. We then see that

$$\sum_j \mathsf{G}_{ij} = \sum_j (\mathsf{A}_j \circ \Phi_i) = \left(\sum_j \mathsf{A}_j\right) \circ \Phi_i = u \circ \Phi_i = \mathsf{T}_i,$$

$$\sum_i \mathsf{G}_{ij} = \sum_i (\mathsf{A}_j \circ \Phi_i) = \mathsf{A}_j \circ \left(\sum_i \Phi_i\right) = \mathsf{A}_j \circ id = \mathsf{A}_j$$

for all $i \in \Omega_\mathsf{T}$ and $j \in \Omega_\mathsf{A}$. Thus, A and T are compatible, and since A was an arbitrary observable, it follows that $\mathsf{T} \in \mathcal{T}_3$. We conclude that

$$\mathcal{T}_1 \subseteq \mathcal{T}_2 \subseteq \mathcal{T}_3 .$$

These three sets and the previous chain of inclusion allows us to give a simple and concise formulation of the two principles: *The no-information-without-disturbance principle means that* $\mathcal{T}_2 = \mathcal{T}_1$, *while the no-free-information principle means that* $\mathcal{T}_3 = \mathcal{T}_1$.

V.   CHARACTERIZATION OF $\mathcal{T}_2$

The aim of this section is to characterize non-disturbing observables and the structure of the state spaces they may exist on. We will have to introduce additional mathematical results to provide the full description of such state spaces.

A.   Direct sum of state spaces

We will introduce a direct sum of state spaces as a generalized description of using only block-diagonal quantum states. Our aim is to mathematically formalize the operational idea of having an ordered pair of weighted states from two different state spaces.

Definition 6. Let $\mathcal{V}_1$, $\mathcal{V}_2$ be real finite-dimensional vector spaces and let $\mathcal{S}_1 \subset \mathcal{V}_1$ and $\mathcal{S}_2 \subset \mathcal{V}_2$ be state spaces. We define a state space $\mathcal{S}_1 \oplus \mathcal{S}_2 \subset \mathcal{V}_1 \times \mathcal{V}_2$ as the set of ordered and weighted pairs of states from $\mathcal{S}_1$ and $\mathcal{S}_2$, i.e.,

$$\mathcal{S}_1 \oplus \mathcal{S}_2 = \{(\lambda x_1, (1-\lambda)x_2) \,|\, x_1 \in \mathcal{S}_1, x_2 \in \mathcal{S}_2, \lambda \in [0,1]\}.$$

Given state spaces $\mathcal{S}_1, \ldots, \mathcal{S}_n$ one can define $\mathcal{S}_1 \oplus \ldots \oplus \mathcal{S}_n$ in a similar fashion as a subset of $\mathcal{V}_1 \times \ldots \mathcal{V}_n$, i.e., one would have

$$\mathcal{S}_1 \oplus \cdots \oplus \mathcal{S}_n = \{(\lambda_1 x_1, \ldots, \lambda_n x_n) \,|\, x_i \in \mathcal{S}_i, \lambda_i \geq 0, \forall i \in \{1, \ldots, n\}, \sum_{i=1}^n \lambda_i = 1\}.$$

In what follows we will present a few basic results about $\mathcal{S}_1 \oplus \mathcal{S}_2$. We will limit only to direct sum of two state spaces for the sake of not drowning in a sea of symbols, but it will be straightforward to see that all of the results hold for any finite direct sum as well.

Proposition 1. $\mathcal{E}(\mathcal{S}_1 \oplus \mathcal{S}_2) = \mathcal{E}(\mathcal{S}_1) \times \mathcal{E}(\mathcal{S}_2)$, *where* $\mathcal{E}(\mathcal{S}_1) \times \mathcal{E}(\mathcal{S}_2) = \{(e_1, e_2) \,|\, e_1 \in \mathcal{E}(\mathcal{S}_1), e_2 \in \mathcal{E}(\mathcal{S}_2)\}$.

Proof. $\mathcal{S}_1 \oplus \mathcal{S}_2 \subset \mathcal{V}_1 \times \mathcal{V}_2$ so we must have $\mathcal{E}(\mathcal{S}_1 \oplus \mathcal{S}_2) \subset \mathcal{V}_1^* \times \mathcal{V}_2^*$. Let $(e_1, e_2) \in \mathcal{V}_1^* \times \mathcal{V}_2^*$ and let $(\lambda x_1, (1-\lambda)x_2) \in \mathcal{S}_1 \oplus \mathcal{S}_2$, then from

$$(e_1, e_2)((\lambda x_1, (1-\lambda)x_2)) = \lambda e_1(x_1) + (1-\lambda)e_2(x_2) \tag{3}$$

it follows that $\mathcal{E}(\mathcal{S}_1) \times \mathcal{E}(\mathcal{S}_2) \subset \mathcal{E}(\mathcal{S}_1 \oplus \mathcal{S}_2)$. Assuming $(e_1, e_2) \in \mathcal{E}(\mathcal{S}_1 \oplus \mathcal{S}_2)$ and setting $\lambda = 0$ and $\lambda = 1$ in (3) we get $e_1 \in \mathcal{E}(\mathcal{S}_1)$ and $e_2 \in \mathcal{E}(\mathcal{S}_2)$. $\qquad\square$

It follows that if $\mathsf{A} \in \mathcal{O}(\mathcal{S}_1 \oplus \mathcal{S}_2)$, then we have $\mathsf{A}_i = (\mathsf{A}_i^1, \mathsf{A}_i^2)$ for some $\mathsf{A}^1 \in \mathcal{O}(\mathcal{S}_1)$, $\mathsf{A}^2 \in \mathcal{O}(\mathcal{S}_2)$.

Proposition 2. *Let* $\mathsf{A}, \mathsf{B} \in \mathcal{O}(\mathcal{S}_1 \oplus \mathcal{S}_2)$, *such that* $\mathsf{A}_i = (\mathsf{A}_i^1, \mathsf{A}_i^2)$, $\mathsf{B}_j = (\mathsf{B}_j^1, \mathsf{B}_j^2)$, *then* $\mathsf{A} \odot \mathsf{B}$ *if and only if* $\mathsf{A}^1 \odot \mathsf{B}^1$ *and* $\mathsf{A}^2 \odot \mathsf{B}^2$.

Proof. If $A^1 \text{ ⊚ } B^1$ and $A^2 \text{ ⊚ } B^2$ then $A \text{ ⊚ } B$ as we can form the joint observable as $(J_{A,B})_k = ((J_{A^1,B^1})_k, (J_{A^2,B^2})_k)$ and apply the respective post-processings to the respective observables, hence $A \text{ ⊚ } B$. Note that to make the observables have the same number of outcomes, we can always pad out one with zero effects corresponding to some extra outcomes that never happen.

If $A \text{ ⊚ } B$, then by restricting the state space only to states of the form $(x_1, 0) \in \mathcal{S}_1 \oplus \mathcal{S}_2$, where $x_1 \in \mathcal{S}_1$ it follows that $A^1 \text{ ⊚ } B^1$ are compatible as we can obtain J_{A^1,B^1} from $J_{A,B}$. $A^2 \text{ ⊚ } B^2$ follows in the same manner. $\square$

This explains our motivational example in Sec. II. One can also prove a similar result for the compatibility of an observable and a channel, but we will leave that for the next section, where we will investigate the conditions for the compatibility of an observable and the identity channel $id : \mathcal{S} \to \mathcal{S}$, where direct sums of state spaces will play a role.

This last result will help us identify the direct sum structure of a state space.

Proposition 3. *Let $\mathcal{S}$ be a state space and let $\mathcal{S}_1, \mathcal{S}_2 \subset \mathcal{S}$ be convex, closed, $\mathrm{conv}(\mathcal{S}_1 \cup \mathcal{S}_2) = \mathcal{S}$ and for every $x \in \mathcal{S}$ there are unique $x_1 \in \mathcal{S}_1$, $x_2 \in \mathcal{S}_2$ and $\lambda \in [0, 1]$ such that $x = \lambda x_1 + (1 - \lambda)x_2$. Then $\mathcal{S} = \mathcal{S}_1 \oplus \mathcal{S}_2$.*

Proof. Let $\mathcal{V}_1$ and $\mathcal{V}_2$ denote the subspaces of $\mathcal{V}$ generated by $\mathcal{S}_1$ and $\mathcal{S}_2$ respectively. Define map $P : \mathcal{S} \to \mathcal{V}_1 \times \mathcal{V}_2$ given for $x \in \mathcal{S}$, $x = \lambda x_1 + (1 - \lambda)x_2$, $x_1 \in \mathcal{S}_1$, $x_2 \in \mathcal{S}_2$ as $P(x) = (\lambda x_1, (1 - \lambda)x_2)$. It follows that we have $P : \mathcal{S} \to \mathcal{S}_1 \oplus \mathcal{S}_2$, moreover one can easily see that P is an affine isomorphism. It follows that $\mathcal{S}$ is affinely isomorphic to $\mathcal{S}_1 \oplus \mathcal{S}_2$, the result follows by simply omitting the isomorphism. $\square$

B. Compatibility of an observable and the identity channel

We are going to derive conditions for an observable to be compatible with the identity channel $id : \mathcal{S} \to \mathcal{S}$. Our results will be similar to the results mentioned in [4, 11], but we will approach the problem from a different angle and with a different objective in mind.

Lemma 2. *An observable A with an n-outcome space Ω_A is compatible with the identity channel $id : \mathcal{S} \to \mathcal{S}$ if and only if there is a channel $\Phi : \mathcal{S} \to \mathcal{S} \dot{\otimes} \mathcal{P}(\Omega_A)$ such that for every extreme point $y \in \mathcal{S}$ we have*

$$\Phi(y) = \sum_{i=1}^{n} A_i(y) y \otimes \delta_i. \tag{4}$$

Proof. Assume that an observable A is compatible with id, then due to Lemma 1 we must have operations $\Phi_1, \ldots, \Phi_n$ such that $id = \sum_{i=1}^{n} \Phi_i$ and $A_i = u \circ \Phi_i$. To prove our claim we will use the defining property of extreme points. We have

$$y = id(y) = \sum_{i} \Phi_i(y)$$

that implies $\Phi_i(y) = \lambda_i(y)y$, where $\lambda_i(y) \in [0, 1]$ may in general depend on i and y. From $A_i = u \circ \Phi_i$ we obtain $\lambda_i(y) = A_i(y)$. For the joint channel Φ of id and A we have

$$\Phi(y) = \sum_{i=1}^{n} \Phi_i(y) \otimes \delta_i = \sum_{i=1}^{n} A_i(y) y \otimes \delta_i.$$

Now assume that for a channel $\Phi : \mathcal{S} \to \mathcal{S} \dot{\otimes} \mathcal{P}(\Omega_A)$ the equation (4) holds. For every extreme point $y \in \mathcal{S}$ we have

$$(u_2 \circ \Phi)(y) = \sum_{i=1}^{n} A_i(y) y = y,$$

$$(u_1 \circ \Phi)(y) = \sum_{i=1}^{n} A_i(y) \otimes \delta_i = A(y).$$

Since this holds for every extreme point of $\mathcal{S}$ it follows that Φ is a joint channel of A and id. $\square$

Proposition 4. *Observable A is compatible with id if and only if there is a set of affinely independent extreme points of $\mathcal{S}$, denote them x_j, $j \in \{1, \ldots, d\}$ such that $\mathcal{S} \subset \mathrm{aff}(\{x_1, \ldots, x_d\})$ and for every extreme point $y \in \mathcal{S}$, $y = \sum_{j=1}^{d} \alpha_j x_j$ it holds that*

$$\alpha_j(A_i(x_j) - A_i(y)) = 0. \tag{5}$$

Proof. Assume that an observable A is compatible with id and let $x_1, \ldots, x_d \in \mathcal{S}$ be a set of affinely independent extreme points, such that $\mathcal{S} \subset \mathrm{aff}(\{x_1, \ldots, x_d\})$. Let $y \in \mathcal{S}$ be an extreme point, then $y = \sum_{j=1}^{d} \alpha_j x_j$, where $\sum_{j=1}^{d} \alpha_j = 1$. According to Lemma 2 there is a channel Φ such that (4) holds. Plugging in the expression $y = \sum_{j=1}^{d} \alpha_j x_j$ we obtain

$$\Phi(y) = \sum_{j=1}^{d} \alpha_j \Phi(x_j) = \sum_{j=1}^{d} \alpha_j \sum_{i=1}^{n} \mathsf{A}_i(x_j) x_j \otimes \delta_i$$

which implies

$$\sum_{i=1}^{n} \mathsf{A}_i(y) y \otimes \delta_i = \sum_{j=1}^{d} \sum_{i=1}^{n} \alpha_j \mathsf{A}_i(x_j) x_j \otimes \delta_i.$$

Since $\delta_1, \ldots, \delta_n$ are linearly independent we must have $\mathsf{A}_i(y) y = \sum_{j=1}^{d} \alpha_j \mathsf{A}_i(x_j) x_j$ which yields

$$\sum_{j=1}^{d} \alpha_j \left(\mathsf{A}_i(x_j) - \mathsf{A}_i(y) \right) x_j = 0.$$

Eq. (5) follows by affine independence of $x_1, \ldots, x_d$.

Assume that (5) holds for an observable A and define a map $\Phi : \mathcal{S} \to \mathcal{S} \dot{\otimes} \mathcal{P}(\Omega_\mathsf{A})$ given for $j \in \{1, \ldots, d\}$ as

$$\Phi(x_j) = \sum_{i=1}^{n} \mathsf{A}_i(x_j) x_j \otimes \delta_i$$

and extended by affinity to all of $\mathcal{S}$. Let $y \in \mathcal{S}$ be an extreme point, then we have $y = \sum_{j=1}^{d} \alpha_j x_j$, $\sum_{j=1}^{d} \alpha_j = 1$ and

$$\Phi(y) = \sum_{j=1}^{d} \alpha_j \Phi(x_j) = \sum_{j=1}^{d} \alpha_j \sum_{i=1}^{n} \mathsf{A}_i(x_j) x_j \otimes \delta_i = \sum_{j=1}^{d} \sum_{i=1}^{n} \alpha_j \mathsf{A}_i(y) x_j \otimes \delta_i = \sum_{i=1}^{n} \mathsf{A}_i(y) y \otimes \delta_i$$

where we have used (5) in the third step. By lemma 2 it follows that A is compatible with id. $\qquad\square$

Note that if $\mathcal{S}$ is a simplex, then the set $\{x_1, \ldots, x_d\}$ is unique and contains all extreme points of $\mathcal{S}$, hence the requirement of Prop. 4 is trivially satisfied.

It is important to note that Prop. 4 provides a condition on the effects A_i, not on A as a whole. Therefore it will be interesting to investigate the set of effects that satisfy the condition (5).

Definition 7. We denote $\mathcal{E}\mathcal{T}_2$ set of effects on a state space $\mathcal{S}$ that satisfy the condition (5), i.e. $f \in \mathcal{E}\mathcal{T}_2$ if there is some set $\{x_1, \ldots, x_d\}$ of affinely independent extreme points of $\mathcal{S}$ such that $\mathcal{S} \subset \mathrm{aff}(\{x_1, \ldots, x_d\})$ and for every extreme point $y \in \mathcal{S}$, $y = \sum_{j=1}^{d} \alpha_j x_j$ it holds that

$$\alpha_j (f(x_j) - f(y)) = 0. \qquad (6)$$

The following is a straightforward.

Lemma 3. *Let $f, g \in \mathcal{E}\mathcal{T}_2$, $0 < \lambda \le 1$ and $0 \le \lambda' \le 1$, then $\lambda f \in \mathcal{E}(\mathcal{S})$ if and only if $f \in \mathcal{E}\mathcal{T}_2$, $f + g \in \mathcal{E}(\mathcal{S})$ if and only if $f + g \in \mathcal{E}\mathcal{T}_2$ and $\lambda' f + (1 - \lambda') g \in \mathcal{E}\mathcal{T}_2$.*

Proof. $\lambda f \in \mathcal{E}\mathcal{T}_2$ and $f + g \in \mathcal{E}\mathcal{T}_2$ follow immediately from linearity of (6). Convexity of $\mathcal{E}\mathcal{T}_2$ follows. $\qquad\square$

Proposition 5. *Let $0 < \lambda \le 1$ and $0 \le \mu \le 1$, then $f \in \mathcal{E}\mathcal{T}_2$ if and only if $\lambda f + (1 - \lambda) \mu u \in \mathcal{E}\mathcal{T}_2$.*

Proof. If $f \in \mathcal{E}\mathcal{T}_2$ then $\lambda f + (1 - \lambda) \mu u \in \mathcal{E}\mathcal{T}_2$ follows by Lemma 3. If $\lambda f + (1 - \lambda) \mu u \in \mathcal{E}\mathcal{T}_2$, then

$$\alpha_j ((\lambda f + (1 - \lambda) \mu)(x_j) - (\lambda f + (1 - \lambda) \mu)(y)) = 0$$

implies $\alpha_j (f(x_j) - f(y)) = 0$. $\qquad\square$

The result of Prop. 5 is non-trivial. As we will see, there are observables that are compatible with all other observables because they are "noisy enough". But according to Prop. 5 this is not the case for compatibility with the identity channel id. Loosely speaking Prop. 5 together with the next result show that the structure of $\mathcal{T}_2$ is more like $\mathcal{T}_1$, than $\mathcal{T}_3$ in the sense that observables in $\mathcal{T}_2$ are in some sense classical; such as was the case in Sec. II.

Corollary 1. *Observable* $\mathsf{A} \in \mathcal{T}_2$ *if and only if* $\mathsf{A}_i \in \mathcal{ET}_2$ *for all* i.

Proof. Follows from Prop. 4. $\square$

Theorem 1. $f \in \mathcal{ET}_2$ *if and only if* $\mathcal{S} = \oplus_{k=1}^N \mathcal{S}_k$ *and* f *is constant on each* $\mathcal{S}_k$.

Proof. If $\mathcal{S}$ is a simplex, then there is only one set $\{x_1, \ldots, x_d\}$ of affinely independent points and we have $\mathcal{S} = \oplus_{j=1}^d x_j$. The claim follows.

Let $x_1, \ldots, x_d$ be a set of affinely independent extreme points of $\mathcal{S}$ and let $y \in \mathcal{S}$ be an extreme point, then we have $y = \sum_{j=1}^d \alpha_j x_j$, $\sum_{j=1}^d \alpha_j = 1$. Assume that $\alpha_{j'} \neq 0$ and $\alpha_{j''} \neq 0$ for some $j', j'' \in \{1, \ldots, d\}$, which holds whenever $\mathcal{S}$ is not a simplex. Eq. (6) implies $f(x_{j'}) = f(y)$ and $f(x_{j''}) = f(y)$, which gives $f(x_{j'}) = f(x_{j''})$.

Denote $\mathcal{S}_c = \operatorname{conv}\left(\{z \in \mathcal{S} : f(z) = c, z \text{ is extreme}\}\right)$. We have just proved that that there is only finite number of the sets $\mathcal{S}_c$, $\mathcal{S}_c \subset \operatorname{aff}\left(\{x_j : f(x_j) = c\}\right)$.

Let $z \in \mathcal{S}$, then we have already proved that we have

$$z = \sum_{c \in [0,1]} \lambda_c y_c, \tag{7}$$

where $0 \leq \lambda_c \leq 1$, $\sum_{c \in [0,1]} \lambda_c = 1$ and $y_c \in \mathcal{S}_c$. Note that y_c is not necessarily an extreme point of $\mathcal{S}$. We will show that the decomposition (7) is unique. Assume there is another decomposition $z = \sum_{c \in [0,1]} \lambda'_c y'_c$, where again $0 \leq \lambda'_c \leq 1$, $\sum_{c \in [0,1]} \lambda'_c = 1$ and $y'_c \in \mathcal{S}_c$. Moreover assume that $\lambda_{c'} \neq 0$, then from $\sum_{c \in [0,1]} \lambda_c y_c = \sum_{c \in [0,1]} \lambda'_c y'_c$ we have

$$y_{c'} = \frac{1}{\lambda_{c'}} \left(\sum_{c \in [0,1]} \lambda'_c y'_c - \sum_{c \in [0,1] \setminus \{c'\}} \lambda_c y_c \right).$$

We can decompose $y_{c'} = \sum_{k=1}^n \mu_k y_{c',k}$, where $0 \leq \mu_k \leq 1$, $\sum_{k=1}^n \mu_k = 1$ and $y_{c',k}$ are extreme points of $\mathcal{S}_{c'}$. Moreover assume that $\mu_{k'} \neq 0$, then we have

$$y_{c',k'} = \frac{1}{\mu_{k'}} \left(\frac{1}{\lambda_{c'}} \left(\sum_{c \in [0,1]} \lambda'_c y'_c - \sum_{c \in [0,1] \setminus \{c'\}} \lambda_c y_c \right) - \sum_{k=1, k \neq k'}^n \mu_k y_{c',k} \right).$$

It follows that the right-hand side must be an affine combination of x_j, $j \in \{1, \ldots, n\}$ such that $f(x_j) = c'$. This implies that for $c \neq c'$ we must have $\lambda_c y_c = \lambda'_c y'_c$ as otherwise the aforementioned result would be violated. We get

$$y_{c',k'} = \frac{1}{\mu_{k'}} \left(\frac{\lambda'_{c'}}{\lambda_{c'}} y'_{c'} - \sum_{k=1, k \neq k'}^n \mu_k y_{c',k} \right).$$

It follows that

$$y_{c'} = \frac{\lambda'_{c'}}{\lambda_{c'}} y'_{c'},$$

hence the two decompositions of z are the same. The result follows from Prop. 3. $\square$

Using Thm. 1 we can easily characterize all two-dimensional state spaces that have observables compatible with the identity channel, i.e. that have information without disturbance. Remember that if a state space is two-dimensional, then $\dim(\mathcal{V}) = 3$ where $\mathcal{V}$ is the vector space containing the cone $\mathcal{V}^+$ which has the base $\mathcal{S}$.

Corollary 2. *Let* $\dim(\mathcal{V}) = 3$, *then* $\mathcal{S} = \mathcal{S}^1 \oplus \mathcal{S}^2$ *if and only if* $\mathcal{S}$ *is the triangle state space.*

Proof. Assume that $\mathcal{S} = \mathcal{S}^1 \oplus \mathcal{S}^2$, then $\mathcal{V} = \mathcal{V}^1 \times \mathcal{V}^2$, where $\mathcal{V}^1$, $\mathcal{V}^2$ are the vector spaces that contain $\mathcal{S}^1$ and $\mathcal{S}^2$ respectively. This implies $\dim(\mathcal{V}^1) + \dim(\mathcal{V}^2) = \dim(\mathcal{V}) = 3$ and we can assume that $\dim(\mathcal{V}^1) = 1$, $\dim(\mathcal{V}^2) = 2$. This implies that $\mathcal{S}^1$ contains only one point and $\mathcal{S}^2$ is a line segment, i.e. it has two extreme points. It then follows that $\mathcal{S}$ must have three extreme points, hence it is a triangle state space, which is a simplex. $\square$

In a similar fashion one can show that every three-dimensional state space that has information without disturbance is pyramid shaped, where the base of the pyramid can be any two-dimensional state space.

VI. CHARACTERIZATION OF $\mathcal{T}_3$

A. Simulability of observables

Simulation of observables is a method to produce a new observable from a given collection of observables by a classical procedure, that is, by mixing measurement settings and post-processing the outcome data [12–15]. For a subset $\mathcal{B} \subseteq \mathcal{O}(\mathcal{S})$, we denote by $\mathfrak{sim}(\mathcal{B})$ the set of observables that can be simulated by using the observables from $\mathcal{B}$, i.e., $\mathsf{A} \in \mathfrak{sim}(\mathcal{B})$ if there exists a probability distribution p, a finite collection of post-processing matrices $\nu^{(i)}$ and observables $\mathsf{B}^{(i)} \in \mathcal{B}$ such that

$$\mathsf{A} = \sum_i p_i \left(\nu^{(i)} \circ \mathsf{B}^{(i)} \right).$$

We will also denote $\mathfrak{sim}(\mathsf{B}) \equiv \mathfrak{sim}(\{\mathsf{B}\})$. Clearly,

$$\mathfrak{sim}(\mathsf{B}) = \{\mathsf{A} \in \mathcal{O}(\mathcal{S}) : \mathsf{B} \to \mathsf{A}\}.$$

We recall from [14] that an observable A is called *simulation irreducible* if for any subset $\mathcal{B} \subset \mathcal{O}$, we have $\mathsf{A} \in \mathfrak{sim}(\mathcal{B})$ only if there is $\mathsf{B} \in \mathcal{B}$ such that $\mathsf{A} \in \mathfrak{sim}(\mathsf{B})$ and $\mathsf{B} \in \mathfrak{sim}(\mathsf{A})$. Thus, a simulation irreducible observable can only be simulated by (essentially) itself. Equivalently, an observable is simulation irreducible if and only if it has indecomposable effects and is post-processing equivalent with an extreme observable. We denote by $\mathcal{O}_{irr}(\mathcal{S})$ the set of simulation irreducible observables. It was shown in [14] that for every observable there exists a finite collection of simulation irreducible observables from which it can be simulated.

It is worth mentioning that simulation irreducible observables are always incompatible, and in fact, a state space is non-classical if and only if there exists at least two inequivalent simulation irreducible observables [14].

B. Intersections of simulation sets

A trivial observable can be simulated by any other observable, and therefore

$$\mathcal{T}_1 = \bigcap_{\mathsf{B} \in \mathcal{O}(\mathcal{S})} \mathfrak{sim}(\mathsf{B}). \tag{8}$$

The following stronger statement is less obvious, although not too surprising.

Proposition 6.

$$\mathcal{T}_1 = \bigcap_{\mathsf{B} \in \mathcal{O}(\mathcal{S}) \smallsetminus \mathcal{T}_1} \mathfrak{sim}(\mathsf{B}). \tag{9}$$

Proof. Since $\mathcal{T}_1 \subseteq \bigcap_{\mathsf{B} \in \mathcal{O}(\mathcal{S})} \mathfrak{sim}(\mathsf{B})$, it is clear that $\mathcal{T}_1 \subseteq \bigcap_{\mathsf{B} \in \mathcal{O}(\mathcal{S}) \smallsetminus \mathcal{T}_1} \mathfrak{sim}(\mathsf{B})$. On the other hand, suppose that the inclusion is strict so that (w.l.o.g.) there exist a dichotomic observable $\mathsf{T} \in \bigcap_{\mathsf{B} \in \mathcal{O}(\mathcal{S}) \smallsetminus \mathcal{T}_1} \mathfrak{sim}(\mathsf{B})$ such that $\mathsf{T} \notin \mathcal{T}_1$. This means that the effects T_+ and T_- are not proportional to the unit effect u so that especially T_+ and u are linearly independent.

We take $\lambda, q \in (0, 1)$ and define another dichotomic observable A by $\mathsf{A} = \lambda\mathsf{T} + (1 - \lambda)\mathsf{Q}$, where $\mathsf{Q} \in \mathcal{T}_1$ is defined as $\mathsf{Q}_+ = qu$ and $\mathsf{Q}_- = (1 - q)u$. Since $\lambda \neq 0$ and $\mathsf{T} \notin \mathcal{T}_1$, we have that $\mathsf{A} \notin \mathcal{T}_1$. Hence, by the definition of T we have that $\mathsf{T} \in \mathfrak{sim}(\mathsf{A})$, i.e. there exists two real numbers $\nu_1, \nu_2 \in [0, 1]$ such that $\mathsf{T}_+ = \nu_1\mathsf{A}_+ + \nu_2\mathsf{A}_-$. When we expand A_+ and A_-, we find that

$$\mathsf{T}_+ = \nu_1(\lambda\mathsf{T}_+ + (1 - \lambda)qu) + \nu_2(\lambda\mathsf{T}_- + (1 - \lambda)(1 - q)u)$$
$$= \lambda(\nu_1 - \nu_2)\mathsf{T}_+ + (1 - \lambda)(\nu_1 - \nu_2)qu + \nu_2 u,$$

where on the second line we have used the fact that $\mathsf{T}_- = u - \mathsf{T}_+$. From the linear independence of u and T_+ it follows that we must have $\lambda(\nu_1 - \nu_2) = 1$, which is a contradiction since $0 < \lambda < 1$ and $\nu_1 - \nu_2 \leq 1$. $\square$

The equations (8) and (9) make one to wonder if the set $\mathcal{O}(\mathcal{S}) \smallsetminus \mathcal{T}_1$ can still be shrunk without altering the intersection property. Remarkably, taking $\mathcal{O}_{irr}(\mathcal{S})$ instead of $\mathcal{O}(\mathcal{S}) \smallsetminus \mathcal{T}_1$ changes the intersection, and leads to the following characterization for the set $\mathcal{T}_3$.

Proposition 7.

$$\mathcal{T}_3 = \bigcap_{\mathsf{B} \in \mathcal{O}_{irr}(\mathcal{S})} \mathfrak{sim}(\mathsf{B}). \tag{10}$$

Proof. Let first $\mathsf{T} \in \mathcal{T}_3$. Since T is compatible with every other observable, it is in particular compatible with every simulation irreducible observable. Thus, for every $\mathsf{B} \in \mathcal{O}_{irr}(\mathcal{S})$ there exists $\mathsf{G}^{\mathsf{B}} \in \mathcal{O}(\mathcal{S})$ such that $\{\mathsf{B}, \mathsf{T}\} \subseteq \mathfrak{sim}(\mathsf{G}^{\mathsf{B}})$. Since B is simulation irreducible it follows from the definition that $\mathsf{B} \leftrightarrow \mathsf{G}^{\mathsf{B}}$ so that $\mathfrak{sim}(\mathsf{B}) = \mathfrak{sim}(\mathsf{G}^{\mathsf{B}})$. Thus, $\mathsf{T} \in \mathfrak{sim}(\mathsf{B})$ for all $\mathsf{B} \in \mathcal{O}_{irr}(\mathcal{S})$.

Now let $\mathsf{A} \in \bigcap_{\mathsf{B} \in \mathcal{O}_{irr}(\mathcal{S})} \mathfrak{sim}(\mathsf{B})$ so that $\mathsf{A} \in \mathfrak{sim}(\mathsf{B})$ for all $\mathsf{B} \in \mathcal{O}_{irr}(\mathcal{S})$. We must show that A is compatible with every other observable. Thus, let $\mathsf{C} \in \mathcal{O}(\mathcal{S})$. For C there exists a finite set of simulation irreducible observables $\mathcal{B} = \{\mathsf{B}^{(i)}\}_{i=1}^n$ such that $\mathsf{C} \in \mathfrak{sim}(\mathcal{B})$. Thus, there exists a probability distribution $(p_i)_{i=1}^n$ and a post-processing $\nu : \{1, \ldots, n\} \times \Omega_{\mathcal{B}} \to \Omega_{\mathsf{C}}$ such that

$$\mathsf{C}_y = \sum_{i,x} p_i \nu_{(i,x)y} \mathsf{B}_x^{(i)} \tag{11}$$

for all $y \in \Omega_{\mathsf{C}}$. If we denote by $\widetilde{\mathsf{B}}$ the (generalized) mixture observable with outcomes set $\{1, \ldots, n\} \times \Omega_{\mathcal{B}}$ defined by $\widetilde{\mathsf{B}}_{(i,x)} = p_i \mathsf{B}_x^{(i)}$ for all $i \in \{1, \ldots, n\}$ and $x \in \Omega_{\mathcal{B}}$, we see that actually $\mathsf{C}_y = (\nu \circ \widetilde{\mathsf{B}})_y$ for all $y \in \Omega_{\mathsf{C}}$ so that $\mathsf{C} \in \mathfrak{sim}(\widetilde{\mathsf{B}})$.

Since $\mathsf{A} \in \mathfrak{sim}(\mathsf{B})$ for all $\mathsf{B} \in \mathcal{O}_{irr}(\mathcal{S})$, we have that $\mathsf{A} \in \mathfrak{sim}(\mathsf{B}^{(i)})$ for all $i = 1, \ldots, n$. Thus, there exists post-processings $\mu^{(i)} : \Omega_{\mathcal{B}} \to \Omega_{\mathsf{A}}$ such that $\mathsf{A} = \mu^{(i)} \circ \mathsf{B}^{(i)}$ for all $i = 1, \ldots, n$. If we use the same probability distribution $(p_i)_i$ as before, we have that for all $z \in \Omega_{\mathsf{A}}$

$$\mathsf{A}_z = \sum_i p_i \mathsf{A}_z = \sum_i p_i \sum_x \mu_{xz}^{(i)} \mathsf{B}_x^{(i)} = \sum_{i,x} \mu_{(i,x)z} p_i \mathsf{B}_x^{(i)} = (\mu \circ \widetilde{\mathsf{B}})_z,$$

where we have defined a new post-processing $\mu : \{1, \ldots, n\} \times \Omega_{\mathcal{B}} \to \Omega_{\mathsf{A}}$ by setting $\mu_{(i,x)z} = \mu_{xz}^{(i)}$ for all $i \in \{1, \ldots, n\}$, $x \in \Omega_{\mathcal{B}}$ and $z \in \Omega_{\mathsf{A}}$. Hence, also $\mathsf{A} \in \mathfrak{sim}(\widetilde{\mathsf{B}})$ so that A and C are compatible. $\square$

As was shown in Prop. 7, the observables that are compatible with every other observable are exactly those that can be post-processed from every simulation irreducible observable. However, we note that it is enough to consider only post-processing inequivalent simulation irreducible observables since two observables B and B' are post-processing equivalent, $\mathsf{B} \leftrightarrow \mathsf{B}'$, if and only if $\mathfrak{sim}(\mathsf{B}) = \mathfrak{sim}(\mathsf{B}')$. Thus, when we consider the intersection of the simulation sets of simulation irreducible observables, we only need to select some representative for each post-processing equivalence class.

The natural choice for the representative is to take the extreme observable with pairwise linearly independent effects: it has linearly independent indecomposable effects with the minimal number of outcomes in the respective post-processing equivalence class. It was shown [14] that such extreme observable exists in every equivalence class for simulation irreducible observables. We denote the set of extreme simulation irreducible observables by $\mathcal{O}_{irr}^{ext}(\mathcal{S})$ so that

$$\mathcal{T}_3 = \bigcap_{\mathsf{B} \in \mathcal{O}_{irr}(\mathcal{S})} \mathfrak{sim}(\mathsf{B}) = \bigcap_{\mathsf{B} \in \mathcal{O}_{irr}^{ext}(\mathcal{S})} \mathfrak{sim}(\mathsf{B}).$$

Corollary 3. *An observable $\mathsf{A} \in \mathcal{O}(\mathcal{S})$ on a state space $\mathcal{S}$ is included in $\mathcal{T}_3$ if and only if*

$$\mathsf{A}_y \in \bigcap_{\mathsf{B} \in \mathcal{O}_{irr}^{ext}(\mathcal{S})} \mathrm{cone}\left(\{\mathsf{B}_x\}_{x \in \Omega_{\mathsf{B}}}\right) \quad \forall y \in \Omega_{\mathsf{A}}. \tag{12}$$

Proof. Let first $\mathsf{A} \in \mathcal{T}_3$. By Prop. 7 for all $\mathsf{B} \in \mathcal{O}_{irr}^{ext}(\mathcal{S})$ there exists a post-processing ν^{B} such that $\mathsf{A} = \nu^{\mathsf{B}} \circ \mathsf{B}$, i.e.,

$$\mathsf{A}_y = \sum_{x \in \Omega_{\mathsf{B}}} \nu_{xy}^{\mathsf{B}} \mathsf{B}_x \tag{13}$$

for all $y \in \Omega_{\mathsf{A}}$. Since $\nu_{xy}^{\mathsf{B}} \geq 0$ for all $x \in \Omega_{\mathsf{B}}$, $y \in \Omega_{\mathsf{A}}$ for all $\mathsf{B} \in \mathcal{O}_{irr}^{ext}(\mathcal{S})$, we have that

$$\mathsf{A}_y \in \mathrm{cone}\left(\{\mathsf{B}_x\}_{x \in \Omega_{\mathsf{B}}}\right) \tag{14}$$

for all $\mathsf{B} \in \mathcal{O}_{irr}^{ext}(\mathcal{S})$ for all $y \in \Omega_{\mathsf{A}}$, which proves the necessity part of the claim.
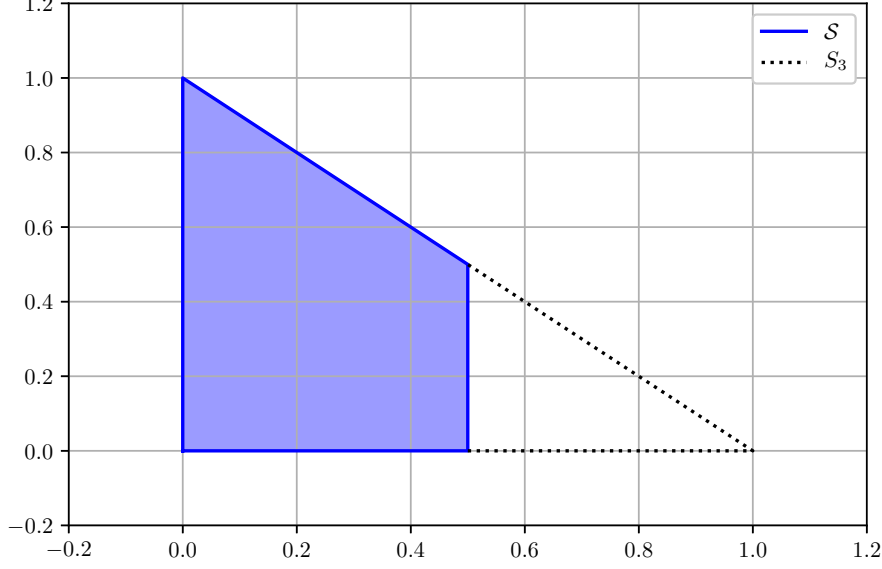
FIG. 2. The blue solid line is the boundary of the state space $\mathcal{S}$ used in the example. The black dotted line shows that $\mathcal{S}$ can be considered as a subset of the simplex S_3.

Then let Eq. (12) hold. Thus, for each $\mathsf{B} \in \mathcal{O}_{irr}^{ext}(\mathcal{S})$ there exists positive numbers $\mu_{xy}^\mathsf{B} \geq 0$ such that

$$\mathsf{A}_y = \sum_{x \in \Omega_\mathsf{B}} \mu_{xy}^\mathsf{B} \mathsf{B}_x$$

for all $y \in \Omega_A$. From the normalization of observables A and B it follows that

$$\sum_{x \in \Omega_\mathsf{B}} \mathsf{B}_x = u = \sum_{y \in \Omega_\mathsf{A}} \mathsf{A}_y = \sum_{x \in \Omega_\mathsf{B}} \left(\sum_{y \in \Omega_\mathsf{A}} \mu_{xy}^\mathsf{B} \right) \mathsf{B}_x. \tag{15}$$

Since each $\mathsf{B} \in \mathcal{O}_{irr}^{ext}(\mathcal{S})$, we have that each B consists of linearly independent effects B_x [14], so that $\sum_{y \in \Omega_\mathsf{A}} \mu_{xy}^\mathsf{B} = 1$ for all $x \in \Omega_\mathsf{B}$. Thus, we can define post-processings μ^B for each $\mathsf{B} \in \mathcal{O}_{irr}^{ext}(\mathcal{S})$ with elements μ_{xy}^B so that $\mathsf{A} \in \mathfrak{sim}(\mathsf{B})$ for all $\mathsf{B} \in \mathcal{O}_{irr}^{ext}(\mathcal{S})$. $\qquad \square$

C.   Example showing that $\mathcal{T}_2 \neq \mathcal{T}_3$

We will present an example of a two-dimensional state space $\mathcal{S}$, such that there is an observable $\mathsf{A} \in \mathcal{O}(\mathcal{S})$ with $\mathsf{A} \in \mathcal{T}_3$ but $\mathsf{A} \notin \mathcal{T}_2$.

Let

$$\mathcal{S} = \mathrm{conv}\left(\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0.5 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0.5 \\ 0.5 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right),$$

where the z-coordinate is used to identify $\mathcal{S}$ with a base of a cone. Let

$$S_3 = \mathrm{conv}\left(\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right)$$

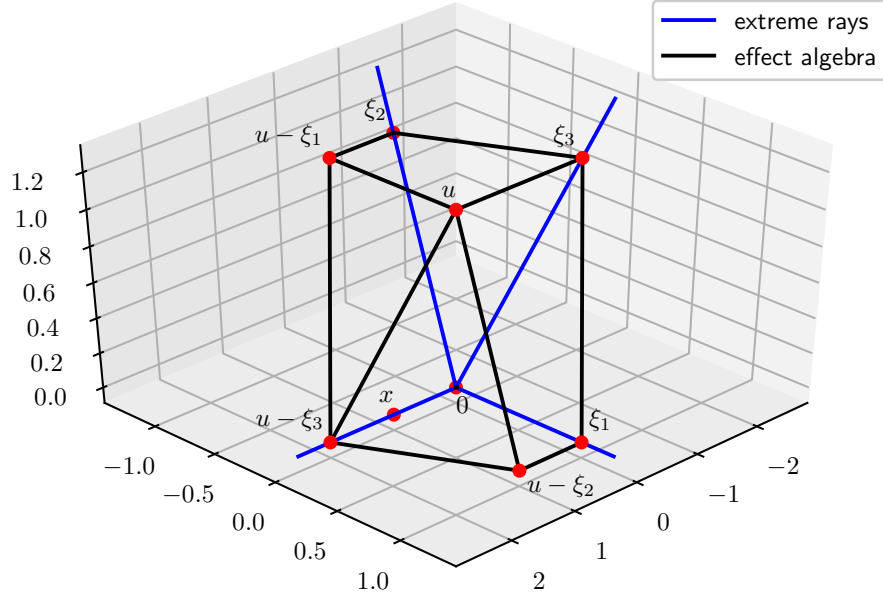be a simpex, then we have $\mathcal{S} \subset S_3$ as shown in Fig. 2.

FIG. 3. The effect algebra $\mathcal{E}(\mathcal{S})$ used in the example. The black lines represent the wireframe model of $\mathcal{E}(\mathcal{S})$, the blue lines are the extreme rays of the cone of positive functions and the red dots denote the effects that we are using in the example (with $\xi_1 = y$).

Let us define functionals x, y, u given as

$$x = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad\qquad y = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \qquad\qquad u = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

The points are shown in Fig. 3.

According to Prop. 16 from appendix A there are 4 indecomposable effects corresponding to the 4 maximal faces of $\mathcal{S}$. They are ξ_1, ξ_2, ξ_3 and $u - \xi_3$, where

$$\xi_1 = y, \qquad\qquad \xi_2 = u - x - y, \qquad\qquad \xi_3 = u - 2x.$$

It was shown in [14, Corollary 1] that simulation irreducible observables must consists of indecomposable effects. We are going to find all simulation irreducible observables on $\mathcal{S}$ as we know that $\mathsf{A} \in \mathcal{T}_3$ if and only if A is simulable by every simulation irreducible observable; see Prop. 7.

Assume that there would be a simulation irreducible observable with the effects $\alpha_1 \xi_1$, $\alpha_2 \xi_2$, $\alpha_3 \xi_3$ and $\alpha_3'(u - \xi_3)$, where $\alpha_1, \alpha_2, \alpha_3, \alpha_3' \in \mathbb{R}$, then we must have

$$\alpha_1 \xi_1 + \alpha_2 \xi_2 + \alpha_3 \xi_3 + \alpha_3'(u - \xi_3) = u$$

which yields

$$-\alpha_2 - 2\alpha_3 + 2\alpha_3' = 0,$$
$$\alpha_1 - \alpha_2 = 0,$$
$$\alpha_2 + \alpha_3 = 1.$$

Since the effects of simulation irreducible observables must be linearly independent, we know that at least one of the coefficients must be equal to zero.

Assuming $\alpha_1 = 0$, we get $\alpha_2 = 0$ and $\alpha_3 = \alpha_3' = 1$ and we obtain a dichotomic observable B with effects

$$B_1 = \xi_3,$$
$$B_2 = u - \xi_3.$$

Assuming $\alpha_2 = 0$ yields $\alpha_1 = 0$ and $\alpha_3 = \alpha_3' = 1$, i.e. the same observable B. Assuming $\alpha_3 = 0$ gives $\alpha_1 = \alpha_2 = 1$ and $\alpha_3' = \frac{1}{2}$ and gives us a three-outcome observable C with effects

$$C_1 = \xi_1,$$
$$C_2 = \xi_2,$$
$$C_3 = x.$$

Finally assuming $\alpha_3' = 0$ leads to a contradiction.

Let A be a dichotomic observable given as

$$A_1 = x,$$
$$A_2 = u - x.$$

Note that we have $x = \frac{1}{2}(u - \xi_3)$, which shows that A is simulable by B and we have $x = u - \xi_1 - \xi_2$, which shows that A is simulable by C. This shows that A $\in \mathcal{T}_3$.

We are going to use Prop. 5 to see that A $\notin \mathcal{T}_2$. Assume that A $\in \mathcal{T}_2$, then $A_1 \in \mathcal{ET}_2$, which by Prop. 5 implies also $u - \xi_3 \in \mathcal{ET}_2$ as $u - \xi_3 = 2x$. This would imply that B would be compatible with every other observable, but it is straightforward to see that B is incompatible with C as they are the only two simulation irreducible observables and if they would be compatible, then all of the observables on $\mathcal{S}$ would be compatible. This would in turn yield that $\mathcal{S}$ would have to be simplex [16] which it clearly is not.

An insight into how we obtained this example is provided by the simplex S_3: ξ_1, ξ_2 and x are effects on the simplex S_3 so that the compatibility of A and C follows. Moreover, the fact that $u - \xi_3 = 2x \geq x$ gives the compatibility of A and B.

VII.   STATE SPACES SATISFYING $\mathcal{T}_1 = \mathcal{T}_2 = \mathcal{T}_3$

Next we will consider conditions under which the no-information-without-disturbance principle ($\mathcal{T}_2 = \mathcal{T}_1$) and the no-free-information principle ($\mathcal{T}_3 = \mathcal{T}_1$) hold and when they do not. First we note that, as was mentioned earlier, in general we have that $\mathcal{T}_1 \subseteq \mathcal{T}_2 \subseteq \mathcal{T}_3$ so that if the no-free-information principle holds, and therefore we have that $\mathcal{T}_3 = \mathcal{T}_1$, it follows that also $\mathcal{T}_2 = \mathcal{T}_1$ so that the no-information-without-disturbance principle must hold as well.

A.   Conditions for $\mathcal{T}_1 = \mathcal{T}_3$

With the help of Prop. 7 we can show the following.

Proposition 8. *The following conditions are equivalent:*

i) $\mathcal{T}_1 = \mathcal{T}_3$

ii) $\bigcap_{B \in \mathcal{O}_{irr}(\mathcal{S})} \mathrm{cone}\left(\{B_x\}_{x \in \Omega_B}\right) = \mathrm{cone}\left(u\right)$

iii) $\bigcap_{B \in \mathcal{O}_{irr}(\mathcal{S})} \mathrm{conv}\left(\{\{B_x\}_{x \in \Omega_B}, o, u\}\right) = \mathrm{conv}\left(\{o, u\}\right).$

Proof. i) $\Rightarrow$ *iii)*: It is clear that $\mathrm{conv}\left(\{o, u\}\right) \subseteq \bigcap_{B \in \mathcal{O}_{irr}(\mathcal{S})} \mathrm{conv}\left(\{\{B_x\}_x, o, u\}\right)$. Now take

$$e \in \bigcap_{B \in \mathcal{O}_{irr}(\mathcal{S})} \mathrm{conv}\left(\{\{B_x\}_x, o, u\}\right)$$

and define a dichotomic observable E with effects $E_+ = e$ and $E_- = u - e$. Since $E_+ \in \mathrm{conv}\left(\{\{B_x\}_x, o, u\}\right)$ for all B $\in \mathcal{O}_{irr}(\mathcal{S})$, it follows from Prop. 8 in [14] that E $\in \mathfrak{sim}(B)$ for all B $\in \mathcal{O}_{irr}(\mathcal{S})$. From Prop. 7 it follows that E $\in \mathcal{T}_3 = \mathcal{T}_1$ so that actually $e \in \mathrm{conv}\left(\{o, u\}\right)$.

iii) ⇒ ii): It is clear that $\operatorname{cone}(u) \subseteq \bigcap_{\mathsf{B} \in \mathcal{O}_{irr}(\mathcal{S})} \operatorname{cone}(\{\mathsf{B}_x\}_x)$. Now take $g \in \bigcap_{\mathsf{B} \in \mathcal{O}_{irr}(\mathcal{S})} \operatorname{cone}(\{\mathsf{B}_x\}_x)$ so that for all $\mathsf{B} \in \mathcal{O}_{irr}(\mathcal{S})$ there exists positive real numbers $(\alpha_x^{\mathsf{B}})_x \subset \mathbb{R}_+$ such that $g = \sum_x \alpha_x^{\mathsf{B}} \mathsf{B}_x$. We denote $\alpha = \sup_{\mathsf{B} \in \mathcal{O}_{irr}(\mathcal{S})} \sum_x \alpha_x^{\mathsf{B}}$. If $\alpha = 0$, then $g = o \in \operatorname{cone}(u)$; otherwise we define an effect $f \in \mathcal{E}(\mathcal{S})$ by $f = \frac{1}{\alpha} g$. Now

$$f \in \bigcap_{\mathsf{B} \in \mathcal{O}_{irr}(\mathcal{S})} \operatorname{conv}(\{\{\mathsf{B}_x\}_x, o\}) \subseteq \bigcap_{\mathsf{B} \in \mathcal{O}_{irr}(\mathcal{S})} \operatorname{conv}(\{\{\mathsf{B}_x\}_x, o, u\}) = \operatorname{conv}(\{o, u\}) \tag{16}$$

so that $f = pu$ for some $p \in (0,1]$. Thus, $g = \alpha p u \in \operatorname{cone}(u)$.

ii) ⇒ i): As noted before, we always have $\mathcal{T}_1 \subseteq \mathcal{T}_3$ so that it suffices to show that $\mathcal{T}_3 \subseteq \mathcal{T}_1$. Thus, take $\mathsf{A} \in \mathcal{T}_3$. By Prop. 7, $\mathsf{A} \in \mathfrak{sim}(\mathsf{B})$ for all $\mathsf{B} \in \mathcal{O}_{irr}(\mathcal{S})$ so that for each $\mathsf{B} \in \mathcal{O}_{irr}(\mathcal{S})$ there exists a post-processing $\nu^{\mathsf{B}} : \Omega_{\mathsf{B}} \to \Omega_{\mathsf{A}}$ such that $\mathsf{A}_y = \sum_{x \in \Omega_{\mathsf{B}}} \nu_{xy}^{\mathsf{B}} \mathsf{B}_x$ for all $y \in \Omega_{\mathsf{A}}$. Since all the post-processing elements are positive for each $\mathsf{B} \in \mathcal{O}_{irr}(\mathcal{S})$, we have that $\mathsf{A}_y \in \operatorname{cone}(\{\mathsf{B}_x\}_{x \in \Omega_{\mathsf{B}}})$ for all $y \in \Omega_{\mathsf{A}}$ and $\mathsf{B} \in \mathcal{O}_{irr}(\mathcal{S})$. Thus,

$$\mathsf{A}_y \in \bigcap_{\mathsf{B} \in \mathcal{O}_{irr}(\mathcal{S})} \operatorname{cone}(\{\mathsf{B}_x\}_{x \in \Omega_{\mathsf{B}}}) = \operatorname{cone}(u) \tag{17}$$

for all $y \in \Omega_{\mathsf{A}}$ from which it follows that $\mathsf{A} \in \mathcal{T}_1$.

$\square$

Proposition 9. *Let $\mathcal{S}$ be a d-dimensional state space. If $|\mathcal{O}_{irr}^{ext}(\mathcal{S})| < \infty$ and all the extreme simulation irreducible observables have $d + 1$ outcomes, then $\mathcal{T}_1 \neq \mathcal{T}_3$.*

Proof. Since $\mathcal{S}$ is d-dimensional (i.e. $\dim(\operatorname{aff}(\mathcal{S})) = d$), the effect space is contained in a $d + 1$-dimensional vector space. Suppose that, on the contrary $\mathcal{T}_1 = \mathcal{T}_3$. From Prop. 8 it follows then that

$$\bigcap_{\mathsf{B} \in \mathcal{O}_{irr}^{ext}(\mathcal{S})} \operatorname{cone}(\{\mathsf{B}_x\}_x) = \bigcap_{\mathsf{B} \in \mathcal{O}_{irr}(\mathcal{S})} \operatorname{cone}(\{\mathsf{B}_x\}_{x \in \Omega_{\mathsf{B}}}) = \operatorname{cone}(u) .$$

Since $\dim(\mathcal{V}^*) = d + 1$ and each extreme simulation irreducible observable consists of $d + 1$ linearly independent effects, it follows that $\operatorname{cone}(\{\mathsf{B}_x\}_x)$ has a non-empty interior, denoted by $\operatorname{int}(\operatorname{cone}(\{\mathsf{B}_x\}_x))$, in $\mathcal{V}^*$ for all $\mathsf{B} \in \mathcal{O}_{irr}(\mathcal{S})$. In particular, $u \in \operatorname{int}(\operatorname{cone}(\{\mathsf{B}_x\}_x))$ for all $\mathsf{B} \in \mathcal{O}_{irr}^{ext}(\mathcal{S})$, so that

$$\varnothing = \operatorname{int}(\operatorname{cone}(u)) = \operatorname{int}\left(\bigcap_{\mathsf{B} \in \mathcal{O}_{irr}^{ext}(\mathcal{S})} \operatorname{cone}(\{\mathsf{B}_x\}_x) \right) = \bigcap_{\mathsf{B} \in \mathcal{O}_{irr}^{ext}(\mathcal{S})} \operatorname{int}(\operatorname{cone}(\{\mathsf{B}_x\}_x)) \neq \varnothing \tag{18}$$

which is a contradiction. $\square$

Proposition 10. *If there exist at least two post-processing inequivalent dichotomic simulation irreducible observables on $\mathcal{S}$, then $\mathcal{T}_1 = \mathcal{T}_2 = \mathcal{T}_3$.*

Proof. By the assumption there exist two dichotomic observables $\mathsf{E}, \mathsf{F} \in \mathcal{O}_{irr}(\mathcal{S})$ such that $\mathsf{E} \not\leftrightarrow \mathsf{F}$. Take $\mathsf{A} \in \mathcal{T}_3$ so that by Prop. 7 we have that $\mathsf{A} \in \mathfrak{sim}(\mathsf{E})$ and $\mathsf{A} \in \mathfrak{sim}(\mathsf{F})$. From Prop. 11 in [14] it follows that $\mathsf{A}_x \in \operatorname{conv}(\{\mathsf{E}_+, \mathsf{E}_-, o, u\})$ and $\mathsf{A}_x \in \operatorname{conv}(\{\mathsf{F}_+, \mathsf{F}_-, o, u\})$ for all $x \in \Omega_{\mathsf{A}}$. Since E and F are inequivalent, it follows that the set $\{u, \mathsf{E}_+, \mathsf{F}_+\}$ is linearly independent, so that $\mathsf{A}_x \in \operatorname{conv}(\{\mathsf{E}_+, \mathsf{E}_-, o, u\}) \cap \operatorname{conv}(\{\mathsf{F}_+, \mathsf{F}_-, o, u\}) = \operatorname{conv}(\{o, u\})$ for all $x \in \Omega_{\mathsf{A}}$. Thus, $\mathsf{A} \in \mathcal{T}_1$ so that $\mathcal{T}_1 = \mathcal{T}_3$. $\square$

With the previous proposition we can show that the no-free-information principle holds in any point-symmetric state space, i.e., in a state space $\mathcal{S}$ where there exists a state s_0 such that for all $s \in \mathcal{S}$ we have that

$$s' := 2s_0 - s \in \mathcal{S}. \tag{19}$$

This means that for each state s there exists another state s' such that s_0 is an equal mixture of s and s', i.e., $s_0 = \frac{1}{2}(s + s')$. Point-symmetric state spaces include the classical bit, the qubit and polygon state spaces with even number of vertices.

One can show that the effect space structure is also symmetric for symmetric state spaces. Firstly, all the non-trivial extreme effects are seen to lie on a single affine hyperplane. Namely, if $e \in \mathcal{E}(\mathcal{S})$ is an extreme effect, $e \neq o, u$, there exists a (pure) state $s \in \mathcal{S}$ such that $e(s) = 0$ [7]. For s, there exists another state s' such that $s_0 = \frac{1}{2}(s + s')$ so that $e(s_0) = \frac{1}{2} e(s')$. Similarly there exists a (pure) state $t \in \mathcal{S}$ such that $e(t) = 1$ [7]. For t, we can find t' such that $e(s_0) = \frac{1}{2}(e(t) + e(t')) = \frac{1}{2}(1 + e(t'))$. Combining these two expressions for $e(s_0)$ we find that $e(s') = 1 + e(t')$ from which it follows that $e(t') = 0$ and $e(s') = 1$ so that $e(s_0) = \frac{1}{2}$ for all extreme effects e. Thus, all the non-trivial extreme effects lie on an affine hyperplane determined by the state s_0.
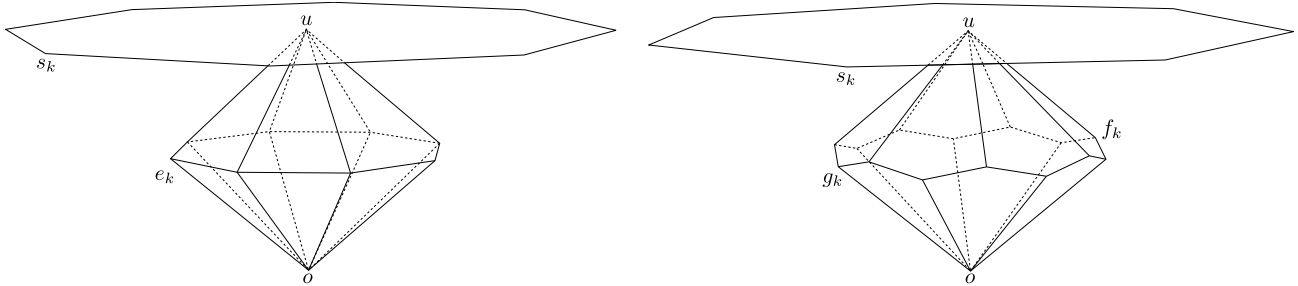
FIG. 4. The even and odd polygon state spaces and their effects spaces.

Secondly, we see that all the non-trivial extreme effects must actually be indecomposable. If $e \in \mathcal{E}(\mathcal{S})$ is an extreme effect, $e \neq o, u$, then we can find some decomposition into indecomposable extreme effects $\{e_i\}_{i=1}^r$ for some $r \in \mathbb{N}$ so that $e = \sum_{i=1}^r \alpha_i e_i$ for some numbers $\{\alpha_i\}_{i=1}^r \subseteq [0,1]$ [7]. Since all extreme effects give probability $\frac{1}{2}$ on the state s_0, we have that $1 = 2e(s_0) = \sum_{i=1}^r \alpha_i$. Since e is extreme, it follows that $r = 1$ so that e is indecomposable.

Thirdly, the convex hull of all the extreme indecomposable effects (that lie on an affine hyperplane) is also point-symmetric: if $e \in \mathcal{E}(\mathcal{S})$ is a non-trivial extreme effect, then $e' \coloneqq u - e$ is also a non-trivial extreme effect so that $e_0 \coloneqq \frac{1}{2}u = \frac{1}{2}(e + e')$ acts as the inversion point of the set.

Corollary 4. *In every non-classical point-symmetric state space $\mathcal{S}$ we have $\mathcal{T}_1 = \mathcal{T}_2 = \mathcal{T}_3$.*

Proof. Since $\mathcal{S}$ is non-classical, there exists two non-trivial extreme effects e and f such that $e, f \neq o, u$, $e \neq f, u - f$. We define two dichotomic observables E and F by setting $\mathsf{E}_+ = e$, $\mathsf{E}_- = u - e$, $\mathsf{F}_+ = f$ and $\mathsf{F}_- = u - f$. Since the state space is point-symmetric, the extreme effects $e, f, u - e$ and $u - f$ are indecomposable so that together with the fact that $\{e, u - e\}$ and $\{f, u - f\}$ are linearly independent sets it follows [14] that E and F are inequivalent dichotomic simulation irreducible observables. The claim follows from Prop. 10. $\qquad \square$

B. Alternative characterization of $\mathcal{T}_1$

Finally, we show that a seemingly different formulation of "free-information" does not lead to a new concept. Consider $\mathsf{T} \in \mathcal{T}_3$ and take an observable $\mathsf{A} \in \mathcal{O}(\mathcal{S})$ such that $\mathsf{A} \notin \mathcal{T}_1$. Since T is compatible with A there exists a joint observable $\mathsf{J}_{\mathsf{A},\mathsf{T}}$ from which both A and T can be post-processed from. Since A is non-trivial and T is compatible with every other observable, we can ask whether measuring the joint observable $\mathsf{J}_{\mathsf{A},\mathsf{T}}$ actually gives us any more information than just measuring A. One way to consider this is to ask whether A is actually post-processing equivalent to $\mathsf{J}_{\mathsf{A},\mathsf{T}}$ so that both can be obtained from each other by classically manipulating their outcomes. If this is the case, there is no "free information" to be gained from measuring the joint observable. Thus, we consider one more set of observables:

$$\mathcal{T}_4 = \{\mathsf{T} \in \mathcal{T}_3 \,|\, \forall \mathsf{A} \in \mathcal{O}(\mathcal{S}) \smallsetminus \mathcal{T}_1 : \; \exists \, \mathsf{J}_{\mathsf{A},\mathsf{T}} \in \mathcal{O}(\mathcal{S}) : \; \mathsf{J}_{\mathsf{A},\mathsf{T}} \leftrightarrow \mathsf{A}\}.$$

We can show the following.

Proposition 11. $\mathcal{T}_1 = \mathcal{T}_4$.

Proof. Since $\mathcal{T}_1 \subseteq \mathcal{T}_4$ it suffices to show that $\mathcal{T}_4 \subseteq \mathcal{T}_1$. Thus, take $\mathsf{T} \in \mathcal{T}_4$ so that for all $\mathsf{A} \in \mathcal{O}(\mathcal{S}) \smallsetminus \mathcal{T}_1$ we have that A is post-processing equivalent with at least one of their joint observables $\mathsf{J}_{\mathsf{A},\mathsf{T}}$. Thus, $\{\mathsf{A}, \mathsf{T}\} \subseteq \mathfrak{sim}(\mathsf{J}_{\mathsf{A},\mathsf{T}})$ and since $\mathsf{A} \leftrightarrow \mathsf{J}_{\mathsf{A},\mathsf{T}}$ it follows that $\mathsf{T} \in \mathfrak{sim}(\mathsf{A})$ for all $\mathsf{A} \in \mathcal{O}(\mathcal{S}) \smallsetminus \mathcal{T}_1$. From Prop. 6 it follows that $\mathsf{T} \in \mathcal{T}_1$. $\qquad \square$

VIII. POLYGON STATE SPACES

A. Characterization of polygons

A regular polygon with n vertices in $\mathbb{R}^2$, or n-gon, is a convex hull of n points $\{\vec{x}_k\}_{k=1}^n$ such that $\|\vec{x}_k\| = \|\vec{x}_j\|$ and $\vec{x}_k \cdot \vec{x}_{k+1} = \|\vec{x}_k\|^2 \cos\left(\frac{\pi}{n}\right)$ for all $j, k = 1, \dots, n$. As a state space $\mathcal{S}_n$, we consider the polygon to be embedded in $\mathbb{R}^3$ on

the $z = 1$ – plane. Thus, we follow the notation of [17] and define the extreme points of $\mathcal{S}_n$ as

$$s_k = \begin{pmatrix} r_n \cos\left(\dfrac{2k\pi}{n}\right) \\ r_n \sin\left(\dfrac{2k\pi}{n}\right) \\ 1 \end{pmatrix}, \quad k = 1, \ldots, n,$$

where we have defined $r_n = \sec\left(\frac{\pi}{n}\right)$.

As the polygons are two-dimensional, the effects can also be represented as elements in $\mathbb{R}^3$. Hence, we can express each $e \in \mathcal{E}(\mathcal{S}_n)$ as a vector $e = (e_x, e_y, e_z)^T \in \mathbb{R}^3$. With this identification we have that $e(s) = e \cdot s$ for all $e \in \mathcal{E}(\mathcal{S}_n)$ and $s \in \mathcal{S}_n$ where $\cdot$ is the Euclidean dot product. Clearly, we now have the zero effect $o = (0, 0, 0)^T$ and the unit effect $u = (0, 0, 1)^T$.

Depending on the parity of n, the state space may or may not have reflective point symmetry around the middle point $s_0 = (0, 0, 1)^T$. As a result of this, the effect space $\mathcal{E}(\mathcal{S}_n)$ has a different structure for odd and even n. For even n, we find that the effect space $\mathcal{E}(\mathcal{S}_n)$ has n non-trivial extreme points

$$e_k = \frac{1}{2} \begin{pmatrix} \cos\left(\dfrac{(2k-1)\pi}{n}\right) \\ \sin\left(\dfrac{(2k-1)\pi}{n}\right) \\ 1 \end{pmatrix}, \quad k = 1, \ldots, n, \tag{20}$$

so that $\mathcal{E}(\mathcal{S}_n) = \mathrm{conv}\left(\{o, u, e_1, \ldots, e_n\}\right)$. All the non-trivial extreme effects lie on a single (hyper)plane determined by those points e such that $e(s_0) = 1/2$.

In the case of odd n, the effect space has $2n$ non-trivial extreme effects

$$g_k = \frac{1}{1+r_n} \begin{pmatrix} \cos\left(\dfrac{2k\pi}{n}\right) \\ \sin\left(\dfrac{2k\pi}{n}\right) \\ 1 \end{pmatrix}, \qquad f_k = u - g_k \tag{21}$$

for $k = 1, \ldots, n$. Now $\mathcal{E}(\mathcal{S}_n) = \mathrm{conv}\left(\{o, u, g_1, \ldots, g_n, f_1, \ldots, f_n\}\right)$ and the non-trivial effects are scattered on two different planes determined by all those points g and f such that $g(s_0) = \sigma_n := \frac{1}{1+r_n}$ and $f(s_0) = 1 - \sigma_n = \frac{r_n}{1+r_n}$. The even and odd polygon state spaces and their respective effect spaces are depicted in Figure 4.

In order to give a simple characterization of polygons, let us define functions $\eta_e^n : \mathbb{R}^2 \to \mathbb{R}$ and $\eta_o^n : \mathbb{R}^2 \to \mathbb{R}$ by

$$\eta_e^n(\vec{x}) = \max_{k \in \{1, \ldots, n\}} r_n \left[\cos\left(\frac{2\pi k}{n}\right) x + \sin\left(\frac{2\pi k}{n}\right) y \right],$$

$$\eta_o^n(\vec{x}) = \eta_e^n\left(R_{\frac{\pi}{n}} \vec{x}\right) = \max_{k \in \{1, \ldots, n\}} r_n \left[\cos\left(\frac{(2k-1)\pi}{n}\right) x + \sin\left(\frac{(2k-1)\pi}{n}\right) y \right],$$

for all $\vec{x} = (x, y)^T \in \mathbb{R}^2$, where

$$R_{\frac{\pi}{n}} = \begin{pmatrix} \cos\left(\dfrac{\pi}{n}\right) & -\sin\left(\dfrac{\pi}{n}\right) \\ \sin\left(\dfrac{\pi}{n}\right) & \cos\left(\dfrac{\pi}{n}\right) \end{pmatrix}$$

is the rotation matrix with a rotation angle π/n around the origin in $\mathbb{R}^2$. We use the notation $\eta_{e/o}^n$ when we consider some properties that hold for both η_e^n and η_o^n.

We see that both $\eta_e^n(\vec{x})$ and $\eta_o^n(\vec{x})$ can be expressed as a maximization over an inner product of $\vec{x}$ and a collection of unit vectors $\vec{b}_{e/o}^{(n,k)}$, i.e.

$$\eta_{e/o}^n(\vec{x}) = r_n \max_{k \in \{1, \ldots, n\}} \vec{x} \cdot \vec{b}_{e/o}^{(n,k)}, \tag{22}$$

where we have defined

$$\vec{b}_e^{(n,k)} = \left(\cos\left(\frac{2\pi k}{n}\right), \sin\left(\frac{2\pi k}{n}\right)\right)^T, \tag{23}$$

$$\vec{b}_o^{(n,k)} = \left(\cos\left(\frac{(2k-1)\pi}{n}\right), \sin\left(\frac{(2k-1)\pi}{n}\right)\right)^T. \tag{24}$$

Thus, both η_e^n and η_o^n are polyhedral convex functions [18].

It is straightforward to see that $\eta_{e/o}^n$ satisfy the following properties for all $\vec{x}, \vec{y} \in \mathbb{R}^2$:

i) $\eta_{e/o}^n(\vec{x}) \geq 0$,

ii) $\eta_{e/o}^n(\vec{x}) = 0 \quad \Leftrightarrow \quad \vec{x} = \vec{0}$,

iii) $\eta_{e/o}^n(\vec{x} + \vec{y}) \leq \eta_{e/o}^n(\vec{x}) + \eta_{e/o}^n(\vec{y})$.

Additionally we see that also the following is satisfied for all $x \in \mathbb{R}^2$:

iv) $\eta_{e/o}^n(\alpha\vec{x}) = \alpha\eta_e^n(\vec{x})$ for all $\alpha \geq 0$.

Thus, both η_e^n and η_o^n almost satisfy the requirements of a norm; the only missing property is the requirement for a reflective point symmetry, i.e. $\eta_{e/o}^n(-\vec{x}) = \eta_{e/o}^n(\vec{x})$ for all $\vec{x} \in \mathbb{R}^2$. For even n, however, it is easy to confirm that both η_e^n and η_o^n are point symmetric so that they are norms on $\mathbb{R}^2$. Similarly for odd n it is easy to see that the point symmetry does not hold.

Even though for general n the functions $\eta_{e/o}^n$ do not define a norm on $\mathbb{R}^2$, we can still use them to define different sized polygons. As continuous polyhedral convex functions, η_e^n and η_o^n have closed polyhedral level sets

$$B_{e/o}^n(r) = \{\vec{x} \in \mathbb{R}^2 \,|\, \eta_{e/o}^n(\vec{x}) \leq r\}$$

which we will show to give rise to the polygons.

First of all, we see that the level sets $B_{e/o}^n(r)$ are bounded so that they actually describe polytopes: When we express $\vec{x} \in \mathbb{R}^2$ in its polar form $\vec{x} = (x,y)^T = \|\vec{x}\| \, (\cos(\theta), \sin(\theta))^T$, we have

$$\eta_e^n(\vec{x}) = r_n \|\vec{x}\| \max_{k \in \{1,\dots,n\}} \cos\left(\frac{2\pi k}{n} - \theta\right), \tag{25}$$

$$\eta_o^n(\vec{x}) = r_n \|\vec{x}\| \max_{k \in \{1,\dots,n\}} \cos\left(\frac{(2k-1)\pi}{n} - \theta\right). \tag{26}$$

Considering η_e^n first, we see that since the angles $\frac{2k\pi}{n}$ are an angle $\frac{2\pi}{n}$ apart from each other for consecutive k's and since the maximization of cosine actually minimizes the angle $\frac{2\pi k}{n} - \theta$, for the $k' \in \{1,\dots,n\}$ which minimizes the angle we have $\frac{2\pi k'}{n} - \theta \leq \frac{\pi}{n}$ so that $\cos\left(\frac{2\pi k'}{n} - \theta\right) \geq \cos\left(\frac{\pi}{n}\right)$. The same arguments hold for η_o^n as well so if $\vec{x} \in B_{e/o}^n(r)$ for some $r > 0$, then

$$\eta_{e/o}^n(\vec{x}) \leq r \quad \Rightarrow \quad \|\vec{x}\| \leq \frac{r}{r_n \cos\left(\frac{\pi}{n}\right)} = r. \tag{27}$$

Hence, the level sets $B_{e/o}^n(r)$ are compact (convex) polytopes for all $r > 0$. Furthermore, each $B_{e/o}^n(r)$ has at most n extreme points since it is an intersection of n closed half-spaces in $\mathbb{R}^2$.

The functions η_e^n and η_o^n have the following connection:

$$\eta_{e/o}^n(\vec{x}) \leq r_n \eta_{o/e}^n(\vec{x}) \tag{28}$$

for all $\vec{x} \in \mathbb{R}^2$ and $r \geq 0$. This can be seen using the expressions from (25) and (26); for example

$$\eta_o^n(\vec{x}) = r_n \|\vec{x}\| \max_{k \in \{1,\dots,n\}} \cos\left(\frac{(2k-1)\pi}{n} - \theta\right) = r_n^2 \|\vec{x}\| \max_{k \in \{1,\dots,n\}} \cos\left(\frac{(2k-1)\pi}{n} - \theta\right) \cos\left(\frac{\pi}{n}\right)$$

$$= \frac{r_n^2 \|\vec{x}\|}{2} \max_{k \in \{1,\dots,n\}} \left[\cos\left(\frac{2(k-1)\pi}{n} - \theta\right) + \cos\left(\frac{2k\pi}{n} - \theta\right)\right]$$

$$\leq \frac{r_n}{2} \left[r_n \|\vec{x}\| \max_{k \in \{1,\dots,n\}} \cos\left(\frac{2(k-1)\pi}{n} - \theta\right) + r_n \|\vec{x}\| \max_{k \in \{1,\dots,n\}} \cos\left(\frac{2k\pi}{n} - \theta\right)\right]$$

$$= r_n \eta_e^n(\vec{x}).$$

Let us consider the specific level set $B_o^n(r_n)$. For each $k \in \{1, \ldots, n\}$, we define $\vec{s}_k = \left(r_n \cos\left(\frac{2k\pi}{n}\right), r_n \sin\left(\frac{2k\pi}{n}\right)\right)^T$ so that $s_k = (\vec{s}_k, 1)^T$. It is easy to see that $\eta_o^n(\vec{s}_k) = r_n$ so that $\vec{s}_k \in B_o^n(r_n)$ for all $k = 1, \ldots, n$. Furthermore, we have that $\|\vec{s}_k\| = r_n$ for all k so that each $\vec{s}_k$ lies on a circle of radius r_n centered at the origin. This shows that $\vec{s}_k$ is extreme in $B_o^n(r_n)$ for all $k = 1, \ldots, n$, since a non-trivial convex decomposition for $\vec{s}_k$ would contradict the fact that $\|\vec{x}\| \leq r_n$ for all $\vec{x} \in B_o^n(r_n)$. This, combined with the fact that $B_o^n(r_n)$ has at most n extreme points, shows that the extreme points of $B_o^n(r_n)$ are exactly the vectors $\vec{s}_k$ for all $k = 1, \ldots, n$. Hence, $s = (\vec{s}, 1) \in \mathcal{S}_n$ if and only if $\vec{s} \in B_o^n(r_n)$.

By similar arguments, we see that also $B_e^n(r)$ is a regular polygon whose extreme points are rotated and scaled from $\vec{s}_k$. For example, in the case of even n, we see that the effects lying on the hyperplane that contains all the non-trivial extreme effects can be characterized in terms of $B_e^n(r)$; namely, $e = \left(\vec{e}, \frac{1}{2}\right)^T \in \text{conv}\left(\{e_1, \ldots, e_n\}\right)$ if and only if $\vec{e} \in B_e^n\left(\frac{1}{2}\right)$. Similarly for odd polygons we have that $g = (\vec{g}, \sigma_n)^T \in \text{conv}\left(\{g_1, \ldots, g_n\}\right)$ if and only if $\vec{g} \in B_o^n(\sigma_n)$.

Hence, we can characterize (both the odd and even) polygon state spaces with the polyhedral functions η_o^n:

$$\mathcal{S}_n = \{(\vec{s}, 1)^T \in \mathbb{R}^3 \,|\, \eta_o^n(\vec{s}) \leq r_n\}. \tag{29}$$

Furthermore, for even n we have that

$$\text{conv}\left(\{e_1, \ldots, e_n\}\right) = \left\{\left(\vec{e}, \frac{1}{2}\right)^T \in \mathbb{R}^3 \,\Big|\, \eta_e^n(\vec{e}) \leq \frac{1}{2}\right\}, \tag{30}$$

and similarly for odd n

$$\text{conv}\left(\{g_1, \ldots, g_n\}\right) = \left\{(\vec{g}, \sigma_n)^T \in \mathbb{R}^3 \,|\, \eta_o^n(\vec{g}) \leq \sigma_n\right\}. \tag{31}$$

In both cases, the above sets serve as a compact bases for the positive dual cones in $\mathbb{R}^3$.

B.   Characterization of $\mathcal{T}_2$

The analysis of $\mathcal{T}_2$ on polygon state spaces is straight-forward. If $n = 3$, then the state space is a simplex and $\mathcal{T}_2 = \mathcal{O}(\mathcal{S}_3)$. In all other cases we have $\mathcal{T}_1 = \mathcal{T}_2$ as a result of Coro. 2.

C.   Characterization of $\mathcal{T}_3$

The post-processing equivalence classes of simulation irreducible observables on polygon state spaces were characterized in [14] where it was found that for an n-gon state space there exists m dichotomic and $\frac{1}{3}m(m-1)(m-2)$ trichotomic extreme simulation irreducible observables when $n = 2m$ for some $m \in \mathbb{N}$ (even polygons) and $\frac{1}{6}m(m+1)(2m+1)$ trichotomic extreme simulation irreducible observables when $n = 2m + 1$ for some $m \in \mathbb{N}$ (odd polygons).

For even polygons with $n = 2m$ where $m \geq 2$, there exists at least two inequivalent dichotomic simulation irreducible observables, so by Prop. 10 the set $\mathcal{T}_3$ coincides with the set of trivial observables.

For odd polygon state spaces we see that the extreme simulation irreducible observables have the same number of outcomes as the dimension of the effect space, so given that there are a finite number of them, it follows from Prop. 9 that $\mathcal{T}_3 \neq \mathcal{T}_1$. We continue to give a characterization of $\mathcal{T}_3$ for the odd polygon state spaces.

Let $\mathcal{S}_n$ be an odd polygon state space so that $n = 2m + 1$ for some $m \in \mathbb{N}$. There are $q_m := \frac{1}{6}m(m+1)(2m+1)$ extreme simulation irreducible observables that generate the cones generated by all the simulation irreducible observables. By using some enumeration $\mathsf{B}^{(1)}, \ldots, \mathsf{B}^{(q_m)}$ for these observables, we have that $\mathcal{O}_{irr}^{ext}(\mathcal{S}_n) = \{\mathsf{B}^{(i)}\}_{i=1}^{q_m}$ so that for an observable $\mathsf{A} \in \mathcal{O}(\mathcal{S}_n)$ we have

$$\mathsf{A} \in \mathcal{T}_3 \quad \Leftrightarrow \quad \mathsf{A}_x \in \bigcap_{j=1}^{q_m} \text{cone}\left(\{\mathsf{B}_x^{(j)}\}_{x \in \Omega_{\mathsf{B}^{(j)}}}\right) \quad \forall x \in \Omega_\mathsf{A}.$$

We can show that there are certain extreme simulation irreducible observables that are enough to characterize the above intersection. Let $\mathsf{B} \in \mathcal{O}_{irr}^{ext}(\mathcal{S})$. Since for all $k \in \{1, 2, 3\}$ the effects B_k are indecomposable, for each $k \in \{1, 2, 3\}$ there exists $0 < c_k \leq 1$ and effect $g_{i_k} \in \{g_1, \ldots, g_{2m+1}\}$ such that $\mathsf{B}_k = c_k g_{i_k}$. We see that we only need to consider the case when $i_k \in \{j, j+m, j+m+1\}$ for all $k \in \{1, 2, 3\}$ for some $j \in \{1, \ldots, 2m+1\}$, where the addition of the indices is taken modulo $2m + 1$.
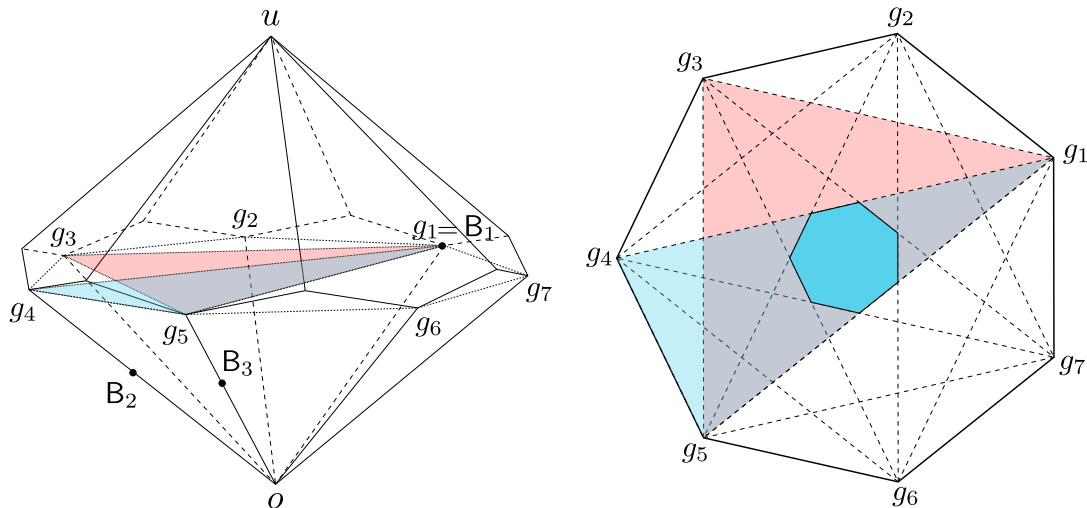
FIG. 5. Simulation irreducible observable B on the heptagon state space with $\mathsf{B}_1 = g_1$, $\mathsf{B}_2 = 0.555g_4$ and $\mathsf{B}_3 = 0.555g_5$. The base of the cone generated by the effects of B forms a (blue) triangle on the base of the positive cone (left). The intersection of the bases of all the simulation irreducible observables forms another polygon (right).

Proposition 12. *An observable* $\mathsf{A} \in \mathcal{O}(\mathcal{S}_{2m+1})$ *on an odd polygon state space* $\mathcal{S}_{2m+1}$ *is in* $\mathcal{T}_3$ *if and only if*

$$\mathsf{A}_x \in \bigcap_{i=1}^{2m+1} \mathrm{cone}\left(\{g_i, g_{i+m}, g_{i+m+1}\}\right) \quad \forall x \in \Omega_\mathsf{A}.$$

The complete proof of the proposition can be found in the appendix but one can easily convince oneself by looking at Fig. 5 which shows the case of heptagon effect space. For each $\mathsf{B} \in \mathcal{O}_{irr}^{ext}(\mathcal{S}_n)$ we can consider the base of the cone $\mathrm{cone}\left(\{\mathsf{B}_1, \mathsf{B}_2, \mathsf{B}_3\}\right)$ on the plane containing the indecomposable extreme effects $\{g_i\}_{i=1}^n$, where the base takes the form of a triangle that contains the middle point $\sigma_n u$. We can see that in order to characterize the intersection of such cones, it is enough to consider the intersection of their respective bases, or triangles containing $\sigma_n u$, equivalently. In the left of Fig. 5, the bases (coloured as blue and red) of two extreme simulation irreducible observables are shown with the whole effects space. On the right is depicted all the triangles (formed by dashed lines) of all the bases on the plane with the blue and red bases from the left figure also shown on the right. We see that the base of the intersection of the cones (darker blue area) is characterized by triangles with vertices g_i, g_{i+m} and g_{i+m+1} (like the blue triangle) so that their intersection is always contained in the intersection of other triangles (like the red triangle).

We are going to proceed with finding the base of the cone $\bigcap_{i=1}^{2m+1} \mathrm{cone}\left(\{g_i, g_{i+m}, g_{i+m+1}\}\right)$ by identifying the extreme points of the base $\bigcap_{i=1}^{2m+1} \mathrm{conv}\left(\{g_i, g_{i+m}, g_{i+m+1}\}\right)$. Let us denote

$$L_i = \mathrm{conv}\left(\{g_i, g_{i+m}\}\right)$$

and

$$C_m = \bigcap_{i=1}^{2m+1} \mathrm{conv}\left(\{g_i, g_{i+m}, g_{i+m+1}\}\right).$$

We will approach the problem as follows: at first, we will identify that C_m must be a polygon itself by looking at its relation with the line segments L_i. Then we will find the form of the extreme points of C_m and in the end we will identify them. During the calculations we will work only in the 2-dimensional vector space given by $\mathrm{aff}\left(\{g_i\}_{i=1}^{2m+1}\right)$.

It is very useful to realize that L_i generate hyperplanes in $\mathbb{R}^2$ and that C_m is an intersection of the halfspaces corresponding to the hyperplanes L_i that contain the point 0. It follows that we must have $L_i \cap C_m \neq \emptyset$, $\forall i \in \{1, \ldots, 2m+1\}$, otherwise there would be hyperplanes separating L_i and C_m, which is a contradiction with C_m being given as an intersection of halfspace corresponding to L_i. Since there are only $2m+1$ different line segments L_i it follows that C_m must have exactly $2m+1$ edges and from the symmetry it also follows that C_m must be a polygon. Now the only thing we need to do is to identify the extreme points of C_m.

Since the line segments L_i must intersect C_m it follows that the extreme points of C_m must correspond to the intersections of these line segments. Let us denote
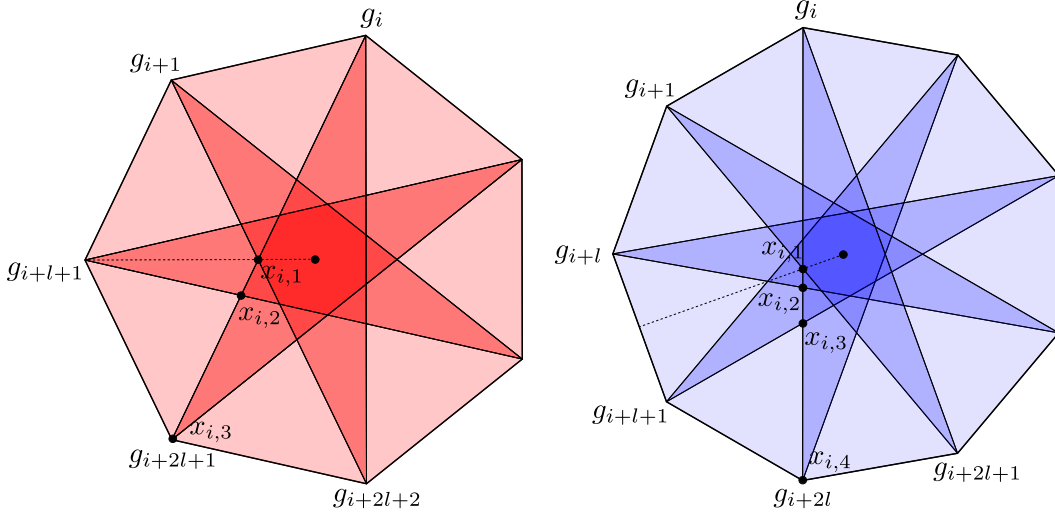
$$x_{i,j} = L_i \cap L_{i+j}$$

FIG. 6. The points $\{x_{i,j}\}_{j=1}^m$ for a fixed i and the orientation of the inner polygon for odd (left, $m = 3$) and even (right, $m = 4$) m.

where $j \in \{1, \ldots, m\}$, where if $i + j \geq 2m + 1$, then we take $(i + j) \mod (2m + 1)$. Also note that considering $j \geq m + 1$ would be redundant. The next key step is to characterize the relation of $x_{i,j}$ and C_m. We can show the following.

Lemma 4. $x_{i,1}$ *are the extreme points of* C.

Again, the complete proof of the lemma can be found in the appendix, but one can easily convince oneself by looking at Fig. 6, where the points $\{x_{i,j}\}_{j=1}^m$ are depicted for a fixed i in the case of a heptagon (left) and nonagon (right) state space.

We are almost ready to move on to the complete characterization of $\mathcal{T}_3$ in odd polygon theories in terms of the previously defined $\eta_{e/o}^n$ functions. We will still make a few remarks on the inner polygons C_m.

Let $n = 2m + 1$. We will consider separately, although analogously, the cases of even and odd m. This is because of the orientation of the inner polygon C_m with respect to the outer polygon conv $(\{g_1, \ldots, g_n\})$. To show the difference between even and odd m, let us consider the intersection point of the boundary of the outer polygon and the half-line through an extreme point $x_{i,1}$ of the inner polygon emanating from the centroid $(0, 0, \sigma_n)^T$. If this intersection point is also an extreme point of the outer polygon, then both the inner and outer polygons are similarly oriented; otherwise they are differently oriented.

As $x_{i,1} = L_i \cap L_{i+1} = \text{conv}\,(\{g_i, g_{i+m}\}) \cap \text{conv}\,(\{g_{i+1}, g_{i+m+1}\})$, it is clear that the half-line through $x_{i,1}$ that emanates from the centroid meets the boundary of the outer polygon at some of the line segments conv $(\{g_{i+1}, g_{i+2}\})\,, \ldots,$ conv $(\{g_{i+m-1}, g_{i+m}\})$.

For even m, i.e., for $m = 2l$ for some $l \in \mathbb{N}$, there exists an even number $2(l-1)$ of vertices g_j between the vertices g_{i+1} and g_{i+m} so that there is an odd number of such edges. From the symmetry it follows that for even m, the intersection point must lie in the middle of the midmost edge conv $(\{g_{i+l}, g_{i+l+1}\})$. Thus, for even m, the inner polygon C_m is differently oriented with respect to the outer polygon conv $(\{g_1, \ldots, g_n\})$.

By contrast, for odd m, i.e., for $m = 2l + 1$ for some $l \in \mathbb{N}$, there exists an even number of such edges, which together with the symmetry of the situation tells us that now the intersection point is exactly one of the vertices of the outer polygon, namely g_{i+l+1}. Thus, for odd m, the inner polygon is similarly oriented to the outer polygon. The orientations of the inner polygon for odd and even m are depicted in Fig. 6.

As we saw in the beginning of the section, the orientation of the polygon can also be characterized with the $\eta_{e/o}^n$ functions. Thus, in the characterization of $\mathcal{T}_3$ we must use either η_e^n or η_o^n depending on the parity of m.

Proposition 13. *An observable* $\mathsf{A} \in \mathcal{O}(\mathcal{S}_{2m+1})$ *with effects* $\mathsf{A}_x = \alpha_x(\vec{a}_x, \sigma_{2m+1})^T$ *for all* $x \in \Omega_\mathsf{A}$ *is compatible with every other observable if and only if for all* $x \in \Omega_\mathsf{A}$

$$\eta_e^n(\vec{a}_x) \leq \sigma_n r_n \sin\left(\frac{\pi}{2n}\right)$$

if $m = 2l$ *for some* $l \in \mathbb{N}$, *or*

$$\eta_o^n(\vec{a}_x) \leq \sigma_n r_n \sin\left(\frac{\pi}{2n}\right)$$

if $m = 2l + 1$ for some $l \in \mathbb{N}$.

Proof. By Prop. 6 it follows that $\mathsf{A} \in \mathcal{T}_3$ if and only if $(\vec{a}_x, \sigma_n)^T \in C_n$ for all $x \in \Omega_A$, and from Lemma 4 we know that the $x_{i,1} = (\vec{x}_{i,1}, \sigma_n)^T$ are the extreme points of C_n. Thus, if we show that $\|\vec{x}_{i,1}\| = \eta_{e/o}^n(\vec{x}_{i,1}) = \sigma_n r_n \sin\left(\frac{\pi}{2n}\right)$, it follows that $C_n = \left\{ (\vec{x}, \sigma_n)^T \in \mathbb{R}^3 \mid \eta_{e/o}^n(\vec{x}) \le \sigma_n r_n \sin\left(\frac{\pi}{2n}\right) \right\}$ which will prove the claim.

From $x_{i,j} = L_i \cap L_{i+j}$ we have that $x_{i,1} = \lambda_i g_i + (1 - \lambda_1) g_{i+m}$, where $\lambda_1 = 1 - \frac{1}{2} r_n = \frac{3\sigma_n - 1}{2\sigma_n}$. By using (rather a lot of) trigonometric identities we find that

$$
x_{k,1} = \begin{pmatrix} -\dfrac{\sin\left(\frac{\pi}{2n}\right)}{1 + \cos\left(\frac{\pi}{n}\right)} \sin\left(\dfrac{(4k+1)\pi}{2n}\right) \\[2mm] \dfrac{\sin\left(\frac{\pi}{2n}\right)}{1 + \cos\left(\frac{\pi}{n}\right)} \cos\left(\dfrac{(4k+1)\pi}{2n}\right) \\[2mm] \sigma_n \end{pmatrix},
$$

so that

$$
\vec{x}_{k,1} = \sigma_n r_n \sin\left(\frac{\pi}{2n}\right) \begin{pmatrix} -\sin\left(\dfrac{(4k+1)\pi}{2n}\right) \\[2mm] \cos\left(\dfrac{(4k+1)\pi}{2n}\right) \end{pmatrix},
$$

from which it is easy to see that $\|\vec{x}_{k,1}\| = \sigma_n r_n \sin\left(\frac{\pi}{2n}\right)$ for all $k \in \{1, \dots, n\}$.

We also see that (the simplified expressions of) $\eta_e^n(\vec{x}_{k,1})$ and $\eta_o^n(\vec{x}_{k,1})$ then read as

$$
\eta_e^n(\vec{x}_{k,1}) = \sigma_n r_n^2 \sin\left(\frac{\pi}{2n}\right) \max_{j \in \{1, \dots, n\}} \sin\left(\frac{(4j - 4k - 1)\pi}{2n}\right) \tag{32}
$$

and

$$
\eta_o^n(\vec{x}_{k,1}) = \sigma_n r_n^2 \sin\left(\frac{\pi}{2n}\right) \max_{j \in \{1, \dots, n\}} \sin\left(\frac{(4j - 4k - 3)\pi}{2n}\right). \tag{33}
$$

In both cases the maximum is attained when the expression inside the sine is closest to $\pi/2$. Now depending on the parity of m, this happens for different values of j resulting in different expressions. For $m = 2l$ for some $l \in \mathbb{N}$, we find that the maximum in Eq. (32) is attained for $j \in \{k + l, k + l + 1\}$ and similarly the maximum in Eq. (33) is attained for $j = k + l + 1$ so that for this case we have

$$
\eta_e^{4l+1}(\vec{x}_{k,1}) = \sigma_{4l+1} r_{4l+1}^2 \sin\left(\frac{\pi}{2(4l+1)}\right) \sin\left(\frac{(4l-1)\pi}{2(4l+1)}\right)
$$

$$
= \sigma_{4l+1} r_{4l+1} \sin\left(\frac{\pi}{2(4l+1)}\right) = \|\vec{x}_{k,1}\|.
$$

However, for $m = 2l + 1$ for some $l \in \mathbb{N}$ we have that the maximum in Eq. (32) is attained for $j = k + l + 1$ and similarly the maximum in Eq. (33) is attained for $j \in \{k + l, k + l + 1\}$ so that for this case we have

$$
\eta_o^{4l+3}(\vec{x}_{k,1}) = \sigma_{4l+3} r_{4l+3}^2 \sin\left(\frac{\pi}{2(4l+3)}\right) \sin\left(\frac{(4l+1)\pi}{2(4l+3)}\right)
$$

$$
= \sigma_{4l+3} r_{4l+3} \sin\left(\frac{\pi}{2(4l+3)}\right) = \|\vec{x}_{k,1}\|.
$$

$\square$

D. Noise content

The noise content $w(\mathsf{A}; \mathcal{N})$ of an observable $\mathsf{A} \in \mathcal{O}(\mathcal{S})$ on a state space $\mathcal{S}$ with respect to a noise set $\mathcal{N} \subset \mathcal{O}(\mathcal{S})$ is defined [9] as

$$
w(\mathsf{A}; \mathcal{N}) = \sup\{\lambda \in [0, 1] \mid \mathsf{A} = \lambda \mathsf{N} + (1 - \lambda)\mathsf{B} \text{ for some } \mathsf{N} \in \mathcal{N} \text{ and } \mathsf{B} \in \mathcal{O}(\mathcal{S})\}.
$$

When describing noisy observables, the noise is most commonly added externally to an observable, but the noise content describes the amount of noise that an observable already has intrinsically. Usually the noise set is taken to be the set of trivial observables $\mathcal{T}_1$.

Examining Prop. 13 more closely, the set $\mathcal{T}_3$ seems to be quite noisy in the sense that the effects of observables in $\mathcal{T}_3$ are scattered quite closely to the trivial effects on the line segment $\mathrm{conv}(\{o, u\})$. Our aim is to show this remark quantitatively by showing that an observable that is compatible with every other observable must have a quite high noise content with respect to the trivial observables. We also show that an observable with a high enough noise content is indeed compatible with every other observable on odd polygon state spaces.

For the noise set $\mathcal{N} = \mathcal{T}_1$, the noise content of an observable $\mathsf{A} \in \mathcal{O}(\mathcal{S})$ takes a rather simple form [9]:

$$w(\mathsf{A}; \mathcal{T}_1) = \sum_{x \in \Omega_\mathsf{A}} \min_{s \in \mathcal{S}} \mathsf{A}_x(s), \tag{34}$$

and furthermore if the state space is a polytope (as is in the case of polygons), we have that

$$w(\mathsf{A}; \mathcal{T}_1) = \sum_{x \in \Omega_\mathsf{A}} \min_{s \in \mathcal{S}^{ext}} \mathsf{A}_x(s), \tag{35}$$

where $\mathcal{S}^{ext}$ denotes the set of extreme points of $\mathcal{S}$.

We start by making a connection between $\min_{s \in \mathcal{S}} \mathsf{A}_x(s)$ and $\eta_o^n(\vec{a}_x)$. As before, for each effect A_x there exists $\alpha_x > 0$ such that $\mathsf{A}_x = \alpha_x a_x$ for some $a_x = (\vec{a}_x, \sigma_n)^T$, where $\vec{a}_x \in \mathbb{R}^2$. Since $a_x \in \mathrm{conv}(\{g_1, \ldots, g_n\})$ for all $x \in \Omega_\mathsf{A}$, we have that for all $x \in \Omega_\mathsf{A}$ there exists $\lambda_x \in [0, 1]$ such that $a_x = \lambda_x h_x + (1 - \lambda_x)\sigma_n u$ for some

$$h_x \in \partial\mathrm{conv}(\{g_1, \ldots, g_n\}) = \{(\vec{g}, \sigma_n)^T \in \mathrm{conv}(\{g_1, \ldots, g_n\}) \mid \eta_o^n(\vec{g}) = \sigma_n\}.$$

We note that since h_x lies on the boundary of the convex hull of the indecomposable effects, for all $x \in \Omega_\mathsf{A}$, there exists $i_x \in \{1, \ldots, n\}$ such that $h_x \in \mathrm{conv}(\{g_{i_x}, g_{i_x+1}\})$. Since g_{i_x} and g_{i_x+1} are indecomposable, by Prop. 16 they give zero for some maximal faces G_{i_x} and G_{i_x+1} of $\mathcal{S}_n$. Furthermore, it is easy to see that they must be adjacent maximal faces so that there exists an extreme state $s_{i_x} \in \mathcal{S}_n$ such that $h_x(s_{i_x}) = 0$. Thus,

$$\min_{s \in \mathcal{S}_n^{ext}} \mathsf{A}_x(s) = \alpha_x \min_{s \in \mathcal{S}_n^{ext}} [\lambda_x h_x(s) + (1 - \lambda_x)\sigma_n u(s)] = \alpha_x \lambda_x \min_{s \in \mathcal{S}_n^{ext}} h_x(s) + \alpha_x(1 - \lambda_x)\sigma_n$$

$$= \alpha_x \lambda_x h_x(s_{i_x}) + \alpha_x(1 - \lambda_x)\sigma_n = \alpha_x(1 - \lambda_x)\sigma_n$$

for all $x \in \Omega_\mathsf{A}$. If we denote $h_x = (\vec{h}_x, \sigma_n)^T$, we then see that $\vec{a}_x = \lambda_x \vec{h}_x$ and

$$\eta_o^n(\vec{a}_x) = \lambda_x \eta_o^n(\vec{h}_x) = \lambda_x \sigma_n = \sigma_n - \frac{1}{\alpha_x} \min_{s \in \mathcal{S}_n^{ext}} \mathsf{A}_x(s)$$

for all $x \in \Omega_\mathsf{A}$. Thus, $\min_{s \in \mathcal{S}^{ext}} \mathsf{A}_x(s) = \alpha_x [\sigma_n - \eta_o^n(\vec{a}_x)]$ for all $x \in \Omega_\mathsf{A}$.

We can now show the following.

Proposition 14. *Let $\mathsf{A} \in \mathcal{O}(\mathcal{S}_n)$ be an observable on an odd polygon state space $\mathcal{S}_n$ with effects $\mathsf{A}_x = \alpha_x(\vec{a}_x, \sigma_n)$ for all $x \in \Omega_\mathsf{A}$. If $\mathsf{A} \in \mathcal{T}_3$, then*

$$w(\mathsf{A}; \mathcal{T}_1) \geq 1 - r_n \sin\left(\frac{\pi}{2n}\right) \tag{36}$$

if $n = 4l + 3$ for some $l \in \mathbb{N}$, or

$$w(\mathsf{A}; \mathcal{T}_1) \geq 1 - r_n^2 \sin\left(\frac{\pi}{2n}\right) \tag{37}$$

if $n = 4l + 1$ for some $l \in \mathbb{N}$.

Proof. As was established above, we have that $\min_{s \in \mathcal{S}^{ext}} \mathsf{A}_x(s) = \alpha_x(\sigma_n - \eta_o^n(\vec{a}_x))$.

For $n = 4l + 3$, we have from Prop. 13 that $\eta_o^n(\vec{a}_x) \leq r_n \sigma_n \sin\left(\frac{\pi}{2n}\right)$ for all $x \in \Omega_\mathsf{A}$ so that

$$w(\mathsf{A}; \mathcal{T}_1) = \sum_{x \in \Omega_\mathsf{A}} \min_{s \in \mathcal{S}^{ext}} \mathsf{A}_x(s) = \sum_{x \in \Omega_\mathsf{A}} \alpha_x(\sigma_n - \eta_o^n(\vec{a}_x))$$

$$\geq \sum_{x \in \Omega_\mathsf{A}} \alpha_x \sigma_n \left(1 - r_n \sin\left(\frac{\pi}{2n}\right)\right)$$

$$= 1 - r_n \sin\left(\frac{\pi}{2n}\right),$$

n	3	5	7	9	11	13	15	17	19	$\cdots$	$\to \infty$
R.H.S. of (36)	0	–	0.753	–	0.852	–	0.893	–	0.916	$\cdots$	$\to 1$
R.H.S. of (37)	–	0.528	–	0.803	–	0.872	–	0.905	–	$\cdots$	$\to 1$

TABLE I. The lower bounds of Eq. (36) and (37) for the noise contents of observables in $\mathcal{T}_3$ for the first few odd polygons and the limit $n \to \infty$.

where on the last line we have used the fact that $\sum_{x \in \Omega_\mathsf{A}} \alpha_x = 1/\sigma_n$ which follows from the normalization of A.

For $n = 4l + 1$, we have from Prop. 13 that $\eta_e^n(\vec{a}_x) \leq r_n \sigma_n \sin\left(\frac{\pi}{2n}\right)$ for all $x \in \Omega_\mathsf{A}$. From Eq. (28) we get that $\eta_o^n(\vec{a}_x) \leq r_n \eta_e^n(\vec{a}_x)$ for all $x \in \Omega_\mathsf{A}$ so that from similar calculation as above we get that $w(\mathsf{A}; \mathcal{T}_1) \geq 1 - r_n^2 \sin\left(\frac{\pi}{2n}\right)$. $\square$

The lower bounds of the noise content from the previous proposition for the first few polygons are presented in Table I. We see that for $n = 3$, i.e., when the state space is classical, Eq. (36) gives the trivial lower bound zero, but already for the pentagon ($n = 5$) Eq. (37) shows that the noise content of an observable in $\mathcal{T}_3$ must be more than $1/2$. We see that as the number of vertices in the polygons increase, so does the noise content of observables in $\mathcal{T}_3$ for both Eq. (36) and (37). In the limit where $n \to \infty$ the right hand sides (R.H.S.) of both equations give the limit 1, so that the observables in $\mathcal{T}_3$ become trivial. Indeed, as the number of vertices approaches infinity, the state space becomes shaped like a disc, which is seen to be a point-symmetric state space so that by Cor. 4 we have $\mathcal{T}_1 = \mathcal{T}_3$.

From the other point of view, we can ask if sufficiently noisy observables are necessarily compatible with every other observable. For that, let us consider the binarizations of an observable $\mathsf{A} \in \mathcal{O}(\mathcal{S}_n)$, i.e., binary observables $\hat{\mathsf{A}}^{(x)}$ with effects $\hat{\mathsf{A}}_+^{(x)} = \mathsf{A}_x$ and $\hat{\mathsf{A}}_-^{(x)} = u - \mathsf{A}_x$ for all $x \in \Omega_\mathsf{A}$. The noise content for these binarizations then read as

$$w(\hat{\mathsf{A}}^{(x)}; \mathcal{T}_1) = \min_{s \in \mathcal{S}_n^{ext}} \mathsf{A}_x(s) + \min_{s \in \mathcal{S}_n^{ext}} (u - \mathsf{A}_x)(s) = 1 + \min_{s \in \mathcal{S}_n^{ext}} \mathsf{A}_x(s) - \max_{s \in \mathcal{S}_n^{ext}} \mathsf{A}_x(s)$$

for all $x \in \Omega_x$.

Denoting the extreme points of the state space $\mathcal{S}_{2m+1}$ by $s_k = (\vec{s}_k, 1)^T$, from the definition of η_e^n we see that

$$\eta_e^n(\vec{a}_x) = \max_{k \in \{1, \ldots, n\}} \vec{a}_x \cdot \vec{s}_k = \frac{1}{\alpha_x} \max_{k \in \{1, \ldots, n\}} \mathsf{A}_x(s_k) - \sigma_n = \frac{1}{\alpha_x} \max_{s \in \mathcal{S}_n^{ext}} \mathsf{A}_x(s) - \sigma_n$$

for all $x \in \Omega_\mathsf{A}$. Hence, together with the previous expressions for $\min_{x \in \Omega_\mathsf{A}} \mathsf{A}_x(s)$, we have shown the following for the binarizations $\hat{\mathsf{A}}^{(x)}$ of an observable A:

$$w(\hat{\mathsf{A}}^{(x)}; \mathcal{T}_1) = 1 - \alpha_x \left[\eta_e^n(\vec{a}_x) + \eta_o^n(\vec{a}_x) \right] \tag{38}$$

for all $x \in \Omega_\mathsf{A}$. We can now show that observables that have a high enough noise content are indeed included in $\mathcal{T}_3$.

Proposition 15. *Let $\mathsf{A} \in \mathcal{O}(\mathcal{S}_n)$ be an observable on an odd polygon state space $\mathcal{S}_n$ with effects $\mathsf{A}_x = \alpha_x(\vec{a}_x, \sigma_n)$ for all $x \in \Omega_\mathsf{A}$. If*

$$\frac{1 - w(\mathsf{A}; \mathcal{T}_1)}{\min_{x \in \Omega_\mathsf{A}} \alpha_x} \leq \sin\left(\frac{\pi}{2n}\right), \tag{39}$$

then A is compatible with every other observable on $\mathcal{S}_n$.

Proof. From the previous expression for the noise contents of the binarizations $\hat{\mathsf{A}}^{(x)}$ of A, and by using Eq. (28), we have that

$$\eta_{e/o}^n(\vec{a}_x) = \frac{1 - w(\hat{\mathsf{A}}^{(x)}; \mathcal{T}_1)}{\alpha_x} - \eta_{o/e}^n(\vec{a}_x) \leq \frac{1 - w(\hat{\mathsf{A}}^{(x)}; \mathcal{T}_1)}{\alpha_x} - \frac{\eta_{e/o}^n(\vec{a}_x)}{r_n}.$$

Since $\mathcal{T}_1$ is closed under post-processing and since $\hat{\mathsf{A}}^{(x)}$ is clearly a post-processing of A for each $x \in \Omega_x$, we have by the basic properties of the noise content [9] that $w(\hat{\mathsf{A}}^{(x)}; \mathcal{T}_1) \geq w(\mathsf{A}; \mathcal{T}_1)$ for all $x \in \Omega_\mathsf{A}$. Thus, by rearranging the previous expression we have that

$$\eta_{e/o}^n(\vec{a}_x) \leq \left(1 + \frac{1}{r_n}\right)^{-1} \frac{1 - w(\hat{\mathsf{A}}^{(x)}; \mathcal{T}_1)}{\alpha_x} = \sigma_n r_n \left(\frac{1 - w(\hat{\mathsf{A}}^{(x)}; \mathcal{T}_1)}{\alpha_x}\right)$$

$$\leq \sigma_n r_n \left(\frac{1 - w(\mathsf{A}; \mathcal{T}_1)}{\alpha_x}\right) \leq \sigma_n r_n \left(\frac{1 - w(\mathsf{A}; \mathcal{T}_1)}{\min_{x \in \Omega_\mathsf{A}} \alpha_x}\right)$$

for all $x \in \Omega_\mathsf{A}$, where we have noticed that $(1 + 1/r_n)^{-1} = \sigma_n r_n$. Now, if Eq. (39) holds, from Prop. 13 it then follows that $\mathsf{A} \in \mathcal{T}_3$. $\square$

ACKNOWLEDGMENTS

[1] Howard Barnum, Jonathan Barrett, Matthew Leifer, and Alexander Wilce, "Cloning and Broadcasting in Generic Probabilistic Theories," (2006), arXiv:0611295.

[2] Howard Barnum, Jonathan Barrett, Matthew Leifer, and Alexander Wilce, "Generalized No-broadcasting theorem," Phys. Rev. Lett. **99**, 240501 (2007), arXiv:0707.0620.

[3] Giulio Chiribella, Giacomo Mauro D'Ariano, and Paolo Perinotti, "Probabilistic theories with purification," Phys. Rev. A **81**, 062348 (2010), arXiv:0908.1583.

[4] Howard Barnum and Alexander Wilce, "Information processing in convex operational theories," Electr. Notes Theor. Comput. Sci. **270**, 3–15 (2011), arXiv:0908.2352.

[5] Teiko Heinosaari and Mário Ziman, *The Mathematical Language of Quantum Theory. From Uncertainty to Entanglement* (Cambridge University Press, 2012).

[6] G. C. Wick, A. S. Wightman, and E. P. Wigner, "The intrinsic parity of elementary particles," Phys. Rev. **88**, 101–105 (1952).

[7] Gen Kimura, Koji Nuida, and Hideki Imai, "Distinguishability measures and entropies for general probabilistic theories," Rep. Math. Phys. **66**, 175 – 206 (2010).

[8] Isaac Namioka and Robert Phelps, "Tensor products of compact convex sets," Pacific J. Math. **31**, 469–480 (1969).

[9] Sergey N. Filippov, Teiko Heinosaari, and Leevi Leppäjärvi, "Necessary condition for incompatibility of observables in general probabilistic theories," Phys. Rev. A **95**, 032127 (2017), arXiv:1609.08416.

[10] Teiko Heinosaari and Takayuki Miyadera, "Incompatibility of quantum channels," J. Phys. A. Math. Gen. **50**, 135302 (2017), arXiv:1608.01794.

[11] Jonathan Barrett, Noah Linden, Serge Massar, Stefano Pironio, Sandu Popescu, and David Roberts, "Nonlocal correlations as an information-theoretic resource," Phys. Rev. A **71**, 022101 (2005), arXiv:0404097.

[12] Leonardo Guerini, Jessica Bavaresco, Marcelo Terra Cunha, and Antonio Acín, "Operational framework for quantum measurement simulability," J. Math. Phys. **58**, 092102 (2017), arXiv:1705.06343.

[13] Michał Oszmaniec, Leonardo Guerini, Peter Wittek, and Antonio Acín, "Simulating Positive-Operator-Valued Measures with Projective Measurements," Phys. Rev. Lett. **119**, 190501 (2017), arXiv:1609.06139.

[14] Sergey N. Filippov, Teiko Heinosaari, and Leevi Leppäjärvi, "Simulability of observables in general probabilistic theories," Phys. Rev. A **97**, 062102 (2018), arXiv:1803.11006.

[15] Michał Oszmaniec, Filip B. Maciejewski, and Zbigniew Puchała, "All quantum measurements can be simulated using projective measurements and postselection," (2018), arXiv:1807.08449.

[16] Martin Plávala, "All measurements in a probabilistic theory are compatible if and only if the state space is a simplex," Phys. Rev. A **94**, 042108 (2016), arXiv:1608.05614.

[17] Peter Janotta, Christian Gogolin, Jonathan Barrett, and Nicolas Brunner, "Limits on nonlocal correlations from the structure of the local state space," New J. Phys. **13**, 063024 (2011), arXiv:1012.1215.

[18] Ralph Tyrrell Rockafellar, *Convex Analysis*, Princeton landmarks in mathematics and physics (Princeton University Press, 1997).

Appendix A: Some results on the structure of GPTs

In this appendix we will prove several minor results about the structure of general probabilistic theories that were needed in the calculations. We will denote the interior of the set $\mathcal{S}$ by $\text{intr}(\mathcal{S})$.

Definition 8. Let $\mathcal{S}$ be a state space and let $E \subset \mathcal{S}$ be a convex subset. We say that E is a face of $\mathcal{S}$ if $z \in E$ and $z = \lambda x + (1-\lambda)y$ for some $x, y \in \mathcal{S}$ and $\lambda \in [0,1]$ implies $x, y \in E$.

Definition 9. Let $E \subset \mathcal{S}$, we say that E is a maximal face if E is a face and for any $x \in \mathcal{S} \smallsetminus E$ we have $\text{conv}(E \cup \{x\}) \cap \text{intr}(\mathcal{S}) \neq \varnothing$.

If $\mathcal{S}$ is d-dimensional and has a finite number of extreme points, then maximal faces are the $(d-1)$-dimensional faces of $\mathcal{S}$. From a geometrical perspective their special properties all follow from the requirement that $\text{conv}(E \cup \{x\}) \cap \text{intr}(\mathcal{S}) \neq \varnothing$.

Lemma 5. *Let $\mathcal{S}$ be a state space, let $e \in \mathcal{E}(\mathcal{S})$ and let $E_0 = \{x \in \mathcal{S} : e(x) = 0\}$. If E_0 is a maximal face, then e is indecomposable.*

Proof. Let $f \in \mathcal{E}(\mathcal{S})$, denote $F_0 = \{x \in \mathcal{S} : f(x) = 0\}$ and assume $e \geq f$. It follows that we must have $E_0 \subset F_0$ and since E_0 is maximal face it follows that either $F_0 = \mathcal{S}$ or $F_0 = E_0$.

If $F_0 = \mathcal{S}$ then $f = 0$. If $F_0 = E_0$, then pick $x \in \mathcal{S}$ such that $x \notin E_0$. Both e and f are uniquely defined by the values $e(x)$ and $f(x)$, because E_0 is a maximal face. This implies that we have $f = \frac{f(x)}{e(x)} e$, which shows that e is indecomposable. $\qquad\square$

Proposition 16. *Assume that $\mathcal{S}$ has only a finite number of extreme points. Let $e \in \mathcal{E}(\mathcal{S})$ and let $E_0 = \{x \in \mathcal{S} : e(x) = 0\}$, then e is indecomposable if and only if E_0 is a maximal face.*

Proof. Assume that $E_0 = \{x \in \mathcal{S} : e(x)\}$ is not a maximal face, then there is a maximal face F_0 such that $E_0 \subset F_0$ [16]. Moreover let $f \in \mathcal{E}(\mathcal{S})$ be such that $F_0 = \{x \in \mathcal{S} : f(x) = 0\}$ and denote $G = \{y \in \mathcal{S} : y \notin E_0, y \text{ is extreme}\}$ and

$$m = \min_{y \in G} e(y).$$

Clearly $m > 0$. We will show that $e \geq mf$; let $z \in \mathcal{S}$ be an extreme point, then either $z \in E_0$ or $z \in G$. If $z \in E_0 \subset F_0$, then $e(z) = 0 \geq 0 = mf(z)$. If $z \in G$, then $e(z) \geq m \geq mf(z)$. $\qquad\square$

Appendix B: Proof of Prop. 12

We recall from [14] that the extreme simulation irreducible observables are characterized by triangles on the base $\mathrm{conv}(\{g_1, \ldots, g_n\})$ with vertices from the set $\{g_1, \ldots, g_{2m+1}\}$ such that $\sigma_n u$ is included in the triangles. We show that such triangles are in one-to-one correspondence with the extreme simulation irreducible observables.

To see this, first let $\mathsf{B} \in \mathcal{O}_{irr}^{ext}(\mathcal{S}_{2m+1})$ so that $\Omega_\mathsf{B} = \{1, 2, 3\}$. Since for all $k \in \{1, 2, 3\}$ the effects B_k are indecomposable, for each $k \in \{1, 2, 3\}$ there exists $0 < c_k \leq 1$ and effect $g_{i_k} \in \{g_1, \ldots, g_{2m+1}\}$ such that $\mathsf{B}_k = c_k g_{i_k}$. From the normalization of B it follows that

$$u = c_1 g_{i_1} + c_2 g_{i_2} + c_3 g_{i_3}$$

so that from the z–components of the vectors we get a requirement that $c_1 + c_2 + c_3 = \frac{1}{\sigma_n}$. Thus, if we denote the sum $c_1 + c_2 + c_3$ by c, we see that

$$\sigma_n u = \frac{c_1}{c} g_{i_1} + \frac{c_2}{c} g_{i_2} + \frac{c_3}{c} g_{i_3} \tag{B1}$$

which shows that the vertices $\{g_{i_1}, g_{i_2}, g_{i_3}\}$ form a triangle $\mathrm{conv}(\{g_{i_1}, g_{i_2}, g_{i_3}\})$ on the base $\mathrm{conv}(\{g_1, \ldots, g_{2m+1}\})$ such that $\sigma_n u \in \mathrm{conv}(\{g_{i_1}, g_{i_2}, g_{i_3}\})$.

To see the contrary, let j_1, j_2, j_3 be any three indices from the set $\{1, \ldots, 2m+1\}$ such that $\sigma_n u \in \mathrm{conv}(\{g_{j_1}, g_{j_2}, g_{j_3}\})$. Thus, there exists convex coefficients $\tilde{d}_1, \tilde{d}_2, \tilde{d}_3 \in [0, 1]$, $\tilde{d}_1 + \tilde{d}_2 + \tilde{d}_3 = 1$, such that $\sigma_n u = \tilde{d}_1 g_{j_1} + \tilde{d}_2 g_{j_2} + \tilde{d}_3 g_{j_3}$. If we denote $d_k = \tilde{d}_k / \sigma_n \in (0, 1]$ and $\mathsf{B}_k' = d_k g_{j_k}$ for all $k \in \{1, 2, 3\}$, we find that $\{\mathsf{B}_1', \mathsf{B}_2', \mathsf{B}_3'\}$ is a set of linearly independent indecomposable effects such that $\mathsf{B}_1' + \mathsf{B}_2' + \mathsf{B}_3' = u$, which shows that an observable B' defined with these effects is an extreme simulation irreducible observable.

Since the set $\mathrm{conv}(\{g_1, \ldots, g_{2m+1}\})$ is a base for the positive cone of the effects, for each effect A_y of an observable $\mathsf{A} \in \mathcal{O}(\mathcal{S}_{2m+1})$ there exists $\alpha_y > 0$ and $a_y \in \mathrm{conv}(\{g_1, \ldots, g_{2m+1}\})$ such that $\mathsf{A}_y = \alpha_y a_y$. Similarly, for each $j \in \{1, \ldots, q_m\}$ we have that $\mathsf{B}_k^{(j)} = c_k^{(j)} g_{i_k^{(j)}}^{(j)}$ for some $c_k^{(j)} \in (0, 1]$ and $i_k^{(j)} \in \{1, \ldots, n\}$ for all $k \in \{1, 2, 3\}$. We then see that in order to characterize the intersection of the cones generated by the extreme simulation irreducible observables, i.e. essentially $\mathcal{T}_3$, we need to only consider the intersection of the respective triangles on the base.

Lemma 6. *Observable $\mathsf{A} \in \mathcal{O}(\mathcal{S}_{2m+1})$ with effects $\mathsf{A}_y = \alpha_y a_y$ such that $a_y \in \mathrm{conv}(\{g_1, \ldots, g_{2m+1}\})$ for all $y \in \Omega_\mathsf{A}$ is in $\mathcal{T}_3$ if and only if*

$$a_y \in \bigcap_{j=1}^{q_m} \mathrm{conv}\left(\left\{g_{i_1^{(j)}}^{(j)}, g_{i_2^{(j)}}^{(j)}, g_{i_3^{(j)}}^{(j)}\right\}\right) \quad \forall y \in \Omega_\mathsf{A}.$$

Proof. By Cor. 3 we see that we need to show that

$$\mathsf{A}_y \in \bigcap_{j=1}^{q_m} \mathrm{cone}\left(\left\{g_{i_1^{(j)}}^{(j)}, g_{i_2^{(j)}}^{(j)}, g_{i_3^{(j)}}^{(j)}\right\}\right) \tag{B2}$$

FIG. 7. Lemma 7

if and only if

$$a_y \in \bigcap_{j=1}^{q_m} \text{conv}\left(\left\{g_{i_1^{(j)}}^{(j)}, g_{i_2^{(j)}}^{(j)}, g_{i_3^{(j)}}^{(j)}\right\}\right) \tag{B3}$$
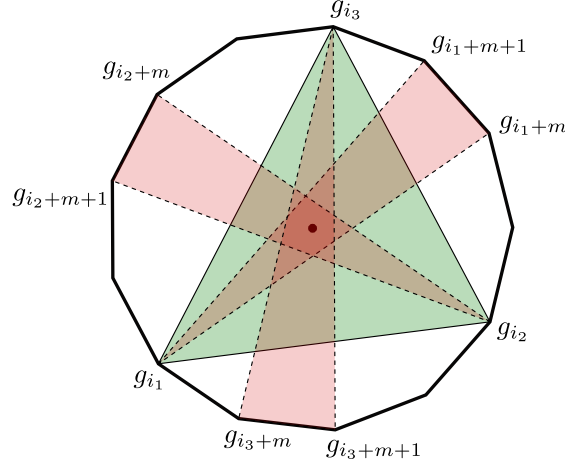
for all $y \in \Omega_\mathsf{A}$.

First let A_y be in the intersection of cones, which itself is a cone, in (B2) for some $y \in \Omega_A$. Since $\mathsf{A}_y = \alpha_y a_y$ for some $\alpha_y > 0$, it follows that also a_y is included in the same intersection of cones. Thus, a_y can be expressed as a positive linear combination of $g_{i_1^{(j)}}^{(j)}, g_{i_2^{(j)}}^{(j)}, g_{i_3^{(j)}}^{(j)}$ for all $j \in \{1, \ldots, q_m\}$. Since all the vectors $a_y, g_{i_1^{(j)}}^{(j)}, g_{i_2^{(j)}}^{(j)}, g_{i_3^{(j)}}^{(j)}$ lie on the same $z = \sigma_n$ –plane for all j, it follows that the positive linear combination must actually be a convex combination which shows (B3).

Let then a_y be included in the intersection of the convex hulls in (B3) for some $y \in \Omega_\mathsf{A}$. Since a convex hull is just a special case of a conic hull, we see that a_y is also included in the intersection of cones in (B2). By multiplying a_y by α_y we see that then (B2) holds. $\qquad\square$

The smallest such triangles to contain the centroid $(0, 0, \sigma_n)^T$ have vertices g_i, g_{i+m} and g_{i+m+1} for $i = 1, \ldots, 2m+1$, where the addition is modulo $2m+1$. We will show that the intersection of these smallest triangles gives us the whole intersection of all the triangles that represent the extreme simulation irreducible observables. We start with a small Lemma (see Fig. 7).

Lemma 7. *For an extreme simulation irreducible observable* B *such that* $\mathsf{B}_k = c_k g_{i_k}$ *for* $k \in \{1, 2, 3\}$ *we have that*

$$\bigcap_{k=1}^{3} \text{conv}\left(\{g_{i_k}, g_{i_k+m}, g_{i_k+m+1}\}\right) \subseteq \text{conv}\left(\{g_{i_1}, g_{i_2}, g_{i_3}\}\right).$$

Proof. To see this, suppose that, contrary to this, there exists a point $x \in \bigcap_{k=1}^{3} \text{conv}\left(\{g_{i_k}, g_{i_k+m}, g_{i_k+m+1}\}\right)$ such that $x \notin \text{conv}\left(\{g_{i_1}, g_{i_2}, g_{i_3}\}\right)$. Without loss of generality we assume that $i_1 < i_2 < i_3$.

If we consider a fixed vertex g_{i_k} for some $k \in \{1, 2, 3\}$, it is clear that the indices $i_k + m$ and $i_k + m + 1$ are contained in the set of indices $\{i_{k+1}, i_{k+1} + 1, \ldots, i_{k+2} - 1, i_{k+2}\}$ (Fig. 7). This is because otherwise they would be contained in either $\{i_{k+2}, i_{k+2} + 1, \ldots, i_k - 1, i_k\}$ or $\{i_k, i_k + 1, \ldots, i_{k+1} - 1, i_{k+1}\}$ so that

$$\text{conv}\left(\{g_{i_k}, g_{i_k+m}, g_{i_k+m+1}\}\right) \subset \text{conv}\left(\{g_{i_{k+2}}, \ldots, g_{i_k}\}\right)$$

or

$$\text{conv}\left(\{g_{i_k}, g_{i_k+m}, g_{i_k+m+1}\}\right) \subset \text{conv}\left(\{g_{i_k}, \ldots, g_{i_{k+1}}\}\right)$$

both of which would contradict the fact that $\sigma_n u \in \text{conv}\left(\{g_{i_k}, g_{i_k+m}, g_{i_k+m+1}\}\right)$.

Since now $x \in \text{conv}\left(\{g_{i_k}, g_{i_k+m}, g_{i_k+m+1}\}\right)$ for all $k \in \{1,2,3\}$ but $x \notin \text{conv}\left(\{g_{i_1}, g_{i_2}, g_{i_3}\}\right)$, we must have for all $k' \in \{1,2,3\}$ that

$$x \notin \text{conv}\left(\{g_{i_{k'}}, g_{i_{k'}+m}, g_{i_{k'}+m+1}\}\right) \bigcap \text{conv}\left(\{g_{i_1}, g_{i_2}, g_{i_3}\}\right).$$

We have by the above statement about the indices that $\{i_k + m, i_k + m + 1\} \subseteq \{i_{k+1}, i_{k+1}+1, \ldots, i_{k+2}-1, i_{k+2}\}$ so that it then follows that

$$\text{conv}\left(g_{i_k}, g_{i_k+m}, g_{i_k+m+1}\right) \subseteq \text{conv}\left(\{g_{i_1}, g_{i_2}, g_{i_3}\}\right) \bigcup \text{conv}\left(\{g_{i_{k+1}}, g_{i_{k+1}+1} \ldots, g_{i_{k+2}}\}\right)$$

which is a disjoint union for all $k \in \{1,2,3\}$. Thus, $x \in \text{conv}\left(\{g_{i_{k+1}}, g_{i_{k+1}+1} \ldots, g_{i_{k+2}}\}\right)$ for all $k \in \{1,2,3\}$ which is a contradiction since the sets do not intersect. $\qquad\square$

Proof of Proposition 12. From Cor. 3 it is clear that in order to prove the statement we need to show that

$$\bigcap_{j=1}^{q_m} \text{cone}\left(\{\mathsf{B}_x^{(j)}\}_{x \in \Omega_{\mathsf{B}^{(j)}}}\right) = \bigcap_{i=1}^{2m+1} \text{cone}\left(\{g_i, g_{i+1}, g_{i+m+1}\}\right).$$

The above statement is about cones but by Lemma 6 we can equivalently consider it in terms of the triangles that represent the observables in $\mathcal{O}_{irr}^{ext}(\mathcal{S}_{2m+1})$. By using the previously introduced notation for the effects of the extreme simulation irreducible observables, the above statement about the triangles then reads as

$$\bigcap_{j=1}^{q_m} \text{conv}\left(\{g_{i_1^{(j)}}^{(j)}, g_{i_2^{(j)}}^{(j)}, g_{i_3^{(j)}}^{(j)}\}\right) = \bigcap_{i=1}^{2m+1} \text{conv}\left(\{g_i, g_{i+m}, g_{i+m+1}\}\right).$$

The inclusion "$\subseteq$" is clear since among the q_m triangles that represent the extreme simulation irreducible observables the triangles with vertices g_i, g_{i+m} and g_{i+m+1} for $i = 1, \ldots, 2m+1$ are included.

For the inclusion "$\supseteq$", we use Lemma 7 for observables $\{\mathsf{B}^{(j)}\}_{j=1}^{q_m}$ which states that

$$\bigcap_{k=1}^{3} \text{conv}\left(\left\{g_{i_k^{(j)}}^{(j)}, g_{i_k^{(j)}+m}^{(j)}, g_{i_k^{(j)}+m+1}^{(j)}\right\}\right) \subseteq \text{conv}\left(\left\{g_{i_1^{(j)}}^{(j)}, g_{i_2^{(j)}}^{(j)}, g_{i_3^{(j)}}^{(j)}\right\}\right)$$

for all $j \in \{1, \ldots, q_m\}$. By taking the intersection of all $j \in \{1, \ldots, q_m\}$ we get

$$\bigcap_{i=1}^{2m+1} \text{cone}\left(\{g_i, g_{i+1}, g_{i+m+1}\}\right) = \bigcap_{j=1}^{q_m} \bigcap_{k=1}^{3} \text{conv}\left(\left\{g_{i_k^{(j)}}^{(j)}, g_{i_k^{(j)}+m}^{(j)}, g_{i_k^{(j)}+m+1}^{(j)}\right\}\right) \subseteq \bigcap_{j=1}^{q_m} \text{conv}\left(\left\{g_{i_1^{(j)}}^{(j)}, g_{i_2^{(j)}}^{(j)}, g_{i_3^{(j)}}^{(j)}\right\}\right)$$

which proves the statement. $\qquad\square$

Appendix C: Proof of Lemma 4

Proof of Lemma 4. We first see that either $x_{i,j}$ is an extreme point of C_m or $x_{i,j} \notin C_m$. Namely, assume that $x_{i,j} \in C_m$ but it is not an extreme point of C_m, then there exists some open line segment M, such that $x_{i,j} \in M$ and $M \subset C_m$. We must have $M \subset L_i$ since if M would intersect L_i, then we would get a contradiction with $M \subset C_m$. But then we must also have $M \subset L_{i+j}$ which is a contradiction with $L_i \neq L_{i+j}$.

Next fix $i \in \{1, \ldots, 2m+1\}$. From $x_{i,j} = L_i \cap L_{i+j}$ we get $x_{i,j} = \lambda_j g_i + (1 - \lambda_j) g_{i+m}$, where

$$\lambda_j = \frac{\cos\left(\frac{(2j+1)\pi}{4m+2}\right)}{2\cos\left(\frac{j\pi}{2m+1}\right)\cos\left(\frac{\pi}{4m+2}\right)} = \frac{1}{2}\left[1 - \tan\left(\frac{\pi}{4m+2}\right)\tan\left(\frac{j\pi}{2m+1}\right)\right].$$

Since $j\pi/(2m+1) \in [0, \pi/2]$ for all $j \in \{1, \ldots, m\}$, we have that $\tan\left(\frac{j\pi}{2m+1}\right)$ is an increasing function of j so that $\lambda_j \leq \lambda_1$. We then see that

$$x_{i,j} = \lambda_j g_i + (1 - \lambda_j) g_{i+m} = \frac{\lambda_j}{\lambda_1} x_{i,1} + \left(1 - \frac{\lambda_j}{\lambda_1}\right) g_{i+m}$$

for all $j \in \{1, \ldots, m\}$, where now $\frac{\lambda_j}{\lambda_1} \in [0, 1]$ so that $x_{i,j} \in \text{conv}\left(\{x_{i,1}, g_{i+m}\}\right)$. Since

$$x_{i,1}, g_{i+m} \in \text{conv}\left(\{g_{i+1}, g_{i+2}, \ldots, g_{i+m+1}\}\right),$$

it follows that also $x_{i,j} \in \text{conv}\left(\{g_{i+1}, g_{i+2}, \ldots, g_{i+m+1}\}\right)$ for all $j \in \{1, \ldots, m\}$.

Clearly $x_{i,j} \in L_i$ for all $j \in \{1, \ldots, m\}$ but $x_{i,j} \notin L_{i+1}$ for all $j \in \{2, \ldots, m\}$, where L_{i+1} can be expressed as

$$\text{conv}\left(\{g_{i+1}, \ldots, g_{i+m+1}\}\right) \cap \text{conv}\left(\{g_{i+1}, g_{i+m+1}, g_{i+m+2}\}\right),$$

so that $x_{i,j} \notin \text{conv}\left(\{g_{i+1}, g_{i+m+1}, g_{i+m+2}\}\right)$ for $j \in \{2, \ldots, m\}$. Thus, it follows that $x_{i,j} \notin C_m$ for $j \in \{2, \ldots, m\}$.

The only candidates for the extreme points of C_m are then $x_{i,1}$ for all $i \in \{1, \ldots, 2m+1\}$. From the symmetry it follows that all $x_{i,1}$ indeed must be extreme since if $x_{i',1}$ is not extreme for some $i' \in \{1, \ldots, 2m+1\}$ it would follow that $x_{i,1}$ is not extreme for any $i \in \{1, \ldots, 2m+1\}$. Hence, the claim follows. $\qquad\square$

On the properties of spectral effect algebras

Anna Jenčová* and Martin Plávala†

Mathematical Institute, Slovak Academy of Sciences, Štefánikova 49, Bratislava, Slovakia

The aim of this paper is to show that there can be either only one or uncountably many contexts in any spectral effect algebra, answering a question posed in [S. Gudder, Convex and Sequential Effect Algebras, (2018), arXiv:1802.01265]. We also provide some results on the structure of spectral effect algebras and their state spaces and investigate the direct products and direct convex sums of spectral effect algebras. In the case of spectral effect algebras with sharply determining state space, stronger properties can be proved: the spectral decompositions are essentially unique, the algebra is sharply dominating and the set of its sharp elements is an orthomodular lattice. The article also contains a list of open questions that might provide interesting future research directions.

I. INTRODUCTION

Effect algebras were defined in [1] as a generalization of the set of projectors on a Hilbert space. Since then a variety of results has been developed in the field, it has been used in the so called general probabilistic theories and it has attracted the interest of both mathematicians and physicist alike.

We present an answer to the open question presented in [2] about the possible number of contexts contained in a spectral effect algebra. The article has slightly overgrown from a simple answer to a question into a broader look on some of the properties of the spectral effect algebras with its own set of open questions.

The article is organized as follows: in Sec. II we present the basic definitions and results. In Sec. III we introduce spectral effect algebras and prove some basic results on properties of spectral decompositions and the structure of the state space. In Sec. IV we prove the main result that a spectral effect algebra may contain only either one or uncountably many contexts. In Sec. V we inspect two standard constructions with convex effect algebras: the direct product and the direct convex sum and we show that while the direct product of spectral effect algebras is again a spectral effect algebra, the direct convex sum of spectral effect algebras is not spectral. In Sec. VI we look at a special case of spectral effect algebras that have sharply determining state space and we show that in such setting a result analogical to [3, Proposition 18] holds, moreover, the algebra is sharply dominating and the set of its sharp elements is an orthomodular lattice. The Sec. VII contains the conclusions and open questions.

II. PROPERTIES OF CONVEX EFFECT ALGEBRAS

In this section we are going to present definition and properties of convex effect algebras. We closely follow the definitions used in [2] with a slightly different notation, more natural for linear effect algebras that are closely related to general probabilistic theories.

Definition 1. An effect algebra is a system $(E, 0, 1, +)$, where E is a set containing at least one element, $0, 1 \in E$ and $+$ is a partial binary operation on E. Let $a, b \in E$, then we write $a + b \in E$ whenever $a + b$ is defined (and hence yields an element of E). Moreover we require that $(E, 0, 1, +)$ satisfies the following conditions:

(E1) if $a + b \in E$ then $b + a \in E$ and $a + b = b + a$,

(E2) if $a + b \in E$ and $(a + b) + c \in E$, then $a + (b + c) \in E$, $a + (b + c) = (a + b) + c$,

(E3) for every $a \in E$ there is unique $a' \in E$ such that $a + a' = 1$, we usually denote $a' = 1 - a$,

(E4) if $a + 1 \in E$, then $a = 0$.

Definition 2. An effect algebra E is convex if for every $a \in E$ and $\lambda \in [0, 1] \subset \mathbb{R}$ there is an element $\lambda a \in E$ such that for all $\lambda, \mu \in [0, 1]$ and $a, b \in E$ we have

(C1) $\mu(\lambda a) = (\lambda \mu)a$,

---

* jenca@mat.savba.sk

† martin.plavala@mat.savba.sk

(C2) if $\lambda + \mu \leq 1$, then $\lambda a + \mu a \in E$ and $(\lambda + \mu)a = \lambda a + \mu a$,

(C3) if $a + b \in E$, then $\lambda a + \lambda b \in E$ and $\lambda(a + b) = \lambda a + \lambda b$,

(C4) $1a = a$.

Definition 3. Let E, F be effect algebras. A map $\phi : E \mapsto F$ is called additive if for $a, b \in E$, $a + b \in E$ we have: $\phi(a) + \phi(b) \in F$ and $\phi(a + b) = \phi(a) + \phi(b)$. An additive map such that $\Phi(1) = 1$ is called a morphism. A morphism is an isomorphism if it is surjective and for $a, b \in E$, $\phi(a) + \phi(b) \in F$ implies $a + b \in E$. If E and F are convex effect algebras, then a morphism Φ is affine if

$$\Phi(\lambda a) = \lambda \Phi(a), \qquad a \in E, \lambda \in [0, 1].$$

Definition 4. A state on an effect algebra E is a morphism $s : E \mapsto [0, 1] \subset \mathbb{R}$. The set of states on an effect algebra will be denoted $\mathfrak{S}(E)$.

It was proved in [4] that any state on a convex effect algebra is affine. The set of states is also referred to as state space of the effect algebra and it will play an important role in later constructions.

Let V be a real vector space with a pointed convex cone P, that is $P \cap (-P) = \{0\}$ where 0 denotes the zero vector. For $v, w \in V$ we define $v \geq w$ if and only if $v - w \in P$. Then $\leq$ is a partial order in V and (V, P) is an ordered vector space. Let $u \in P$, then the set

$$[0, u] = \{v \in V : 0 \leq v \leq u\}$$

is an effect algebra with the operation $+$ defined as the sum of the vectors and for $a, b \in [0, u]$ we have $a + b \in [0, u]$ if and only if $a + b \leq u$ which is exactly why we have chosen such unusual notation in Def. 1. Also note that in this case u is the unit of the effect algebra $[0, u]$, i.e. we have $1 = u$.

Definition 5. A linear effect algebra is an effect algebra of the form $[0, u]$ for some ordered vector space (V, P) and $u \in P$.

The following is an important result.

Proposition 1. *Every convex effect algebra is affinely isomorphic to a linear effect algebra.*

Proof. See [5]. $\qquad\square$

From now on we are going to assume that all of the effect algebras we will work with are convex. Below, we omit the isomorphism and identify convex effect algebras with the linear effect algebras they are isomorphic to. Moreover, we may and will assume that the interval $[0, u]$ generates the ordered vector space (V, P), so that u is an order unit in (V, P), [6, Lemma 3.1].

Definition 6. Let $f \in E$. We say that

- f is one-dimensional if $f \neq 0$ and for $g \in E$ we have that $f \geq g$ implies $g = \lambda f$ for some $\lambda \in [0, 1]$;

- f is sharp if $f \geq g$ and $1 - f \geq g$ implies $g = 0$;

- f is extremal if $f = \lambda g_1 + (1 - \lambda)g_2$ for some $\lambda \in (0, 1)$ implies that $f = g_1 = g_2$.

The set of sharp elements will be denoted by $S(E)$. The set of sharp one-dimensional elements will be denoted by $S_1(E)$.

It was shown in [6, Lemma 4.4] that any extremal element is sharp, but the converse is not necessarily true. Moreover, for any sharp effect f there are states $s_0, s_1 \in \mathfrak{S}(E)$ such that $s_0(f) = 0$, $s_1(f) = 1$. Note that in general the states s_0 and s_1 do not have to be unique. In general probabilistic theories [7] one-dimensional effects are called indecomposable. In what follows we will be interested in the properties of sharp one-dimensional effects.

Proposition 2. *Let E be a convex effect algebra, then a one-dimensional effect is sharp if and only if it is extremal.*

Proof. As noted above, every extremal element is sharp. Conversely, let $f \in S_1(E)$ and assume that we have $f = \lambda g_1 + (1 - \lambda)g_2$ for some $g_1, g_2 \in E$ and $\lambda \in (0, 1)$. It follows that we have $f \geq \lambda g_1$ and $f \geq (1 - \lambda)g_2$ and since f is one-dimensional we must have $g_1 = \mu_1 f$ and $g_2 = \mu_2 f$, i.e. f must be a convex combination of its multiples. Let now $\mu \geq 1$ be such that $\mu f \in E$. Let $\eta = \min\{\frac{1}{\mu}, 1 - \frac{1}{\mu}\}$ then we have

$$f \geq \eta f$$

$$1 - f = \frac{1}{\mu}(1 - \mu f) + (1 - \frac{1}{\mu})1 \geq \eta 1 \geq \eta f.$$

Since f is sharp, this implies that $\eta = 0$ and hence $\mu = 1$. It follows that we must have $g_1 = g_2 = f$ and f is extremal. $\qquad\square$

III.   CONTEXTS AND SPECTRAL EFFECT ALGEBRAS

In this section we are going to introduce contexts and spectral effect algebras and provide some results on their structure.

Definition 7. A context in a convex effect algebra E is a finite collection $A = \{a_1, \ldots, a_n\} \subset S_1(E)$ such that $\sum_{i=1}^n a_i = 1$.

Let $A = \{a_1, \ldots, a_n\}$ be a context in a convex effect algebra E and let $\hat{a}_i = \{s_i \in \mathfrak{S}(E) : s_i(a_i) = 1\}$. By the remarks below Definition 6, the sets $\hat{a}_i$ are nonempty and it is easy to see that

$$s_i(a_j) = \delta_{ij}, \qquad \forall s_i \in \hat{a}_i, \ i, j = 1, \ldots, n,$$

where δ_{ij} is the Kronecker delta.

Let us denote by $\bar{A}$ the convex effect subalgebra generated by A, that is

$$\bar{A} = \left\{ \sum_i \mu_i a_i, \ \mu_i \in [0, 1] \right\}.$$

Then $\bar{A}$ is the interval $[0, u_A]$ in the ordered vector space $V(A) := \{\sum_i t_i a_i, \ t_i \in \mathbb{R}\}$, with an obvious positive cone and order unit $u_A = \sum_i a_i$.

Definition 8. We say that a convex effect algebra E is spectral if for every $f \in E$ there is a context $A \subset E$ such that $f \in \bar{A}$.

Specifically, any $f \in E$ has the form

$$f = \sum_i \mu_i a_i \tag{1}$$

for some $\mu_i \in [0, 1]$ and some context $A = \{a_1, \ldots, a_n\}$. Any expression of the form (1) will be called a spectral decomposition of f.

The most important examples are the algebras of finite dimensional classical and quantum effects, that were characterized in [2]. It is also easy to see that the algebra of effects over a finite dimensional real Hilbert space is spectral. Note that it is not assumed that the number of elements is the same in each context or that the space V generated by E is finite dimensional, although all the above examples have these properties.

We next discuss some properties of spectral effect algebras and their state spaces. The following will be a useful tool.

Lemma 1. Let $f = \sum_{i=1}^n \mu_i a_i$ be a spectral decomposition of f. Then

$$\max_{s \in \mathfrak{S}(E)} s(f) = \max\{\mu_1, \ldots, \mu_n\},$$

$$\min_{s \in \mathfrak{S}(E)} s(f) = \min\{\mu_1, \ldots, \mu_n\}.$$

Proof. Let $s \in \mathfrak{S}(E)$, then we have

$$s(f) = \sum_{i=1}^n \mu_i s(a_i) \leq \max\{\mu_1, \ldots, \mu_n\} \sum_{i=1}^n s(a_i) = \max\{\mu_1, \ldots, \mu_n\}.$$

To show that the bound is tight let $\max\{\mu_1, \ldots, \mu_n\} = \mu_{i'}$ for some $i' \in \{1, \ldots, n\}$ and let $s \in \hat{a}_{i'}$. Then we have

$$s(f) = \mu_{i'} = \max\{\mu_1, \ldots, \mu_n\}.$$

The proof for the minimum is analogical. □

We next show that the elements of the embedding ordered vector space V have spectral decompositions as well.

Lemma 2. Let E be spectral and let (V, P) be the ordered vector space with an order unit $u \in P$ such that $E = [0, u]$. Then for any $v \in V$ there is a context A such that $v \in V(A)$.

Proof. Since u is a order unit, we have $-\lambda u \le v \le \lambda u$ for some $\lambda > 0$. Then $0 \le v + \lambda u \le 2\lambda u$, so that $a := \frac{1}{2\lambda}v + \frac{1}{2}u \in E$. Let $a = \sum_i \nu_i a_i$ for some context A, then $v = \lambda(2a - u) = \sum_i \lambda(2\nu_i - 1)a_i$. $\square$

Proposition 3. *Any spectral effect algebra E has an order determining set of states, that is, $s(a) \le s(b)$ for all $s \in \mathfrak{S}(E)$ implies that $a \le b$.*

Proof. Let $a, b \in E$ be such that $s(a) \le s(b)$ for all states s. By Lemma 2, there is some context A and real numbers μ_i such that $v = b - a = \sum_i \mu_i a_i$. By the assumption, $\mu_i = s_i(v) \ge 0$ for $s_i \in \hat{a}_i$, so that $b - a \ge 0$ and $a \le b$. $\square$

Let us remark that it was proved in [6, Theorem 3.6] that a convex effect algebra E has an order determining set of states if and only if it is Archimedean. In this case, the order unit seminorm

$$\|v\| := \inf\{\lambda > 0, \ -\lambda u \le v \le \lambda u\} = \sup_{s \in \mathfrak{S}(E)} |s(v)|$$

is a norm. Using Lemma 1, we obtain for $f \in E$ with spectral decomposition $f = \sum_i \mu_i a_i$ that

$$\|f\| = \max\{\mu_1, \ldots, \mu_n\}, \qquad \|1 - f\| = 1 - \min\{\mu_1, \ldots, \mu_n\}. \tag{2}$$

More generally, if $v = \sum_i \alpha_i a_i$ is a spectral decomposition, then we have for $s_i \in \hat{a}_i$

$$\|v\| = \sup_{s \in \mathfrak{S}(E)} |s(v)| = \max_i |\alpha_i| =: |\alpha_{i_{max}}| = |s_{i_{max}}(v)|.$$

An element $f \in E$ does not have to have a unique spectral decomposition. The following result is inspired by [3] and is immediate from (2).

Proposition 4. *Let $f \in E$ have two spectral decompositions*

$$\sum_{i=1}^n \mu_i a_i = f = \sum_{j=1}^m \nu_j b_j$$

Then $\max\{\mu_1, \ldots, \mu_n\} = \max\{\nu_1, \ldots, \nu_m\}$ and $\min\{\mu_1, \ldots, \mu_n\} = \min\{\nu_1, \ldots, \nu_m\}$.

Up to now it is not clear whether the sets $\hat{a}_i$ consist of a single state. Clearly, these sets are faces of $\mathfrak{S}(E)$. More generally, for any element $f \in E$, the set $\hat{f} := \{s \in \mathfrak{S}(E), s(f) = 1\}$ is a face of $\mathfrak{S}(E)$ (note that this also may be empty). A face of this form will be called E-exposed. If $\hat{f} = \{s\}$, we say that s is and E-exposed point of $\mathfrak{S}(E)$, in this case, we write $\hat{f} = s$. If $a \in S_1(E)$, $\hat{a}$ is a nonempty E-exposed proper face of $\mathfrak{S}(E)$. Note that if $A = \{a_1, \ldots, a_n\}$ is a context, the faces $\hat{a}_i$ are affinely independent. We next show a property of the E-exposed points of $\mathfrak{S}(E)$.

Proposition 5. *Every E-exposed point of $\mathfrak{S}(E)$ has the form $\hat{a}$ for some $a \in S_1(E)$.*

Proof. Let $s \in \mathfrak{S}(E)$ be an E-exposed point, so that $s = \hat{f}$ for some $f \in E$. Then $1 = s(f) = \max_{s' \in \mathfrak{S}(E)} s'(f)$ and s is the unique point where this maximum is attained. We have already showed in the proof of Lemma 1 that every effect in E attains its maximum at some $s \in \hat{a}$ with $a \in S_1(E)$. The result follows. $\square$

IV. NUMBER OF CONTEXTS IN A SPECTRAL EFFECT ALGEBRA

In this section we are going to prove the main result that answers the open question from [2] about the possible number of contexts in a spectral effect algebra.

Proposition 6. *Assume that every pair $a, a' \in S_1(E)$ is summable. Then there is only one context in E.*

Proof. Assume that there are two contexts $A \ne B$ in E. Then there is at least one effect a such that $a \in A$ but $a \notin B$. Let $B = \{b_1, \ldots b_n\}$ and let $s \in \hat{a}$. For every b_i, $i \in \{1, \ldots, n\}$, we must have $a + b_i \in E$. This implies

$$s(a + b_i) \le 1$$

which gives $s(b_i) = 0$ for all $i \in \{1, \ldots, n\}$ which is a contradiction with $\sum_{i=1}^n b_i = 1$. $\square$

Proposition 7. *Every spectral effect algebra E contains either one context or uncountably many contexts.*

Proof. Assume that E contains at least two contexts. As a result of Prop. 6, there are $a, b \in S_1(E)$, $a \neq b$, that are not summable. Let $\lambda \in [0, 1]$ and denote

$$c_\lambda = \lambda a + (1 - \lambda) b.$$

We will show that every c_λ must belong to a different context, hence E we must contain uncountably infinite number of contexts.

Since E is spectral we have that for every $\lambda \in [0, 1]$ there is a context C_λ such that $c_\lambda \in \bar{C}_\lambda$. Assume that for some $\lambda \neq \mu$, $\mu \in [0, 1]$ we have $C_\lambda = C_\mu$. Then we have $a, b \in V(C_\lambda)$ as $c_\lambda, c_\mu \in V(C_\lambda)$ and we can express a and b as linear combinations of c_λ and c_μ.

Let $C_\lambda = \{c_1, \dots, c_n\}$, then we must have

$$a = \sum_{i=1}^{n} \alpha_i c_i$$

for some $\alpha_i \in \mathbb{R}$. Moreover, let $s_i \in \hat{c}_i$, then $\alpha_i = s_i(a) \in [0, 1]$ for all i implies that $a \in \bar{C}_\lambda$, which yields $a \in C_\lambda$ as a is sharp and one-dimensional. In a similar fashion we get $b \in C_\lambda$, which is a contradiction with the assumption that a and b are not summable. □

V. COMPOSITION OF SPECTRAL EFFECT ALGEBRAS

In the study of spectral effect algebras, it is a natural question whether spectrality is preserved by some constructions over convex effect algebras. In this section, we study the direct products and direct convex sums, note that these are the product and coproduct in the category of convex effect algebras.

Definition 9. Let E_1, E_2 be effect algebras, then their direct product is an effect algebra $E_1 \times E_2$ given as $E_1 \times E_2 = \{(f_1, f_2) : f_1 \in E_1, f_2 \in E_2\}$. The partial binary operation $+$ is given for $f_i, f_i' \in E_i$, such that $f_i + f_i' \in E_i$, $i \in \{1, 2\}$ as $(f_1, f_2) + (f_1', f_2') = (f_1 + f_1', f_2 + f_2')$ and the unit of $E_1 \times E_2$ is $(1, 1)$. If E_1, E_2 are convex effect algebras, then we can define a convex structure on $E_1 \times E_2$ by $\lambda(f_1, f_2) = (\lambda f_1, \lambda f_2)$, for $\lambda \in [0, 1]$. In this way, the direct product of convex effect algebras is a convex effect algebra.

Proposition 8. *The direct product of spectral effect algebras is a spectral effect algebra.*

Proof. It is easy to see that an element $(f_1, f_2) \in E_1 \times E_2$ is sharp if and only if both f_1 and f_2 are sharp. Moreover, since $(f_1, f_2) = (f_1, 0) + (0, f_2)$, such an element is one-dimensional if and only if one of the elements is one-dimensional and the other is 0. Let now $A = \{a_1, \dots, a_n\} \subset E_1$ and $B = \{b_1, \dots, b_m\} \subset E_2$ be contexts. It is straightforward to see that $(a_1, 0) + \dots + (a_n, 0) + (0, b_1) + \dots + (0, b_m) = (1, 1)$, hence

$$\{(a_1, 0), \dots, (a_n, 0), (0, b_1), \dots, (0, b_m)\} \subset E_1 \times E_2 \tag{3}$$

is a context. Moreover, any context in $E_1 \times E_2$ is of this form.

Finally let $(f_1, f_2) \in E_1 \times E_2$ be any element. Since E_1 and E_2 are spectral there are contexts $A \subset E_1$ and $B \subset E_2$ such that $f_1 \in \bar{A}$ and $f_2 \in \bar{B}$. It follows that we have

$$(f_1, f_2) = \sum_{i=1}^{n} \mu_i (a_i, 0) + \sum_{j=1}^{m} \nu_j (0, b_j)$$

which shows that $E_1 \times E_2$ is spectral effect algebra. □

The definition of the convex direct sum is a bit more involved.

Definition 10. Let E_1, E_2 be convex effect algebras, then they are affinely isomorphic to the intervals $[0, u_1] \subset C_1 \subset V_1$ and $[0, u_2] \subset C_2 \subset V_2$ respectively, where for $i \in \{1, 2\}$ we have that V_i are vectors spaces, C_i are pointed cones and $u_i \in C_i$. Take the vector space $V_1 \times V_2$ that corresponds to the coproduct of the respective vector spaces and define a relation of equivalence $\approx$ by $(u_1, 0) = (0, u_2)$, i.e. $(x_1, y_2) \approx (x_2, y_2)$ if and only if we have

$$(x_1, y_1) = (x_2, y_2) + \alpha(1, -1)$$

for some $\alpha \in \mathbb{R}$. It is clear that $\approx$ is reflexive, symmetric and transitive. Let V be the quotient space $V = (V_1 \times V_2)/_\approx$ and define the cone $C = (C_1 \times C_2)/_\approx \subset V$ as the set of equivalence classes containing an element of $C_1 \times C_2$. Clearly, C

is a convex cone containing $u := [(u_1, 0)]_\approx = [(0, u_2)]_\approx$. To see that C is pointed, simply note that $[(x, y)]_\approx \in C \cap (-C)$ if and only if we can choose $x \in C_1$, $y \in C_2$ and for some $\alpha \in \mathbb{R}$ we have $x + \alpha u_1 \in (-C_1)$ and $y - \alpha u_2 \in (-C_2)$. This implies $0 \le x \le -\alpha u_1$ in (V_1, C_1) so that $\alpha \le 0$. At the same time, $0 \le y \le \alpha u_2$ in (V_2, C_2), so that $\alpha \ge 0$. This implies $\alpha = 0$ and consequently $(x, y) = (0, 0)$.

Now we can define the direct convex sum of effect algebras E_1 and E_2 as

$$E_1 \oplus E_2 := [0, u] \subset C.$$

We can see that $E_1 \oplus E_2$ is a set of equivalence classes of the form $[(\lambda f_1, (1 - \lambda) f_2)]_\approx$, where $f_1 \in E_1$, $f_2 \in E_2$ and $\lambda \in [0, 1]$. Note that $E_1 \oplus E_2$ is a convex effect algebra by construction as it is an interval in an ordered vector space.

In the less general circumstances of the general probabilistic theories the direct convex sum of the effect algebras can be introduced as the effect algebra corresponding to the direct product of state spaces [8, Definition 6].

Proposition 9. *The direct convex sum of spectral effect algebras is not a spectral effect algebra.*

Proof. In a sense we are going to mimic the proof of Prop. 8 with the only difference that now we will show that there are two types of contexts on $E_1 \oplus E_2$: they are either of the form $\{[(a_1, 0)]_\approx, \ldots, [(a_n, 0)]_\approx\}$ where $A = \{a_1, \ldots a_n\} \subset E_1$ is a context, or $\{[(0, b_1)]_\approx, \ldots, [(0, b_m)]_\approx\}$ where $B = \{b_1, \ldots b_m\} \subset E_2$ is a context.

Let $f = [(\lambda f_1, (1 - \lambda) f_2)]_\approx \in S_1(E_1 \oplus E_2)$, then from

$$[(\lambda f_1, (1 - \lambda) f_2)]_\approx = \lambda[(f_1, 0)]_\approx + (1 - \lambda)[(0, f_2)]_\approx$$

we see that both $\lambda[(f_1, 0)]_\approx$ and $(1 - \lambda)[(0, f_2)]_\approx$ must be multiples of f. This implies that there are some $t, \alpha \in [0, 1]$ such that

$$t\lambda f_1 = \alpha u_1, \qquad (1 - t)(1 - \lambda) f_2 = \alpha u_2.$$

Assuming that both f_1 and f_2 are nonzero, this implies that either both are multiples of identity or $\lambda \in \{0, 1\}$. In both cases, f is of the form $f = [g_1, 0]_\approx$ for some $g_1 \in E_1$ or $f = [0, g_2]_\approx$ for some $g_2 \in E_2$. It is clear that $[(g_1, 0)] \in S_1(E_1 \oplus E_2)$ if and only if $g_1 \in S_1(E_1)$ and similarly for elements of the form $[(0, g_2)]_\approx$. From the definition, we can see that in this case, $[(g_1, 0)]_\approx$ and $[(0, g_2)]_\approx$ are not summable in $E_1 \oplus E_2$, hence these cannot belong to the same context.

It follows that all contexts on $E_1 \oplus E_2$ are of the above two types. It is straightforward to see that if neither of f_1 or f_2 is a multiple of identity and $\lambda \in (0, 1)$, the element $[(\lambda f_1, (1 - \lambda) f_2)]_\approx$ cannot be given as $\sum_{i=1}^n \mu_i[(a_i, 0)]_\approx$ or $\sum_{j=1}^m \nu_j[(0, b_j)]_\approx$. Hence $E_1 \oplus E_2$ is not spectral. $\square$

VI. SHARPLY DETERMINING STATE SPACES

Here we consider a special case of spectral effect algebras, for which stronger properties can be proved.

Definition 11. We say that the state space $\mathfrak{S}(E)$ is sharply determining if for any sharp $f \in E$ and any $g \in E$ such that $g \not\geq f$ there is a state $s \in \mathfrak{S}(E)$ such that $s(f) = 1 > s(g)$.

We first obtain a way stronger version of Prop. 4 that is similar to [3, Proposition 18].

Proposition 10. *Let E be a spectral effect algebra such that $\mathfrak{S}(E)$ is sharply determining. Let $A, B \subset E$ be contexts, $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots b_m\}$ such that for some $f \in E$ we have $f \in \bar{A} \cap \bar{B}$, specifically*

$$\sum_{i=1}^n \mu_i \left(\sum_{j_i=1}^{n_i} a_{i_j} \right) = f = \sum_{k=1}^m \nu_k \left(\sum_{l_k=1}^{m_k} b_{l_k} \right)$$

where $\mu_1 > \ldots > \mu_n > 0$ and $\nu_1 > \ldots > \nu_m > 0$. Then we have $n = m$, $\mu_i = \nu_i$ and $\sum_{j_i=1}^{n_i} a_{i_j} = \sum_{l_i=1}^{m_i} b_{l_i}$.

Proof. Denote $a_i' = \sum_{j_i=1}^{n_i} a_{i_j}$ and $b_j' = \sum_{l_k=1}^{m_k} b_{l_k}$. As a result of Prop. 4 we already know that $\mu_1 = \nu_1$. We will only show that $a_1' = b_1'$, the result will follow by repeating the same procedure for $f - \mu_1 a_1' = f - \nu_1 b_1'$.

As first note that a_i' and b_i' are sharp [6, Theorem 4.8]. Assume that $a_1' \not\geq b_1'$ then there is $s \in \mathfrak{S}(E)$ such that $s(b_1') = 1 > s(a_1')$. We have $s(f) = \sum_{k=1}^m \nu_k s(b_k') = \nu_1 = \mu_1$ as well as $s(f) = \sum_{i=1}^n \mu_i s(a_i') < \mu_1$ which is a contradiction. Hence we must have $a_1' \ge b_1'$ and by the same logic we must have also $b_1' \ge a_1'$ which together yields $a_1' = b_1'$. $\square$

It was proved in [6, Theorem 4.8] that if $\mathfrak{S}(E)$ is sharply determining, then all sharp elements are extremal, moreover, the sum of sharp elements, if it exists, is sharp. If E is also spectral, then it is clear that the sharp elements are precisely the finite sums of one-dimensional sharp elements. The next results show that in this case E is sharply dominating, see [9] for a definition, and the set $S(E)$ of all sharp elements with ordering induced from E is an orthomodular lattice. We will need the following reformulation of the condition in Definition 11. For the proof, it is enough to realize that f is sharp if and only if $1 - f$ is sharp.

Lemma 3. *Let E be a convex effect algebra. Then $\mathfrak{S}(E)$ is sharply determining if and only if for any sharp $f \in E$ and any $g \in E$ such that $g \not\leq f$ there is a state $s \in \mathfrak{S}(E)$ such that $s(f) = 0 < s(g)$.*

Proposition 11. *Let E be a spectral effect algebra with a sharply determining state space and let $a = \sum_i \mu_i a_i$ be a spectral decomposition. Then $a^0 := \sum_{i, \mu_i > 0} a_i$ is the smallest sharp element larger than a.*

Proof. It is clear that a^0 is a sharp element and $a \leq a^0$. Assume next that b is a sharp element such that $a \leq b$. Then for any $s \in \mathfrak{S}(E)$ such that $s(b) = 0$ we have $s(a) = \sum_i \mu_i s(a_i) = 0$, so that $s(a_i) = 0$ whenever $\mu_i > 0$. It follows that $s(a^0) = 0$. By Lemma 3, $a^0 \leq b$. $\square$

Proposition 12. *Let E be a spectral effect algebra with a sharply determining state space. Let $f, g \in E$ be sharp. Then $(\lambda f + (1 - \lambda)g)^0$ does not depend on $\lambda \in (1, 0)$ and we have $(\lambda f + (1 - \lambda)g)^0 = f \vee g$ in $S(E)$. Moreover, we have $f \wedge g = 1 - ((1 - f) \vee (1 - g))$.*

Proof. Let us first observe that for any $f \in E$ and $s \in \mathfrak{S}(E)$, $s(f) = 0$ iff $s(f^0) = 0$, this is easy to see from the definition of f^0. Let f, g be sharp and for $\lambda \in (0, 1)$ let $p_\lambda := (\lambda f + (1 - \lambda)g)^0$. Then for $s \in \mathfrak{S}(E)$, $s(p_{1/2}) = 0$ iff $s(\frac{1}{2}(f + g)) = 0$ iff $s(f) = s(g) = 0$ iff $s(\lambda f + (1 - \lambda)g) = 0$ iff $s(p_\lambda) = 0$. By Lemma 3 this implies that $p_\lambda = p_{1/2}$ for all $\lambda \in (0, 1)$. If $h \in E$ is any sharp element such that $f, g \leq h$, then $s(h) = 0$ implies $s(f) = s(g) = 0$ so that $s(p_{1/2}) = 0$, hence $p_{1/2} \leq h$. By [6, Corollary 4.10], $h - p_{1/2} \in S(E)$, this shows that $p_\lambda = p_{1/2} = f \vee g$ in $S(E)$. The last assertion follows by de Morgan laws. $\square$

Corollary 1. *Let E be a spectral effect algebra with a sharply determining state space. Then $S(E)$ is an orthomodular lattice.*

Proof. By [6, Corollary 4.9], $S(E)$ is a sub-effect algebra in E that is an orthoalgebra and by Proposition 12, $S(E)$ is a lattice. By [10, Prop. 1.5.8], any lattice ordered orthoalgebra is an orthomodular lattice. $\square$

VII. CONCLUSIONS AND OPEN QUESTIONS

In this article we have proved that there can be either only one or uncountably many contexts contained in a spectral effect algebra as well as few other results concerning spectral effect algebras. There are still quite a few open questions that will be left for future research.

1. What are the extreme points of a state space of a spectral effect algebra? In Prop. 5 we have shown that all exposed points of the effect space of a spectral effect algebra have the form $s = \hat{a}$ for $a \in S_1(E)$. It is an open question whether one can show a similar result for the extreme points of the state space.

2. Let $A = \{a_1, \ldots, a_n\}$ be a context and let $\hat{A} = \text{conv}(\cup_i \hat{a}_i)$. Since $\hat{a}_i$ are affinely independent, $\hat{A} = \oplus_c \hat{a}_i$ is their convex direct sum. Let I be a set that indexes all of the contexts A_α of an effect algebra E; does it hold that $\mathfrak{S}(E) = \cup_{\alpha \in I} \hat{A}_\alpha$? If $\hat{a}$ are E-exposed points for all $a \in S_1(E)$, this would be even a stronger result than the one proposed in the question above.

3. Is the cardinality of contexts always the same? It is straightforward to see that all contexts in the effect algebras used in quantum theory have the same number of elements. It would be very interesting to know whether this is true in general or whether there is a counter-example. As we have seen in Prop. 9 there are effect algebras that have contexts with different numbers of elements, but the given example is not a spectral effect algebra.

4. Is it possible to extend the results of Prop. 4 to a result similar to Prop. 10? The property in question is to show that $\sum_{i=1}^{n} \mu_i a_i = f = \sum_{j=1}^{m} \nu_j b_j$ where $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_m\}$ are contexts would imply that for every μ_i there is ν_j such that $\mu_i = \nu_j$. A stronger version of said result would be to show that Prop. 10 holds for all spectral effect algebras; this can be either done by showing that the state space of every spectral algebra is sharply determining or by other means.

5. For a given context $A = \{a_1, \ldots, a_n\}$ is the set of states $\{\hat{a}_1, \ldots, \hat{a}_n\}$ unique? The existence of the states $\hat{a}_i$ is simply implied by the fact that the effects a_i are sharp one-dimensional but it is rather easy to find convex effect algebras where the set $\{\hat{a}_1, \ldots, \hat{a}_n\}$ is not unique.

6. Does spectrality of the effect algebra implies any kind of weak duality between the effect algebra and the state space? It is tempting to define a map $T : a_i \mapsto \hat{a}_i$ but it is of question whether the map would be well defined and whether the map would be affine. This would be a very strong result to prove or to at least find some conditions for when it holds. Still all of the examples of spectral effect algebras that we know of have said property.

7. Are there examples of spectral effect algebras that are not operator algebras nor classical? It is straightforward to see that the effect algebras for finite-dimensional real or complex quantum theory are spectral. It would be very interesting and possibly helpful to have other examples of spectral effect algebras.

ACKNOWLEDGMENTS

[1] D. J. Foulis and M. K. Bennett, Found. Phys. **24**, 1331 (1994).
[2] S. Gudder, "Convex and Sequential Effect Algebras," (2018), arXiv:1802.01265.
[3] J. van de Wetering, "Reconstruction of quantum theory from universal filters," (2018), arXiv:1801.05798.
[4] S. Gudder, Int. J. Theor. Phys. **38**, 3179 (1999).
[5] S. Gudder and S. Pulmannová, Comment. Math. Univ. Carolinae **39**, 645 (1998).
[6] S. Gudder, S. Pulmannová, S. Bugajski, and E. Beltrametti, Reports Math. Phys. **44**, 359 (1999).
[7] G. Kimura, K. Nuida, and H. Imai, Reports Math. Phys. **66**, 175 (2010).
[8] T. Heinosaari, L. Leppäjärvi, and M. Plávala, "No-free-information principle in general probabilistic theories," arXiv:1808.07376.
[9] S. Gudder, Tatra Mt. Math. Publ. **15**, 23 (1998).
[10] A. Dvurecenskij and S. Pulmannová, *New Trends in Quantum Structures*, Mathematics and Its Applications (Springer Netherlands, 2000).