Journal of Physics A: Mathematical and Theoretical

Decomposing all multipartite non-signalling channels via quasiprobabilistic mixtures of local channels in generalised probabilistic theories

OPEN ACCESS
IOP Publishing
Journal of Physics A: Mathematical and Theoretical
J. Phys. A: Math. Theor. **55** (2022) 404001 (28pp)
https://doi.org/10.1088/1751-8121/ac8ea4

Decomposing all multipartite non-signalling channels via quasiprobabilistic mixtures of local channels in generalised probabilistic theories

Paulo J Cavalcanti[1,*], John H Selby[1], Jamie Sikora[2] and Ana Belén Sainz[1]

[1] International Centre for Theory of Quantum Technologies, University of Gdańsk, 80-309 Gdańsk, Poland
[2] Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, United States of America

E-mail: paulo.cavalcantidevasconcelosfilho@phdstud.ug.edu.pl

Received 26 April 2022, revised 25 August 2022
Accepted for publication 2 September 2022
Published 20 September 2022

Abstract

Non-signalling quantum channels—relevant in, e.g., the study of Bell and Einstein–Podolsky–Rosen scenarios—may be decomposed as an affine combinations of local operations in bipartite scenarios. Moreover, when these channels correspond to stochastic maps between classical variables, such a decomposition is possible even in multipartite scenarios. These two results have proven useful when studying the properties of these channels, such as their communication and information processing power, and even when defining measures of the non-classicality of physical phenomena (such as Bell non-classicality and steering). In this paper we show that such useful quasi-stochastic characterizations of channels may be unified and applied to the broader class of multipartite non-signalling channels. Moreover, we show that this holds for non-signalling channels in quantum theory, as well as in a larger family of generalised probabilistic theories. More precisely, we prove that channels are non-signalling if and only if they can be decomposed as an affine combinations of corresponding local operations, provided that the underlying physical theory is locally

*Author to whom any correspondence should be addressed.

tomographic—a property that quantum theory satisfies. Our results then can be viewed as a generalisation of references (*Phys. Rev. Lett.* **111** 170403) and (2013 *Phys. Rev.* A **88** 022318) to the multipartite scenario for arbitrary tomographically local generalised probabilistic theories (including quantum theory). Our proof technique leverages Hardy's duotensor formalism, highlighting its utility in this line of research.

Keywords: generalised probabilistic theories, duotensors, non-signalling channels, quasiprobabilities

(Some figures may appear in colour only in the online journal)

1. Introduction

Quantum operations are at the core of communication and information processing tasks, and how well we can perform at the latter may depend on the properties of the quantum operations that we have at hand. One particular set of operations of interest is that of *non-signalling quantum channels* [1], i.e., those that cannot be used by two distant parties to exchange information in a way that is against the laws of relativity theory. Bipartite non-signalling quantum operations have been extensively studied, specially since they play a central role in Bell [2] and Einstein–Podolsky–Rosen 'steering' [3, 4] scenarios, which in turn underpin cryptographic protocols [5, 6]. In addition, the simulation of bipartite non-signalling quantum channels via affine combinations of local operations has provided valuable insight on the exploration of the advantage they provide for communication and information processing tasks [7, 8].

In recent years it has become fruitful to study quantum theory from the 'outside', that is, by placing it as one theory within a broad landscape of logically consistent theories. This allows one to understand *why* quantum theory has particular features, and also its possibilities and limitations for various applications. The framework of generalised probabilistic theories [9, 10] (GPTs) has become the preferred tool for such studies, for example, shedding light on matters pertaining to: cryptography [11–16]; computation [17–23]; interference [24–30]; thermodynamics [31–36]; contextuality [37–40]; nonlocality [41–47]; steering [48–51]; decoherence [52–55]; information processing [10, 56–61]; incompatibility [62–66]; uncertainty [67–70]; as well as providing a foundational view of the primitive structures of physical theories [71–87]. For a comprehensive introduction to the field see references [47, 88, 89].

In this work we investigate no-signalling channels in GPTs. In particular, we prove a useful technical result, namely that multipartite channels in locally-tomographic GPTs [9] are non-signalling if and only if they can be decomposed as an affine combinations of product (local) channels (theorem 5.1). Our results can be viewed as a generalisation of those of reference [7] and of reference [8, lemma 1] to arbitrary tomographically local GPTs: the former applies only to multipartite non-signalling stochastic maps on classical variables, while the latter applies to bipartite non-signalling quantum channels.

Our proofs leverage the convenient duotensor formalism of reference [90] with a slight twist based on reference [91] which allows us to directly lift the result of reference [7] (using a generalisation of lemma 2 in reference [8]) to this more general setting. We believe that this way of lifting structural properties of stochastic maps to properties of channels in arbitrary tomographically local GPTs via the duotensor formalism [90] may be a useful tool in future research.

Table 1. Elements that define a generalised probabilistic theory, and how they are defined for the particular case of quantum and classical theories viewed as GPTs.

Elements of a GPT	Quantum theory	Classical theory
Systems	Hilbert spaces	Finite sets
States	Density operators	Probability distributions
Effects	POVM element	$[0, 1]-$valued functions
Discarding effect	(Partial) trace	Marginalisation
Transformations	CPTNI linear maps	Substochastic maps
Composition rule	Tensor product	Cartesian product

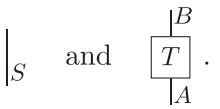
2. Generalised probabilistic theories: the basics

The framework of GPTs can be used to define arbitrary physical theories. The simplicity of the framework enables various alternative theories to be formulated and explored while allowing at the same time a deep study of the probabilistic and compositional aspects of such theories. It is based on the tenet that a minimal requirement of any physical theory is that it must make probabilistic predictions about the outcomes of experiments. Whilst this is conceptually extremely minimal, the mathematical consequences of this lead to a rich formal structure known as a GPT.

Because physical theories describe predictions about measurement outcomes in experiments, a few elements are necessarily present in all of them. Namely, these theories need to talk about types of systems, possible states for each of them, possible measurement outcomes, transformations, and the operation of discarding a system (see table 1). In quantum theory, these elements are, respectively, the Hilbert spaces, the density operators on them, positive operators upper-bounded by the identity, completely positive trace-non-increasing (CPTNI) linear maps, and the (partial) trace operation.

Having those elements present, although necessary, is not sufficient to express the full form of a physical theory. Some structure relating them are implied by the way that experiments are performed. Abstractly speaking, a notion of connectivity between those elements must also be present because, in experiments, we perform actions on systems, that is, we subject them to processes, and these processes can happen in parallel (independently) or in sequence. This motivates a notion of compositionality of processes.

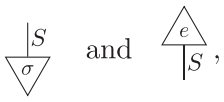
From this notion of how the experimental processes connect, or compose, a convenient diagrammatic notation can be defined so as to capture the entire structure of the GPTs. We can represent any process by a box, and encode the type of system on which it happens as a labelled input wire at the bottom of it. (Hence, we have also implied that systems are represented by wires.) Additionally, since the type of a system can change after a process, we denote the output type of a process by a labelled wire on top of its box. In this notation, then, a system type S, and a transformation T from a system type A to a system type B, respectively, appear as



(1)

A state of a system can be conceptualised as some *preparation procedure*, which, abstractly speaking, is also a process. Hence we can represent it as a box that has no input wire but has as output the wire corresponding to the type of that system. Similarly, an effect, or measurement outcome, is a box with input wire corresponding to the system where it can be observed, and

no output wire. We follow the convention that states and effects are represented by triangular boxes, so a state σ and an effect e of a system S appear as

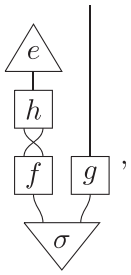


(2)

respectively. Because of this, the discarding operation, since it has an input but no output, appears as a special effect in the theory. This effect is sometimes called the deterministic effect and is unique for each system type[3]. In this notation, it is represented by

(3)

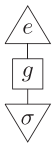
These diagrammatic pieces can be connected when the input/output wire types match. This represents the sequential composition of processes. When processes are instead drawn side by side, we are representing their parallel composition. By connecting boxes, therefore, we can then construct more complex diagrams, i.e. complex processes, such as



(4)

where we omit the wire labels for simplicity, but it should be clear that only matching types can be connected.

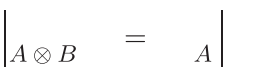
When a diagram has no loose wires, they are interpreted as numbers, which in the case of GPTs are the probabilities generated by the theory. For instance,



$= \mathsf{Prob}(e|g,\sigma)$ (5)

denotes the probability that the outcome associated to effect e is observed when the system is prepared in state σ and a transformation g is applied to it.

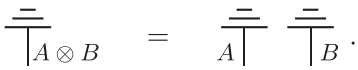
Of course, we might need to describe systems that are composed by simpler parts—multipartite systems—so we can emphasize that some system is composite by drawing the wires of its parts side by side



(6)

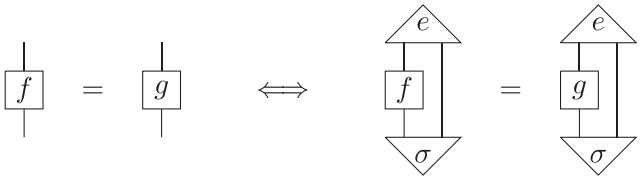
[3] The uniqueness of this discarding effect means that we are dealing with so-called causal GPTs [92].

When we represent bipartite composite systems by the two wires together, its deterministic effect is represented by the composition of the deterministic effects of its parts:



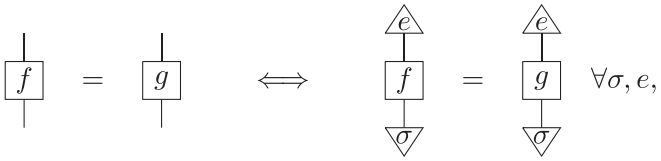
(7)

With what we have, we can represent simple experimental processes, composite processes, and probabilities of outcomes in those experiments. To reason about them, we need now a notion of equality of processes, or, in other words, a notion of tomography. We say that two processes are equal if they give the same probabilities in all situations, so



(8)

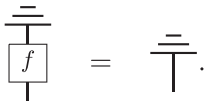
In fact, in this paper we will work with a special class of GPTs which satisfy the principle of *tomographic locality* [9, 90]. This means that:



(9)

that is, in a tomographically local theory we can do process tomography without a side channel.

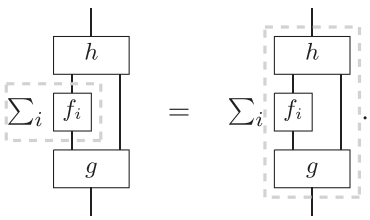
An important type of processes is that of those that are *discard-preserving*, which means they satisfy the following:



(10)

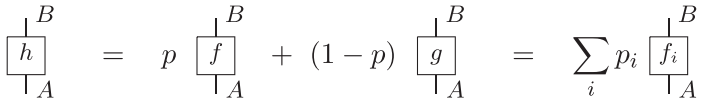
In the case of quantum theory, these correspond to the trace-preserving maps. Physically-realisable discard-preserving processes in a GPT are known as *channels*. Discard-preservation also defines a notion of causality for processes [92, 93]: a process is said to be causal if it is discarding preserving. This is so because this condition ensures compatibility with relativistic causal structure [94].

A final ingredient in the GPT formalism is the possibility to represent convex mixtures of processes. This stems from the requirement that in an experiment we can always decide to perform f with probability p or g with probability $(1-p)$, at least, provided that f and g have the same input and output systems. This is introduced through the definition of a sum of processes that distributes over diagrams:



(11)

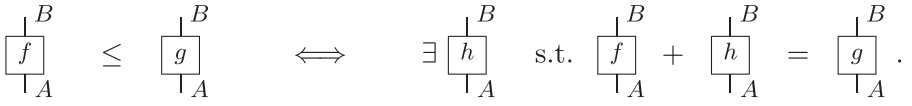
This definition implies that we can only sum processes with the same input/output types. From this, since a probability p can be a number (diagram without loose wires) of the theory, we can write



(12)

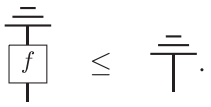
to describe convex mixtures of processes.

At this point a notion of order can be defined for processes:



(13)

This order allows us to define discard-non-increasing processes. A process f is said to be discard-non-increasing if and only if



(14)

In any GPT all (physically-realisable) processes must be discard-nonincreasing; this corresponds to the constraint in quantum theory that processes are trace-nonincreasing. In particular, this means that for any effect in the theory, there must be another effect such that they sum to the deterministic effect. Note that this is important for the definition of measurements. In quantum theory, for example, the deterministic effect is the trace operation, or multiplication by identity followed by the trace, and it is required that the POVM elements forming a measurement sum to identity, so each of them is less than or equal to the deterministic effect.

Since in this work we focus on the class of GPTs that are tomographically local, we can moreover use the particular duotensor notation of reference [90]. Next we will present the basics of this notation.

3. Duotensor basics

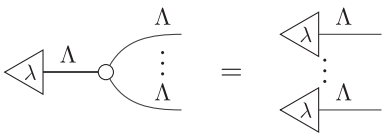
Here we present an adaptation to the duotensor formalism where, in addition to the GPT systems of the previous section, we also have classical systems representing measurement outcomes and control systems. In order to distinguish these two kinds of systems, the classical ones will be drawn horizontally. We will also label them by finite sets, Λ:

(15)

The physical processes transforming between these classical systems are (sub)stochastic maps between these finite sets. We draw these as white boxes, such as:

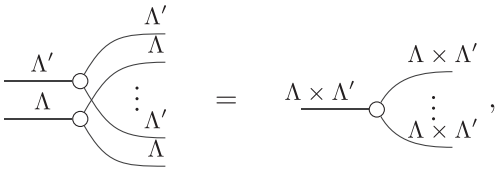
(16)

A particularly useful example which we will make use of in this work is the copy map, which we draw as a white dot and is defined by:



$$\forall \lambda \in \Lambda. \tag{17}$$

The copy map satisfies:



(18)

that is, copying the components of a system is the same as copying the composite system.

In contrast to the physical (sub)stochastic maps, we will draw mathematically well defined but (potentially) unphysical processes as black boxes such as:

(19)

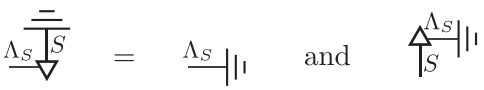
which, in this case, would be a linear map from Λ to Λ' which is not (sub)stochastic, e.g., it may have negative coefficients.

In contrast to the approach of reference [90], rather than labelling horizontal systems by black and white dots, we instead label the processes as being either black or white. This is equivalent but more convenient for us as, on the one hand, we can interpret the color as representing whether or not a process is physical, and, on the other hand, it takes us to a more standard category-theoretic notation. Indeed, categorically there is no distinction between the horizontal and vertical wires, it is simply a convenient way to label the different objects, at which point it is clear that all of the processes that we draw below live inside the category of real linear maps.

For each system S in the GPT we define a particular minimal informationally-complete state preparation and measurement. We call these the fiducial preparation and fiducial measurement. A state preparation is a box which has a classical input and a GPT output where the classical input controls which state is prepared, whilst a measurement is a box which has a GPT input and a classical output where the classical output encodes the result of the measurement. We can therefore denote the fiducial preparation and fiducial measurement for a system S as:

(20)

where without loss of generality we take Λ_S to index both the fiducial set of states and the fiducial set of effects. Moreover, all of the fiducial states are normalised and the fiducial effects sum to the unit effect, such that:



(21)

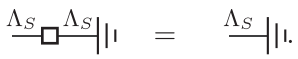
Note that here we follow the convention of reference [91] rather than reference [90], as the former demands that the fiducial effects form a measurement whilst the latter does not. This

does not constitute a loss of generality as a minimal informationally-complete measurement can be shown to exist for any GPT (see section 4.2).

Now, for each system S, define the *fiducial transition matrix* by

(22)

and note that equation (21) implies that these fiducial transition matrices are stochastic maps, such that:

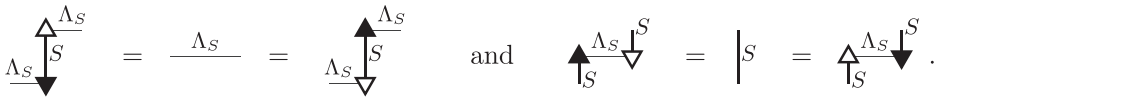


(23)

Now, the fact that the fiducial preparation and measurement are informationally-complete means that they are invertible linear maps. Importantly, however, these inverses are *not* typically physical transformations. We therefore denote them as:

(24)

such that:

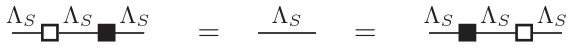


(25)

Where we are again using our convention that processes that are filled in black represent mathematical processes which may not be physical. In particular, we think of the inverse of the fiducial preparation as a (potentially) unphysical measurement, and the inverse of the fiducial measurement as a (potentially) unphysical state preparation. We can then moreover define

(26)

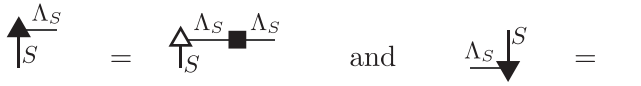
which can easily be seen using equation (25) to be the inverse of the fiducial transition matrix. Hence:



(27)

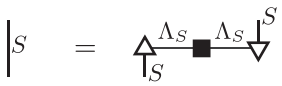
The fiducial transition matrix and its inverse (the white and black squares respectively) are known as *hopping metrics* in the terminology of Hardy.

It is also easy to see from these conditions that:



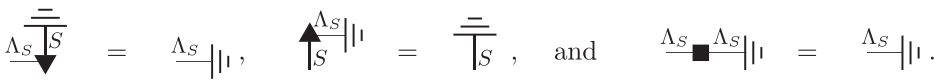
(28)

which, in particular, means that:



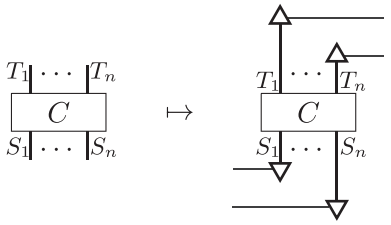
(29)

Moreover, it is also easy to show that:



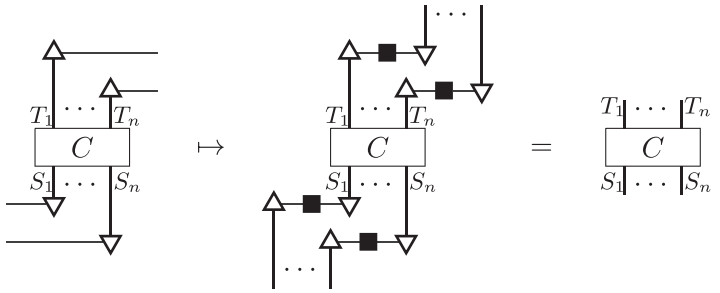
(30)

The key use of all of this for us, is that it allows us to map any GPT channel to a stochastic map and back again as follows: a GPT channel is mapped to a stochastic map via



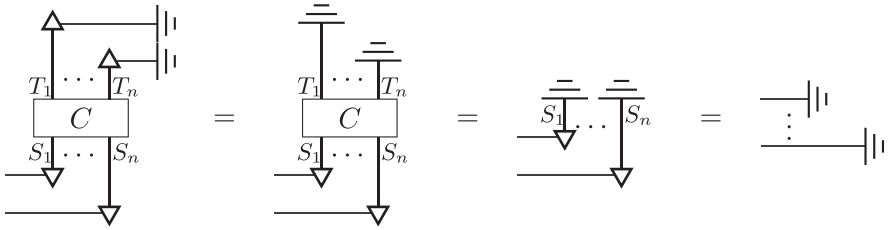
(31)

and the stochastic map associated to the GPT channel can be mapped back to the GPT channel via



(32)

It is clear that the rhs of equation (31) is indeed stochastic as it is positive (since it is composed out of physically realisable GPT transformations) and satisfies:



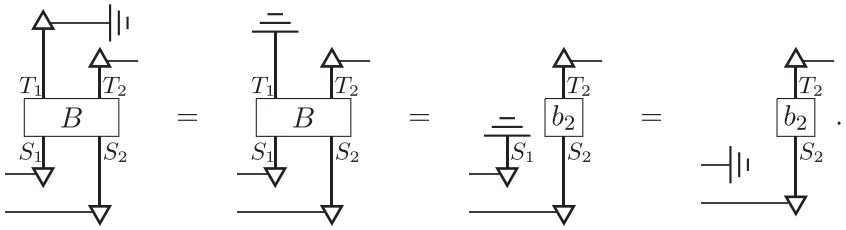
(33)

where the second equality holds because C is a GPT channel rather than a generic GPT process. Similar arguments imply that if C satisfies certain no-signalling conditions then so too will the associated stochastic map.

For example, a bipartite channel B is said to be non-signalling if:

(34)

from which it is easy to show that the associated stochastic map will also be non-signalling, for example:



(35)

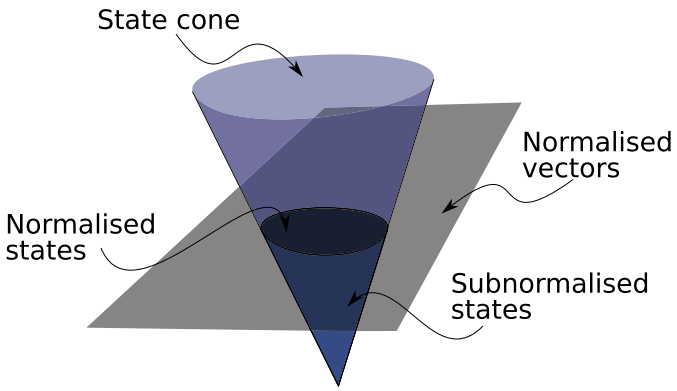
This straightforwardly generalises to multipartite GPT channels, and also to the case where only some of the no-signalling conditions hold. That is, the non-signalling structure of the channel and of the associated stochastic map are the same.

4. Geometry of transformations

In this section we present a geometric perspective on some of the processes discussed above, as well as on particular types of channels.

4.1. States

First let us start by discussing the geometry of the state space for some system S. Schematically this looks like:

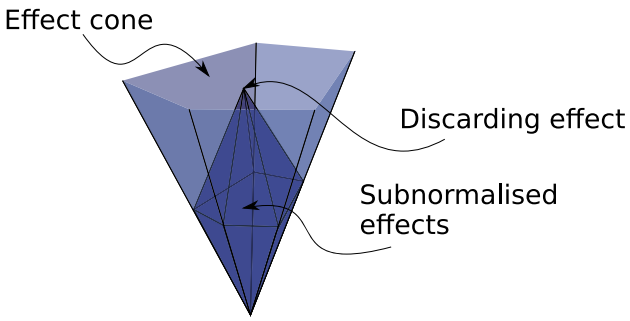


(36)

Formally, we have some real vector space V_S which contains a convex cone of states, $\mathcal{T}^S$, which is closed, pointed, and full dimensional, with an intersecting hyperplane which defines the normalised vectors. The intersection of this hyperplane and the state cone defines the normalised state space, Ω_S. A subnormalized state s is a vector in the cone such that there exists $\alpha \geqslant 1$ such that αs is normalized. In particular, the convex set of subnormalised states spans the vector space and, moreover, there exists at least one normalized state which is interior to the cone.

4.2. Effects

Next let us consider the geometry of the effect space for some system S. Schematically this looks like:



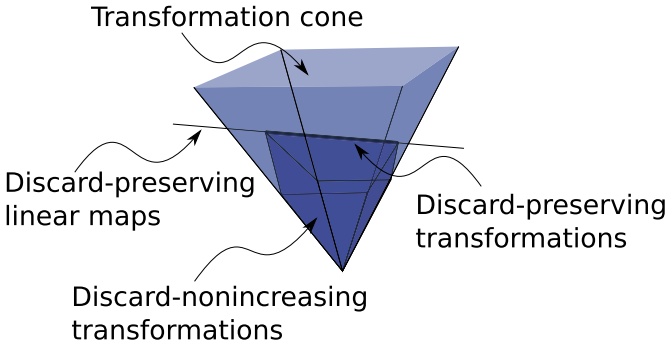
(37)

Formally, the effect space of S lives inside the dual of the vector space of states, V_S^*, and consists of a convex cone of effects, $\mathcal{T}_S$, which is closed, pointed and full dimensional. The unique 'normalised' effect (the *discarding effect*) $\overline{\top}$, is the unique linear functional that evaluates to 1 on the intersecting hyperplane defining the normalised states. This must be in the interior of the effect cone such that it is an order unit for the cone. That is, we have that every effect e in the cone can be rescaled to an effect αe, for some $\alpha > 0$, such that there exists some e' in the cone which satisfies $\alpha e + e' = \overline{\top}$. The set of subnormalised effects, $\mathcal{E}_S$, can be defined as those that satisfy this condition for some $\alpha \geqslant 1$. In particular, this ensures that the convex set of subnormalised effects spans the dual space and that $\overline{\top}$ is in the interior of the effect cone.

This lets us justify our earlier claim that a minimal fiducial measurement exists for every system. Consider a minimally spanning set of vectors $\{\overline{\top} - \sum_i e_i, e_i\}$ where $e_i \in \mathcal{E}_S$. Then we can re-scale the $e_i \mapsto \lambda_i e_i$ where $\lambda_i > 0$ such that $\overline{\top}$. In this case $\{\overline{\top} - \sum_i \lambda_i e_i\} \cup \{\lambda_i e_i\}_i$ is a valid measurement in the theory[4].

4.3. Physical transformations

Finally, we turn to our main focus which is the geometry of transformations within a tomographically local GPT. Schematically this looks like:



(38)

[4] This simple argument leverages a form of the no-restriction hypothesis [92] that says that every collection of effects that sums to $\overline{\top}$ is a measurement in the theory. A less-simple yet more-general argument that does not use this assumption is presented in appendix B.

As we are assuming tomographic locality, the transformations from S to T live inside the vector space of linear maps from V_S to V_T, which we denote as $\mathcal{L}(V_S, V_T)$. The geometric picture that we present here is not as standard in the literature as it is for the state and effect cases, and so we now explain how this structure arises.

In this picture we have a convex set of normalised transformations which are defined by the intersection of an affine set (namely, the discard-preserving linear maps) and a convex cone (namely, the cone of transformations, $\mathcal{T}_S^T$). We can then view this as a positive cone such that the discard-nonincreasing transformations are those that are 'underneath' the discard-preserving transformations in the associated partial order.

As there exists a set of states which span V_T and a set of effects which span V_S^*, then, using the fact that $\mathcal{L}(V_S, V_T) \cong V_S^* \otimes V_T$, we have that

$$\mathcal{L}(V_S, V_T) \cong \text{span}\left\{ \left. \begin{array}{c} T \\ \boxed{s} \\ \\ \boxed{e} \\ S \end{array} \right| s \in \Omega_T, e \in \mathcal{T}_S \right\}. \tag{39}$$

This means that any linear map in $\mathcal{L}(V_S, V_T)$ can be written as

$$\left(\sum_{i=1}^{k} \begin{array}{c} T \\ \boxed{s_i} \\ \\ \boxed{e_i} \\ S \end{array} \right) - \left(\sum_{i'=1}^{k'} \begin{array}{c} T \\ \boxed{s'_{i'}} \\ \\ \boxed{e'_{i'}} \\ S \end{array} \right) \tag{40}$$

for some finite values of k and k', where $s_1, \ldots, s_k, s'_1, \ldots, s'_{k'}$ are normalised states, and $e_1, \ldots, e_k, e'_1, \ldots, e'_{k'}$ are in the effect cone.

With this in mind, we define the following convex subcone $\mathcal{K}$ of the cone of transformations $\mathcal{T}_S^T$ which is useful in our analysis:

$$\mathcal{K} := \left\{ \left. \sum_{i=1}^{k} \begin{array}{c} T \\ \boxed{s_i} \\ \\ \boxed{e_i} \\ S \end{array} \right| s_i \in \Omega_T, e_i \in \mathcal{T}_S, k \text{ finite} \right\} \subseteq \mathcal{T}_S^T \subset \mathcal{L}(V_S, V_T). \tag{41}$$

Equipped with this definition, we can express $\mathcal{L}(V_S, V_T)$ neatly as

$$\mathcal{L}(V_S, V_T) = \mathcal{K} - \mathcal{K} := \{\phi_1 - \phi_2 | \phi_1, \phi_2 \in \mathcal{K}\}, \tag{42}$$

which means that $\mathcal{K}$ spans $\mathcal{L}(V_S, V_T)$.

4.4. Measure-and-prepare transformations and discard-preserving channels

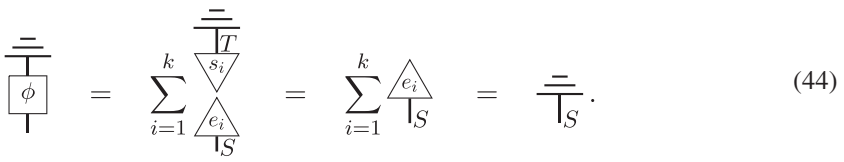
Of particular interest is the set of physical transformations referred to as *measure-and-prepare*, which we denote as MP:

$$\text{MP} := \left\{ \left. \sum_{i=1}^{k} \begin{array}{c} T \\ \boxed{s_i} \\ \\ \boxed{e_i} \\ S \end{array} \right| s_1, \ldots, s_k \in \Omega_T, e_1, \ldots, e_k \in \mathcal{E}_S, \sum_{i=1}^{k} e_i = \mathrel{\overline{\top}}, k \text{ finite} \right\}, \tag{43}$$

recalling that $\overline{\top}$ is the discarding effect.

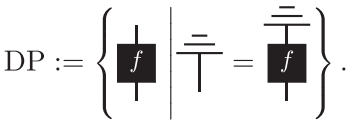
Since these are physically possible in any GPT, they are a subset[5] of the valid transformations, that is, they live inside the convex cone $\mathcal{T}_S^T$ and, in fact, MP $\subseteq \mathcal{K}$.

A measure-and-prepare transformation $\phi \in$ MP has the additional property of being discard-preserving:



(44)

We denote the set of discard-preserving linear maps as DP and define it formally as



(45)

We again use black boxes to represent these maps as they may not be physical transformations. The set of discard-preserving maps forms an affine space (see appendix A for definitions relating to 'affine' concepts), which is easily proven given the definition above.

Note that DP may contain non-physical transformations. However, from the above discussion, it does contain the measure-and-prepare transformations. Neatly, we have that MP $\subseteq$ DP $\cap\, \mathcal{K}$. Perhaps surprisingly, this containment is not strict, as shown in lemma 1 (see the appendix A.2).

The sets DP, MP, and $\mathcal{K}$ allow us to get a useful characterization of the discard-preserving linear maps, which we now discuss.

4.5. A useful characterization of discard-preserving linear maps

The following theorem characterizes the set of discard-preserving maps in terms of those that are also measure-and-prepare.

Theorem 4.1. *Any discard-preserving linear map can be written as an affine combination of measure-and-prepare transformations and any affine combination of measure-and-prepare transformations is a discard-preserving linear map. More formally,*

$$\text{DP} = \text{Aff}(\text{MP}), \tag{46}$$

where Aff *denotes the affine hull operation.*

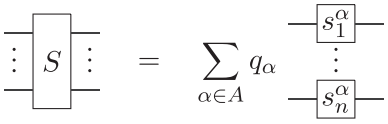
We provide a proof of theorem 4.1 preceded by a background on convex geometry in appendix A.

5. A characterisation of no-signalling GPT channels

Critical to our result is that of reference [7]. In the duotensor formalism presented in the previous section, the result of reference [7] is: if we have some non-signalling stochastic map S it

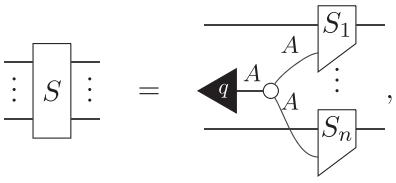
[5] In quantum theory, for example, these are a proper subset of all quantum channels and are known as entanglement-breaking channels.

can be written as an affine combination of product stochastic maps:



(47)

where n is the number of input/output system, $q_\alpha \in \mathbb{R}$, $\sum_{\alpha\in A} q_\alpha = 1$, and the s_i^α are stochastic maps. That is the q_α define a quasiprobability distribution q over the set A. We can therefore equivalently draw this as:



(48)

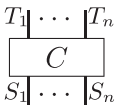
where the white dot is the copy operation, the quasiprobability distribution q is a black triangle because it is not physically realisable as it can have negative coefficients, and the S_i are stochastic maps controlled by the variable A.

The duotensor formalism of reference [90], together with the above understanding of the geometry of GPT transformations, allow us to easily lift this result to arbitrary no-signalling channels in arbitrary tomographically local GPTs.

Theorem 5.1. *A GPT channel C in a tomographically local GPT $\mathcal{G}$ is non-signalling if and only if it can be written as an affine combination of product channels.*

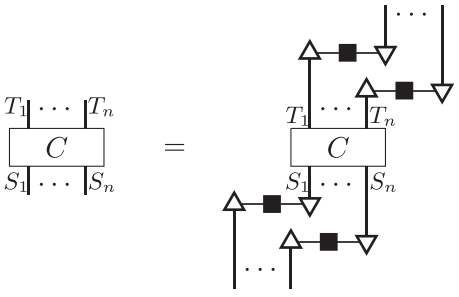
Proof. The if direction is trivial, hence we will focus on the only if direction here.

Consider some n-partite non-signalling channel C in a tomographically local GPT:



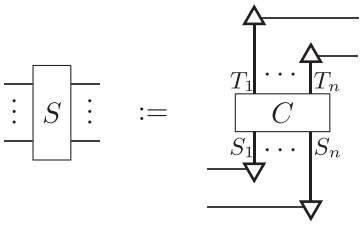
(49)

By decomposing the input and output identities using equation (29), we obtain:



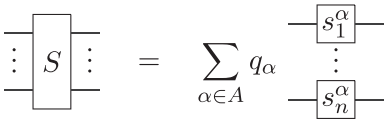
(50)

We can then observe that:



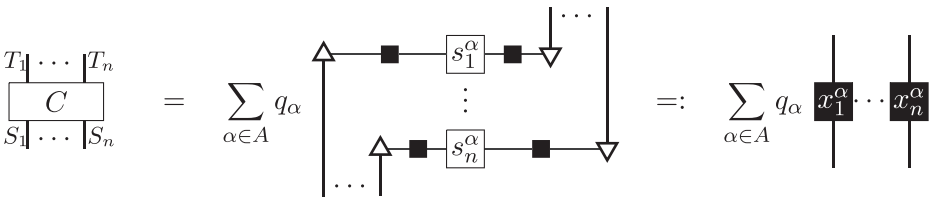
(51)

is a non-signalling stochastic map, and hence we can apply the result of reference [7] to obtain:



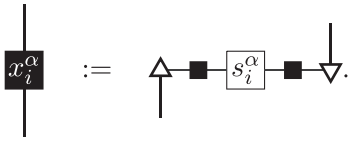
(52)

where s_i^α are stochastic maps, $q_\alpha \in \mathbb{R}$, and $\sum_{\alpha\in A} q_\alpha = 1$. By substituting this back in equation (50), we obtain:



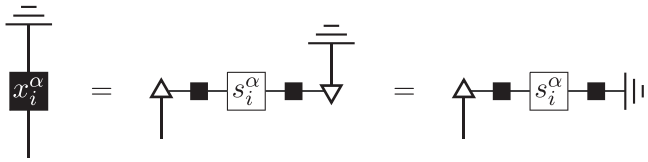
(53)

where

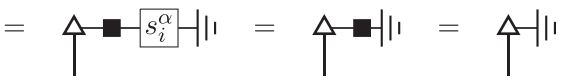


(54)

It is then easy to check that the x_i^α are discard preserving:



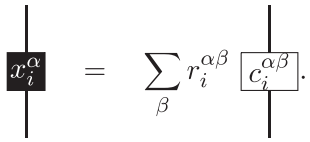
(55)



(56)

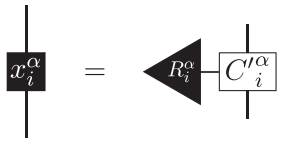
(57)

We can then use theorem 4.1 to write each x_i^α as an affine combination of GPT channels:



$$= \sum_\beta r_i^{\alpha\beta} \; c_i^{\alpha\beta}. \tag{58}$$

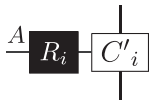
We can write this instead as:



$$= \tag{59}$$

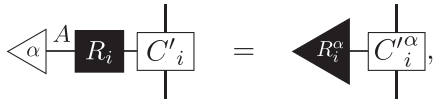
where C'^α_i is a classically controlled channel and R_i^α is a quasidistribution.

Now, let us define:



(60)

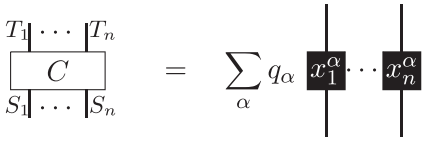
such that



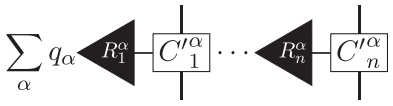
(61)

where C'_i is a classically controlled channel and R_i is a quasistochastic map.

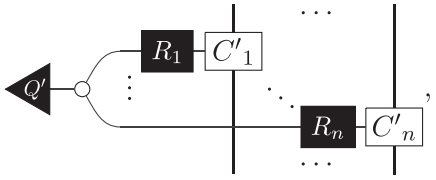
Putting this together with equation (52) we find that:



$$= \sum_\alpha q_\alpha \; x_1^\alpha \cdots x_n^\alpha \tag{62}$$

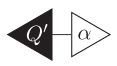


(63)



(64)

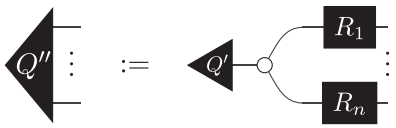
where Q' is the quasidistribution defined by the q_α, i.e.:



$$:= \; q_\alpha \tag{65}$$

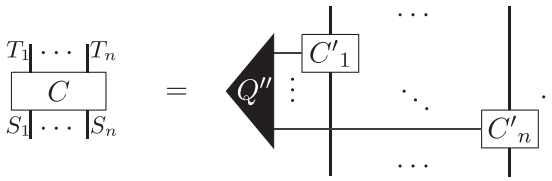
for all α.

Next, let us define Q'' by



(66)

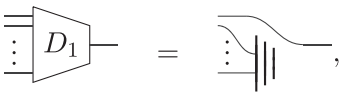
such that we can now combine this with equation (64) to write our channel as



(67)

Equation (67) gives us a quasiprobability distribution over a set of variables, one for each C'_i. In the remainder of this proof we show that this can be rewritten as a quasidistribution over a single variable, which is then copied to each of the C'_i's. Diagrammatically, this means that the copy operation should be the last operation prior to the C'_i's. It is then this quasidistribution over a single variable which defines our affine combination of product channels.

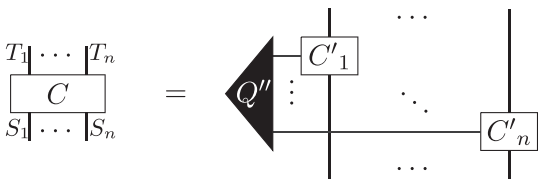
Now, define 'all but system i' marginalisation maps, D_i as:



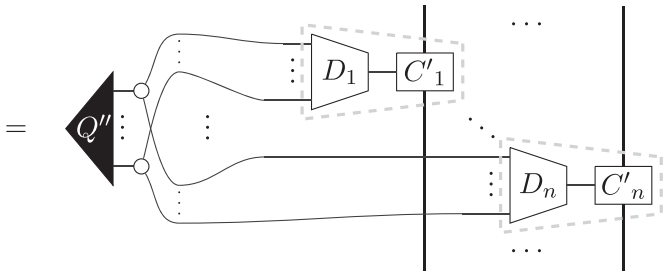
(68)

where the case $i \neq 1$ follows similarly.

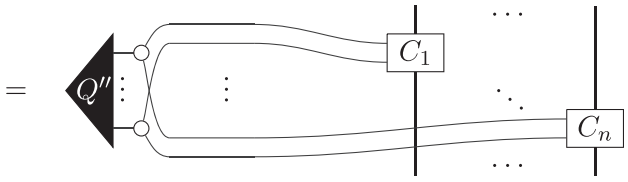
We can then write:



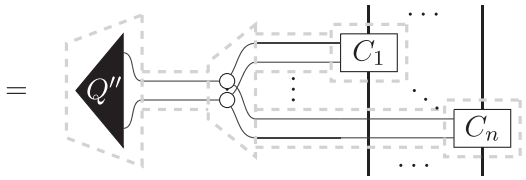
(69)



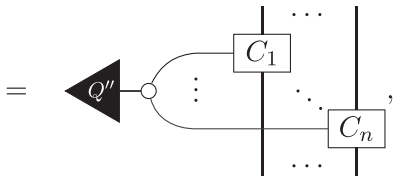
(70)



(71)



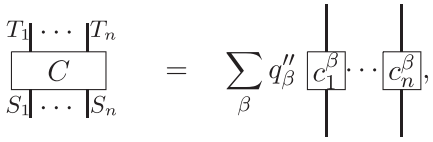
(72)



(73)

where in the last step we have simply merged together parallel wires into a single composite wire, whilst using equation (18) to write the composite of copies as a copy of the composite.

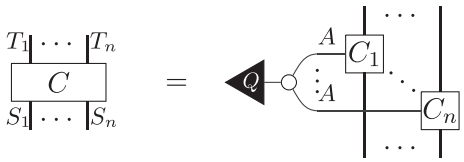
By decomposing the quasidistribution Q'' we can equivalently write this as:



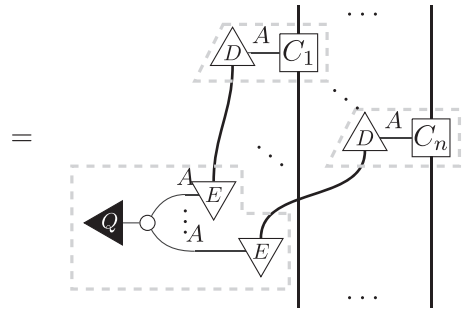
(74)

where c_i^β are GPT channels and q''_β is a quasidistribution. That is, any no-signalling GPT channel can be written as an affine combination of product GPT channels. □

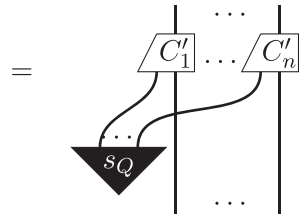
If we have a GPT, such as quantum theory, in which one can always reversibly encode classical data into a GPT system, then we can rewrite this as:



(75)



(76)



(77)

where E is the encoding map, D the decoding map, s_Q is some vector which is not necessarily a physical GPT state, and where the C_i' are GPT channels.

6. Outlook

In this work we have provided a characterisation of multipartite non-signalling channels in arbitrary locally-tomographic theories: these channels can always be represented as affine combinations of local channels. In the case where the input and output system types are classical, i.e., where the channel is a multipartite non-signalling stochastic map, we recover the result of reference [7]. In the case of bipartite non-signalling channels whose inputs and outputs are quantum systems, we in turn recover the results of reference [8].

The application of the results in references [7, 8] spark many interesting directions for future work based on the generalisation we have presented here. In particular, the result of reference [7] has been widely used in the study of non-local and contextual correlations, and therefore our generalisation to arbitrary tomographically local GPTs may well prove useful to studying generalisations of these phenomena (e.g., steering scenarios). In contrast, the result of reference [8] was central to the development of the field of quantumly-indefinite causal order. An important direction for future study is therefore to see whether this result opens the door to exploring the possibilities of GPT-indefinite causal order.

Beyond the particular result that we present here, our proof technique highlights the usefulness of the duotensor formalism [90], and we hope this will motivate its use throughout the quantum community. In particular, we show how it can be used to lift properties of multipartite stochastic maps, to arbitrary tomographically local GPTs. This motivates the question as to which other properties of stochastic maps can be similarly lifted?

Acknowledgments

PJC, JHS, and ABS acknowledge support by the Foundation for Polish Science (IRAP project, ICTQT, Contract No. 2018/MAB/5, co-financed by EU within Smart Growth Operational Programme). All of the diagrams within this manuscript were prepared using TikZit.

Data availability statement

No new data were created or analysed in this study.

Appendix A. A bit of convex geometry and a proof of theorem 4.1

This appendix aims to arrive at the proof of theorem 4.1. In order to do so, a background on convex geometry is provided, and the concepts presented are used to prove lemmas 1–4. Then, lemmas 1, 3 and 4 are directly used to prove theorem 4.1, while lemma 2 is used to prove lemma 3. All the sets DP, MP, and $\mathcal{K}$ that are referred to here are defined in section 4.

A.1. Convex geometry

Given a real vector space V and a (finite) set of vectors $v_1, \ldots, v_k \in V$, we define an *affine combination* of those vectors to be of the form

$$\sum_{i=1}^{k} q_i v_i, \tag{78}$$

where $q_i \in \mathbb{R}$ and $\sum_{i=1}^{k} q_i = 1$. Note the distinction between an affine combination and a convex combination, where the latter also requires that each q_i is nonnegative[6]. Given a set of vectors C, its *affine hull* is the set of all affine combinations of vectors in C and is denoted Aff(C). Lastly, if a set is equal to its affine hull, that is, it contains all of its affine combinations, then we say that the set is affine (or an affine space). Geometrically, one can view an affine space as a subspace translated by a fixed single vector.

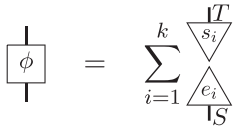
Conceptually, a convex set contains all the line segments between all pairs of points in the set. An affine set contains all the lines that extend beyond the endpoints of the line segments. This brings us to the definition of the *core* of a set. Thinking of line segments, $x \in S$ is in the core of a set S if for all $z \in V$, there exists a $t_z > 0$ such that $x + tz \in S$ for all $t \in [0, t_z]$. Conceptually, this means that given x, you can start drawing a line in *any* direction and stay within the set S. Formally,

$$\text{core}(S) := \{x \in S | \forall\, z \in V, \exists t_z > 0, \text{ such that } x + t\, z \in S, \quad \text{for all } t \in [0, t_z]\}. \tag{79}$$

A.2. Lemmas and proof of theorem 4.1

Lemma 1. $\text{MP} = \text{DP} \cap \mathcal{K}$.

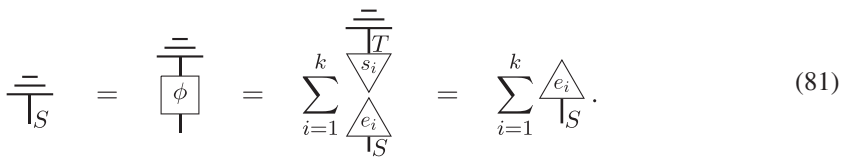
Proof. Since $\text{MP} \subseteq \text{DP} \cap \mathcal{K}$, all that remains to show is the opposite containment. Let $\phi \in \text{DP} \cap \mathcal{K}$ be a fixed, arbitrary vector. Since $\phi \in \mathcal{K}$, we can write it as



(80)

where k is finite, $s_1, \ldots, s_k \in \Omega_T$ are normalized states, and $e_1, \ldots, e_k \in \mathcal{T}_S$ are in the effect cone.

It remains to show that $e_1, \ldots, e_k$ sum to $\overline{\top}$. Since $\phi \in \text{DP}$, we have that



(81)

since the states s_i are normalised. This is the desired equality we seek. Finally, by the partial order given in section 4.2, for all e_i in the sum above, we have $e_i \in \mathcal{E}_S$, so ϕ satisfies all requirements for membership in MP. □

Lemma 2. *Suppose S is a set and*

$$K = \left\{ \sum_{i=1}^{n} \alpha_i s_i \,\middle|\, \alpha_i \geqslant 0, s_i \in S, n \text{ finite} \right\} \tag{82}$$

[6] This difference is analogous to the difference between quasiprobability distributions and (proper) probability distributions.

is a full-dimensional cone, i.e., $V = K - K$. Suppose for $x \in K$, we have that for all $s \in S$, there exists $t_s > 0$ such that $x - ts \in K$ for all $t \in [0, t_s]$. Then $x \in \mathrm{core}(K)$.

The only difference between the definition of core(K) and the condition above is that the vectors in the statement above are not arbitrary but rather belong to a set which generates a full-dimensional cone.

Proof of lemma 2. Since $V = K - K$, for any arbitrary $v \in V$ we can write $v = y - z$ where $y, z \in K$. Then for $x \in K$, $t \geqslant 0$, we can write the following

$$x + tv = x + ty - tz. \tag{83}$$

So, for x to be in core(K), it suffices to find a $t_v > 0$ such that $x + tv \in K$ for all $t \in [0, t_v]$. To do that, we have two cases to analyse, $z = 0$ and $z \neq 0$. Note that if $z = 0$, $x + tv = x + ty \in K$ for all $t \geqslant 0$ since $x, y \in K$, and K is cone, so this case is trivial. Suppose $z \in K$ is nonzero, then we can write it as $\sum_i \alpha_i s_i$ where $\alpha_i > 0$ and $s_i \in S$ and the sum is finite. Then we have

$$x + tv = x + ty - \sum_i \alpha_i t s_i. \tag{84}$$

For brevity, define $a = \sum_i \alpha_i > 0$. By hypothesis, let $t_i > 0$ be such that

$$x - a t s_i \in K \tag{85}$$

for all $t \in [0, t_i]$. This exists since a is positive and by assumption there is some $t_i' = a t_i$ such that $x - t s_i \in K$ for all $t \in [0, t_i']$.

We now have

$$\begin{aligned} x + tv &= x + ty - tz && (86)\\ &= ty + x - \sum_i \alpha_i t s_i && (87)\\ &= ty + \frac{1}{a}\left(\sum_i \alpha_i x - \sum_i a \cdot \alpha_i t s_i\right) && (88)\\ &= ty + \frac{1}{a}\sum_i \alpha_i (x - a t s_i) && (89) \end{aligned}$$

which is in K for all $t \in [0, t_v]$ where $t_v := \min_i\{t_i\}$ (which is positive since there are finitely many indices i). This concludes the proof. □

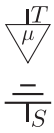
We use this particular case characterization of a core element to prove the following lemma which is helpful in our proof of theorem 4.1.

Lemma 3. *Let $\mu \in \mathrm{int}(\mathcal{T}^T)$ be a normalised state. Then*

$$\begin{array}{c} \mathbb{1}^T \\ \overline{\mu} \\ = \\ \overline{\mathbb{1}_S} \end{array} \in \mathrm{DP} \cap \mathrm{core}(\mathcal{K}), \tag{90}$$

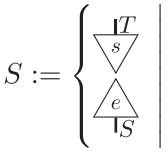
where, recall, $\overline{=} \in \mathrm{int}(E_S)$ is the discarding effect.

Proof. Clearly

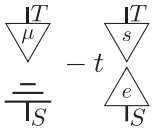

$\in \mathrm{DP},$ (91)

as μ is normalised, so all that remains to show is that it is in core($\mathcal{K}$).

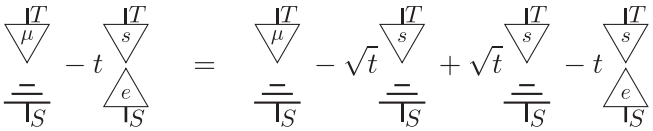
Define

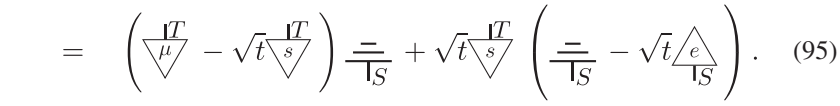

(92)

Since $\mathcal{K}$ (as defined in equation (41)) is the convex hull of S (as defined in equation (82)) and $\mathcal{L}(V_S, V_T) = \mathcal{K} - \mathcal{K}$, by lemma 2, it suffices to show that for a fixed $s \in \mathcal{T}^T$ and $e \in \mathcal{T}_S$, there exists $\hat{t} > 0$ such that

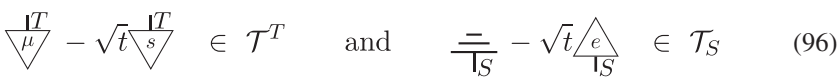

(93)

for all $t \in [0, \hat{t}]$. For $t > 0$, we can write

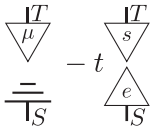

(94)


(95)

Note that


(96)

for all sufficiently small $t > 0$ as μ and $\frac{=}{\top}$ are interior in their respective cones. Therefore,

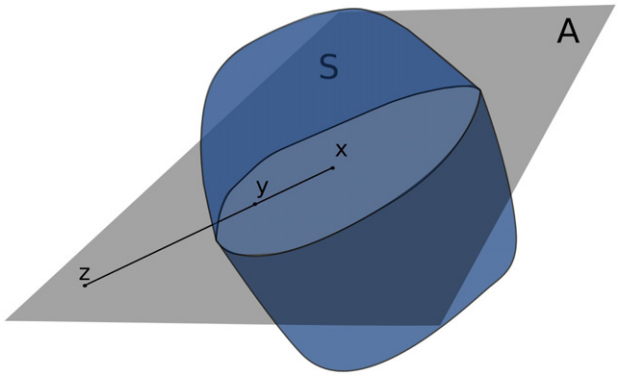

(97)

is in $\mathcal{K}$ for all $t > 0$ sufficiently small. This concludes the proof. □

Lemma 4. *Given a real vector space V, let $S \subseteq V$ be a subset and let $A \subseteq V$ be an affine space. If* core$(S) \cap A \neq \emptyset$, *then* Aff$(S \cap A) = A$.

Before diving into the proof, we explain the idea first since the proof is actually quite simple to picture geometrically, but the proof we give here is algebraic. We start with a point

$x \in \text{core}(S) \cap A$ and draw the line segment between that point to some other arbitrary fixed point $z \in A$. Since $x \in \text{core}(S)$, there is a point on that line segment, call it y, such that it is still in S. And since A is affine, y is in A as well. The proof concludes by noting that z is an affine combination of x and y.



(98)

Proof of lemma 4. Since $S \cap A \subseteq A$, we have that $\text{Aff}(S \cap A) \subseteq \text{Aff}(A) = A$. Therefore, all that remains to show is the reverse containment. To this end, let $z \in A$ be a fixed, arbitrary vector and let $x \in \text{core}(S) \cap A$ (which exists by hypothesis). Define

$$y = (1 - t)x + tz, \tag{99}$$

for some $t > 0$ which we define momentarily. Notice that y is an affine combination of x and z, both of which are in A, and thus $y \in A$ as well. We now want to show that $y \in S$. Note that y can be rewritten as

$$y = x + t(z - x). \tag{100}$$

Since $x \in \text{core}(S)$, there exists a $t \in (0, 1)$ such that $y \in S$. Thus, we have x and y both belonging to $S \cap A$. Notice that

$$z = \left(\frac{1}{t}\right)y + \left(\frac{t-1}{t}\right)x. \tag{101}$$

Since z is an affine combination of x and y, both belonging to $S \cap A$, the result follows. □

Lemmas 1–4 together allow us to write a short proof for theorem 4.1.

Proof of theorem 4.1. We want to show that $\text{DP} = \text{Aff}(\text{MP})$.

By lemma 1, we know

$$\text{MP} = \text{DP} \cap \mathcal{K}. \tag{102}$$

Now, lemma 3 tells us that

$$\text{DP} \cap \text{core}(\mathcal{K}) \neq \emptyset. \tag{103}$$

If we set $\text{A} := \text{DP}$ and $S := \mathcal{K}$, the assumptions in lemma 4 are satisfied and we can use it to conclude that

$$\text{DP} = \text{Aff}(\mathcal{K} \cap \text{DP}) = \text{Aff}(\text{DP} \cap \mathcal{K}) = \text{Aff}(\text{MP}). \tag{104}$$

□

Appendix B. Minimal fiducial measurements

We now prove the following proposition.

Proposition B.1. *In any generalised probabilistic theory, there exists a minimal fiducial measurement for each system.*

Proof. The effect space $\mathcal{E}_S$ for a given system S spans the dual vector space V_S^*. Hence, we can consider a set of effects

$$\{\dot{=}\} \cup \{e_i\}_{i=1}^{\dim[V_S^*]-1} \tag{105}$$

which is a basis for V_S^*.

Recall that every effect e_i must correspond to some outcome of some measurement N_i. Therefore, by performing classical post-processing of N_i, one can construct for each e_i a two-outcome measurement M_i of which e_i can be taken to be its zeroth outcome. For each $i \in \{1, \ldots, \dim[V_S^*] - 1\}$, M_i is hence explicitly given by $M_i = \{e_i, \bar{e}_i\}$, where the effect $\bar{e}_i$ is such that $e_i + \bar{e}_i = \dot{=}$. We also define $M_0 = \{\dot{=}, 0\}$ which is a binary outcome measurement such that the $\dot{=}$ outcome will always occur.

Next we can define a classically-controlled measurement M with a setting variable of the form $C = \{0, \ldots, \dim[V_S^*] - 1\}$ such that when setting i is chosen, measurement M_i is implemented. Note that M is also a binary outcome measurement.

We can now define our minimal fiducial measurement M_F via a suitable processing of the classical input and output of M,[7] which we explicitly present next. First, sample a classical variable $i \in \{0, \ldots, \dim[V_S^*] - 1\}$ according to some probability distribution $\lambda_i := p(i)$ with full support—that is $\lambda_i > 0$ for all i and $\sum_i \lambda_i = 1$. Next, keep a record of the variable i, and use it as the input of the device M. Let $o \in \{0, 1\}$ be the output of M. The outcome at this stage of this process is hence the pair (o, i). Finally, to generate the outcome o_f of the measurement M_F post-process the pair (o, i) as follows:

$$(0, i) \mapsto i, \tag{106}$$

$$(1, i) \mapsto 0. \tag{107}$$

Obtaining an outcome $o_f = i > 0$ after this post-processing corresponds to the effect $\lambda_i e_i$, and obtaining outcome $o_f = 0$ after this post-processing corresponds to the effect

$$\lambda_0 \dot{=} + \sum_{i=1}^{\dim[V_S^*]-1} \lambda_i \bar{e}_i = \dot{=} - \sum_{i=1}^{\dim[V_S^*]-1} \lambda_i e_i. \tag{108}$$

Therefore, the collection of effects presented below is a valid measurement and also forms a basis of V_S^*:

$$\left\{ \dot{=} - \sum_{i=1}^{\dim[V_S^*]-1} \lambda_i e_i \right\} \bigcup \{\lambda_i e_i\}_{i=1}^{\dim[V_S^*]-1}. \tag{109}$$

It follows that the set of effects in equation (109) forms a minimal fiducial measurement. □

[7] In the language of reference [95], M_F is a flag-convexification of the multimeter M.

ORCID iDs

John H Selby https://orcid.org/0000-0002-4596-7501
Ana Belén Sainz https://orcid.org/0000-0003-3123-8436

References

[1] Beckman D, Gottesman D, Nielsen M A and Preskill J 2001 Causal and localizable quantum operations *Phys. Rev.* A **64** 052309
[2] Bell J S 1964 On the Einstein Podolsky Rosen paradox *Physics* **1** 195–200
[3] Albert E, Podolsky B and Rosen N 1935 Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **47** 777
[4] Schrödinger E 1935 Discussion of probability relations between separated systems *Math. Proc. Camb. Phil. Soc.* **31** 555–63
[5] Brunner N, Cavalcanti D, Pironio S, Scarani V and Wehner S 2014 Bell nonlocality *Rev. Mod. Phys.* **86** 419–78
[6] Uola R, Ana C, Costa S, Nguyen H C and Gühne O 2020 Quantum steering *Rev. Mod. Phys.* **92** 015001
[7] Al-Safi S W and Short A J 2013 Simulating all nonsignaling correlations via classical or quantum theory with negative probabilities *Phys. Rev. Lett.* **111** 170403
[8] Chiribella G, D'Ariano G M, Perinotti P and Benoit V 2013 Quantum computations without definite causal structure *Phys. Rev.* A **88** 022318
[9] Hardy L 2001 Quantum theory from five reasonable axioms (arXiv:quant-ph/0101012)
[10] Barrett J 2007 Information processing in generalized probabilistic theories *Phys. Rev.* A **75** 032304
[11] Sikora J and Selby J 2018 Simple proof of the impossibility of bit commitment in generalized probabilistic theories using cone programming *Phys. Rev.* A **97** 042302
[12] Selby J H and Sikora J 2018 How to make unforgeable money in generalised probabilistic theories *Quantum* **2** 103
[13] Sikora J and Selby J H 2019 On the impossibility of coin-flipping in generalized probabilistic theories via discretizations of semi-infinite programs (arXiv:1901.04876)
[14] Howard B, Dahlsten O C O, Leifer M and Toner B 2008 Nonclassicality without entanglement enables bit commitment *2008 IEEE Information Theory Workshop (ITW'08)* (Piscataway, NJ: IEEE) pp 386–90
[15] Lami L, Palazuelos C and Winter A 2018 Ultimate data hiding in quantum mechanics and beyond *Commun. Math. Phys.* **361** 661–708
[16] Barrett J, Hardy L and Kent A 2005 No signaling and quantum key distribution *Phys. Rev. Lett.* **95** 010503
[17] Barrett J, de Beaudrap N, Hoban M J and Lee C M 2017 The computational landscape of general physical theories (arXiv:1702.08483)
[18] Lee C M and Barrett J 2015 Computation in generalised probabilistic theories *New J. Phys.* **17** 083001
[19] Krumm M and Müller M P 2018 Quantum computation is an island in theoryspace (arXiv:1804.05736)
[20] Lee C M and Selby J H 2016 Generalised phase kick-back: the structure of computational algorithms from physical principles *New J. Phys.* **18** 033023
[21] Lee C M and Selby J H 2016 Deriving Grover's lower bound from simple physical principles *New J. Phys.* **18** 093047
[22] Howard B, Lee C M and Selby J H 2018 Oracles and query lower bounds in generalised probabilistic theories *Found. Phys.* **48** 954–81
[23] Müller M P and Ududec C 2012 Structure of reversible computation determines the self-duality of quantum theory *Phys. Rev. Lett.* **108** 130401
[24] Ududec C, Barnum H and Emerson J 2011 Three slit experiments and the structure of quantum theory *Found. Phys.* **41** 396–405
[25] Lee C M and Selby J H 2017 Higher-order interference in extensions of quantum theory *Found. Phys.* **47** 89–112
[26] Garner A J P 2018 Interferometric computation beyond quantum theory *Found. Phys.* **48** 886–909

[27] Howard B, Müller M P and Ududec C 2014 Higher-order interference and single-system postulates characterizing quantum theory *New J. Phys.* **16** 123029
[28] Dakić B, Paterek T and Brukner Č 2014 Density cubes and higher-order interference theories *New J. Phys.* **16** 023028
[29] Howard B, Lee C M, Scandolo C M and Selby J H 2017 Ruling out higher-order interference from purity principles *Entropy* **19** 253
[30] Horvat S and Dakić B 2020 Interference as an information-theoretic game (arXiv:2003.12114)
[31] Short A J and Wehner S 2010 Entropy in general physical theories *New J. Phys.* **12** 033023
[32] Howard B, Barrett J, Clark L O, Leifer M, Spekkens R, Stepanik N, Wilce A and Wilke R 2010 Entropy and information causality in general probabilistic theories *New J. Phys.* **12** 033024
[33] Chiribella G and Scandolo C M 2015 Entanglement and thermodynamics in general probabilistic theories *New J. Phys.* **17** 103027
[34] Kimura G, Ishiguro J and Fukui M 2016 Entropies in general probabilistic theories and their application to the Holevo bound *Phys. Rev.* A **94** 042113
[35] Chiribella G and Scandolo C M 2017 Microcanonical thermodynamics in general physical theories *New J. Phys.* **19** 123043
[36] Howard B, Barrett J, Krumm M and Müller M P 2015 Entropy, majorization and thermodynamics in general probabilistic theories (arXiv:1508.03107)
[37] Schmid D, Selby J H, Pusey M F and Spekkens R W 2020 A structure theorem for generalized-noncontextual ontological models (arXiv:2005.07161)
[38] Schmid D, Selby J, Wolfe E, Kunjwal R and Spekkens R W 2019 The characterization of noncontextuality in the framework of generalized probabilistic theories (arXiv:1911.10386)
[39] Shahandeh F 2019 Contextuality of general probabilistic theories and the power of a single resource (arXiv:1911.11059)
[40] Chiribella G and Yuan X 2014 Measurement sharpness cuts nonlocality and contextuality in every physical theory (arXiv:1404.3348)
[41] Gross D, Müller M, Colbeck R and Dahlsten O C O 2010 All reversible dynamics in maximally nonlocal theories are trivial *Phys. Rev. Lett.* **104** 080402
[42] Czekaj Ł, Horodecki M and Tylec T 2018 Bell measurement ruling out supraquantum correlations *Phys. Rev.* A **98** 032117
[43] Barnum H, Beigi S, Boixo S, Elliott M B and Wehner S 2010 Local quantum measurement and no-signaling imply quantum correlations *Phys. Rev. Lett.* **104** 140401
[44] Czekaj Ł, Sainz A B, Selby J and Horodecki M 2020 Correlations constrained by composite measurements (arXiv:2009.04994)
[45] Joe H, Lal R and Pusey M F 2014 Theory-independent limits on correlations from generalized Bayesian networks *New J. Phys.* **16** 113043
[46] Weilenmann M and Colbeck R 2020 Analysing causal structures in generalised probabilistic theories *Quantum* **4** 236
[47] Lami L 2018 Non-classical correlations in quantum mechanics and beyond (arXiv:1803.02902)
[48] Barnum H, Gaebler C P and Wilce A 2013 Ensemble steering, weak self-duality, and the structure of probabilistic theories *Found. Phys.* **43** 1411–27
[49] Plávala M 2017 Conditions for the compatibility of channels in general probabilistic theory and their connection to steering and bell nonlocality *Phys. Rev.* A **96** 052127
[50] Banik M 2015 Measurement incompatibility and Schrödinger–Einstein–Podolsky–Rosen steering in a class of probabilistic theories *J. Math. Phys.* **56** 052101
[51] Cavalcanti P J, Selby J H, Sikora J, Galley T D and Sainz A B 2021 Witworld: a generalised probabilistic theory featuring post-quantum steering (arXiv:2102.06581)
[52] Richens J G, Selby J H and Al-Safi S W 2017 Entanglement is necessary for emergent classicality in all physical theories *Phys. Rev. Lett.* **119** 080503
[53] Lee C M and Selby J H 2018 A no-go theorem for theories that decohere to quantum mechanics *Proc. R. Soc.* A **474** 20170732
[54] Scandolo C M, Salazar R, Korbicz J K and Horodecki P 2018 Is it possible to be objective in every physical theory? (arXiv:1805.12126)
[55] Selby J and Coecke B 2017 Leaks: quantum, classical, intermediate and more *Entropy* **19** 174
[56] Bae J, Kim D-G and Kwek L-C 2016 Structure of optimal state discrimination in generalized probabilistic theories *Entropy* **18** 39
[57] Jenčová A and Plávala M 2017 Conditions on the existence of maximally incompatible two-outcome measurements in general probabilistic theory *Phys. Rev.* A **96** 022113

[58] Barnum H, Barrett J, Leifer M and Wilce A 2007 Generalized no-broadcasting theorem *Phys. Rev. Lett.* **99** 240501
[59] Barnum H and Wilce A 2011 Information processing in convex operational theories *Electron. Notes Theor. Comput. Sci.* **270** 3–15
[60] Barnum H, Barrett J, Leifer M and Wilce A 2012 Teleportation in general probabilistic theories *Proc. Symp. Applied Mathematics* vol 71 pp 25–47
[61] Heinosaari T, Leppäjärvi L and Plávala M 2019 No-free-information principle in general probabilistic theories *Quantum* **3** 157
[62] Plávala M 2016 All measurements in a probabilistic theory are compatible if and only if the state space is a simplex *Phys. Rev.* A **94** 042108
[63] Jenčová A 2018 Incompatible measurements in a class of general probabilistic theories *Phys. Rev.* A **98** 012133
[64] Filippov S N, Heinosaari T and Leppäjärvi L 2017 Necessary condition for incompatibility of observables in general probabilistic theories *Phys. Rev.* A **95** 032127
[65] Kuramochi Y 2020 Compatibility of any pair of two-outcome measurements characterizes the choquet simplex *Positivity* **24** 1479–86
[66] Bluhm A, Jenčová A and Nechita I 2020 Incompatibility in general probabilistic theories, generalized spectrahedra, and tensor norms (arXiv:2011.06497)
[67] Dahlsten O C O, Garner A J P and Vedral V 2014 The uncertainty principle enables non-classical dynamics in an interferometer *Nat. Commun.* **5** 4592
[68] Takakura R and Miyadera T 2020 Preparation uncertainty implies measurement uncertainty in a class of generalized probabilistic theories *J. Math. Phys.* **61** 082203
[69] Takakura R and Miyadera T 2021 Entropic uncertainty relations in a class of generalized probabilistic theories *J. Phys. A: Math. Theor.* **54** 315302
[70] Sun L L, Zhou X, Kwek L C and Yu S 2020 No disturbance without uncertainty under generalized probability theory https://doi.org/10.21203/rs.3.rs-106927/v1
[71] Masanes L, Galley T D and Müller M P 2019 The measurement postulates of quantum mechanics are operationally redundant *Nat. Commun.* **10** 1361
[72] Galley T D and Masanes L 2017 Classification of all alternatives to the Born rule in terms of informational properties *Quantum* **1** 15
[73] Galley T D and Masanes L 2018 Any modification of the Born rule leads to a violation of the purification and local tomography principles *Quantum* **2** 104
[74] Galley T D and Masanes L 2020 How dynamics constrains probabilities in general probabilistic theories (arXiv:2002.05088)
[75] Chiribella G, D'Ariano G M and Perinotti P 2011 Informational derivation of quantum theory *Phys. Rev.* A **84** 012311
[76] Chiribella G 2014 Dilation of states and processes in operational-probabilistic theories (arXiv:1412.8539)
[77] Chiribella G 2014 Distinguishability and copiability of programs in general process theories (arXiv:1411.3035)
[78] Barnum H and Wilce A 2014 Local tomography and the Jordan structure of quantum theory *Found. Phys.* **44** 192–212
[79] Branford D, Dahlsten O C O and Garner A J P 2018 On defining the Hamiltonian beyond quantum theory *Found. Phys.* **48** 982–1006
[80] Wilce A 2009 Four and a half axioms for finite dimensional quantum mechanics (arXiv:0912.5530)
[81] Wilce A 2018 A royal road to quantum theory (or thereabouts) *Entropy* **20** 227
[82] Barnum H, Graydon M A and Wilce A 2016 Composites and categories of Euclidean Jordan algebras (arXiv:1606.09331)
[83] Barnum H, Duncan R and Wilce A 2013 Symmetry, compact closure and dagger compactness for categories of convex operational models *J. Phil. Logic* **42** 501–23
[84] Wilce A 2011 Symmetry and composition in probabilistic theories *Electron. Notes Theor. Comput. Sci.* **270** 191–207
[85] Masanes L and Müller M P 2011 A derivation of quantum theory from physical requirements *New J. Phys.* **13** 063001
[86] Masanes L, Müller M P, Augusiak R and Pérez-García D 2013 Existence of an information unit as a postulate of quantum theory *Proc. Natl Acad. Sci. USA* **110** 16373–7
[87] Müller M P and Masanes L 2013 Three-dimensionality of space and the quantum bit: an information-theoretic approach *New J. Phys.* **15** 053040

[88] Plávala M 2021 General probabilistic theories: an introduction (arXiv:2103.07469)
[89] Müller M 2021 Probabilistic theories and reconstructions of quantum theory *SciPost Phys. Lect. Notes* 28
[90] Hardy L 2011 Reformulating and reconstructing quantum theory (arXiv:1104.2066)
[91] van de Wetering J 2017 Quantum theory is a quasi-stochastic process theory (arXiv:1704.08525)
[92] Chiribella G, D'Ariano G M and Perinotti P 2010 Probabilistic theories with purification *Phys. Rev.* A **81** 062348
[93] Coecke B 2016 Terminality implies no-signallingand much more than that *New Gener. Comput.* **34** 69–85
[94] Kissinger A, Hoban M and Coecke B 2017 Equivalence of relativistic causal structure and process terminality (arXiv:1708.04118)
[95] Selby J H, Schmid D, Wolfe E, Sainz A B, Kunjwal R and Spekkens R W 2021 Accessible fragments of generalized probabilistic theories, cone equivalence, and applications to witnessing nonclassicality (arXiv:2112.04521)