Common-cause structure of non-signalling phenomena in generalised theories

Paulo José Cavalcanti de Vasconcelos Filho

Supervisor: Dr. Hab. Ana Belén Sainz

UNIVERSITY OF GDAŃSK
Gdańsk, Poland
June 24, 2024

Abstract

Many features of quantum theory such as Bell nonlocality and steering have been shown to have no classical counterpart. While both nonlocality and steering are defined in experimental setups that assume space-like separation, the no-signalling constraints imposed by relativity are not enough to single out either classical or quantum correlations. In this work, we use the generalized probabilistic theory (GPT) framework, which contains classical and quantum theories as particular cases, as well as more general ones, to investigate the relationship between the no-signalling constraints and the aforementioned phenomena. We define a GPT and use it to show that there exists GPTs that can model so-called post-quantum assemblages as common-cause processes. We further use that GPT to investigate the information processing consequences of post-quantum steering. Next, we prove that non-signalling channels in locally tomographic theories (a class that includes classical and quantum theories) are affine combinations of product channels. Finally, we use this parametrization of the non-signalling channels to show that given any causal locally tomographic GPT, there exists a second GPT that can realize all non-signalling channels of the first theory as common-cause processes.

Abstrakt

Wiele cech teorii kwantowej, takich jak nielokalność Bella i sterowanie kwantowe, nie ma klasycznego odpowiednika. Choć zarówno nielokalność, jak i sterowanie są zdefiniowane w układach eksperymentalnych zakładających oddalone od siebie laboratoria, zasada braku sygnalizacji wynikająca z teorii względności nie jest wystarczająca do wyodrębnienia korelacji klasycznych ani kwantowych. W tej pracy wykorzystujemy formalizm uogólnionej teorii probabilistycznej (GPT), który zawiera teorie klasyczne i kwantowe jako szczególne przypadki, a także bardziej ogólne teorie, do zbadania związku między zasadą braku sygnalizacji a wyżej wymienionymi zjawiskami. Definiujemy GPT i używamy ich do pokazania, że istnieją GPT, które mogą modelować tak zwane post-kwantowe asamblaże jako procesy wspólnej przyczyny. Następnie wykorzystujemy tę GPT do zbadania zastosowania post-kwantowego sterowania do przetwarzania informacji. Dalej udowadniamy, że kanały bez sygnalizacji w teoriach lokalnie tomograficznych (klasa obejmująca teorie klasyczne i kwantowe) są kombinacjami afinicznymi kanałów produktowych. Na koniec, używamy tej parametryzacji kanałów bez sygnalizacji, aby pokazać, że dla dowolnej przyczynowej lokalnie tomograficznej GPT istnieje druga GPT, która może realizować wszystkie kanały bez sygnalizacji pierwszej teorii jako procesy wspólnej przyczyny.

Acknowledgments

I wish express my primary ackowledgement to my supervisor Ana Belén Sainz, whose trust and guidance made this entire work possible. I also thank my collaborators John Selby, Jamie Sikora and Thomas D. Galley. My gratitude also goes to my friends Anubhav Chaturvedi, Giovanni Scala, Sumit Rout for the interesting discussions, all colleagues in ICTQT for the nice time at work, and personally to my girlfriend Elena Kemoklidze, my parents Paulo Vasconcelos and Monica Nascimento, and my sister Maria Eduarda Nascimento for always supporting my efforts.

This work was supported by the Foundation for Polish Science (IRAP project, ICTQT, Contract No. 2018/MAB/5, co-financed by EU within Smart Growth Operational Programme).

Publications included in the dissertation

[First paper] Cavalcanti, P. J., Selby, J. H., Sikora, J., Galley, T. D., Sainz, A. B. (2022). Post-quantum steering is a stronger-than-quantum resource for information processing. npj Quantum Information, 8(1), 76.

[Second paper] Cavalcanti, P. J., Selby, J. H., Sikora, J., Sainz, A. B. (2022). Decomposing all multipartite non-signalling channels via quasiprobabilistic mixtures of local channels in generalised probabilistic theories. Journal of Physics A: Mathematical and Theoretical, 55(40), 404001.

[Third paper] Cavalcanti, P. J., Selby, J. H., Sainz, A. B. (2024). Every nonsignaling channel is common-cause realizable. Physical Review A, 109(4), 042211.

Contents

List of Figures

Chapter 1

Introduction

Already in the early days of its development, quantum theory evoked surprise from physicists. Famously for its inherent randomness[1], and notably for how this randomness motivated arguments claiming that something could be missing from the theory [2] or that the theory could allow for a distant system to steer the state of another one despite of no influence propagating between the two [3], this trend exists to this day.

A crucial development of the thought about these features was given by Bell with his famous theorem [4], and his argument was later given a cleaner form by Clauser, Horne, Shimony and Holt [5]. Their works concern the sets of correlations that can be modelled with classical or quantum theories in experimental setups where distant parties perform space-like separated measurements on their parts a shared system. Those results established a mathematical inequivalence between the classical way of computing probabilities and the quantum way. More specifically, under constraints rooted in the causal structure of the experimental setup [6, 7, 8, 9], they showed that some *correlations* in an experimental setup that we now call a Bell scenario [10, 11] are forbidden by classical theory but allowed by quantum theory. That showed that attempts to reduce quantum theory to classical models should fail, unless the price of rejecting some natural-sounding assumptions is paid. To the property of quantum theory of realizing those correlations the name was given of *Bell nonlocality*, and to the conditions defining that property the name of Bell inequalities. We may also refer to this as the classical (hidden) variable problem.

More than a mere mathematical fact, this gap between the two theories, i.e. the existence of nonlocality, provided an experimental criterion to decide which of the two theories was physically correct [5]. When those classically forbidden but quantumly allowed correlations were finally observed in experiments with quantum system, [12, 13, 14, 15, 16] a confirmation of the necessity of the quantum formalism was given. Moreover, the same statistical properties that initially raised suspicion on quantum theory have been shown to provide computational capabilities beyond what is classically possible[17, 18, 19], sprouting entire lines of research dedicated to the mission of exploiting such capabilities.

Concurrently to the technological efforts [20, 21, 22, 23, 24, 25, 26, 27, 28, 29] aimed at extracting the most of the information processing capabilities from quantum theory, the door

opened by Bell with his seminal result still motivated foundational questions. After all, a gap between the two theories was established, but no explanation besides their mathematical formulations was provided. That result deepened the interpretational difficulties of the theory that defies the classical intuition.

Another interesting, also relevant to us, development on the problem of hidden variables was given when Popescu and Rohrlich [30] imposed a different kind of constraint on the probabilities of the Bell scenario setup, if compared to the ones used in Bell inequalities. Namely, instead of requiring a specific form for the causal structure of the experimental setup, they required only that the setup, however organized it may be, cannot be used to send information between the parties, i.e. that it obeys a constraint imposed by relativistic considerations. The surprising result thereby obtained was that such a constraint allowed for correlations that are forbidden not only by classical, but also by quantum theory. The most extreme of those form what is commonly call the PR-box (or Popescu-Rohrlich box) [30] correlations. This result showed that, at least at this level of analysis, quantum bounds on the achievable correlations concern more than merely the causal connections in the experiment.

Likewise for correlations, quantum steering [31, 32, 33], which we also call here EPR inference [34], was also shown to display some kind of nonlocality with analogous properties. Here, we deal with an experimental setup type that can be seen as a variant of the Bell scenario where measurements are not performed in all parts of the relevant system, retaining the assumption that the non-measured parts of the system remain as quantum systems. This difference makes for the fundamental mathematical object of study not being the familiar correlation, but the so-called *assemblage* [35], which plays nevertheless an analogous role. These assemblages can also be constrained in multiple ways. Under the no-signalling condition, the resulting set of non-signalling assemblages has been shown [36] to contain elements which cannot be realized by quantum theory by means of parties sharing a resource prepared in their common past, but that nevertheless can't be used to signal information, much like the PR-box did for correlations. Also, some assemblages that can be prepared when the common-cause is a quantum state cannot be prepared when the common-cause is a classical state. Again, there are boundaries defined by classical means of preparation, by quantum means, and by what is allowed by the no-signalling condition. We classify the non-signalling assemblages as classical, quantum or post-quantum, depending on whether of not they are realizable by classical or quantum means,or not realizable by either.

Many questions can be asked about those facts. For instance, if some non-signalling correlations can't be realized within quantum theory, are there other theories where they can (like the PR-box is realized in Boxworld), and if so, are they physical? Or, alternatively, if quantum theory is the right framework for inferences about physical systems, what constraints we need to add to the no-signalling ones to explain its behavior?

The first of those questions is related to the fields of the operational/generalized probabilitic theories (OPTs/GPTs) [37, 38, 39, 40, 41, 42, 43, 44, 45], which investigate the properties of the mathematical structures that can be interpreted as theories in the same way as in classical and quantum theories. We can say these are the candidates for mathematical frameworks for probabilistic predictions about physical experiments. For instance,

a GPT that now is called Boxworld was defined in such a way that it can realize the PR-box correlations in a common cause scenario, in a manner structurally similar to how quantum measurements on a Bell state realizes correlations displaying Bell nonlocality. The second question is related, for instance, to axiomatizations of quantum theory [41, 46, 44, 47, 48] which attempt to give quantum theory some grounding in considerations more humanly interpretable than its bare complete mathematical formalism.

The works presented in this thesis investigate the interplay between these two angles on the contraints imposable on Bell scenarios and alike. Namely, the angle from which we analyse experiments like the one proposed by Bell through the causality lens, or through the information signalling lens. The results we obtained concern how non-signalling phenomena such as nonlocality and Einstein-Podolsky-Rosen steering are instances of common-cause processes in generalized theories. This theme connects the three articles presented here, where special attention is given to non-signalling phenomena that cannot be realized in quantum theory.

In this thesis, we start by looking at post-quantum steering, which, previously to the first publication here included, was defined and shown to (mathematically) exist [36]. Inspired by the fact that the post-quantum correlations introduced by the PR-box have been shown to be realizable in a generalized probabilistic theory (GPT) which is now known as Boxworld [49, 50, 51, 52], we define a new GPT, which we call Witworld [53], that can model the post-quantum assemblages defined in the literature previous to our work, with the specific form of the common-cause decomposition. This GPT is then used to investigate the information processing properties of post-quantum steering. In the second work [54] we study the non-signalling channels within the GPT framework, generalizing some results about their mathematical form. Finally, in the third article [55], we use the knowledge from the second article to prove the existence of a map that always connects a causal locally tomographic GPT, such as quantum or classical theory, to another one that realizes its non-signalling channels by means of common-cause decomposable channels. This final result consists of a proof that the results we obtained in the first work are an instance of a more general connection. More specifically, these results constitute a bridge between the non-signalling and the (GPT) causal constraints on GPT channels, of which correlations and assemblages are particular cases.

For our mathematical framework, additionally to the traditional linear algebraic formalism that are normally used for quantum and classical theories, we use a diagrammatic calculus [56, 57, 58, 59] whose validity is guaranteed by results in category theory [60, 61]. Although this calculus is equivalent to the linear algebraic one, and can be regarded as only notation, it provides a very clear interpretability which we find useful in the developtment of the results. Nevertheless, since this is not an usual tool in quantum theory research, we provide a chapter containing all relevant definitions from category theory that are needed to ground our calculation techniques on solid results that guarantee their coherence. Our approach to the GPT formalism relies both on geometric (from linear algebra) and on compositional concepts (from category theory).

Therefore, we structure this work as follows. Firstly, we provide a traditional presentation

of the non-signalling principle, nonlocality, and quantum steering, using only the traditional linear algebraic notation. Next, a section containing from the basic definition of category, all the way to our class of categories of interest, the symmetric monoidal categories, and the basic properties of its diagrammatic calculus, which we make heavy use of. Finally, we define operational and generalized probabilistic theories based on that framework, and explain how non-signalling and nonlocality are expressed within this framework, which puts all our concepts of interest in a unifying, clear notation.

These preliminaries together provide enough background to enable the reading of the articles contaning our results in a relatively self-contained manner. These articles are then included in this text in a contextualized way.

Chapter 2

Preliminaries

In this chapter, we provide definitions and an overview of the mathematical formalism in the study of nonlocality. We first take a look at the no-signalling principle and nonlocality in classical and quantum treatments in their traditional formalisms. Later, after an overview of the category theoretical definitions and notation, we provide a generalised probabilistic theoretic (GPT) view of those scenarios.

Using the GPT approach, we can provide definitions that put the previous considerations under a unifying framework, making similarities and differences clear to the eye. Moreover, with the use of the diagrammatic notation, provided by results in category theory, we are able to give exact and general definitions with a much less cumbersome notation for certain concepts in multipartite scenarios. For example, we don't need to clutter the view with many swap maps that would be required for exact definitions in standard linear algebraic notation. In summary, given this diagrammatic notation and an operational interpreration of the category theoretical structures, constituting the definition of GPT, our problems, results and methods related to no-signalling and nonlocality should be clearly understood.

2.1 The No-Signalling Principle and Locality

We discuss the concepts of locality and no-signalling in the context of quantum and classical theories. The purpose of this is to see what the structures we will be talking about under the GPT framework are, and how this latter approach helps seeing all of them in the same footing. For this reason, we provide definitions in the simplest case, the bipartite one, but keep in mind that generalisations are straightforward and will be provided when we give the GPT treatment of the problem. A more complete treatment of nonlocality can be found in reviews or textbooks [10, 11].

2.1.1 Bell Scenarios

The so-called Bell nonlocality of quantum theory was noticed first in the famous Einstein-Podolsky-Rosen paper [2], where the authors make the case that there is an incompleteness

of the theory on the basis that its predictions allow for correlations that would require superluminal signalling to be properly explained. In this context, "incompleteness" and "properly" have meanings specific to their argument. The authors proceed to point out that such a signal would violate the principles of special relativity, and therefore was taken as an evidence of a problem with quantum theory. The EPR argument, despite being powerful, left space for discussion because it relied only on physical arguments. Later, a formalization of the problem that helped prove the gap between classical and quantum explanations was found by Bell [4], and later made more clear with the Clauser-Horn-Shimony-Hold (CHSH) work [5] and their famous inequality. In this section, we present the mathematical definitions used in this context to establish a gap between classical and quantum models in terms of Bell nonlocality.

Bell nonlocality is a feature that quantum theory demonstrates in the context of experimental setups which we call Bell scenarios. Here, two parties, which we call Alice and Bob, choose and perform measurements on some shared system and record the output values. If Alice and Bob can respectively choose between M_A and M_B measurements, each measurement with m_A and m_B possible outputs each, we define the $(M_A, m_A; M_B, m_B)$ Bell scenario. Calling Alice's and Bob's choices x and y, and outputs a and b, respectively, this experiment defines the observed behavior

$$p(a, b|x, y). \tag{2.1}$$

So we have the following definition:

Definition 1 (Bipartite Behavior)**.** A behavior $\mathcal{P}$ of a bipartite Bell scenario is a conditional probability distribution

$$\mathcal{P}\{p(a, b|x, y)\}_{a\in A, b\in B, x\in X, y\in Y}, \tag{2.2}$$

where A, B are the sets of possible outcomes for Alice and Bob, respectively, and X, Y the sets of measurements that they choose from.

We can then study how the behaviors could possibly be produced by Alice and Bob. It is clear given the generality of the setup that there are many strategies they can use to produce the behavior, which may even rely on different physical systems. We refer, in a generic way, by *process* to the whole of the choice of physical systems and actions that they perform. If Alice and Bob perform this experiment according to processes λ from a set Λ, with probabilities $p(\lambda)$, each of those defines a conditional probability distribution

$$p(a, b|x, y, \lambda) \tag{2.3}$$

so that under the assumption that these processes are prepared independently of the rest of the random variables, the observed behavior is written as

$$p(a, b|x, y) = \int d\lambda\, p(\lambda) p(a, b|x, y, \lambda). \tag{2.4}$$

The behaviors are the fundamental mathematical objects in the study of Bell nonlocality. In this context we can classify them and study what conditions are required for the realization of specific behaviors or classes thereof.

2.1.2 Local Hidden Variable and Non-signalling Behaviors

Given that behaviors are observed in Bell scenarios, we can then classify these distributions with regard to their information flow properties according to the following definitions:

Definition 2 (Non-signalling process)**.** A process λ is non-signalling if each party's output does not depend on the other party's input:

$$p(a|x, y, \lambda) = p(a|x, \lambda) \tag{2.5}$$

$$p(b|x, y, \lambda) = p(b|y, \lambda) \tag{2.6}$$

for all a, b, x, y.

Definition 3 (Product Process)**.** A process λ is a product if it factorizes as follows:

$$p(a, b|x, y, \lambda) = p(a|x, \lambda)p(b|y, \lambda). \tag{2.7}$$

Note that a process being product implies that it is non-signalling but the converse is not true [30].

We also have similar definitions for behaviors.

Definition 4 (Non-signalling Behavior)**.** A behavior $p(a, b|x, y)$ is non-signalling if each party's output does not depend on the other party's input

$$p(a|x, y) = p(a|x) \tag{2.8}$$

$$p(b|x, y) = p(b|y) \tag{2.9}$$

for all a, b, x, y.

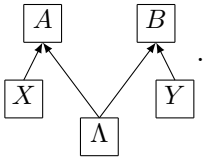
Definition 5 (Local Hidden Variable Behavior)**.** A behavior $p(a, b|x, y)$ has a local hidden variable model if it factorizes as follows:

$$p(a, b|x, y) = \int d\lambda\, p(\lambda)p(a|x, \lambda)p(b|y, \lambda) \tag{2.10}$$

If a behavior has such a model, we say it is local hidden variable.

We can see a process is a product when i can be seen as two marginal, local, processes happening independently at Alice's and Bob's labs. So, a local hidden variable (LHV) behavior is one that is realizable via statistical mixture of such product processes.

Given the formal definitions, we can ask where do they come from. Note that the definition of locality comes not just from considerations about the flow of information, as is the case for no-signalling, but really on the way we model statistics classicaly. To understand that intuitively, consider the following graph:



(2.11)

Here, in a classical description, A, B, X, Y and Λ are random variables encoding Alice's and Bob's observed outcomes, measurement choices, and the choice of process governing each round of the experiment. The direction of the wires encodes the directions where the flow of information is assumed to be possible. By imposing on the joint distribution $p(a, b, x, y, \lambda)$ probabilistic independencies that respect the directions of those arrows, we obtain a factorization that defines a local process:

$$p(a, b, x, y, \lambda) = p(\lambda)p(x)p(y)p(a|x, \lambda)p(b|y, \lambda) \tag{2.12}$$

so

$$p(a, b|x, y, \lambda) = p(a|x, \lambda)p(b|y, \lambda). \tag{2.13}$$

Here, it would be possible to draw different graphs that would result in different factorizations that still respect locality, but this specific one encodes the assumptions of independence of the choices of x, y and λ. What is crucial to realize is that we are assuming that the mathematical object modelling the random variables is a joint probability distribution of their observed values, which is true in the case of classical theory but not in general as we shall see. We will later explicitly see how alternatives to that can be given, first in the example of quantum theory, and then in generalised probabilistic theories, which allow us to put the classical and quantum models in equal footing.

Although we merely use this as an intuitive argument, the correspondence between graph representations of causal structures and the factorizations of the probability distributions describing the experiments is formalized by the theory of (classical) causal modelling [62]. Morever, its connection to Bell inequalities has been previously studied [63, 64]. We don't eleborate further on this approach in this work, as we only need to see that considerations of this kind are enough to show the inequivalence between quantum and classical models, and to motivate the GPT approach.

2.1.3 Bell Nonlocality

We now describe what is meant by saying that quantum theory demonstrates Bell nonlocality.

Definition 6 (Bell nonlocality)**.** A behavior $p(a, b|x, y)$ describing the statistics of a Bell scenario is said to exhibit Bell nonlocality if it is not local as per definition 5.

In order to check whether a behavior is Bell nonlocal, one can use Fine's theorem [65] to charaterize the local set as the convex combinations of deterministic behaviors:

Theorem 2.1.1 (Fine's Theorem)**.** The following two propositions are true and equivalent:

- A behavior $p(a,b|x,y)$ is local if and only if it is a convex mixture of local deterministic processes:

$$p(a,b|x,y) = \sum_{i=1}^{m_A^{M_A}} \sum_{j=1}^{m_B^{M_B}} p_{ij}\delta_{a=f_i(x)}\delta_{b=g_j(y)} \tag{2.14}$$

  where $\sum_{ij} p_{ij} = 1$ and $\forall i,j : p_{ij} \geq 0$.

- A behavior $p(a,b|x,y)$ is local if and only if there exists a joint probability distribution $p(a_1...a_{M_A}, b_1...b_{M_B})$ such that the behavior is given by its marginals:

$$p(a,b|x,y) = \sum_{a_i, i\neq x} \sum_{a_j, j\neq y} p(a_1...a_{M_A}, b_1...b_{M_B}) \tag{2.15}$$

From this it is easy to see that the set of local behaviors forms a polytope, whose extremal points are the local deterministic behaviors. The boundary of this set has facets, which is related to the idea of Bell inequality. In general, a Bell inequality is any criterion that separates some nonlocal behaviors from the set of local behaviors. Nevertheless, usually by Bell inequality one refers to linear Bell inequalities.

Definition 7 ((linear) Bell Inequality)**.** A Bell inequality is a criterion of the form

$$I(p) = \sum_{a,b,x,y} I_{a,b,x,y} p(a,b|x,y) \leq I_L \tag{2.16}$$

that is true for all local behaviors.

Given a behavior, one can use a Bell inequality to certify that it is not local. If one observes a violation of the inequality, then one can be sure that the observed behavior cannot possibly be expressed as in Eq. 2.10. The linear Bell inequality is also interesting because it can be directly evaluated with the experimental data.

We can now use a Bell inequality to separate between quantum and classical models. Specifically, our example is the CHSH inequality. Consider the $(2,2,2,2)$ Bell scenario, and take $x,y \in \{0,1\}$ and $a,b \in \{-1,1\}$. Here, each party can choose between two binary measurements. The following quantity can be defined:

$$S = E(a,b|0,0) + E(a,b|0,1) + E(a,b|1,0) - E(a,b|1,1). \tag{2.17}$$

in terms of the expected values $E(a,b|x,y) = \sum_{a,b} ab * p(a,b|x,y)$. It can be shown that the behavior $p(a,b|x,y)$ being local implies

$$S_{\text{local}} \leq 2. \tag{2.18}$$

It is straightfoward to check that $S \leq 2$ can be written in the form in Def. 7.

If, however, we follow a quantum treatment of the problem, we don't factorize the behavior $p(a,b|x,y)$ as in Eq. 2.10 anymore. Rather, we take the shared system to be modelled by

a bipartite state ρ_λ, and the measurements performed by Alice and Bob by POVMs $\{M_{a|x}\}_a$ and $\{M_{b|y}\}_b$, so we use instead

$$\begin{aligned} p(a,b|x,y) &= \int d\lambda\, p(\lambda)\, \mathrm{tr}\big(M_{a|x} \otimes M_{b|y} \cdot \rho_\lambda\big) \\ &= \mathrm{tr}\big(M_{a|x} \otimes M_{b|y} \cdot \rho\big) \end{aligned} \tag{2.19}$$

with $\rho = \int d\lambda\, \rho_\lambda$. Then, defining S analogously, this alternative form for the behavior implies [66]

$$S_{\text{quantum}} \leq 2\sqrt{2}, \tag{2.20}$$

where equality can be reached for some choices of $\rho, \{M_{a|x}\}_{ax}, \{M_{b|y}\}_{by}$. Hence, we can establish a gap between quantum and classical descriptions: under the same information flow constraints, the two approaches to modelling our probabilities yield difference ranges of possible observed values.

Finally, we note that even though quantum theory allows for a higher value of S than classical theory, it is not achieving the highest value possible for non-signalling correlations. Using the PR-box correlation [30]

$$p_{PR}(a,b|x,y) = \frac{1}{2}\delta_{a\oplus b=xy}, \tag{2.21}$$

we can achieve

$$S_{\text{PR}} = 4, \tag{2.22}$$

the highest possible value for S under the no-signalling constraint. In the GPT section, we will see that it is possible to mathematically define a theory in the same framework as one can specify quantum and classical theories, and that this new theory is able to describe the PR-box under the same constraints that led us to the quantum and classical bounds on S. That shows that non-signalling correlations can be some conceivable and mathematically more general correlations than their classical and quantum counterparts. We may call correlations that are, in this sense, not realizable in quantum theory by post-quantum correlations. Naturally, no physical system has been found to require such a theory for its proper description, and many radical consequences of their existence have been shown to hold. In this work, however, since we are interested in the mathematical structure involved in these considerations, it is worth to keep in mind those facts.

2.1.4 Einstein-Podolsky-Rosen Inference, or Steering

In the Bell nonlocality discussion, we have seen that it is possible to establish a gap between correlations describable by quantum and classical theory under the same constraints. Additionally, we have seen that non-signalling correlations are even more general than the quantum and classical ones. In that discussion, as we want to describe a Bell scenario, and in that case the parties Alice and Bob perform measurements on their systems, we end up by studying behaviors of the form $p(a,b|x,y)$. Now, we can generalise that to a scenario where,

say, Bob does not perform a measurement. In that case, we will be studying a correlation between classical outputs at Alice's side and quantum states at Bob's side. It is this generalised scenario that we call steering [32, 33, 35, 67], or Einstein-Podolsky-Rosen inference, in reference to the famours EPR paper, which considered the original scenario.

Let's start with the biparte scenario. If Alice, upon choosing to perform the measurement $x \in X$, sees outcome $a \in A$ with probability $p(a|x)$, then the state at Bob's side is given by a density matrix $\rho_{a|x}$, labelled by Alice's measurement and outcome. The full description of the scenario, then, is given by a set of pairs

$$\{(p(a|x), \rho_{a|x})\}_{a\in A, x\in X} \tag{2.23}$$

of the conditional states and their probabilities. Since all quantum states ρ satisfy $\operatorname{tr}\rho = 1$, we can encode the probability $p(a|x)$ in the density matrix by considering subnormalized states $\sigma_{a|x} = p(a|x)\rho_{a|x}$. In that case, we can recover the probabilities by taking the traces, and the states by normalizing the subnormalized states.

Definition 8 (Assemblage)**.** Let A and X be sets of labels referring to outcomes and measurements, respectivelly. We call an assemblage a labelled set of subnormalized density matrices

$$\{\sigma_{a|x}\}_{a\in A, x\in X} \tag{2.24}$$

such that for all x,

$$\sum_a \operatorname{tr}\big(\sigma_{a|x}\big) = 1. \tag{2.25}$$

The assemblages play in the study of steering the role of fundamental mathematical object, similarly to the behaviors in the context of Bell nonlocality. So, we can classify the assemblages according to their information properties just as we did to behaviors.

Definition 9 (Non-signalling Assemblage)**.** An assemblage $\{\sigma_{a|x}\}_{a\in A, x\in X}$ is non-signalling if

$$\forall x, x' \in X: \quad \sum_a \sigma_{a|x} = \sum_a \sigma_{a|x'} = \rho_B \tag{2.26}$$

for some normalized quantum state ρ_B, the marginal state of Bob's system.

Again, we give a definition that says the marginal descriptions (probabilisties for Alice, and states for Bob) are well defined and Alice's input cannot be used to influence Bob's side.

Definition 10 (Local Hidden State Assemblage)**.** An assemblage $\{\sigma_{a|x}\}_{a\in A, x\in X}$ has a local hidden state model if it can be written as

$$\sigma_{a|x} = \sum_\lambda p(\lambda)p(a|x,\lambda)\rho_{B,\lambda} \tag{2.27}$$

If an assemblage has such a model, we say that it is local hidden state.

This encodes the idea that the assemblage can be prepared by Alice and Bob if they have only a shared classical random variable available, so that Alice measures it and Bob uses the variable to control the preparation of the state at his lab. It can be shown that here too not all non-signalling assemblages have local hidden variable models.

Definition 11 (Quantum Assemblage)**.** An assemblage $\{\sigma_{a|x}\}$ is a quantum assemblage if it can be written as

$$\sigma_{a|x} = \mathrm{tr}_A(M_{a|x} \otimes \mathbb{1}_B \cdot \rho_{AB}) \tag{2.28}$$

for some bipartite state ρ_{AB} and POVM $\{M_{a|x}\}_{a\in A, x\in X}$.

Continuing the analogy, we see that a quantum assemblage is one that can possibly be prepared if the parties Alice and Bob share a quantum biparte state, and perform local operations on their shares to prepare the assemblage.

Definition 12 (Post-quantum Assemblage)**.** A non-signalling assemblage is called post-quantum if it is not quantum.

Similarly to the existence of the PR-box as a non-signalling correlation that is neither quantum nor classical, it can be shown that non-signalling assemblages exist that cannot be realized by the parties if they have only classical or quantum shared states. To see that example, though, one need to go to a more general scenario [36, 68, 69, 70], either in the tripartite case, with at least two “Alices" making measurements, or by allowing Bob to also have an input to the preparation of the state at his lab.

2.1.5 Non-signalling and Localizable Quantum Channels

After giving the first step of generalising the Bell scenario to allow one of the parties to have a quantum state, rather than the classical input/ouput values, we can continue and consider the case where all parties have quantum input and output to their local procedures. In this case, the object of study will be collections of quantum channels, and we have to give a classification [71, 35, 72] that extends the ideas of no-signalling and locality to those objects.

Definition 13 (Quantum Channel)**.** A quantum channel $\Lambda : A \to A'$ is a completely positive trace-preserving map between two linear spaces of operators on complex Hilbert spaces.

This is the standard definition of quantum channel found in quantum information texts. Note that this requirement just says that a map is a quantum channel if it takes quantum density matrices to quantum density matrices, even in the presence of additional side systems on which it doesn't act.

Now, we can define non-signalling quantum channels in a way analogous to our previous discussion. These channels are also sometimes called causal channels because they preserve the relativistic causal structure. We will give the definitions in the bipartite case, which are analogous to the Bell scenario except that the inputs and outputs are quantum systems. In this way, we will refer to the subsystems either by A and B or also by Alice and Bob.

Definition 14 (Non-Signalling Bipartite Quantum Channel)**.** A quantum channel $\Lambda : AB \to A'B'$ from the bipartite system AB to the bipartite system $A'B'$, is non-signalling from the AA' wing of the experiment to the BB' wing if for any state ρ of the total input system,

$$\mathrm{tr}_A(\Lambda(\rho)) = \Lambda_B(\mathrm{tr}_A(\rho)) \tag{2.29}$$

for some quantum channel $\Lambda_B : B \to B'$. We say that the channel is fully non-signalling if it is both non-signalling from AA' to BB' and from BB' to AA'.

This means that no channel E applied by Alice on her subsystem before the action of the channel Λ can affect Bob's marginal state after Λ and vice versa:

$$\begin{aligned} \mathrm{tr}_A([E \otimes \mathbb{1}_B](\rho)) &= \Lambda_B(\mathrm{tr}_A([E \otimes \mathbb{1}_B]\rho)) \\ &= \Lambda_B(\mathrm{tr}_A(\rho)) \end{aligned} \tag{2.30}$$

since E is trace-preserving.

Next we have the analogous of a local hidden state behavior for the quantum case:

Definition 15 (Localizable Biparte Channel)**.** A quantum channel $\Lambda : AB \to A'B'$ is localizable if there exist auxiliary systems C and D, a bipartite quantum state ρ_{CD} of the system CD channels $E_{AC} : AC \to A'$ and $E_{BD} \to B'$ such that

$$\Lambda = (E_{AC} \otimes E_{BD}) \circ (1_A \otimes \rho_{CD} \otimes 1_B) \tag{2.31}$$

where by 1_A and 1_B we denote the identity channels $1_A : A \to A$ and $1_B : B \to B$ that do not change their inputs.

The analogy to the local hidden state behaviors can be seen if one takes ρ_{CD} to be a separable state.

The definition of localizable channel means that two parties Alice and Bob that cannot communicate would be able to implement the channel Λ if they shared the bipartite state ρ by locally applying the operations E_{AC} and E_{CD} on the composite systems of their part of the input to Λ together with their share of ρ. Notice how this mirrors the idea behind local behaviors except that the shared state λ is taken to be a quantum state, and the operations become quantum channels, so when ρ is a separable state, one recovers the notion of classicaly correlated quantum channel.

2.2 Notions of Category Theory

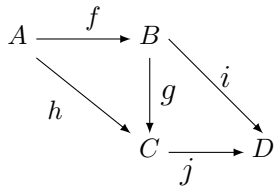
This section concerns the relevant mathematical details of the main structure we need in order to formalize physical theories as abstract mathematical structures. We do not intend to provide a complete introduction to category theory, but we aim to define only the structures of interest to us. This provides more context to our problems and results on generalised probabilistic theories. In particular, we want to understand what a strict symmetrical monoidal category is. A more complete exposure of the concepts presented in this chapter can be found in standard category theory texts [73, 74, 75, 76, 58].

2.2.1 Notation

It is costumary in category theory to use commutative diagrams to express conditions, rather than equations. In such a diagram, a morphism $f : A \to B$ is represented as an arrow between nodes A and B:

$$A \xrightarrow{\ f\ } B \ . \tag{2.32}$$

Moreover, when any two paths have the same starting and ending points, with at least one of them being composite, they must represent equal compositions, so

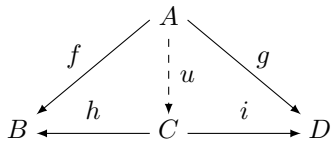


(2.33)

means

$$\begin{aligned} &h = g \circ f, \\ &i = j \circ g, \\ &i \circ f = j \circ h, \\ &i \circ f = j \circ g \circ f, \\ &j \circ h = j \circ g \circ f. \end{aligned} \tag{2.34}$$

Additionally, there are many cases where, if a certain diagram commutes, then one of its arrows is uniquely defined. In that case, we draw that arrow with a dashed line. For example, in



(2.35)

u is the unique arrow $A \to C$ such that the diagram commutes.

2.2.2 Categories

To begin the discussion, let's take a look at a direct definition of category.

Definition 16 (Category)**.** A category $\boldsymbol{C}$ consists of the following elements:

- A class C_0 of objects
$$A, B, C, ... \tag{2.36}$$
- A class C_1 of morphisms
$$f, g, h, ... \tag{2.37}$$

- For each morphism f, two given objects

$$\mathrm{dom}(f), \quad \mathrm{cod}(f), \tag{2.38}$$

  called the domain and codomain of f. We denote $A = \mathrm{dom}(f)$ and $B = \mathrm{cod}(f)$ by $f : A \to B$, and the class of morphisms from A to B by $\mathrm{Hom}_{\boldsymbol{C}}(A, B)$ and call it a *hom-class*.

Moreover, these elements are subject to the following properties:

- Given morphisms $f : A \to B$ and $g : B \to C$, that is, with matching domain and codomain, there is a given morphism

$$g \circ f : A \to C \tag{2.39}$$

  called the composite of f and g.

- For each object A, a morphism

$$1_A : A \to A \tag{2.40}$$

  called the identity morphism of A exists.

Moreover, these elements are required to satisfy the following properties:

- Composition is associative

$$f \circ (g \circ h) = (f \circ g) \circ h \tag{2.41}$$

  for all $f : A \to B$, $g : B \to C$, $h : C \to D$.

- Composition has units

$$f \circ 1_A = 1_B \circ f \tag{2.42}$$

  for all $f : A \to B$.

The morphisms are sometimes called arrows, and we can take these two names as synonymous in this context. Also, we can call the domain and codomain of a morphism by source and target, respectively.

Note that in the definition of category we deliberately don't call the classes sets, as they are allowed to be "too large to be sets" for a given set theory. This is an important distinction in category theory because we often work with collections that are "too big to be sets", such as the class of all sets. The formal treatment of those classes depends on the foundations of mathematics of choice. For instance, one can choose to base the treatment on the von Neumann-Bernays-Gödel set theory (NBG), where classes are primitive and a set is defined to be a class that belongs to some other class, or take the Zermelo-Fraenkel set theory (ZF) and take some set to be a universe, such as a Grothendieck universe, and talk in terms of small sets and classes respectively as elements and subsets of that universe. It is also possible to define categories only in terms of first-order logic and use internal category

theory to define everything else purely in category theoretical terms. Nevertheless, details with regards to foundations of mathematics are beyond the scope of this text, so we merely point that these formalizations exist, and that functions between classes too can be defined, which allows us to give all the definitions that we need using a familiar language.

Given these remarks, we can make the following distinctions that help keep clarity:

Definition 17 (Small Category)**.** A category $\boldsymbol{C}$ is said to be small if both its class C_0 of objects and its class C_1 of morphisms are sets (or small sets).

Definition 18 (Locally Small Category)**.** A category $\boldsymbol{C}$ is said to be locally small if each of its hom-classes are sets (or small sets). In that case we call them hom-sets.

Definition 19 (Large Category)**.** A category $\boldsymbol{C}$ is said to be large if it is not small.

Let's look at a few examples of categories:

Definition 20 (**Sets**)**.** The category of sets, denoted **Sets**, is the category that has the class of all sets as its objects' class, and for a pair of sets A and B, the hom-class $\mathrm{Hom}_{\mathbf{Sets}}(A, B)$ is the set of functions from A to B.

Definition 21 (ℝ**Linear**)**.** The category of real vector spaces, denoted ℝ**Linear**, is the category that has the class of all real vector spaces for its objects' class, and for a pair of real vector spaces V and V', the hom-class $\mathrm{Hom}_{\mathbb{R}\mathbf{Linear}}$ is the set of linear maps from V to V'.

Both **Sets** and ℝ**Linear** are locally small categories. It is useful to have **Sets** in mind to think about categories because one can think of the composition of morphisms as an abstraction of function composition, so in this category these two notions coincide. For us, ℝ**Linear** is relevant because we can take an important class of GPTs as subcategories of it. This fact and the concept of subcategory will be defined later.

There are a few more category theoretical constructions that will be useful for our definitions.

Definition 22 (Isomorphism)**.** A morphism $f : A \to B$ in a category $\boldsymbol{C}$ is called an isomorphism if there is in $\boldsymbol{C}$ an arrow $g : B \to A$ such that

$$g \circ f = 1_A \quad \text{and} \quad f \circ g = 1_B \tag{2.43}$$

It can be shown that isomorphisms, if they exist, are unique for a domain/codomain pair, so we write $g = f^{-1}$. If there is an isomorphism $f : A \to B$, we say that A is isomorphic to B, and denote that by

$$A \cong B. \tag{2.44}$$

The isomorphisms work in category theory intuitively as a notion of "generalised equality". In **Sets** they correspond to bijective functions, and in ℝ**Linear** to vector space isomorphisms.

Definition 23 (Functor)**.** A functor F between categories $\boldsymbol{C}$ and $\boldsymbol{D}$ is a map that takes objects of $\boldsymbol{C}$ to objects of $\boldsymbol{D}$ and morphisms of $\boldsymbol{C}$ to morphisms $\boldsymbol{D}$ such that

- F preserves composition:
$$F(g \circ f) = F(g) \circ F(f) \tag{2.45}$$
- F preserves identities:
$$F(1_A) = 1_{F(A)} \tag{2.46}$$

One can see that a functor is a homomorphism of categories. That is, a functor $F : \boldsymbol{C} \to \boldsymbol{D}$ respects the morphism composition of $\boldsymbol{C}$ and gives what we can call intuitively a representation of $\boldsymbol{C}$ in terms of $\boldsymbol{D}$. Furthermore, functors can be composed in the intuitive way via function composition, so if $F : \boldsymbol{C} \to \boldsymbol{D}$ and $G : \boldsymbol{D} \to \boldsymbol{E}$ are functors, $G \circ F : \boldsymbol{C} \to \boldsymbol{E}$ is defined to be the functor that takes a morphism $f : A \to B$ of $\boldsymbol{C}$ to a morphism $G(F(f)) : G(F(A)) \to G(F(B))$ of $\boldsymbol{E}$.

Similarly to the way that functors preserve the structure of a category, we can define a transformation between functors that preserves their functoriality.

Definition 24 (Natural Transformation)**.** Given two functors $F : \boldsymbol{C} \to \boldsymbol{D}$ and $G : \boldsymbol{C} \to \boldsymbol{D}$, a natural transformation $\alpha : F \to G$ from F to G is a class of morphisms

$$\{\alpha_A : F(A) \to G(A)\}_{A \in C_0} \tag{2.47}$$

indexed by the objects of $\boldsymbol{C}$, such that for every morphism $f : A \to B$ of $\boldsymbol{C}$,

$$\alpha_B \circ F(f) = G(f) \circ \alpha_A. \tag{2.48}$$

Each of the morphisms α_A is called a component of the natural transformation α.

This definition might sound a bit counterintuitive, so for the sake of clarity, let's elaborate on it. Consider the following: first, it is clear that given natural transformations $\alpha : F \to G$ and $\beta : G \to H$, where F, G, H are functors from some category $\boldsymbol{C}$ to some $\boldsymbol{D}$, one can define the composite natural transformation $(\beta \circ \alpha) : F \to H$ by composing their components so that $(\beta \circ \alpha)_A = \beta_A \circ \alpha_A$. Second, since we are using the composition of morphisms in the definition, it is also clear that the composition of natural transformations inherits the associativity property. Finally, with the identity morphisms we can construct an identity natural transformation. Therefore, we can define the following:

Definition 25 (Functor Category)**.** Given two categories $\boldsymbol{C}$ and $\boldsymbol{D}$ the functor category $\mathrm{Func}(\boldsymbol{C}, \boldsymbol{D})$ is the category that has as objects the functors from $\boldsymbol{C}$ to $\boldsymbol{D}$ and as morphisms natural transformations between them.

Definition 26 (Natural Isomorphism)**.** A natural isomorphism is an isomorphism in a functor category.

It is worth pointing out that the definition of natural isomorphism in terms of the functor category is not the only one. For example, one can give a definition in terms of inverse operations. Here, we simply opt for the definition with the functor category. Now, we can use the notion of natural isomorphism to understand why the definition of natural transformation makes sense by looking at the following fact:

Proposition 2.2.1. A natural transformation $\alpha : F \to G$ is a natural isomorphism if and only if each of its components α_A is an isomorphism.

So, because of the notion of isomorphism between functors induced by the natural transformations, we will say that two functors F and G in a functor category are isomorphic when they map isomorphic objects of the source category to isomorphic objects of the target category. In particular, when there is a natural isomorphism $\alpha : F \to G$, then for all objects A of the source category, $F(A) \cong G(A)$. Hence, we can think intuitively that naturally isomorphic functors give representations of the source category in the target category that are equivalent up to isomorphism.

Having established that, we can understand now how to compare categories.

Definition 27 (Isomorphic Categories)**.** Two categories $\boldsymbol{C}$ and $\boldsymbol{D}$ are said to be isomorphic if there are two functors $F : \boldsymbol{C} \to \boldsymbol{D}$ and $G : \boldsymbol{D} \to \boldsymbol{C}$ such that

$$G \circ F = 1_{\boldsymbol{C}} \tag{2.49}$$

$$F \circ G = 1_{\boldsymbol{D}} \tag{2.50}$$

where $1_{\boldsymbol{C}}$ and $1_{\boldsymbol{D}}$ are the identity functors on $\boldsymbol{C}$ and $\boldsymbol{D}$, respectively.

If two categories are isomorphic, going from one to the other via the isomorphism is just renaming of its elements. Note that since we know how to compare functors as well, we can be even more general and relax the equalities in the definition above to be isomorphisms as well. That motivates the definition of equivalence of categories.

Definition 28 (Equivalent Categories)**.** Two categories $\boldsymbol{C}$ and $\boldsymbol{D}$ are said to be equivalent if there are two functors $F : \boldsymbol{C} \to \boldsymbol{D}$ and $G : \boldsymbol{D} \to \boldsymbol{C}$ and two natural isomorphisms $\alpha : (G \circ F) \to 1_{\boldsymbol{C}}$ and $\beta : (F \circ G) \to 1_{\boldsymbol{D}}$, in $\mathrm{Func}[\boldsymbol{C}, \boldsymbol{C}]$ and $\mathrm{Func}[\boldsymbol{D}, \boldsymbol{D}]$, respectively, so that

$$G \circ F \cong 1_{\boldsymbol{C}} \tag{2.51}$$

$$F \circ G \cong 1_{\boldsymbol{D}} \tag{2.52}$$

We denote $\boldsymbol{C}$ and $\boldsymbol{D}$ being equivalent by $\boldsymbol{C} \simeq \boldsymbol{D}$.

With this notion, we can say that categories are equivalent not only when they differ by the names of their elements, but we allow them for having different numbers of copies of isomorphic objects, which makes sense as a less strict notion of equivalence than isomorphism. We can understand this by saying that equivalent categories are isomorphic up to isomorphism. This notion will play a role when we connect monoidal categories to strict

monoidal categories. The usefulness of this concept to us will be apparent when we see that through the process of *strictification* we can perform all of our calculations within a simpler category, and translate our results back to our original, more complicated category.

Next, we have a way to combine two categories into another one.

Definition 29 (Product Category)**.** If $\boldsymbol{C}$ and $\boldsymbol{D}$ are categories, the product category $\boldsymbol{C} \times \boldsymbol{D}$ is defined to have

- Pairs (C, D) of objects from $\boldsymbol{C}$ and $\boldsymbol{D}$ as its objects
- Pairs of morphisms (f, g) with $f : A \to B$ and $g : C \to D$ as morphisms from (A, C) to (B, D).

Moreover, the morphisms of $\boldsymbol{C} \times \boldsymbol{D}$ compose entry-wise:

$$(f, g) \circ (f', g') = (f \circ f', g \circ g') \tag{2.53}$$

whenever $f \circ f'$ and $g \circ g'$ are defined, that is, we have $\mathrm{cod}((f, g)) = \mathrm{dom}((f', g'))$.

We can use the previous definitions to see that there is a sense in which the product of categories produces naturally isomorphic categories. First, we can define a functor

$$\tau : (\boldsymbol{C} \times \boldsymbol{C}) \times \boldsymbol{C} \to \boldsymbol{C} \times (\boldsymbol{C} \times \boldsymbol{C}) \tag{2.54}$$

that takes an object $((A, B), C)$ to an object $(A, (B, C))$ and similarly morphisms $((f, g), h)$ to $(f, (g, h))$. Clearly, this functor has an inverse $\tau^{-1} : \boldsymbol{C} \times (\boldsymbol{C} \times \boldsymbol{C}) \to (\boldsymbol{C} \times \boldsymbol{C}) \times \boldsymbol{C}$. So, we can say that the products of categories with different bracketing orders are isomorphic.

2.2.3 Monoidal Categories

To get one step closer of the definition of strict symmetric monoidal categories, we can now give the definition of monoidal category.

Definition 30 (Monoidal Category)**.** A monoidal category $\boldsymbol{C}$ is a category equipped with

- a functor
$$\otimes : \boldsymbol{C} \times \boldsymbol{C} \to \boldsymbol{C} \tag{2.55}$$
from the product category of $\boldsymbol{C}$ with itself, to itself. We denote
$$\otimes(x, y) \equiv x \otimes y. \tag{2.56}$$
- an object
$$I \in C_0 \tag{2.57}$$
that we call the monoidal identity, or tensor identity.

- a natural isomorphism

$$\alpha : \otimes \circ (\otimes \times 1_{\boldsymbol{C}}) \to \otimes \circ (1_{\boldsymbol{C}} \times \otimes) \circ \tau \tag{2.58}$$

  called the associator, with components $\alpha_{A,B,C} : (A \otimes B) \otimes C \to A \otimes (B \otimes C)$, where $1_{\boldsymbol{C}}$ is the identity functor on $\boldsymbol{C}$.

- natural isomorphisms

$$\lambda : \otimes(I, -) \to 1_{\boldsymbol{C}} \tag{2.59}$$

$$\rho : \otimes(-, I) \to 1_{\boldsymbol{C}} \tag{2.60}$$

  respecitvely called the left and right unitors, with components $\lambda_A : I \otimes A \to A$ and $\rho_A : A \otimes I \to A$.

Additionally, these elements are required to be such that the following diagrams commute:

$$\begin{array}{ccc} (A \otimes I) \otimes B & \xrightarrow{\alpha_{A,I,B}} & A \otimes (I \otimes B) \\ & {\scriptstyle \rho_A \otimes 1_B} \searrow \quad \swarrow {\scriptstyle 1_A \otimes \lambda_B} & \\ & A \otimes B & \end{array} \tag{2.61}$$

$$\begin{array}{ccccc} & & (A \otimes B) \otimes (C \otimes D) & & \\ & {\scriptstyle \alpha_{A\otimes B,C,D}} \nearrow & & \searrow {\scriptstyle \alpha_{A\otimes B,C,D}} & \\ ((A \otimes B) \otimes C) \otimes D & & & & (A \otimes (B \otimes (C \otimes D))) \\ {\scriptstyle 1_A \otimes \alpha_{B,C,D}} \downarrow & & & & \uparrow {\scriptstyle 1_A \otimes \alpha_{B,C,D}} \\ (A \otimes (B \otimes C)) \otimes D & & \xrightarrow{\alpha_{A,B\otimes C,D}} & & A \otimes ((B \otimes C)) \otimes D) \end{array} \tag{2.62}$$

Notice that the monoidal product captures the idea that pairings of objects/morphisms of $\boldsymbol{C}$, i.e. objects/morphisms from the product category $\boldsymbol{C} \times \boldsymbol{C}$, are associated with objects/morphisms of the original category $\boldsymbol{C}$. This notion mirrors what we have in physics when we say that when we consider two systems together, that is the same as considering a single, higher dimensional, system which we call composite. In this analogy, this association between pairings of systems with systems is modeled by the monoidal functor.

At this point, the fact that we are requiring that different bracketings of objects such as $(A \otimes B) \otimes C$ and $A \otimes (B \otimes C)$ are just isomorphic, not necessarily equal, can lead to very messy calculations. Luckily, we will be able to simplify calculatios using the notion of equivalence of categories. Consider the following:

Definition 31 (Strict Monoidal Category)**.** A strict monoidal category $\boldsymbol{C}$ is a monoidal category where the associator and unitor natural isomorphisms are identity natural isomorphisms. That is, in $\boldsymbol{C}$,

$$A \otimes (B \otimes C) = (A \otimes B) \otimes C \tag{2.63}$$

$$f \otimes (g \otimes g) = (f \otimes g) \otimes h \tag{2.64}$$

$$I \otimes A = A \otimes I = A \tag{2.65}$$

$$1_I \otimes f = f \otimes 1_I = f \tag{2.66}$$

for any objects A, B, C and morphisms f, g, h.

The strict monoidal categories can be used to simplify our calculations in monoidal categories via a notion of equivalence. It is clear, though, that monoidal categories have an extra structure as compared to general categories, namely, the tensor product. Here we can introduce a notion of equivalence more specific to monoidal categories.

Definition 32 (Monoidal Functor)**.** A monoidal functor F between monoidal categories $\boldsymbol{C}$ and $\boldsymbol{D}$ is a functor $F : \boldsymbol{C} \to \boldsymbol{D}$ together with coherence maps

- A morphism of $\boldsymbol{D}$

$$\epsilon : I_{\boldsymbol{D}} \to F(I_{\boldsymbol{C}}) \tag{2.67}$$

- A natural transformation

$$\mu : \otimes_{\boldsymbol{D}} \circ (F \times F) \to F \circ \otimes_{\boldsymbol{C}} \tag{2.68}$$

These are additionally required to be such that the following diagrams commute:

$$\begin{array}{ccc}
(F(A) \otimes_{\boldsymbol{D}} F(B)) \otimes_{\boldsymbol{D}} F(C) & \xrightarrow{\alpha_{\boldsymbol{D}}} & F(A) \otimes_{\boldsymbol{D}} (F(B) \otimes_{\boldsymbol{D}} F(C)) \\
\downarrow{\scriptstyle \mu_{A,B} \otimes_{\boldsymbol{D}} 1_{F(C)}} & & \downarrow{\scriptstyle 1_{F(A)} \otimes_{\boldsymbol{D}} \mu_{B,C}} \\
F(A \otimes_{\boldsymbol{C}} B) \otimes_{\boldsymbol{D}} F(C) & & F(A) \otimes_{\boldsymbol{D}} F(B \otimes_{\boldsymbol{C}} C) \\
\downarrow{\scriptstyle \mu_{A \otimes_{\boldsymbol{C}} B, C}} & & \downarrow{\scriptstyle \mu_{A, B \otimes_{\boldsymbol{C}} C}} \\
F((A \otimes_{\boldsymbol{C}} B) \otimes_{\boldsymbol{C}} C) & \xrightarrow[F(\alpha_{\boldsymbol{C}})]{} & F(A \otimes_{\boldsymbol{C}} (B \otimes_{\boldsymbol{C}} C))
\end{array} \tag{2.69}$$

$$\begin{array}{ccc}
F(A) \otimes_{\boldsymbol{D}} I_{\boldsymbol{D}} & \xrightarrow{1_{F(A)} \otimes_{\boldsymbol{D}} \epsilon} & F(A) \otimes_{\boldsymbol{D}} F(I_{\boldsymbol{C}}) \\
\downarrow{\scriptstyle \rho_{\boldsymbol{D}}} & & \downarrow{\scriptstyle \mu_{A, I_{\boldsymbol{C}}}} \\
F(A) & \xrightarrow[F(\rho_{\boldsymbol{C}})]{} & F(A \otimes_{\boldsymbol{C}} I_{\boldsymbol{C}})
\end{array} \quad \text{and} \quad
\begin{array}{ccc}
I_{\boldsymbol{D}} \otimes_{\boldsymbol{D}} F(A) & \xrightarrow{1_{F(A)} \otimes_{\boldsymbol{D}} \epsilon} & F(I_{\boldsymbol{C}}) \otimes_{\boldsymbol{D}} F(A) \\
\downarrow{\scriptstyle \lambda_{\boldsymbol{D}}} & & \downarrow{\scriptstyle \mu_{I_{\boldsymbol{C}}, A}} \\
F(A) & \xrightarrow[F(\lambda_{\boldsymbol{C}})]{} & F(I_{\boldsymbol{C}} \otimes_{\boldsymbol{C}} A)
\end{array} \tag{2.70}$$

where $I_{\boldsymbol{C}}, I_{\boldsymbol{D}}, \alpha_{\boldsymbol{C}}, \alpha_{\boldsymbol{D}}, \rho_{\boldsymbol{C}}, \rho_{\boldsymbol{D}}, \lambda_{\boldsymbol{C}}, \lambda_{\boldsymbol{D}}, \otimes_{\boldsymbol{C}}, \otimes_{\boldsymbol{D}}$ correspond to the monoidal structures of $\boldsymbol{C}$ and $\boldsymbol{D}$.

Note that in this definition, we require the monoidal functor to preserve not only the composition of the category, but also the monoidal product. Here, the first diagram requires that it preserves the associativity of the monoidal product and the second and third diagram requires preservation of the monoidal identity.

Definition 33 (Monoidally Equivalent Categories)**.** Two monoidal categories $\boldsymbol{C}$ and $\boldsymbol{D}$ are said to be monoidally equivalent if there exists monoidal functors $F : \boldsymbol{C} \to \boldsymbol{D}$ and $G : \boldsymbol{D} \to \boldsymbol{C}$ such that

$$G \circ F \cong 1_{\boldsymbol{C}} \tag{2.71}$$

$$F \circ G \cong 1_{\boldsymbol{D}} \tag{2.72}$$

Theorem 2.2.2 (Strictification)**.** Every monoidal category is monoidally equivalent to a strict monoidal category.

This means that we can always work with the strict monoidal category where the equalities hold, and be sure that our results translate to results in the monoidal category for any choice of corresponding isomorphic objects. That is, we if say something for $A \otimes B \otimes C$, then we can take that as true for $A \otimes (B \otimes C)$ in the monoidal category. We will see that there is something similar for symmetric monoidal categories.

2.2.4 Symmetric Monoidal Categories

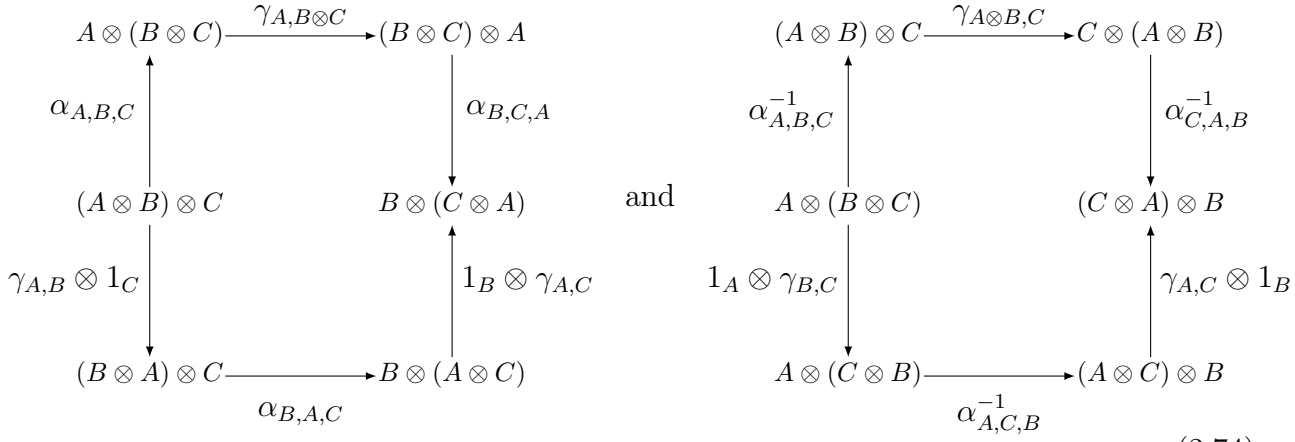
We know how to use categories to model the idea that pairings of systems/processes are systems/processes themselves through the monoidal product in monoidal categories. Another ingredient we need is to model that swaping the order of pairings yields isomorphic systems. A formalization of that idea we can find in symmetric monoidal categories, which are a particular case of braided monoidal categories. Hence we start this section with the following definition, requiring that different orders of pairings be isomorphic:

Definition 34 (Braided Monoidal Category)**.** A braided monoidal category $\boldsymbol{C}$ is a monoidal category equipped with a natural isomorphism

$$\gamma : \otimes \to \overline{\otimes} \tag{2.73}$$

where $\overline{\otimes}(x, y) \equiv \otimes(y, x)$. Morevoer, γ is required be such that the following diagrams

commute:



(2.74)

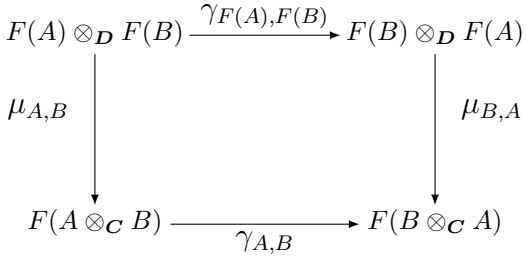
The braiding provides therefore isomorphisms between pairings with different orders, so $A \otimes B \cong B \otimes A$, but if we use it to "switch" the orders again won't necessarily get back the same $A \otimes B$. This is because as per the definition, the braidings don't need to be idempontent in any sense, hence the name braiding. For this reason, to really have something like a swap of the order $A \otimes B$ to $B \otimes A$, we need a stronger requirement.

Definition 35 (Symmetric Monoidal Category)**.** A symmetric monoidal category $\boldsymbol{C}$ is a braided monoidal category such that the braiding natural isomorphism γ satisfies

$$\gamma_{B,A} \circ \gamma_{A,B} = 1_{A\otimes B} \tag{2.75}$$

In a symmetric monoidal category we can therefore use the braiding to change the order of a monoidal product back and forth, recovering the same object rather than some object merely isomorphic to it. Now, we can rely on a strictification theorem similar to what we have with monoidal categories.

Definition 36 (Braided Monoidal Functor)**.** A braided monoidal functor $F : \boldsymbol{C} \to \boldsymbol{D}$ between braided monoidal categories $\boldsymbol{C}$ and $\boldsymbol{D}$ is a monoidal functor such that



(2.76)

commutes.

So a braided monoidal functor is a monoidal functor that also respects the braiding structure of the source category.

Definition 37 (Symmetric Monoidal Functor)**.** A symmetric monoidal functor is a braided monoidal functor between symmetric monoidal categories.

Definition 38 (Symmetric-Monoidaly Equivalent Categories)**.** Two symmetric monoidal categories $\boldsymbol{C}$ and $\boldsymbol{D}$ are symmetric-monoidally equivalent if there are symmetric monoidal functors $F : \boldsymbol{C} \to \boldsymbol{D}$ and $G : \boldsymbol{D} \to \boldsymbol{C}$ such that

$$G \circ F \cong 1_{\boldsymbol{C}} \tag{2.77}$$

$$F \circ G \cong 1_{\boldsymbol{D}} \tag{2.78}$$

Finally we can give the definition of the strict category which we use to perform our calculations and state equivelence theorem.

Definition 39 (Symmetric Strict Monoidal Category)**.** A symmetric strict monoidal category is a symmetric monoidal category whose associator and unitor natural isomorphisms are identity natural isomorphisms.

Definition 40 (Strictification of Symmetric Monoidal Categories)**.** Every symmetric monoidal category is symmetric-monoidally equivalent to a symmetric strict monoidal category.

This gives us what we need. We have at our disposal a formalization of the association of pairings of objects/morphisms and objects/morphisms and, with rebracketing taken as an associative operation and reordering as a commutiative operation, both up to isomorphisms. Moreover, we can perform our calculations in a simpler version of that formalization where the rebracketings are trivial, so we don't need to care about them. We can now discuss the notation we use to perform these simpler calculations.

2.2.5 Diagrammatic Calculus

One of the most useful property to us of the symmetric strict monoidal categories is that they have available a diagrammatic calculus that is very expressive, specially under our intended interpretation of generalised probabilistic theories. In what follows, we will first specify how to denote diagrammatically the general objects and morphisms, and how their compositions behave in this donation. Next, we deal with the monoidal unit and then with the braiding of the symmetric monoidal category.

Notice that since in a symmetric strict monoidal cateogry the associativity of the monoidal product is given with an equality, we can simply write

$$(A \otimes B) \otimes C = A \otimes (B \otimes C) = A \otimes B \otimes C, \tag{2.79}$$

that is, we don't need to use the brackets at all. Now, since the monoidal product is simply given by juxtaposition of the objects, we can represent it as any sequence of things each of which represents an object. In the diagrammatic notation we do so using labelled wires, so

$$A \doteq \Big|_{A} \tag{2.80}$$

and a product $A_1 \otimes ... \otimes A_n$ is represented by

$$A_1 \otimes ... \otimes A_n \doteq \Big|_{A_1} \quad ... \quad \Big|_{A_n} \tag{2.81}$$

Next, we need to represent the morphisms. Since in the categorical operations what is important to represent is the domain and codomain of the morphisms, so as to correctly write the compositions, we only need to give to the representation of a morphism a label, with its name, and a distinction between its source and target. This is done using boxes with wires connected to them, where wires at the bottom are taken as the domain, and wires at the top as the codomain of the morphism:

$$f : A \to B \doteq \begin{array}{c} \Big|_{B} \\ \boxed{f} \\ \Big|_{A} \end{array} . \tag{2.82}$$

Since we already established that the monoidal product of objects is given by wires side by side, we denote

$$f : A_1 \otimes ... \otimes A_n \to B_1 \otimes ...B_m \doteq \begin{array}{c} B_1 \Big| ... \Big| B_m \\ \boxed{f} \\ A_1 \Big| ... \Big| A_n \end{array} \tag{2.83}$$

In the same spirit as with the objects, we can represent the monoidal product of morphisms by drawing them side by side:

$$\begin{array}{c} B \Big| \quad \Big| C \\ \boxed{f \otimes g} \\ A \Big| \quad \Big| C \end{array} = \begin{array}{c} B \Big| \quad \Big| D \\ \boxed{f} \quad \boxed{g} \\ A \Big| \quad \Big| C \end{array} . \tag{2.84}$$

Next, we have to specify how compositions are denoted. Since the category definition requires matching of codomain and domain, we can enforce that by allowing compositions connecting matching wires. So, for $f : A \to B$ and $g : B \to C$, we have

$$\begin{array}{c} \Big|_{C} \\ \boxed{g \circ f} \\ \Big|_{A} \end{array} = \begin{array}{c} \Big|_{C} \\ \boxed{g} \\ \Big|_{B} \\ \boxed{f} \\ \Big|_{A} \end{array} \tag{2.85}$$

and, for $f : A_1 \otimes ... \otimes A_n \to B_1 \otimes B_m$ and $g : B_1...B_n \to C_1 \otimes C_p$, we have

$$g \circ f \;\; \begin{array}{c} C_1 \,|\, \ldots \,|\, C_p \\ \boxed{g \circ f} \\ A_1 \,|\, \ldots \,|\, A_n \end{array} \;\; = \;\; \begin{array}{c} C_1 \,|\, \ldots \,|\, C_p \\ \boxed{g} \\ B_1 \,|\, \ldots \,|\, B_m \\ \boxed{f} \\ A_1 \,|\, \ldots \,|\, A_n \end{array} . \tag{2.86}$$

Now, let's see how to represent the monoidal identity. First, since $A \otimes I = I \otimes A = A$, and $\otimes$ is being represented by juxtaposition, we can take the wire representing I to be the same as the space around the diagrams:

$$I \otimes A = A \doteq \Big|_{A} \tag{2.87}$$

so

$$I \doteq \quad \tag{2.88}$$

Immediately, that means that processes with I as their domain or codomain will have no bottom/top wire. For instance, for $s : I \to A$ and $e : A \to I$ we have

$$s \doteq \begin{array}{c} | A \\ \triangledown \\ s \end{array} \qquad \text{and } e \doteq \begin{array}{c} e \\ \triangle \\ | A \end{array} \tag{2.89}$$

We can also consider the representation of the identity morphisms, whose behavior is analogous to that of the monoidal identity. Since we represent the $\circ$ composition by the connection of wires, and, for and $f : A \to B$ we have $1_B \circ f = f \circ 1_A = f$, we can take 1_A to be present anywhere in the wire that reprents A:

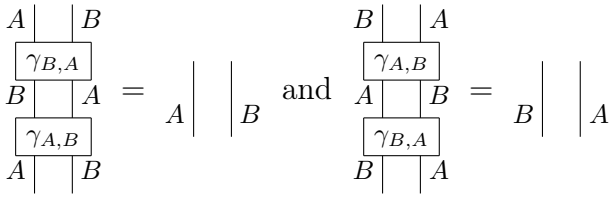
$$1_B \circ f = f \circ 1_A = f \doteq \begin{array}{c} | B \\ \boxed{f} \\ | A \end{array} \tag{2.90}$$

so

$$1_A \doteq \Big|_{A} \tag{2.91}$$

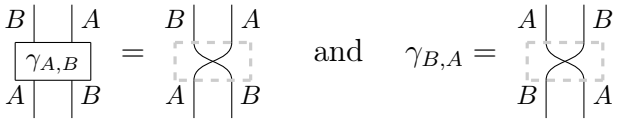
where we use the dotted lines to emphasize the empty space, but normally nothing would be there.

With this, we can represent the monoidal structure of the category and the compositions. It remains to discuss how the braiding of the symmetric monoidal category is represented in the diagrammatic notation. Note that we can use the components of the braiding natural isomorphism to go back and forth from $A \to B$ and from $B \to C$. In diagrams we would have then



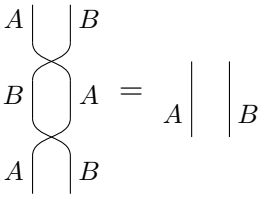
(2.92)

This behavior is very similar to a simple swap of wires, so we can have the following very vivid diagrammatic representations:



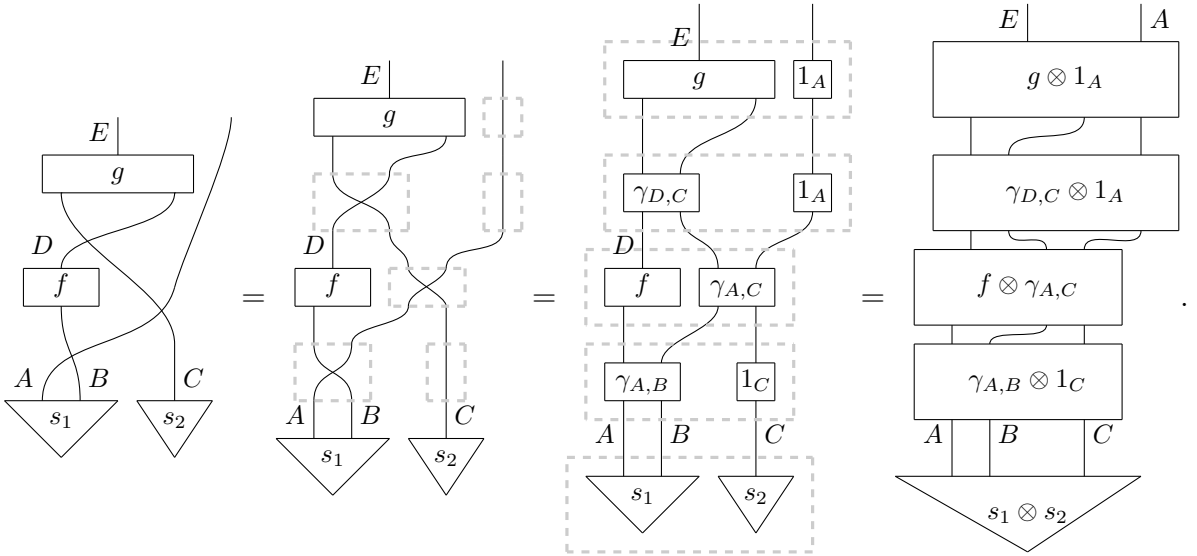
(2.93)

Now, composing one with the other appears in the diagrams as simply



(2.94)

Even more nicely, in general in any diagram, the swaps allow us to deform the wires to any order very much as if they were physical wires (except they don't braid!). One example of all of those properties together can be seen if we take a more complex diagram and interpret it back as a morphism of the category.



$$= (g \otimes 1_A) \circ (\gamma_{D,C} \otimes 1_A) \circ (f \otimes \gamma_{A,C}) \circ (\gamma_{A,B} \otimes 1_C) \circ (s_1 \otimes s_2) \tag{2.95}$$

The only rule we have to keep in mind, therefore, is that any deformation that doesn't change a domain wire to a codomain wire is allowed, and wire connections are allowed as long as the labels match. This is guaranteed by the following theorem:

Theorem 2.2.3. An equation between morphisms in a symmetric monoidal category follows from the axioms if and only if it holds in the diagrammatic calculus.

Some categories though, namely the rigid symmetric monoidal categories, or compact closed categories, have even more nice properties that allow us to bend wires up and down. Nevertheless, for the purposes of this thesis we are interested in the more general case were we can't because this is the mathematical structure that we will take as the abstraction capturing the idea of generalised probabilistic theory.

2.3 Generalised Probabilistic Theories

In this section I describe what generalised probabilistic theories (GPTs) are and show how they naturally accommodate the signalling and locality considerations from the previous sections.

We take generalised probabilistic theories as mathematical abstractions that model theories about experiments that provide probabilistic predictions for given scenarios. We will see that the process theory framework very directly encodes intuitive compositional primitives, and later talk in terms of the operational probabilistic theory framework, which provides a vocabulary more directly connected to probabilistic considerations. In both cases, we will see that the structure of interest is the symmetric monoidal category.

2.3.1 Generalised Probabilistic Theories

Conceptually speaking, the generalised probabilistic theory framework is an abstraction of the idea of theories about experiments that give probabilistic predictions for given procedures. More specifically we will use the category theory framework as a model of systems and procedures, such that given some procedures we can combine them into more complex ones.

In what follows, we will rely on the process theory framework [38, 58] to build some intuition about how to interprete categories in terms of processes, and later define the operational probabilistic theory framework which provides us with a very clear language to talk about GPTs.

The first step for us is to look at the analogy between the compositional structure of the symmetric monoidal categories and the one naturally found in the way we think about processes, so that we give to each of the elements of the former an interpretation in terms of elements of the latter. This parallel is formalized in the process theory framework, which can provide a basis for our generalised probabilistic theories.

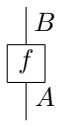
Definition 41 (Process Theory)**.** A process theory is a symmetric strict monoidal category under the following interpretation:

- The objects $A, B, C, ...$ are called system types, or simply systems.
- The monoidal identity I is the trivial system, i.e. no system.
- The morphisms $f : A \to B$ are processes that take an input system of type A to an output system of type B.
- The categorical composition $g \circ f : A \to C$ of processes $f : A \to B$ and $g : B \to C$ is the sequential composition of processes, i.e. the process given by the application of g to the output of f.
- The monoidal product $f \otimes g : A \otimes C \to B \otimes D$ of processes $f : A \to B$ and $g : C \to D$ is the parallel composition of the processes f and g, i.e., the processes f and g occurring independently taken as a single process.

This very general definition has a direct physical motivation, and lends itself to modelling generalised probabilistic theories in a simple way. In the context of experimental setups, we talk about systems of different nature as different system types. For example, in quantum information, a qubit can be seen as a system of type different than that of a qutrit, and therefore processes applicable to one do not apply to the other, which in the quantum case is encoded in the difference between the vector spaces that model them. For the processes, we note that the fact that we can talk about them without talking about specific input states, shows that we should be able to have an abstraction for processes in terms only of their label and the type of their input and output systems. For instance, we can talk about a rotation without talking about what is the initial direction of the rotated object. Next, a state can be seen a preparation process, that starts from a trivial system and produces a system of a certain type. The effects are destructive observation processes where one observes a system of a certain type and ends up with the trivial system. Finally, if we want to make probabilistic predictions, we can take the processes from the trivial system to itself to be numbers in the $[0, 1]$ interval, as they can be seen as prepare-transform-measure procedures. If a certain diagram is a probability, then that is the predicted probability for the entire process represented by it to happen.

Since the process theories are base on symmetric monoidal categories, we know that we can always reason in terms of the diagrammatic calculus available for those categories. Hence, we keep in mind the following dictionary:

Table 2.1: Diagrammatic representation of GPT elements

Systems	Processes
	

These diagrams can be combined normally as per the diagrammatic calculus rules, to form multipartite systems, states and effects, as well as sequential and parallel composition of transformations, states and effects.

Having in mind the interpretation above, we can introduce the operational probabilistic theory (OPT) framework [42] approach directly. Here, rather than thinking of the category theory pieces as general processes, we will specialize to interpreting them as primitives for experimental considerations. The most primitive notions for us will be those of systems, tests and events.

Definition 42 (System)**.** System types, or simply systems, are primitive labels that specify the kinds of systems on which tests/events apply. Diagrammatically we represent them by wires:

$$A \doteq \Big|_{A} \tag{2.96}$$

if A is a system type.

Definition 43 (Trivial System)**.** We denote by I the trivial system, which, in this formalism, denotes the absense of a system. Diagrammatically this is the empty space around a diagram:

$$I \doteq \quad \tag{2.97}$$

Definition 44 (Composite System)**.** The system we obtain by considering two given two systems A and B together as one is the composite system $A \otimes B$. The operation $\otimes$ is associative and has the trivial system I as a left and right unit. Moreover, $A \otimes B$ and $B \otimes A$ are operationally equivalent as per Def. 52. Diagrammatically, we represent $A \otimes B$ by the juxtaposition of their wires:

$$A \otimes B \doteq \Big|_{A}\Big|_{B} \qquad \text{and} \qquad I \otimes A = A \otimes I \doteq \Big|_{A} \tag{2.98}$$

Definition 45 (Tests and Events)**.** A test from an input system A to an output system B is a labelled set of events $\{E_x\}_{x\in X}$ over some set of outcomes X. In diagrammatic notation,

$$\{E_x\}_{x\in X} \doteq \begin{array}{c} \Big|_{B} \\ \boxed{\{E_x\}_{x\in X}} \\ \Big|_{A} \end{array} \quad \text{and} \quad E_x \doteq \begin{array}{c} \Big|_{B} \\ \boxed{E_x} \\ \Big|_{A} \end{array} \tag{2.99}$$

If f is a test/event, we denote $f : A \to B$ to indicate its input and output systems.

Definition 46 (Preparation)**.** A preparation event, or simply preparation, is an event with the trivial system I as input, and a nontrivial system as output. In diagrammatic notation,

$$P : I \to A \doteq \begin{array}{c} \Big|_{A} \\ \bigtriangledown \\ P \end{array}. \tag{2.100}$$

Definition 47 (Observation)**.** An observation event, or simply observation, is an event with output in the trivial system I, and input in a nontrivial system. In diagrammatic notation,

$$O : A \to I \doteq \tag{2.101}$$

If test is composed by preparation/observation events, we can call it a preparation/observation test.

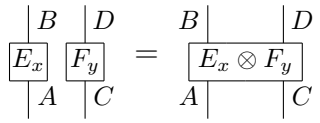
Given those definitions, we can define deterministic tests even before we talk about probabilities.

Definition 48 (Deterministic Test)**.** A test $\{E_x\}_{x\in X}$ is deterministic if it contains only one event, that is $|X| = 1$.

This notion will be important when we talk about the discarding operation.

Definition 49 (Parallel Composition of Tests)**.** Given two tests $\{E_x\}_{x\in X} : A \to B$ and $\{F_y\}_{y\in Y} : C \to D$, their parallel composition is the test $\{E_x \otimes F_y\}_{(x,y)\in X\times Y}$ from $A \otimes C$ to $B \otimes D$. The parallel composition is associative. Diagrammatically we represent it by the juxtaposition of diagrams, so the events of the resulting test are

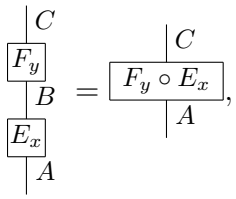
$$E_x \otimes F_y \doteq \tag{2.102}$$



and the respective tests compose the same way.

Definition 50 (Sequential Composition of Tests)**.** Given two tests of matching output and input, $\{E_x\}_{x\in X} : A \to B$ and $\{F_y\}_{y\in Y} : B \to C$, their sequential composition is the test $\{F_y \circ E_x\}_{(x,y)\in X\times Y}$. The sequential composition is associative. Diagrammatically we represent it by the connection of wires, so the events of the resulting test are

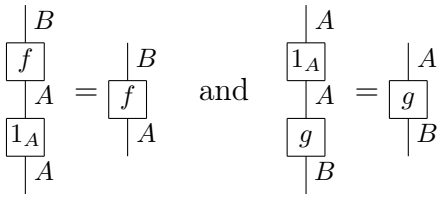
$$F_y \circ E_x \doteq \tag{2.103}$$



and the respective tests compose the same way.

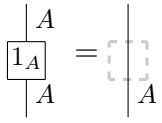
Note that by denoting the events of the compositions by $F_y \circ E_x$ and $E_x \otimes F_y$ we already hint the categorical structure underlying our compositions.

Definition 51 (Identity Test)**.** The identity test on a system A is a test with a single event 1_A such that for all events $f : A \to B$ and $g : C \to A$,



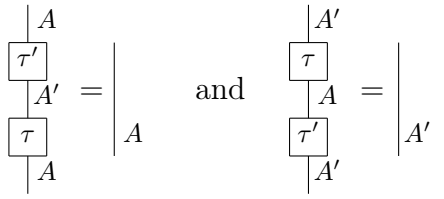
(2.104)

We can represent the identity test by a wire without a box:



(2.105)

Definition 52 (Operationally Equivalent Systems)**.** We say that two systems A and A' are operationally equivalent, denoted $A \cong B$, if there are two deterministic tests, $\{\tau\} : A \to A'$ and $\{\tau'\} : A' \to A$ that are inverses of each other. Diagrammaticaly, that means



(2.106)

Notice that when these elements form a category, as it does for operational theories, this correspond to the objects of the category that are associated to A and A' being isomorphic.

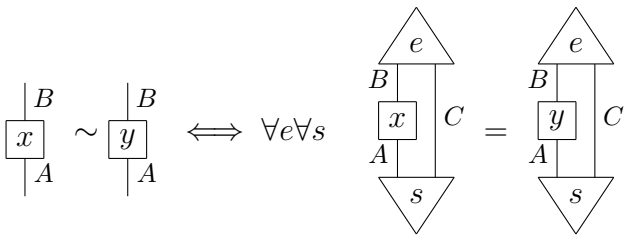
Definition 53 (Operational Theory)**.** An operational theory is a collection of systems that is closed under composition $\otimes$ with the trivial system I as identity, and a collection of tests between those systems that is closed under sequential, $\circ$, and parallel compositions, $\otimes$, having the identity tests 1_A as identities for $\circ$, and 1_I as the identity for $\otimes$.

Note that this definition means that both the tests and the events form symmetric strict monoidal categories. Hence the usage of the diagrammatic notation from the beginning. Now, going towards the definition of generalised probabilistic theory we have to introduce the elements related to the the probabilistic structure of the theory.

Definition 54 (Operational Probabilistic Theory)**.** An operational probabilistic theory (OPT) is an operational theory such that (i) all of its events $p : I \to I$, from the trivial system I to itself, are probabilities, (ii) for any test $\{p_x\}_{x\in X} : I \to I$ satisfies $\sum_{x\in X} p_x = 1$, and (iii) both the sequential and parallel composition of events $I \to I$ are given by their product as probabilities, i.e. $p_x \otimes p_y = p_x \circ p_y = p_x p_y$.

It is possible that two events are not distinguishable from the point of view of the probabilities that they can produce. Since we are interested in probabilistic descriptions, this means we might have a redundant description. This motivates the following definitions:

Definition 55 (Operationally Equivalent Events)**.** Two events $x : A \to B$ and $y : A \to B$ are operationally equivalent, denoted $x \sim y$, if changing one for the other in any composition that forms a probability event does not change that probability. Diagrammatically, we can write



(2.107)

This notion allows us to define the familiar notions of states, effects and transformations from our primitives.

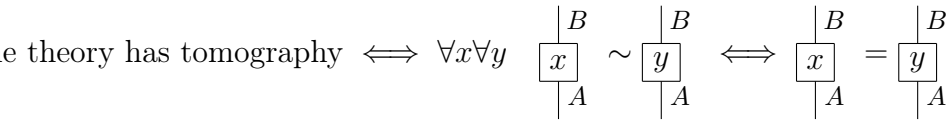
Definition 56 (States)**.** States are equivalence classes of preparation events under operational equivalence.

Definition 57 (Effects)**.** Effects are equivalence classes of observation events under operational equivalence.

Definition 58 (Transformations)**.** Transformations are equivalence classes of tests from $A \neq I$ to $B \neq I$ under operational equivalences.

It is easy to see that we can define compositions for these equivalence classes in terms of some representatives thereof and these compositions will be independent of the choice of representative. So, we can take the category formed by the states, effects, transformations and probabilities as our category of events. The OPT formed with those objects will then have a notion of tomography:

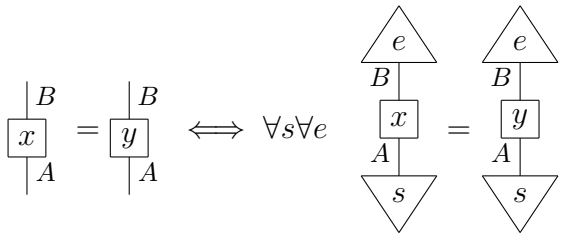
Definition 59 (Tomography)**.** A theory is said to have a notion of tomography if for any two events x and y, if x is operationally equivalent to y, then $x = y$. Diagrammatically:



(2.108)

To put it simply, tomography is a notion of equality in terms of probabilities. This notion will be important for our definition of GPT, more preciselly, we are interested in a strictier notion of tomography:

Definition 60 (Local Tomography)**.** A theory is said to have a notion of local tomography if it has a notion of tomography where the operational equivalence condition simplifies to the following:



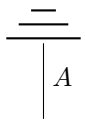
(2.109)

The next ingredient we need to talk about GPTs specifically is the notion of causality.

Definition 61 (Causality)**.** An OPT is said to be causal if for any preparation test with events s_x and any observation test with events e_y, the marginal probability $p_x = \sum_y e_y \circ s_x$ is independent of the set of events $\{e_y\}_{y\in Y}$ of the observation test.

It can be proven that a more directly applicable, though harder to interprete, definition of causality can be given. Namely, the definition above is satisfied by an OPT if and only if it satisfies the property in the definition bellow:

Definition 62 (Causality)**.** An OPT is said to be causal if for every system A there exists a unique deterministic effect u_A. In that case we can denote it diagrammatically as follows:



(2.110)

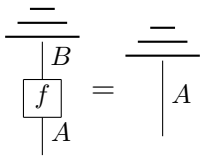
In this case we call the deterministic effect the discard operation.

In quantum theory the discarding effect is the (partial) trace operation. Note that to state this definition we are relying on the notion of determinism provided to tests, which is, in this framework, more primitive than the notion of probability. Now, to make more concrete the reason why this property is called causality, we can simply look at the following theorem, proven in Ref. [42]:

Theorem 2.3.1. In a theory that is causal, it is impossible to have signalling without exchanging systems.

If a theory has causality, then it also has many properties that we expect given our experience with, e.g. quantum theory. To list a few: (i) the discarding effect factorizes so that $u_{A\otimes B} = u_A \otimes u_B$, (ii) marginal states are uniquely defined, (iii) a simple notion of normalized state, that recovers the quantum version, can be defined (iii), a transformation is deterministic if and only if discarding after applying it is the same as discarding before. For those reasons, and the fact that our goal is to arrive at the definition of causal GPT, we will provide many notions for the particular case of causal OPTs only, which is the one of interest to us. These definitions will be given in terms of *the* discarding effect, but we can keep in mind that more general ones, e.g. of normalization, exist for more general OPTs.

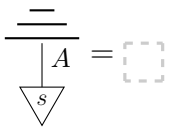
Definition 63 (Normalized Test)**.** A test is normalized if it is discarding preserving, i.e., applying it and then discarding the output is equivalent to discarding its input:



(2.111)

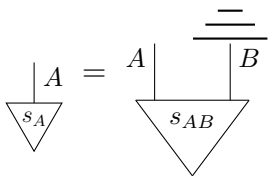
This definition has to important special cases:

Definition 64 (Normalized State)**.** A state s is normalized if discarding it results in the output of the deterministic effect with probability 1:



(2.112)

Definition 65 (Marginal State)**.** The marginal $s_A : I \to A$ of a state $s_{AB} : I \to A \otimes B$ is the result of discarding part of its output:



(2.113)

Definition 66 (Channel)**.** A transformation is a channel if it is discarding preserving.

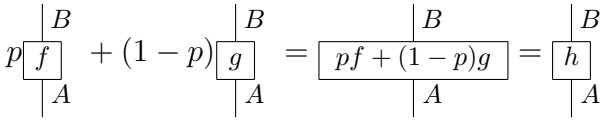
Next, we point that there is a natural way in which we can view all of the objects above as living in vector spaces. Consider the following: since the composition of a state $s : I \to A$ and an effect $e : A \to I$ is a probability $p \in [0, 1] \subset \mathbb{R}$, we can regard effects as functions from the set of states to real numbers, and, conversely, states as functions from the effects to real numbers. This gives a natural way to define linear combinations of those objects, and therefore to consider them as vectors in some vector space, in such a way the the space where the states live is dual to the one where the effects live. We can denote V_A for the vector space containing $s : I \to A$ and V_A^* for the one containing $e : A \to I$. Moreover, any event $f : A \to B$ induces a linear map if we take $f \circ (\sum_i a_i s_i) = \sum_i a_i (f \circ s_i)$ where $s_i \in V_B$, which is well defined [42]. This observation comes in handy as we can define and talk about OPTs using the familiar language of linear algebra, and connects this discussion with our previous knowledge in quantum theory and classical theory.

We can now give the first example of an operational theory: **Stoch**. This is the operational theory of classical stochastic maps, or, in other words, classical theory. This example is particularly important because it will be used in our definition of generalised probabilistic theory.

Definition 67 (**Stoch**)**.** The theory of classical stochastic maps, **Stoch**, is the OPT which has real vector spaces as system types and whose events are substochastic matrices. In particular, they are given as follows: (i) the states $s : I \to A$ are column vectors whose entries are non-negative, $s_i \geq 0$ and sum to $\sum_i s_i \leq 1$, (ii) effects $e : A \to I$ are row vectors whose entries are in the unit interval, $e_i \in [0, 1]$, (iii) transformations are substochastic matrices $T : A \to B$, (iv) sequential composition is matrix multiplication, and (v) parallel composition is the tensor/kronecker product.

Note that **Stoch** has additionally a convex structure: given the matrices x, y associated to any two events with the same input/ouput systems we can form a convex combination $px + (1-p)y$ that is itself an event of that type. This will be important for the definition of GPT.

Definition 68 (Convex Structure)**.** An OPT is said to have a convex structure if for any real number $p \in [0, 1]$ and two tests $f : A \to B$ $g : A \to B$ of arbitrary but matching input and output types, there exists $h : A \to B$ corresponding to the probabilistic mixture of f with probability p and g with probability $(1 - p)$. That is, for all such f, g, p:



(2.114)

With those definitions we are ready to give the definition of GPT as a particular case of OPT:

Definition 69 (Generalised Probabilistic Theory)**.** A generalised probabilistic theory (GPT) is an operational probabilistic theory that satisfies tomography, causality, is closed under convex combinations of all processes, and can be represented in a real vector space.

Looking from another angle, we can also say that a OPT quotiented by operational equivalence, that is, the OPT defined on states, effects and transformations (which we defined as equivalence classes) is a GPT, as it has been shown [43] that such an OPT satisfies the properties required above.

We can define now our main GPTs of interest.

Definition 70 (Classical Theory)**.** Classical theory is the GPT that coincides with **Stoch**. We can view its states as probability distributions and transformations as stochastic maps.

Definition 71 (Quantum Theory)**.** Quantum theory is the GPT whose system types are Hilber spaces $\mathcal{H}_d$, labelled by their dimension d. Its corresponding states live in the real vector spaces of Hermitian operators on $\mathcal{H}_d$ and are unit trace positive semidefinite operators. The effects are trace inner products with POVM elements, with the discarding effects being inner products with the identity matrix. The transformations are all completely positive trace-preserving linear maps on the states. Finally, sequential composition is given by the matrix product, and parallel composition by the tensor product.

Next, we want to define a GPT known as Boxworld. In order to do that more cleanly, we can introduce a few more concepts. Additionally, these concepts will prove themselves useful when we discuss some of the results later. What we will see now is how causal GPTs can be defined *geometrically* [77, 44] in terms of cones in vector spaces. That is in contrast to the compositional approach we took by starting with OPTs.

Consider that a practical way to visualize the state space of a qubit is to use the Bloch sphere. We can use the fact that after imposing $\mathrm{tr}(\rho) = 1$ the density matrix of the qubit has only 3 degrees of freedom to picture its state space as a sphere in a 3 dimensional space. In that case, we are representing only the normalized quantum states, that is, states ρ such that $\mathrm{tr}(\rho) = \mathrm{tr}(\mathbb{1}\rho) = \langle \mathbb{1}, \rho \rangle = 1$. Now, the qubit density matrices, seen as vectors in the space of Hermitian matrices, have 4 real components. So, if we give a step back, we can imagine the Bloch sphere living in a 3 dimensional hyperplane of the 4 dimensional real vector space of Hermitian matrices. If we now consider the vectors given by $\lambda\rho$ for $\lambda \geq 0$, we effectivelly dropped the normalization of the quantum states while keeping their positivity. Crucially, note that the set of all such $\lambda\rho$ defines a *cone* in the 4 dimensional vector space, the cone of Hermitian matrices ρ such that $\rho \geq 0$. The picture we would like to generalize to more GPTs is the following: the set of quantum states is defined by the intersection between the *cone of positive vectors* and the hyperplane of matrices whose inner product with the discarding effect, the identity $\mathbb{1}$, equals 1. This works because the set of normalized quantum states is convex: given two normalized states ρ, ρ' and a probability p, $p\rho + (1-p)\rho'$ is also a normalized state, so we can't produce a hollow cone by dropping only the normalization.

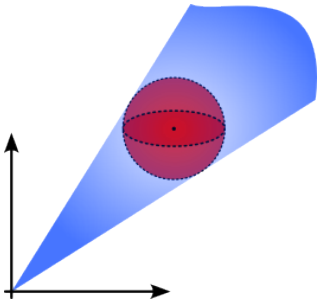
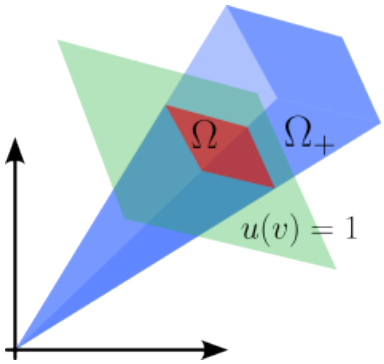


Figure 2.1: Pictorial representation of the cone generated by the Bloch sphere in the real vector space of Hermitian matrices.

With this view in mind, we could alternatively define quantum theory by saying what the *positive cones* of their states spaces are, and what the discarding vector (unique by virtue of the causality property) defining the hyperplane of normalized vectors is. With this, we can recover all normalized states of a given system type. Then, for the transformations we need to consider which of the linear functions that take cones to cones are allowed as transformations, or which positive transformation are allowed. Finally, for the combination of systems we need to specify which cone is the result of the combination of any two given cones.

We now extend these considerations to other GPTs, specifically the ones that, like quan-

Figure 2.2: Pictorial representation of the state space Ω as the intersection between the positive cone Ω_+ and the hyperplane of normalized vectors $\{v \in V : u(v) = 1\}$.



tum and classical theories, satisfy local tomography. Starting from the fact that OPTs satisfying local tomography always have representations as subtheories of $\mathbb{R}$**Linear**, we give a few definitions.

Definition 72 (Positive Cone)**.** In a vector space V, a cone generated by a convex set S is denoted S_+ and defined as follows

$$S_+ = \{\lambda s : s \in S \text{ and } \lambda \geq 0\} \tag{2.115}$$

Definition 73 (Positive Vector)**.** In the vector space V^A containing the set of states Ω^A of a GPT system type A, a vector $v \in V^A$ is said to be positive, denote $v \geq 0$, if $v \in \Omega^A_+$.

Since causal locally tomographic GPTs also have a unique discarding effect u^A for each system type A, their states spaces can be defined by giving the positive cones Ω^A_+ and their discarding effects u^A so that

$$\Omega^A = \{v \in V^A : v \geq 0 \text{ and } u^A(v) = 1\} \tag{2.116}$$

where $v \geq 0 \iff v \in \Omega^A_+$.

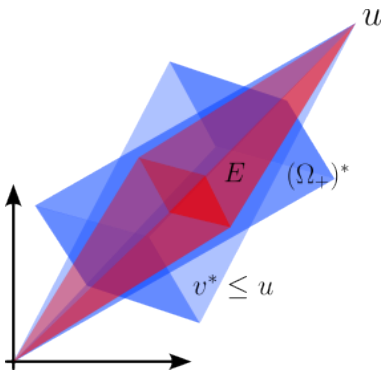
That is in direct parallel with quantum theory, and it works because we defined the cones to be generated by the states. This choice is made for convenience, by making the parallel with quantum theory stronger, as we could have instead defined cones first and required the states be a smaller subset of it.

We have similar considerations for effects in such GPTs. Given the set of effects E^A of a system A, we have $E^A \subset (V^A)^*$, where V^* denotes the dual of V, and for any $e \in E^A$ and any state $s \in \Omega^A$, $e(s) \in [0, 1]$. The requirement that $e(s) \geq 0$ means that e must be an element of $(\Omega^A_+)^*$ where

$$(\Omega^A_+)^* = \{f \in V^* : \forall v \in \Omega^A_+, \quad f(v) \geq 0\}. \tag{2.117}$$

Therefore, the requirement for positive probabilities defines a cone where effects must be contained. Similarly, $e(s) \leq 1$ defines another cone, and the intersection between those two defines the largest set of linear functionals admissible as effects for the states in Ω^A.

Figure 2.3: Pictorial representation of the effect set as a subset of the intersection between the dual cone to the positive cone, and the cone of the dual vectors that are less than the unit effect.



That said, there is no hard requirement [77, 78] on all those functionals being effects, but the property that they are is given a special name:

Definition 74 (No-Restriction Hypothesis)**.** A GPT is said to satisfy the no-restriction hypothesis if

$$E^A = \{e \in (V^A)^* : \forall s \in \Omega^A \quad e(s) \in [0,1]\}. \tag{2.118}$$

Both quantum theory and classical theory satisfy the no-restriction hypothesis.

Using the concepts just defined we talk about combination rules. Since we know the state spaces are the subset of normalized vectors in the cones, we can talk about system combination rules as rules that combine cones. A cone combination rule present in the literature that is relevant to us is the following:

Definition 75 (Maximal Tensor Product)**.** Given two system types A and B of a GPT, the maximal tensor product $\otimes_{max}$ combines their positive cones Ω_+^A and Ω_+^B to a cone $\Omega_+^{A\cdot B} \subset V^A \otimes V^B$ defined as follows:

$$\Omega_+^A \otimes \Omega_+^B = \{v \in V^A \otimes V^B : \forall e^A \in E^A \quad \forall e^B \in E^B \quad (e^A \otimes e^B)(v) \geq 0\}. \tag{2.119}$$

The idea here is that we take the cone of the combined system to be the largest cone in the tensor product space whose vectors are compatible with the separable effects $e^A \otimes e^B$. This is called maximal because the effects e^A and e^B are well defined in the individual systems A and B so we should expect that at least their tensor products will not produce invalid probabilities in the combined system. For this reason, we can think of the set of separable effects as defining the weakest possible set of constraints on the combined coned.

Next we define the analogous of the no-restriction hypothesis to the transformations [53]. We can use our notion of positive vectors to generalize also the notions of positive and completely positive map.

Definition 76 (Positive Transformation)**.** A linear transformation $T : V^A \to V^B$, with A and B systems of a GPT, is said to be positive if it takes positive vectors of V^A to positive vectors of V^B, that is, if

$$T \text{ is positive if } \; s \in \Omega_+^A \implies T(s) \in \Omega_+^B \tag{2.120}$$

Definition 77 (Completely Positive Transformation)**.** A linear transformation $T : V^A \to V^B$, with A and B systems of a GPT, is said to be completely positive if for every system C of the GPT, the operation $T \otimes \mathbb{1} : V^A \otimes V^C \to V^B \otimes V^C$ is positive.

Definition 78 (Trace Non-Increasing Transformation)**.** A transformation $T : V^A \to V^B$, with A and B systems of a GPT, is said to be trace non-increasing if for all $s \in \Omega^A$,

$$u^A(s) \geq u^B(T(s)) \tag{2.121}$$

Notice how this generalizes the notions from matrix algebra to GPTs by using our positive cones as the key idea. Next, let's define a strenghtening of the no-restriction hypothesis.

Definition 79 (Generalized No-Restriction Hypothesis)**.** A GPT is said to satisfy the generalized no-restriction hypothesis if it satisfies the no-restriction hypothesis, and if all completely positive trace-preserving transformations, as defined by its systems' postive cones and discarding effects, are transformations allowed by the theory.

With this, we can define Boxworld cleanly.

Definition 80 (Boxworld)**.** Boxworld is the GPT that satisfies the generalized no-restriction hypothesis with states, discarding effects and compositions given as follows. The states spaces $\mathcal{B}_{n,k}$ are real vector spaces of dimension $n(k-1)+1$, with $n, k \in \mathbb{N}$. Its states are vectors whose last component equals 1, and the rest can be seen as the first $k-1$ probabilities from n probability distributions, stacked one after the other. The unit effects are the inner product with the vector whose only non zero component is the last one, which equals 1. Finally, sequential composition is the matrix product, and parallel composition is given by the maximal tensor product.

We could define quantum and classical theory in a similar fashion, as they both satisfy the generalized no-restriction hypothesis, but we chose to give the more familiar definitions earlier for the sake of continuity.

2.3.2 Realizability

With the definition of GPT in hands, we can discuss the concept of main interest to us: realizability.

Consider the following generic question: if we have a given resource, say, a Bell experiment setup with Alice and Bob sharing a pair of entangled qubits, can it be modelled inside a given

GPT? We know that certain correlations, e.g. in the Bell scenario we just mentioned, cannot be explained by classical models unless we include in them some information communication which is then used to coordinate the outcomes observed at different parts of the experimental setup. Such a signal might be inconsistent with some extra assumptions that we might have about the setup [79, 80]: for instance, if Alice and Bob measurements are space-like separated, our trust in relativity theory requires any model of the experiment to include no signalling between those two events.

Those considerations show that when we are here asking those types of questions about realizability, we should do so under an additional constraint. What we really want to know is whether the model under the given GPT can be expressed using only non-signalling resources. That is, Alice and Bob can coordinate only by means of a shared past. This shared past is what we call here a common-cause.

What we show next is how, within the GPT framework, we can establish neat connections between what we know about Bell Nonlocality, EPR Inference, the no-signalling principle and localizability of quantum channels, how all of this can be expressed cleanly with the diagrammatic notation, and use those insights to make new questions that in some sense include all of those.

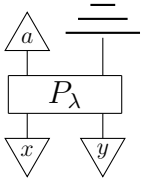
2.3.3 Bell Nonlocality, EPR Inference and Non-Signalling Channels

We will now see how our concepts of interest, are represented within the diagrammatic formalism. Here, we will use simple bipartite examples, as they are easily readable and easily generalizable. By the end of this section, the formal relationship between those three concepts shall be apparent given their representations.

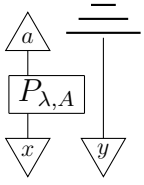
Our classifications, namely, for correlations of having local hidden variables or not, and assemblages of having local hidden states or not, concern onstraints at least as strong as the no-signalling conditions, we start by expressing the latter in diagrammatic notation.

If a process λ is non-signalling, then

$$p(a|xy\lambda) = \sum_b p(ab|xy\lambda) =$$



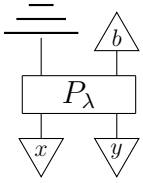
$$= p(a|x\lambda) \sum_y p(y) =$$



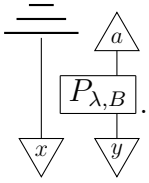
(2.122)

and

$$p(b|xy\lambda) = \sum_a p(ab|xy\lambda) =$$



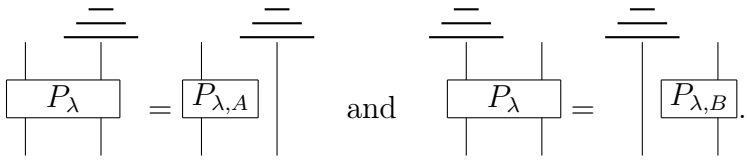
$$= \sum_x p(x)p(b|y\lambda) =$$



. (2.123)

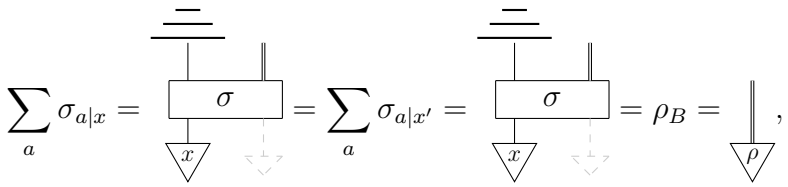
for some valid classical maps $p_{\lambda,A}$ and $p_{\lambda,B}$. Since classical theory satisfies local tomography

and the above are true for arbitrary a,x,b,y, that means we can equivalently take the inner bipartite classical channel p_λ to satisfy



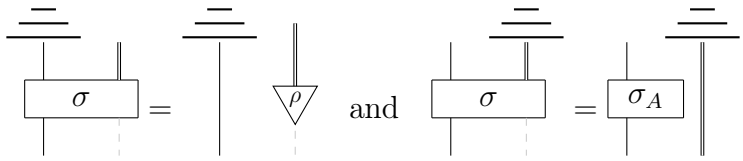
(2.124)

Next, for a non-signalling assemblage $\sigma_{a|x}$, we can take a/x as classical inputs/ouputs to a process that prepares a quantum state at Bob's lab, and hence represent the no-signalling conditions as follows:



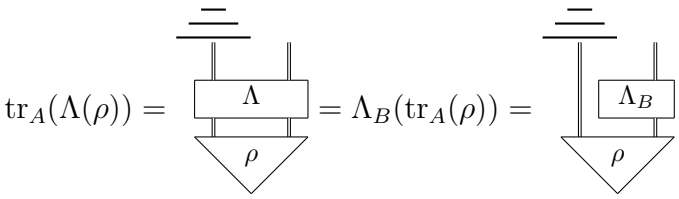
(2.125)

for any x, x', where we the ghost system at the right-side input was is the trivial system and was drawn only to emphasize the formal similarity between the diagrams we are drawing. Again by local tomography, what we see that the above means $\sigma_{a|x}$ is non-signalling if the inner channels satisfies



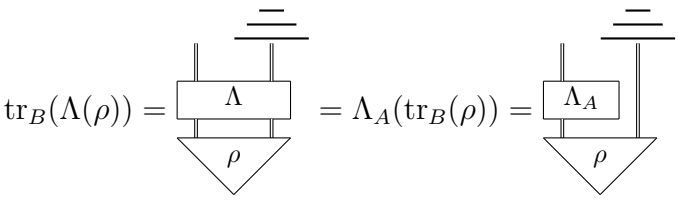
(2.126)

Similarly, if a quantum bipartite channel Λ is non-signalling, then it satisfies



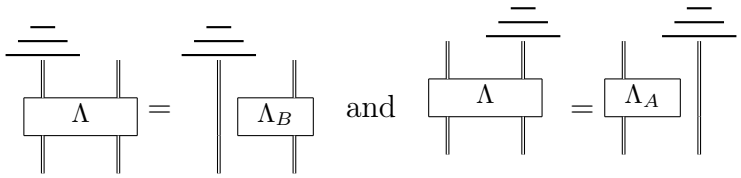
(2.127)

and



(2.128)

which in turn, by local tomography of quantum theory, means that the diagrammatic representation of of the channel Λ satisfies



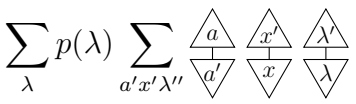
(2.129)

We can see that the assemblages are non-signalling precisely when the underlying channel [35] that models it can be taken to be a non-signalling channel. The same works for local hidden variable correlations, if we consider only the form of the diagrams and not the wire types. Later we will use this to generalize those concepts to more arbitrary GPTs.

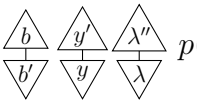
Let's now turn our attention to the Bell nonlocality and steering scenarios. The first step will be to see what form Bell scenario correlations will take. We can start with a conditional probability distribution $p(ab|xy)$ and see that, if it has a local hidden variable model, it can be represented as follows:

$$p(ab|xy) = \sum_\lambda p(\lambda)p(a|x\lambda)p(b|y\lambda) = \sum_\lambda p(\lambda) \sum_{a'x'\lambda'} \delta_{aa'}\delta_{xx'}\delta_{\lambda\lambda'}p(a'|x'\lambda') \sum_{b'y'\lambda''} \delta_{bb'}\delta_{yy'}\delta_{\lambda\lambda'}p(b'|y'\lambda'')$$

$$= \sum_\lambda p(\lambda) \sum_{a'x'\lambda''}$$

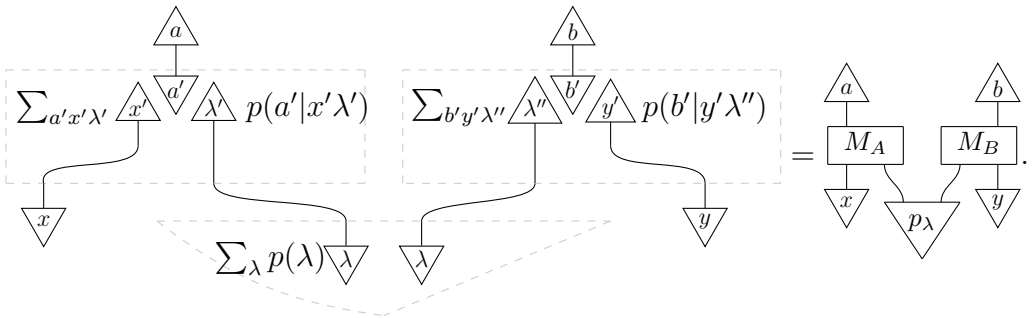


$$p(a'|x'\lambda') \sum_{b'y'\lambda''}$$



$$p(b'|y'\lambda'')$$

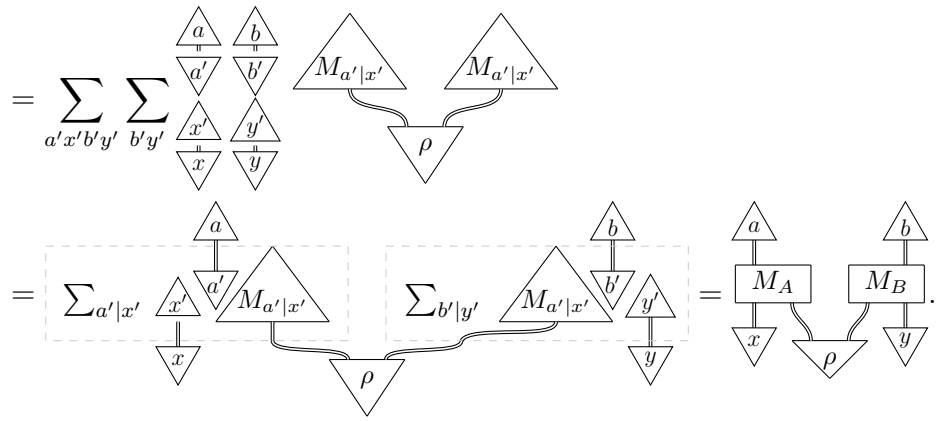
$$=$$



(2.130)

Here, we have taken the states/effects labelled by $a, a', b, b', x, x', y, y', \lambda, \lambda', \lambda''$ to be elements of orthonormal bases of the vector spaces to which they belong. This is possible since these systems are classical, and as members of **Stoch**, they are allowed to represent deterministic probability distributions, which happen to be vectors whose all columns are zero except for one, which is 1. Effectively, we can read this diagram as Alice and Bob performing local classical operations M_A and M_B, respectively, on a shared bipartite classical system that follows a distribution p_λ. A quantum correlation in a Bell scenario, on the other hand, becomes the following diagram:

$$p(ab|xy) = \mathrm{tr}\big(M_{a|x} \otimes M_{b|y} \cdot \rho\big) = \sum_{a'x'b'y'} \delta_{aa'}\delta_{xx'}\delta_{bb'}\delta_{yy'}\, \mathrm{tr}\big(M_{a'|x'} \otimes M_{b'|y'} \cdot \rho\big)$$



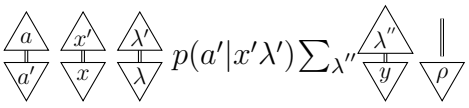
(2.131)

Again, we can think of orthonormal vectors and interprete the M operations as Alice and Bob performing local operations on their shares of the system. In this case, we took those operations to have quantum inputs and outputs, but we could as well take a, b, x, y (but not ρ!) to be classical systems if we would like to model it as quantum theory with a classical interface. Notice how the diagrams we've seen for Bell scenarios have the same general shape, and in going from quantum correlations to classical correlations we are simply changing the types of the wires.

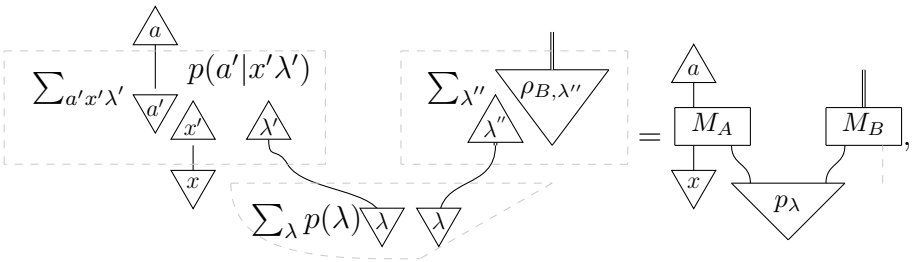
Let's now do a similar procedure for assemblages. Starting from the previous discussion, we have that an assemblage has a local hidden state model if all its elements $\sigma_{a|x}$ can be represented as follows:

$$\sigma_{a|x} = \sum_{\lambda} p(\lambda)p(a|x\lambda)\rho_{B,\lambda} = \sum_{\lambda} p(\lambda) \sum_{a'x'\lambda'} \delta_{aa'}\delta_{xx'}\delta_{\lambda\lambda'} p(a'|x'\lambda') \sum_{\lambda''} \delta_{\lambda\lambda''}\rho_{B,\lambda''}$$

$$= \textstyle\sum_\lambda p(\lambda) \sum_{a'x'\lambda'}$$



$$\textstyle p(a'|x'\lambda') \sum_{\lambda''}$$



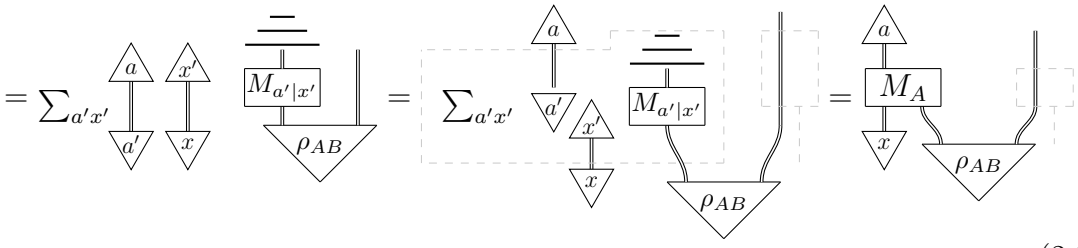
(2.132)

where we used a gray dashed edge to represent the (normally not explicitly represented) identity system as one of the inputs to M_B with the sole purpose of stressing the similarity in the shape of the diagrams we are now writing. Notice we can interprete it again as Alice

and Bob sharing a classical state p_λ and doing local operations on their systems. Contrasting it against the Bell scenario diagrams, we see that the difference is that Bob never performs a measurement, but prepares a quantum system in a manner dependent of his share of the classical system.

Let's repeat the same procedure for quantum assemblages:

$$\sigma_{a|x} = \mathrm{tr}_A(M_{a|x} \otimes \mathbb{1}_B \cdot \rho_{AB}) = \sum_{a'x'} \delta_{aa'}\delta_{xx'} \,\mathrm{tr}_A(M_{a'|x'} \otimes \mathbb{1}_B \cdot \rho_{AB})$$

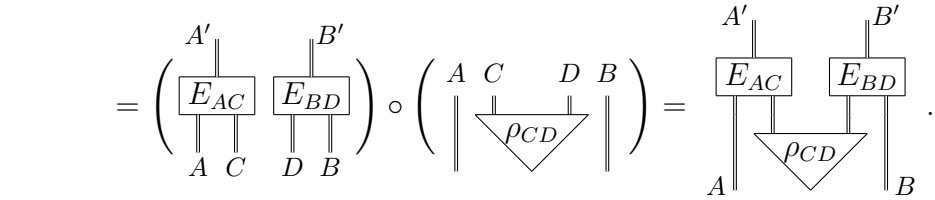


(2.133)

Again, we are going from the local hidden state model to the quantum model by writing another diagram of the same shape with different wire types.

Finally, let's write the diagrammatic representation for localizable channels. If $\Lambda : A \otimes B \to A' \otimes B'$ is a localizable channel, then we know that

$$\Lambda = (E_{AC} \otimes E_{BD}) \circ (\mathbb{1}_A \otimes \rho_{CD} \otimes \mathbb{1}B)$$



(2.134)

This diagram has the same clear interpretation as the previous examples: it means that the Λ channel could be realized by two parties operating locally when they share a bipartite quantum state.

Importantly, we can see that all of the previous examples, namely the quantum and local hidden variable correlations, and the quantum and local hidden state assemblages can be see as being constructed from particular cases of the localizable maps [35].

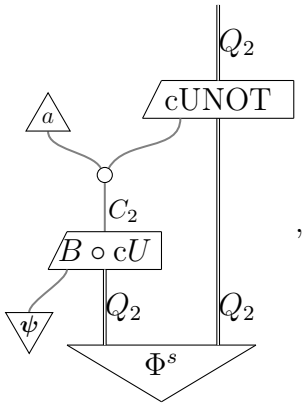
Chapter 3

Summary of dissertation

3.1 GPT Common-Cause Realizability of Post-Quantum Assemblages

The first article [53] investigates the properties of Einstein-Podolsky-Rosen inference, here called EPR steering, in the context of the compositional apparatus of operational theories. Specifically, we start from the realization that EPR steering is defined in terms of non-signalling assemblages and needs no considerations about how such assemblages are constructed. From there, we define a new GPT, called Witworld, and use it to analyse the problem.

The first point is whether post-quantum steering is realizable (in a common-cause scenario) within any GPT at all. The answer is given by providing an example, constructed within Witworld, showing that can it can indeed provide such scheme. One example is the assemblage whose elements are defined as follows

$\sigma_{a|\psi} =$ 

, (3.1)

where B is a (specific) quantum measurement, cU is a controlled unitary map, cUNOT is a controlled universal not operation on quantum states (note that this operation is not allowed by quantum theory but it is valid within Witworld), Φ^s a quantum state and a and ψ classical states.

Furthermore, not only Witworld can realize post-quantum steering, but we show that this is true for the explicitly defined post-quantum assemblages in the literature previous to the article.

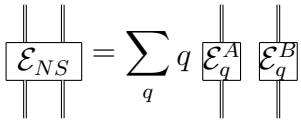
The combination rule for states chosen for Witworld is the maximal tensor product. One consequence of this, which names the theory, is that all so-called Gleason assemblages are realizable in Witworld, because the state spaces of Witworld contain all entanglement witnesses as states. That follows because replacing quantum states with entanglement witnesses is what is required to express matematically the Gleason assemblages, some of which are post-quantum assemblages.

We prove a lemma stating that if a theory combines systems with the maximal tensor product, then the sets of positive and completely positive maps (as per our generalized definition of positivity) coincide. This means that when the agents, say, Alice and Bob, combine quantum systems within Witworld, they have more local operations available than what quantum theory would allow them. These extra operations can be seen at least partially responsible for the possibility of preparing assemblages beyond what is allowed by quantum theory. Namely, they have available positive but not completely positive operations on local quantum systems.

Lastly, we use Witworld's GPT apparatus to investigate the computational properties of post-quantum steering. By expressing diagrammatically a protocol for remote state preparation, we prove by inspection that the realization of a post-quantum assemblage can provide a stronger-than-quantum performance for said task. We used the assemblage from equation 3.1 and we show that this assemblage allows for a deterministic sucess performance to be achieved with only 1 classical bit of communication, less than the 2 bits required by the optimal quantum strategy.

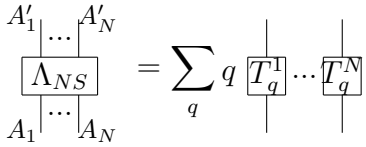
3.2 Non-signalling Channels and Affine Combinations

The second article [54] generalizes to multipartite maps of locally tomographic GPTs a fact that was known to be true for multipartite classical stochastic maps and bipartite quantum maps. Namely, any non-signalling channels in the aforementioned class can be mathematically expressed as affine combinations of product channels. To give some context, it was known that bipartite non-signalling quantum states can always be represented as follows:



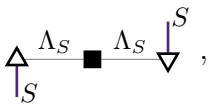
(3.2)

where E_q^A, E_q^B are quantum channels and $\sum_q q = 1$. We generalize this to the following form, expressed as a diagram in the formalism of tomographically local GPTs:



(3.3)

Our proof consists of developing a scheme to lift the result from the GPT of classical stochastic maps to more general GPTs. In doing so, we employ a few extra notational features from the duotensor framework to improve clarity. Namely, we use diagram boxes filled in black color to represent, within the familiar diagrams, mathematical objects that are not considered valid states/transformations/effects in the GPT of interest. In our calculations, the diagrams as a whole are GPT objects, but we decompose them in non GPT mathematical objects in intermediate steps, while keeping track of them with the color coding. One example of such a diagram is as follows:



(3.4)

where the triangles represent some fiducial preparation and measurement, and the black square represents a map that his mathematically defined but not physically allowed in the GPT.

This notation, complemented by an auxiliary theorem that we prove, stating that the affine hull of the measure-and-prepare and the set of discard preserving transformations coincide, is what allows us to lift the classical result to more general GPTs in a relatively clean manner. Furthermore, the nature of our proof tecnique opens the possibility of lifting more results from classical theory to the more general class of locally tomographic GPTs.

3.3 Common-Cause Realizability of Nonsignalling Channels

The third article [55] investigates a more general form of the problem tackled in the first one. While in the first article we examined questions related to the realizability of EPR assemblages, which can be viewed as a specific type of non-signalling channels, in the third article we ask similar questions about general non-signalling channels. In doing so, we also address as a particular case one of the open problems posed in the first article, namely, whether there exists a GPT that can realize all post-quantum assemblages.

More specifically, we answer the question of whether a GPT common-cause decomposition is possible for any non-signalling channel of causal and locally tomographic GPTs such as classical and quantum theory. For context, recall how in Bell scenarios some correlations (which are classical stochastic maps) cannot be modelled by only classical common-cause states shared by the parties, but can if we use quantum states, rather than classical. Similarly, some non-signalling correlations cannot be realized in that fashion with quantum states, but

with Boxworld states they can. With that in mind, we can ask whether there always exists a GPT where one can model the non-signalling channels of a source GPT (with those being classical, quantum, or channels of another causal locally tomographic GPT) as a common-cause process.

By employing the main theorem of the second article, we define a map that takes a locally tomographic causal GPT, here denoted source, to a target causal GPT such that all the non-signalling channels of the source are realizable in a common-cause scenario of in the target GPT.

This result provides a bridge between the two constraints on channels that we discussed: the decomposability into a common-cause form, and the impossibility of signalling information. This has an interesting consequence for resource theory considerations, as in there the so-called free resources are often taken to be the non-signalling resources, which are very simply mathematically chacterized, or the common-cause ones, which is grounded on causality considerations but not so straightforwardly characterized.

Finally, another contribution of the third article is that it presents a way of defining GPTs that differs from the standard approach in the GPT literature. While normally one defines GPTs by referencing the geometry of the sets of states, effects and transformations, we focus on the compositional aspects of the GPT for our constructions.

Chapter 4

Outlook

We investigated two important notions from quantum information research, that of non-signalling channels and that of nonlocality, from the perspective of the generalized probabilistic theory framework. In doing so, the scope of the questions that regard differences between quantum and classical theory widens to a wide variety of mathematical structures interpretable as theories, which includes for instance Boxworld, a GPT of central importance to our three articles.

Within this large class of theories, we find more commonalities between nonlocality and non-signalling conditions, evidenced by the forms of the diagrams defining them, which we can establish through our results on realizability. The first result we obtained gives by explicit construction common-cause decompositions of post-quantum non-signalling assemblages in a GPT. We furthermore use that GPT to learn about the information processing consequences of post-quantum steering, specifically for the task of remote state preparation. In that task, we show that post-quantum steering is a stronger-than-quantum resource. Previously to the aforementioned results, those assemblages, as quantum objects, were known to exist but no GPT was known where they could be represented in common-cause scenarios. Despite of finding such constructions for many examples, by that point the question of whether this is a general fact remained opened.

Continuing learning about non-signalling channels, we prove a mathematical property for all such maps in locally tomographic GPTs. Namely, in those GPTs, the non-signalling channels can always be represented as affine combinations of product channels. That result generalizes others from classical and quantum theories to a larger class of GPTs. Here, not just the result but the proof technique itself can be seen as interesting because it works by lifting previously known results from one GPT (classical) to an entire class (locally tomographic GPTs).

Finally, this additional knowledge we obtained about the non-signalling channels in locally tomographic GPTs is used to tackle the remaining question that stemmed from the first work. We take the point of view that assemblages are a subset of the quantum channels. So, if it is the case that non-signalling channels in causal locally tomographic GPTs can always be represented in common-cause scenarios in some other GPT, then we give a

positive answer to the particular case of the question where the channels are assemblages. Indeed, after giving a GPT diagrammatic definition of multipartite non-signalling channels and applying the affine combination decomposition theorem to it, we were able to obtain a scheme that proves that to be true. Namely, our common-cause completion map is proved to take any causal locally tomographic GPT to a causal, possibily locally tomographic, GPT that can realize all of the non-signalling channels of the original GPT in common-cause scenarios, just like Boxworld does that to classical theory. Importantly, this result concerns only the existence of the common-cause completion but does not yield concretely workable theories such as the one constructed for our first result about assemblages. For instance, if one applies that procedures which define the common-cause completion map to classical theory, the resulting theory is not Boxworld, despite the fact that Boxworld is a common-cause completion of classical theory. Nevertheless, with the proof of existence established, we show that, at least for the class of locally tomographic theories, the concepts of non-signalling and common-cause realizability are not as distinct as previously one could suspect.

Bibliography

[1] E. Schrödinger, Naturwissenschaften **23**, 807 (1935), URL https://doi.org/10.1007/BF01491891.

[2] A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. **47**, 777 (1935), URL https://link.aps.org/doi/10.1103/PhysRev.47.777.

[3] E. Schrödinger, Mathematical Proceedings of the Cambridge Philosophical Society **32**, 446–452 (1936).

[4] J. S. Bell, Phys. Phys. Fiz. **1**, 195 (1964), URL https://link.aps.org/doi/10.1103/PhysicsPhysiqueFizika.1.195.

[5] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Physical review letters **23**, 880 (1969).

[6] R. Chaves, R. Kueng, J. B. Brask, and D. Gross, Physical review letters **114**, 140403 (2015).

[7] C. J. Wood and R. W. Spekkens, New Journal of Physics **17**, 033002 (2015), URL https://dx.doi.org/10.1088/1367-2630/17/3/033002.

[8] E. Wolfe, D. Schmid, A. B. Sainz, R. Kunjwal, and R. W. Spekkens, Quantum **4**, 280 (2020).

[9] M. Weilenmann and R. Colbeck, Quantum **4**, 236 (2020).

[10] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Rev. Mod. Phys. **86**, 419 (2014), URL https://link.aps.org/doi/10.1103/RevModPhys.86.419.

[11] V. Scarani, *Bell nonlocality* (Oxford University Press, 2019).

[12] A. Aspect, P. Grangier, and G. Roger, Physical review letters **47**, 460 (1981).

[13] A. Aspect, P. Grangier, and G. Roger, Physical review letters **49**, 91 (1982).

[14] B. Hensen et al., Nature **526**, 682 (2015), URL https://doi.org/10.1038/nature15759.

[15] L. K. Shalm et al., Phys. Rev. Lett. **115**, 250402 (2015), URL https://link.aps.org/doi/10.1103/PhysRevLett.115.250402.

[16] M. Giustina et al., Phys. Rev. Lett. **115**, 250401 (2015), URL https://link.aps.org/doi/10.1103/PhysRevLett.115.250401.

[17] C. H. Bennett and G. Brassard, Theoretical computer science **560**, 7 (2014).

[18] D. Deutsch and R. Jozsa, Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences **439**, 553 (1992).

[19] P. W. Shor, in *Proceedings 35th annual symposium on foundations of computer science* (Ieee, 1994), pp. 124–134.

[20] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani, in *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, USA, 1999), STOC '99, p. 376–383, ISBN 1581130678, URL https://doi.org/10.1145/301250.301347.

[21] S. Wiesner, SIGACT News **15**, 78–88 (1983), ISSN 0163-5700, URL https://doi.org/10.1145/1008908.1008920.

[22] A. M. Childs, Quantum Info. Comput. **5**, 456–466 (2005), ISSN 1533-7146.

[23] J. Barrett, L. Hardy, and A. Kent, Phys. Rev. Lett. **95**, 010503 (2005), URL https://link.aps.org/doi/10.1103/PhysRevLett.95.010503.

[24] A. Acín, N. Gisin, and L. Masanes, Phys. Rev. Lett. **97**, 120405 (2006), URL https://link.aps.org/doi/10.1103/PhysRevLett.97.120405.

[25] U. Vazirani and T. Vidick, Phys. Rev. Lett. **113**, 140501 (2014), URL https://link.aps.org/doi/10.1103/PhysRevLett.113.140501.

[26] R. Colbeck and R. Renner, Nat. Phys. **8**, 450 (2012), URL https://doi.org/10.1038/nphys2300.

[27] S. Pironio et al., Nature **464**, 1021 (2010), URL https://doi.org/10.1038/nature09008.

[28] J. I. de Vicente, J. Phys. A **47**, 424017 (2014).

[29] L.-M. Duan, M. D. Lukin, J. I. Cirac, and P. Zoller, Nature **414**, 413 (2001).

[30] S. Popescu and D. Rohrlich, Found. Phys. **24**, 379 (1994), URL https://doi.org/10.1007/BF02058098.

[31] H. M. Wiseman, S. J. Jones, and A. C. Doherty, Phys. Rev. Lett. **98**, 140402 (2007), URL https://link.aps.org/doi/10.1103/PhysRevLett.98.140402.

[32] R. Uola, A. C. S. Costa, H. C. Nguyen, and O. Gühne, Rev. Mod. Phys. **92**, 015001 (2020), URL https://link.aps.org/doi/10.1103/RevModPhys.92.015001.

[33] D. Cavalcanti and P. Skrzypczyk, Rep. Prog. Phys. **80**, 024001 (2016), URL https://doi.org/10.1088/1361-6633/80/2/024001.

[34] B. Zjawin, D. Schmid, M. J. Hoban, and A. B. Sainz, Quantum **7**, 926 (2023), ISSN 2521-327X, URL https://doi.org/10.22331/q-2023-02-16-926.

[35] M. J. Hoban and A. B. Sainz, New J. Phys. **20**, 053048 (2018).

[36] Ana Belén Sainz, Nicolas Brunner, Daniel Cavalcanti, Paul Skrzypczyk, and Tamás Vértesi, Phys. Rev. Lett. **115**, 190403 (2015), URL https://link.aps.org/doi/10.1103/PhysRevLett.115.190403.

[37] J. Barrett, Physical Review A **75**, 032304 (2007).

[38] B. Coecke, Deep beauty: understanding the quantum world through mathematical innovation pp. 129–186 (2011).

[39] M. Plávala, arXiv preprint arXiv:2103.07469 (2021).

[40] C. M. Lee and J. Barrett, New Journal of Physics **17**, 083001 (2015).

[41] M. P. Müller, arXiv preprint arXiv:2011.01286 (2020).

[42] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. A **81**, 062348 (2010), URL https://link.aps.org/doi/10.1103/PhysRevA.81.062348.

[43] D. Schmid, J. H. Selby, M. F. Pusey, and R. W. Spekkens, arXiv preprint arXiv:2005.07161 (2020).

[44] P. Janotta and H. Hinrichsen, Journal of Physics A: Mathematical and Theoretical **47**, 323001 (2014).

[45] J. Barrett, N. de Beaudrap, M. J. Hoban, and C. M. Lee, npj Quantum Inf. **5**, 1 (2019), URL https://doi.org/10.1038/s41534-019-0156-9.

[46] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. A **84**, 012311 (2011).

[47] L. Hardy, Preprint at https://arxiv.org/abs/1104.2066 (2011).

[48] L. Hardy, Preprint at https://arxiv.org/abs/quant-ph/0101012 (2001).

[49] H. Barnum and A. Wilce, ENTCS **270**, 3 (2011).

[50] D. Gross, M. Müller, R. Colbeck, and O. C. O. Dahlsten, Physical Review Letters **104** (2010), ISSN 1079-7114, URL http://dx.doi.org/10.1103/PhysRevLett.104.080402.

[51] A. J. Short and J. Barrett, New Journal of Physics **12**, 033034 (2010), URL `https://doi.org/10.1088/1367-2630/12/3/033034`.

[52] S. W. Al-Safi and A. J. Short, Journal of Physics A: Mathematical and Theoretical **47**, 325303 (2014), ISSN 1751-8121, URL `http://dx.doi.org/10.1088/1751-8113/47/32/325303`.

[53] P. J. Cavalcanti, J. H. Selby, J. Sikora, T. D. Galley, and A. B. Sainz, npj Quantum Information **8**, 1 (2022).

[54] P. J. Cavalcanti, J. H. Selby, J. Sikora, and A. B. Sainz, Journal of Physics A: Mathematical and Theoretical (2022).

[55] P. J. Cavalcanti, J. H. Selby, and A. B. Sainz, Physical Review A **109**, 042211 (2024).

[56] E. Patterson, D. Spivak, and D. Vagner, Electronic Proceedings in Theoretical Computer Science **333**, 49 (2021).

[57] S. Abramsky and B. Coecke, Handbook of quantum logic and quantum structures **2**, 261 (2009).

[58] B. Coecke and E. O. Paquette, New structures for physics pp. 173–286 (2011).

[59] B. Coecke and A. Kissinger, in *International Conference on Theory and Application of Diagrams* (Springer, 2018), pp. 28–31.

[60] A. Joyal and R. Street, Advances in mathematics **88**, 55 (1991).

[61] P. Selinger, in *New structures for physics* (Springer, 2010), pp. 289–355.

[62] J. Pearl, *Causality* (Cambridge university press, 2009).

[63] C. J. Wood and R. W. Spekkens, New Journal of Physics **17**, 033002 (2015).

[64] J.-M. A. Allen, J. Barrett, D. C. Horsman, C. M. Lee, and R. W. Spekkens, Physical Review X **7**, 031021 (2017).

[65] A. Fine, Physical Review Letters **48**, 291 (1982).

[66] B. S. Cirel'son, Letters in Mathematical Physics **4**, 93 (1980).

[67] D. Cavalcanti, P. Skrzypczyk, G. Aguilar, R. Nery, P. S. Ribeiro, and S. Walborn, Nat. Commun. **6**, 1 (2015).

[68] A. B. Sainz, M. J. Hoban, P. Skrzypczyk, and L. Aolita, Phys. Rev. Lett. **125**, 050404 (2020), URL `https://link.aps.org/doi/10.1103/PhysRevLett.125.050404`.

[69] Ana Belén Sainz, Leandro Aolita, Marco Piani, Matty J. Hoban, and Paul Skrzypczyk, New J. Phys. **20**, 083040 (2018).

[70] V. P. Rossi, M. J. Hoban, and A. B. Sainz, Journal of Physics A: Mathematical and Theoretical **55**, 264002 (2022).

[71] D. Beckman, D. Gottesman, M. A. Nielsen, and J. Preskill, Phys. Rev. A **64**, 052309 (2001), URL `https://link.aps.org/doi/10.1103/PhysRevA.64.052309`.

[72] G. Chiribella, G. M. D'Ariano, P. Perinotti, and B. Valiron, Physical Review A **88**, 022318 (2013).

[73] C. Heunen and J. Vicary, *Categories for quantum theory: An introduction* (2015).

[74] F. Borceux, *Handbook of categorical algebra: volume 1, Basic category theory*, vol. 1 (Cambridge University Press, 1994).

[75] F. Borceux, *Handbook of Categorical Algebra: Volume 2, Categories and Structures*, vol. 2 (Cambridge University Press, 1994).

[76] S. Awodey, *Category theory*, vol. 52 (OUP Oxford, 2010).

[77] P. Janotta and R. Lal, Phys. Rev. A **87**, 052131 (2013).

[78] A. B. Sainz, Y. Guryanova, A. Acín, and M. Navascués, Physical review letters **120**, 200402 (2018).

[79] B. Coecke, New Generation Computing **34**, 69 (2016).

[80] A. Kissinger, M. Hoban, and B. Coecke, arXiv preprint arXiv:1708.04118 (2017).

ARTICLE OPEN

Check for updates

Post-quantum steering is a stronger-than-quantum resource for information processing

Paulo J. Cavalcanti[1], John H. Selby[1], Jamie Sikora[2]✉, Thomas D. Galley[3] and Ana Belén Sainz[1]

We present the first instance where post-quantum steering is a stronger-than-quantum resource for information processing – remote state preparation. In addition, we show that the phenomenon of post-quantum steering is not just a mere mathematical curiosity allowed by the no-signalling principle, but it may arise within compositional theories beyond quantum theory, hence making its study fundamentally relevant. We show these results by formulating a new compositional general probabilistic theory – which we call Witworld – with strong post-quantum features, which proves to be a intuitive and useful tool for exploring steering and its applications beyond the quantum realm.

npj Quantum Information (2022)8:76 ; https://doi.org/10.1038/s41534-022-00574-8

INTRODUCTION

A striking property of nature is that it is non-classical. Entanglement[1,2], Bell nonlocality[3], and steering[4–6] are examples of quantum phenomena that can be observed experimentally[7–12] and which cannot be explained by classical physics. Besides their foundational relevance, with the advent of quantum information theory we learned a valuable lesson: these seemingly bizarre quantum features can be exploited to process information more efficiently, even in ways that could never be possible with classical resources alone[13–15].

A ubiquitous framework in which the scope of quantum advantage in information processing is explored is the so-called device-independent framework, where the parties executing protocols only rely on the classical inputs and outputs with which they operate their shared (and possibly quantum) devices. Such a framework is particularly well suited to the necessarily paranoid perspective on cryptographic tasks[13], and is almost ubiquitously underpinned by a Bell nonlocality setup. A special milestone in the research of non-classical resources for device-independent information processing was the realisation that there exist correlations beyond what is quantumly admissible (a.k.a., post-quantum correlations), but which nonetheless are consistent with special relativity[16]. These so-called no-signalling correlations, moreover, were shown to be consistent with alternative theories of nature, confirming the necessity of their study. Exploring these general no-signalling correlations enabled, for example, the design of quantum cryptographic protocols that are robust against powerful adversaries that are not bounded by the laws of quantum theory[17,18], and also the formulation of physical principles that a quantum world must satisfy[16,19–25]. These post-quantum correlations are hence studied beyond philosophical motivations, and from the perspective of the resources they provide for operational tasks.

Device-independent frameworks for information processing, however, are substantially demanding to implement experimentally. Indeed, for practical purposes, even if cryptographically secure, device-independent protocols are yet to move beyond 'proof of principle' applications into scalable and easily-accessible technologies. There are situations, however, where one may argue that the quantum description of some of the parties involved can be leveraged in the protocols: in the simplest case where two parties are involved and a single party is assigned a quantum description this is usually referred to as a one-way device-independent framework[14,15]. In such scenarios, the non-classical phenomenon providing quantum advantage is steering rather than Bell nonlocality. Recently, it has been shown that steering beyond that which quantum theory allows, whilst still consistent with special relativity, may exist[26,27], which opens a new plethora of questions, such as (i) can post-quantum steering provide an advantage beyond what is possible with quantum theory for some information processing task?; (ii) is post-quantum steering just a mathematical curiosity, or may it emerge within alternative physical theories?

In this work, we tackle those two questions. First, we show that there are alternative theories beyond quantum which feature post-quantum steering, making the phenomenon physically relevant for post-quantum information processing and motivating its exploration. Second, we find a task for which post-quantum steering is a stronger-than-quantum resource: remote state preparation. Remote state preparation (RSP) is a task similar in spirit to teleportation: the goal is to transmit quantum states from one party to another distant party using only shared entanglement and classical communication. Unlike teleportation, though, in RSP the sender has knowledge of the transmitted state, which makes RSP protocols more economical than teleportation in terms of resources (e.g. classical communication) needed to succeed at the task[28,29]. In addition, RSP protocols do not necessarily require the ability to experimentally implement Bell (entangling) measurements, which makes them potentially more feasible experimentally[30]. The kind of RSP protocols that we focus on are so-called oblivious – namely those where no information about the state is leaked to the receiver, apart from the state itself, something that is relevant for certain applications such as blind quantum computation[31]. RSP is indeed an insightful task to explore from both a fundamental and applied viewpoint.

In order to prove our results, we define a generalised probabilistic theory (GPT) that we name Witworld, given its strong connection to entanglement witnesses. Witworld combines system types of three well-known GPTs (classical, quantum,

[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, USA. [3]Perimeter Institute for Theoretical Physics, 31 Caroline St. N, Waterloo, ON N2L 2Y5, Canada. ✉email: jamiesikora@gmail.com

and Boxworld) in a simple mathematical way, via the so-called max tensor product. Remarkably, even though Witworld cannot reproduce all the phenomenology of quantum theory, it does realise all quantum predictions for Bell and steering experiments. Hence, we can learn about the limitations of quantum advantage in one-sided and fully device-independent protocols by exploring the performance of Witworld. Despite its simplicity, Witworld displays powerful post-quantum features: not only can it realise all non-signalling correlations in Bell experiments, but it also displays post-quantum steering. As Witworld is a fully compositional theory, it comes equipped with an intuitive diagrammatic calculus[32,33]. This provides a convenient toolkit for exploring other applications of post-quantum phenomena for information processing.

The paper is organised as follows. In what remains of this section we present a brief introduction to the three main topics of this paper: generalised probabilistic theories, Bell nonlocality, and steering. Section "Witworld" presents the definition of Witworld, assuming basic knowledge of GPTs – the reader who is not familiar with them may consult Section A of the Supplementary Material. Section "Post-quantum phenomena: Bell non-classicalityand steering" discusses the post-quantum properties of Witworld with a focus on Bell and steering experiments, while section "Post-quantum advantage for informationprocessing" discusses how Witworld outperforms quantum theory at certain information processing tasks. Section "Post-quantum advantage for informationprocessing" also presents the first task where post-quantum steering outperforms quantum steering as a resource. Technicalities, as well as a brief review on steering, are included in the appendices.

Generalised probabilistic theories

The framework of generalised probabilistic theories[34,35] provides tools with which to explore the operational features of candidate theories in a unified fashion. Classical theory, as well as quantum theory, may be recast within the language of GPTs[34,35], which enables their unified and comparative study. The GPT framework has been proven useful not only from a foundational perspective (e.g. for developing axiomatic reconstructions of quantum theory[33,35–38]), but also when exploring the quantum information capabilities of post-quantum theories, such as their computational power[39–45] or cryptographic security[34,46–49].

Bell nonlocality

One example of a non-classical phenomenon of foundational and applied relevance is Bell nonlocality. Bell experiments are ubiquitous in the fields of quantum foundations and quantum information processing. On the one hand, Bell's Theorem[3] established a precise sense in which quantum theory requires a departure from a classical worldview, and violations of Bell inequalities provide a means for certifying the nonclassicality of nature. On the other hand, the correlations observed in a Bell test have become a resource for certain tasks[13], and the violation of so-called Bell inequalities by these correlations has become a standard certification tool for security in cryptographic protocols[17,18,50]. In brief, a Bell scenario consists of a set of distant parties that perform space-like separated actions on their part of a physical system, and the objects of study are the correlations they observe among their measurement outcomes. In the case of a bipartite scenario, let $x \in \mathbb{X}$ and $a \in \mathbb{A}$ denote the classical variables that label the measurement choices and produced outcomes, respectively, corresponding to the first party (hereon, Alice), and, respectively, $y \in \mathbb{Y}$ and $b \in \mathbb{B}$ those for the second party (hereon, Bob). The correlations observed in this bipartite Bell experiment are captured by the conditional probability distribution $\{p(ab|xy)\}$. It is therefore natural to ask ourselves which possible $\{p(ab|xy)\}$ may be generated, and at what cost. Given the space-like separation constraints, the largest set of correlations observable in a Bell scenario corresponds to those that satisfy the No-Signalling Principle, and it is known that correlations allowed by quantum theory are a strict subset of those correlations. Notably, a GPT colloquially referred to as Boxworld[51,52] has been defined[34], which can realise all the correlations compatible with the no-signalling principle via its bipartite states and local measurements.

Steering

Steering is another non-classical phenomenon of foundational and practical relevance, which was identified back in the 1930s[4] but, unlike Bell nonlocality, only recently caught the attention of the quantum information community[5,6]. Steering captures the idea that Alice seemingly remotely 'steers' the state of a distant Bob, in a way which has no classical explanation. A main feature of a steering experiment is the asymmetric role that the parties play, which makes it particularly suitable as a resource for certain asymmetric information processing tasks[14,15]. In brief, the simplest steering experiment consist of two distant parties – Alice and Bob – which perform local actions on their part of a physical system. Unlike in a Bell experiment, though, the parties here perform different types of transformations in their labs: Alice performs a measurement, labelled by $x \in \mathbb{X}$, on her system, and obtains a classical outcome $a \in \mathbb{A}$, whereas Bob performs full tomography of the quantum system and so describes it via a density matrix $\rho^{\mathrm{B}}_{a|x}$ that is effectively prepared in his lab after Alice's actions. In this way, the object of study in these experiments are the ensembles of ensembles (a.k.a. *assemblages*[53]) given by $\{\{\sigma_{a|x}\}_{a\in\mathbb{A}}\}_{x\in\mathbb{X}}$, where $\mathrm{tr}(\sigma_{a|x}) = p(a|x)$ and $\sigma_{a|x} = p(a|x)\,\rho^{\mathrm{B}}_{a|x}$. While nonclassical properties of steering within quantum theory have been considerably explored, not much is known about steering beyond quantum theory[26,27,54,55]. One main obstacle for this is the complexity of capturing fundamentally what could be post-quantum about an assemblage of quantum states. An operational recast of the steering phenomenon has been recently put forward[26,27], which facilitates a way to articulate the concept of post-quantum assemblages. The study of post-quantum steering has only just begun, and, unlike for Bell nonlocality, important fundamental and practical questions are yet to be answered. One such question is: does there exist a GPT that realises all these post-quantum assemblages?

RESULTS

Witworld

In this section, we provide a simple and concise introduction to Witworld, which should enable the understanding of the subsequent results. We moreover provide a detailed formal definition in Section B of the Supplementary Material.

In Witworld, there are three types of basic systems, which can be composed to construct more general system types. The basic systems are classical systems, quantum systems, and Boxworld systems[34]. (One could easily modify the theory to allow for further system types. However, it is not clear that this will provide any further benefit to the study of steering). Systems that are of one of those three types are called *atomic*. Witworld features a composition rule (which we define shortly) by which these simple system types can form new ones that are neither classical, quantum, nor Boxworld. We denote the atomic types diagrammatically with different types of wires by:

$$\left|_{\mathcal{C}_v}\right., \quad \left\|_{\mathcal{Q}_d}\right., \quad \left\|\right|_{\mathcal{B}_{n,k}}\left.\right., \tag{1}$$

where $\mathcal{C}_v$ denotes a classical system of dimension v, $\mathcal{Q}_d$ denotes a quantum system of dimension d, and $\mathcal{B}_{n,k}$ denotes a Boxworld system of dimension (n, k) (These two integers relate to the input/

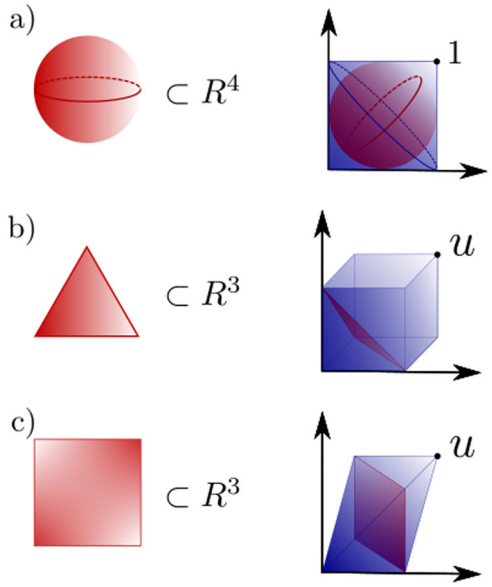
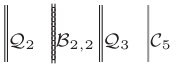


Fig. 1 The geometry of states and effects. The geometry of the set of states (to the left) and effects (to the right) for (**a**) atomic quantum systems of dimension $d = 2$, (**b**) atomic classical systems of dimension $v = 3$, and (**c**) atomic Boxworld systems of dimension $(n, k) = (2, 2)$.

output cardinality of the correlations in Bell scenarios that the system is tailored at[34]). Moreover, when we need to use a generic system type (which can be either simple or composite), we denote this by

(2)

We can also explicitly denote the components of a composite system by using parallel wires, for example:



(3)

corresponds to a system composed of a qubit, a (2, 2) Boxworld system, a qutrit, and a classical system of dimension 5.

The state space of a given system type is represented by a convex set living inside some real vector space. For instance, an atomic quantum system $\mathcal{Q}_2$ has states living inside a Bloch sphere in a 4-dimensional real vector space, an atomic classical state $\mathcal{C}_3$ has states living inside a triangle in a 3-dimensional vector space, and an atomic Boxworld system $\mathcal{B}_{2,2}$ has states inside a square in a 3-dimensional vector space. These examples are depicted in Fig. 1. Diagrammatically we denote a state σ of a system S by

(4)

Regarding the effects, Witworld includes all the elements of the dual of the vector space which evaluate to valid probabilities for every state. That is, Witworld satisfies the no-restriction hypothesis (NRH)[32]. For example, for a system of the type $\mathcal{Q}_2$, the effects correspond to POVM elements and are represented as a particular region of $(\mathbb{R}^4)^*$. This region can be defined as the intersection of the cone of linear functionals which evaluate to positive reals on the set of state vectors with the set of linear functionals which evaluate to less than 1 for all state vectors. For $\mathcal{C}_3$, the effects live in a cube, and for $\mathcal{B}_{2,2}$, the effects live in an octahedron. A pictorial representation of these can be seen in Fig. 1. An effect e for a system S is diagrammatically denoted by

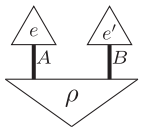
(5)

Since effects belong to the dual vector space, when we compose them with a state we obtain a real number, which, by assumption, must give a valid probability. That is, for all states σ^S and effects e^S we have $e^S(\sigma^S) \in [0, 1]$. Diagrammatically this is written as

$\in \quad [0, 1]$. (6)

As mentioned previously, we define Witworld to satisfy the NRH. This, however, is not the only simplifying assumption that we make in this construction. Additionally, we define the composition of systems to be via the so-called max tensor product[46] and, hence, that the theory is locally tomographic[35]. Moreover, we demand that Witworld satisfies the generalised no-restriction hypothesis (GNRH). Intuitively, the GNRH is the NRH together with the requirement that every transformation that takes every element of a valid state space to an element of another valid state space is a valid transformation in the theory, that is, every completely positive transformation is considered valid.

The max tensor product (see Definition A.8 in the Supplementary Material) is a composition rule that assigns as valid states of a composite system $A \cdot B$ any vector in the product vector space $(V^{A\cdot B} = V^A \otimes V^B)$ that is consistent with the separable effects. Formally, ρ^{AB} is a valid state of the composite system AB if and only if, for every effect e^A of A and e'^B of B, $e^A \otimes e'^B(\rho^{AB}) \in [0, 1]$. Diagrammatically we express this condition as:

$\forall e, e'$ 

$\in [0, 1]$. (7)

From an intuitive point of view, the max tensor product gives rise to a GPT that somehow maximises the set of states that the system can be prepared in, whilst strongly restricting the set of measurements that one may perform on it. As a matter of fact, even though Witworld might appear to be a more general theory than quantum theory, these two are actually incomparable: quantum theory allows for measurements that Witworld systems cannot be acted upon with (with the latter having a more restricted set of measurements on collections of quantum atomic systems types), whilst Witworld allows for more states on which the composition of quantum atomic system types can be prepared (Witworld allows for two qubits to be prepared on a state mathematically corresponding to a quantum entanglement witness, whereas in quantum theory this is not an allowed state of a two-qubit system).

The fact that we have defined composition via the max tensor product and are demanding that the theory satisfies the GNRH, means that when we define the atomic states, we define the whole theory, since from the atomic states and max tensor product every possible state is defined, and from the states and GNRH the effects and transformations are also determined.

Finally, as mentioned above, the max tensor product is tomographically local, that is to say that its states can be uniquely determined by the information obtained from performing local measurements on its parts. Using the example above, this means that ρ^{AB} is completely determined by a set of values $e_i^A \otimes e_j^B(\rho^{AB})$.

At this point, it is worth mentioning some further consequences of our definitions. The first one is that the use of the max tensor product to compose systems implies that every effect in Witworld

is separable (see Lemma A.14 in the Supplementary Material). Therefore, an important feature of Witworld is that it does not contain entangling effects.

A second important fact about Witworld is that the combination of two atomic quantum systems yields systems whose states spaces are larger than the joint state space obtained from the standard quantum composition rule (see Theorems B.5 and B.6 in the Supplementary Material). For example, in the bipartite case, we have that the composite of $\mathcal{Q}_n$ and $\mathcal{Q}_m$, denoted by $\mathcal{Q}_n \cdot \mathcal{Q}_m$, has as its state space the set of entanglement witnesses (including density matrices[56]) for the quantum bipartite states, which strictly contains the set of bipartite quantum states $\mathcal{Q}_{nm}$. Therefore, whilst Witworld does contain arbitrary quantum systems, quantum theory is not a compositional subtheory within Witworld. Note that if one were to allow quantum systems in Witworld (where these quantum systems do not have the additional dynamics of Witworld quantum systems) to be composable both according to the standard quantum rule as well as to the Witworld composition rule, one could construct a protocol giving negative probabilities[57]. As such one cannot extend Witworld in such a manner as to contain quantum theory as a compositional subtheory.

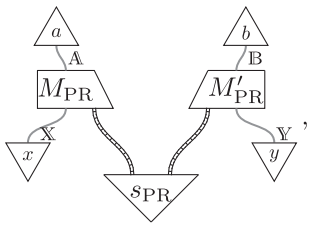
A third important feature is that Boxworld-type (resp. classical-type) systems in Witworld compose exactly as they do in Boxworld (resp. classical theory). Therefore, both Boxworld and classical theory are indeed full subtheories of Witworld. Here, by *full subtheory* we mean that you can recover Boxworld or classical theory from Witworld by suitably restricting it to a particular collection of system types. This restriction recovers all and only the states, effects, transformations, and the composition rules of Boxworld or classical theory.

Finally, another important feature of Witworld is that because of the combination of GNRH and max tensor product, there is no difference between positive and completely positive maps (see Lemma A.15 in Supplementary Material). Of course, for systems that are not quantum, a more general notion (relative to that of positive operators in quantum theory) of positivity must be used in order to make that statement (see Definition A.6 in the Supplementary Material). Now, in the case of atomic quantum systems, this means that the valid Witworld transformations correspond to positive, but not necessarily completely positive, quantum transformations. Hence, this means that in Witworld there are more transformations available to local agents (i.e. to Alice and Bob) than would be available in quantum theory.

Post-quantum phenomena: Bell non-classicality and steering

In this section we explore the non-classical features that Witworld displays, starting with the case of Bell scenarios. One can readily see that Witworld can realise all non-signalling correlations in arbitrary Bell scenarios (see Fig. 2), since Boxworld is a full subtheory of Witworld. Therefore, one can leverage the Boxworld realisations of any non-signalling correlation, and translate them straightforwardly to a realisation within Witworld. For example, take the case of Popescu-Rohrlich (PR) correlations, which read $p_{PR}(ab|xy) = \frac{1}{2}\delta_{a\oplus b=xy}$ with $a, b, x, y \in \{0, 1\}$ and $\oplus$ denoting modulo-2 addition; these correlations can be realised in Witworld as follows:

$$p_{\mathrm{PR}}(ab|xy) = $$



(8)

with the state s_{PR} and controlled measurements M_{PR} and M'_{PR} as introduced in Ref. [34], whose explicit form we present in Eqs. (C8),

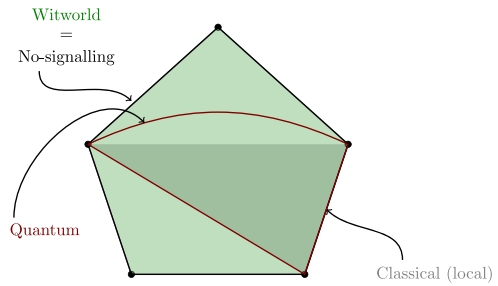


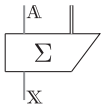
Fig. 2 Pictorial representation of the sets of correlations in Bell scenarios. The polytope of classical (Local Hidden State) correlations is depicted by the grey triangle. The convex set of quantum correlations is depicted by the red-bordered convex region. The set of Witworld correlations is depicted by the green transparent region. The polytope of no-signalling correlations is depicted by the black-bordered pentagon. The sets of no-signalling and Witworld correlations are equivalent. Witworld correlations strictly contain the quantum set, which strictly contains the classical set.

(C3), and (C6) in the Supplementary Material. Note that for simplicity of notation we will often label the classical systems by an outcome or setting variable such as $\mathbb{X}$, in this case the relevant GPT system is $\mathcal{C}_{|\mathbb{X}|}$.

The situation slightly changes when we instead focus on the non-classical phenomenon of steering (see Section D of the Supplementary Material for a comprehensive introduction). In brief, a traditional bipartite steering experiment consists of two distant parties, Alice and Bob, who share a physical system, perform space-like separated actions, and, unlike in Bell experiments, play asymmetric roles in the experiment. On the one hand, Alice (sometimes referred to as the black-box party in the steering literature) chooses a measurement labelled by $x \in \mathbb{X}$ to perform on her share of the system, and obtains a classical outcome $a \in \mathbb{A}$ with probability $p(a|x)$. Bob, on the other hand, merely characterises the quantum state $\rho^B_{a|x}$ to which his subsystem is steered. The information collected in this experiment (Alice's probabilities and Bob's conditional states) is expressed concisely as an assemblage[53]:

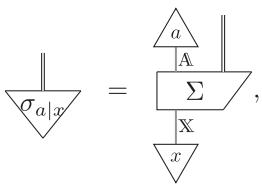
$$\boldsymbol{\Sigma}_{\mathbb{A}|\mathbb{X}} = \left\{\sigma^B_{a|x}\right\}_{a\in\mathbb{A}, x\in\mathbb{X}} \text{ where } \sigma^B_{a|x} := p(a|x)\,\rho^B_{a|x}. \tag{9}$$

Note that in Ref. [54] it is shown how assemblages can be equivalently represented by so-called 'causal' channels. With a slight abuse of notation we therefore diagrammatically represent the assemblage by the causal classical-quantum channel:



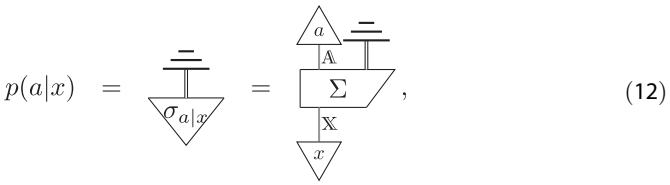
. (10)

To see that this is indeed a good representation, note that we can extract the elements of the assemblage, i.e. the subnormalised steered states as:



, (11)

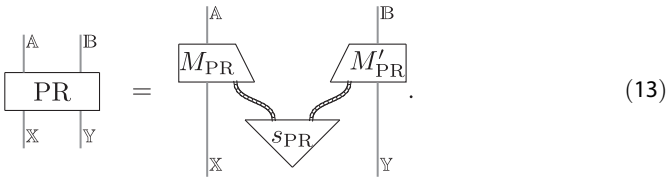
and then the probabilities as:



(12)

where

denotes the so-called unit effect (see Eq. (A11) and its preceding paragraph in the Supplementary Material), which in quantum theory corresponds to the partial trace of the relevant subsystem. Analogously, we can view non-signalling correlations as particular channels, in this case channels with classical inputs and outputs which correspond to stochastic maps. Using this we can rewrite Eq. (8) as



(13)

Beyond the traditional scenario, one may have steering experiments with more black-box parties also in a space-like separated configuration[26,58], or even situations where Bob may influence the state preparation of his system by choosing a classical variable y (Bob-with-input scenarios)[27].

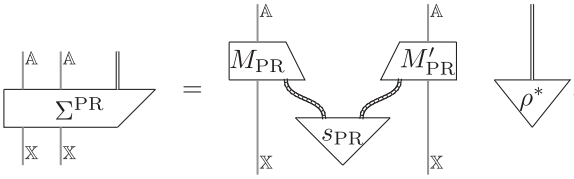
In a similar fashion to Bell non-classicality, one can define what "classical" (a.k.a. LHS), quantum, and non-signalling assemblages are[26,27]. Notice that the differences in all these kinds of steering are not related to the type of system prepared in Bob's lab, but rather to the types of shared resources that are used to prepare those quantum systems in Bob's lab. From the point of view of Witworld, then, an assemblage in a steering experiment is produced by the parties performing local operations in a shared arbitrary composite multipartite system, which may include classical, quantum, and Boxworld systems. One fascinating property of Witworld is that it not only features all LHS and quantum assemblages (see Definitions D.5 and D.6, respectively, in the Supplementary Material), but may also realise post-quantum assemblages. That is, Witworld features post-quantum steering. In this section, we present a few key examples of this. Whether Witworld can realise *all* non-signalling assemblages is still an open question (see Fig. 3).

The first example we present is in a tripartite steering scenario, since in traditional bipartite steering scenarios post-quantum steering is forbidden by the Gisin[59] and Hughston, Josza and Wootters[60] theorems. In a tripartite scenario, it is enough to consider the simplest setup with two black-box parties choosing among two dichotomic measurements each, so $\mathbb{X} = \mathbb{A} = \{0, 1\}$, and where Bob's subsystem is a qubit. The particular assemblage we present is the PR-box assemblage, defined by

$$\boldsymbol{\Sigma}^{\mathrm{PR}}_{\mathrm{AA}|\mathrm{XX}} = \left\{\sigma^{*B}_{a_1a_2|x_1x_2}\right\}_{a_j\in\mathbb{A},x_j\in\mathbb{X}j\in\{1,2\}}, \tag{14}$$

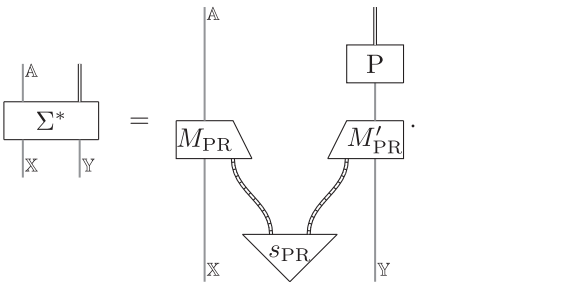
$$\text{with } \sigma^{*B}_{a_1a_2|x_1x_2} = p_{\mathrm{PR}}(a_1a_2|x_1x_2)\,\frac{\mathbb{I}}{2}. \tag{15}$$

This assemblage cannot be realised by the three parties sharing quantum resources[26], i.e. it is post-quantum. $\boldsymbol{\Sigma}^{\mathrm{PR}}_{\mathrm{AA}|\mathrm{XX}}$ can however be realised within Witworld when the parties share the following mutipartite system: a bipartite Boxworld system of dimension (2, 2) on a PR state shared by the black-box parties, composed in parallel with a quantum state $\rho^{*B} = \frac{\mathbb{I}}{2}$ for Bob. Leveraging the realisation of PR-box correlations as in Eq. (8), the assemblage $\boldsymbol{\Sigma}^{\mathrm{PR}}_{\mathrm{AA}|\mathrm{XX}}$ can be realised by:



(16)

The second example we present is in a bipartite Bob-with-input steering scenario, where Alice has $\mathbb{X} = \mathbb{A} = \{0, 1\}$, Bob's subsystem is a qubit, and Bob's input is dichotomic (i.e. $y \in \mathbb{Y} = \{0, 1\}$). The particular post-quantum assemblage $\boldsymbol{\Sigma}^{*}_{\mathrm{A}|\mathrm{XY}}$ we consider has elements defined by $\sigma^{*B}_{a|xy} = \frac{1}{2}(|a\rangle\langle a|\delta_{xy=0} + |a\oplus 1\rangle\langle a\oplus 1|\delta_{xy=1})$[27]. This assemblage can be realised in Witworld by Alice and Bob sharing a bipartite Boxworld system of dimension (2, 2) prepared in a PR state, and implementing the following protocol. On the one hand, Alice performs the measurement M_{PR} of Eq. (8) controlled on her classical input x, and obtains the output a. On the other hand, here the state preparation of Bob's system further depends on Bob's choice of a classical variable y which he inputs in a device. In this protocol, this device has a two-stage process: first it implements the measurement M'_{PR} of Eq. (8) on the Boxworld system, conditioned on y; second, there is a controlled state preparation P which prepares the quantum state $|b\rangle\langle b|$ conditioned on b, the classical output of M'_{PR}. Diagrammatically, the whole protocol reads:



(17)

One can readily see that Eq. (17) indeed holds, since the assemblage elements of $\boldsymbol{\Sigma}^{*}_{\mathrm{A}|\mathrm{XY}}$ can be rewritten as $\sigma^{*B}_{a|xy} = \frac{1}{2}|a\oplus xy\rangle\langle a\oplus xy|$, and PR-box correlations satisfy $b = a\oplus xy$ and $\frac{1}{2} = \sum_b p_{\mathrm{PR}}(ab|xy)$.

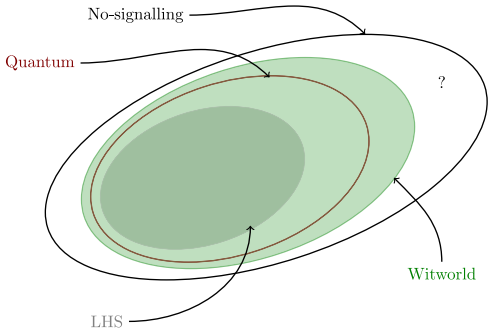
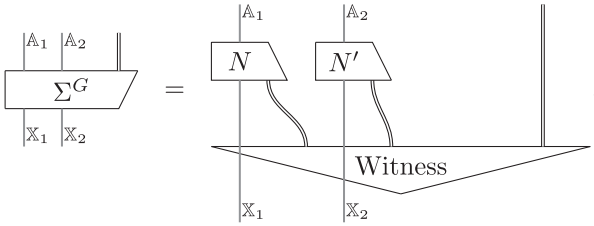


Fig. 3 Pictorial representation of the sets of assemblages in steering scenarios. The set of classical (Local Hidden State) assemblages is depicted by the grey ellipse. The set of quantum assemblages is depicted by the red-bordered region. The set of Witworld assemblages is depicted by the green transparent region. The set of no-signalling assemblages is depicted by the black-bordered ellipse. The set of Witworld assemblages strictly contains the quantum set, which strictly contains the LHS set. The set of no-signalling assemblages contains the Witworld set, and an open question (pictorially depicted by the question mark) is whether this inclusion is strict.

The third example we present is that of Gleason assemblages[55]. In short, Gleason assemblages are those that can be mathematically expressed in the language of quantum theory by having the parties measure a shared system whose state preparation is represented by a normalised quantum entanglement witness. Gleason assemblages are particularly useful, since there are constructions that yield provably post-quantum Gleason assemblages. More importantly, the post-quantumness of some Gleason assemblages is not implied by post-quantum Bell non-locality, which renders post-quantum steering as a genuinely new effect[55]. Witworld readily provides realisations of any Gleason assemblage, by noticing two facts: (i) any quantum entanglement witness is a valid state of composite quantum-type systems in Witworld (Theorem B.6), and (ii) in Witworld, any local quantum measurement is a valid Witworld measurement (Lemma 8.14). The explicit diagram for a Witworld realisation of a generic Gleason assemblage $\boldsymbol{\Sigma}^{G}_{\mathbb{A}_1\mathbb{A}_2|\mathbb{X}_1\mathbb{X}_2}$ in a tripartite scenario with two black-box parties is:

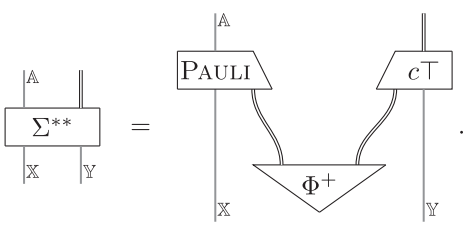


(18)

The fourth example that we present is in a bipartite Bob-with-input steering scenario, with $\mathbb{A}=\mathbb{Y}=\{0,1\}$ and $\mathbb{X}=\{1,2,3\}$. The particular post-quantum assemblage $\boldsymbol{\Sigma}^{**}_{\mathbb{A}|\mathbb{X}\mathbb{Y}}$ here has elements defined by Sainz et al.[27]:

$$\sigma^{**B}_{a|xy}=\frac{1}{4}\left(\mathbb{I}+(-1)^{a+\delta_{x,2}\delta_{y,1}}\Sigma_x\right), \tag{19}$$

where $(\Sigma_1,\Sigma_2,\Sigma_3)$ are the Pauli X, Y, and Z operators, respectively. $\boldsymbol{\Sigma}^{**}_{\mathbb{A}|\mathbb{X}\mathbb{Y}}$ is the first assemblage found in the Bob-with-input scenario whose post-quantumness cannot be proven directly from leveraging post-quantum Bell non-locality, which renders this type of post-quantum steering as a genuinely new effect. To see that Witworld can realise this assemblage, first notice that its elements can be mathematically written as $\sigma^{**B}_{a|xy}=(\overline{\sigma}^{B}_{a|xy})^{\mathsf{T}^y}$, where $\overline{\sigma}^{B}_{a|xy}=\frac{1}{4}(\mathbb{I}+(-1)^a\Sigma_x)$ are the elements of a quantum assemblage (see Definition D.6 in the Supplementary Material), and T_y is the identity operator for $y=0$ and the Transpose operator (denoted T) for $y=1$. The final step is to observe that all of these mathematical objects are acceptable physical operations in Witworld: (i) the maximally entangled quantum state that realises $\{\overline{\sigma}^{B}_{a|xy}\}$ is a valid Witworld state preparation for Alice and Bob by Theorem B.6, (ii) the quantum Pauli measurements for Alice that realise $\{\overline{\sigma}^{B}_{a|xy}\}$ are valid Witworld measurements by Lemma A.14, where we denote the classically controlled Pauli measurement by PAULI, and (iii) the operations $\{\mathsf{T}_y\}$, which are positive quantum maps, are valid Witworld transformations by Theorem A.15, and, hence, so too is the controlled transformation $c\mathsf{T}$ which implements T conditioned on a classical input system. Diagrammatically, this is represented as:

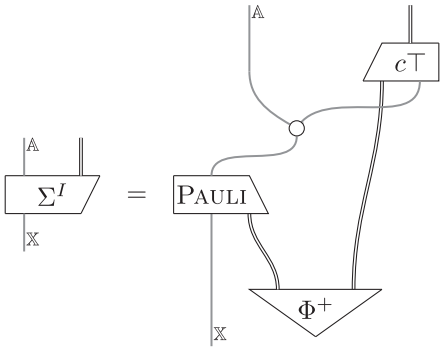


(20)

The final example that we consider is steering in the instrumental scenario[27]. This can be seen as an adaptation of the Bob-with-input scenario, in which Alice's output a determines the setting y for Bob. The particular example of post-quantum steering in this scenario that we present here is given by modifying our previous example, by wiring Alice's output to Bob's input. That is, it can be shown that the assemblage

$$\sigma^{IB}_{a|x}=\frac{1}{4}\left(\mathbb{I}+(-1)^{a+\delta_{x,2}\delta_{a,1}}\Sigma_x\right), \tag{21}$$

which is obtained by setting $y=a$ in Eq. (19), is post-quantum[27]. It is then a simple modification of Eq. (20) to see that this too can be realised in Witworld:



(22)

where the small white circle splitting the classical system is the copy operation.

With this, we see that Witworld features a variety of non-classical and post-quantum properties, both in Bell and steering scenarios, and hence is the first GPT that has been shown to display post-quantum steering.

Post-quantum advantage for information processing

Post-quantum resources may outperform quantum ones for information processing tasks[13,17,18,50,61,62]. A natural question then is whether the post-quantum features of Witworld enable this theory to be more powerful than quantum theory in this respect. First, one can focus on device-independent information processing tasks, such as quantum cryptography[17,18,50], which rely on the use of correlations in Bell scenarios. Here, it is known that Boxworld may outperform quantum theory, since it realises any non-signalling correlation. Since Boxworld is a subtheory of Witworld, then, the latter inherits these properties; that is, Witworld outperforms quantum theory in those device-independent information processing tasks. A more relevant question then is whether such advantage persists when moving on from device-independent tasks. Hence in this section, we investigate whether Witworld provides an advantage for tasks that go beyond the processing of Bell-type correlations.

There are two features of Witworld that go beyond Boxworld which are noteworthy when looking for an information task where Witworld is resourceful. One is the fact that Witworld has quantum systems as atomic system types, and the other is the fact that positive (but not necessarily completely positive) quantum operations are allowed physical operations in Witworld. Using these two facts we first show that Witworld outperforms Quantum Theory in the task of Remote State Preparation, and then we show that the resource underlying this advantage is post-quantum steering.

Remote State Preparation (RSP)[28,29] is a protocol with a similar flavour to state teleportation. A main difference between teleportation and RSP is that in the former, Alice can send to Bob a state she knows nothing about, whereas in the latter she may require a complete classical description of $|\psi\rangle$. We denote this complete classical description by $\boldsymbol{\psi}$. In both cases, the main goal is for Alice to deterministically prepare a state $|\psi\rangle$ in Bob's lab, such that he gets no additional information about $|\psi\rangle$. In RSP (see Fig. 4), however, Alice does not need to perform experimentally challenging entangling measurements (as in a full Bell-state

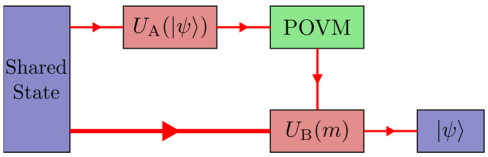
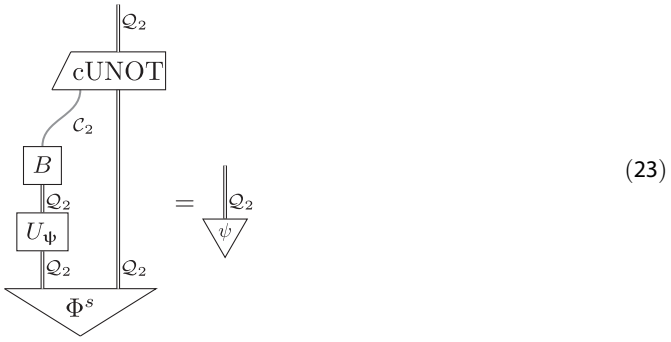


Fig. 4 Pictorial representation of the remote state preparation protocol. Alice performs $U_A(|\psi\rangle)$ on her share of a physical system – a unitary operation that depends on $|\psi\rangle$ – and then a measurement (POVM). She sends a classical message m to Bob, who, in turn, performs an unitary transformation (which depends on m) on his share of the system. The outcome of the protocol is a quantum state on state $|\psi\rangle$ on Bob's lab.

analysis)[30]. Instead, she can directly encode the information about the state she wishes to send onto her share of an entangled state shared with Bob. When Alice and Bob use quantum resources, the minimum amount of classical information that she needs to send him for the protocol to succeed is $2\log d$ bits of information, where d is the dimension of the Hilbert space containing $|\psi\rangle$[29]. Here we present a protocol using Witworld resources which may prepare an arbitrary qubit state in Bob's lab using only 1 bit (instead of 2) of classical communication.

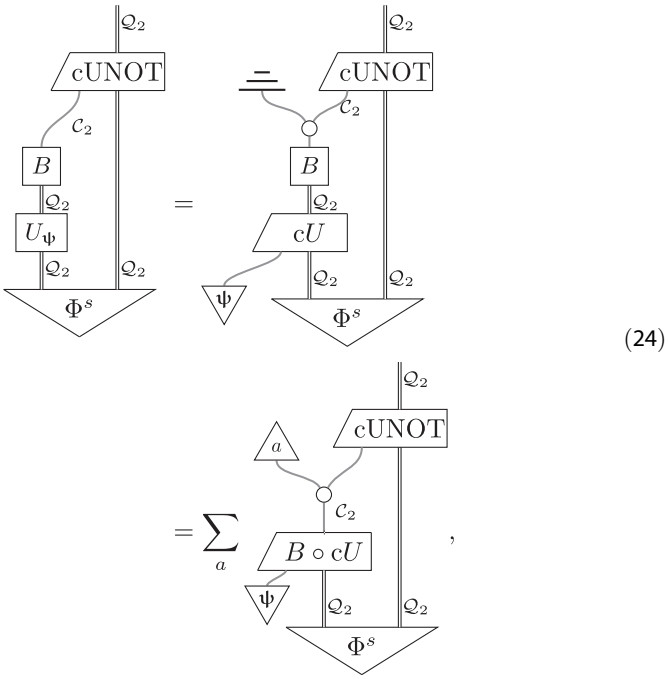
Consider the following protocol in Witworld. Alice and Bob share the two qubit state $|\Phi^s\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$. Alice performs the unitary $U_{\boldsymbol{\psi}} = |0\rangle\langle\psi^\perp| + |1\rangle\langle\psi|$ on her qubit, which encodes the state $|\psi\rangle$ to be sent. This effectively applies $U^\dagger_{\boldsymbol{\psi}}$ to Bob's half of the state (the singlet state $|\Phi^s\rangle$ transforms trivially under $U\otimes U$; implying that $(U_{\boldsymbol{\psi}}\otimes\mathbb{I})|\Phi^s\rangle = (\mathbb{I}\otimes U^\dagger_{\boldsymbol{\psi}})|\Phi^s\rangle$). Next, she performs the measurement given by $B = \{|0\rangle\langle 0|, |1\rangle\langle 1|\}$, whose outcome consists of one classical bit a which indicates exactly whether Bob now has the post measured state $-|\psi\rangle$ (if $a=0$) or $|\psi^\perp\rangle$ (if $a=1$). Then, Alice sends a to Bob, who now knows whether he is holding $-|\psi\rangle$ or $|\psi^\perp\rangle$. The task can be completed if Bob has access to a universal-NOT operation, which maps an arbitrary input $|\phi\rangle$ into an orthogonal state to it (which is unique up to global phases for qubits). The universal-NOT operation is not valid in quantum theory since it is a positive transformation, but not a completely positive transformation. However, in Witworld, this is an allowable transformation. Thus in Witworld Bob can apply the universal-NOT gate when $a=1$, leaving him with a perfect copy of $|\psi\rangle$ (up to a physically irrelevant global phase). Diagrammatically, this protocol is represented as follows:



(23)

where cUNOT is the controlled-universal-NOT operation. The diagrammatic manipulations that prove that Eq. (23) holds are presented in Section E of the Supplementary Material. Through this protocol, Witworld performs RSP of a qubit deterministically with the transmission of only one classical bit from Alice to Bob, outperforming quantum theory at the task.

We now move on to unveiling what the critical resource is underlying the success of RSP in Witworld. For this, it is convenient to rewrite the diagram in the left hand side of Eq. (23) as:



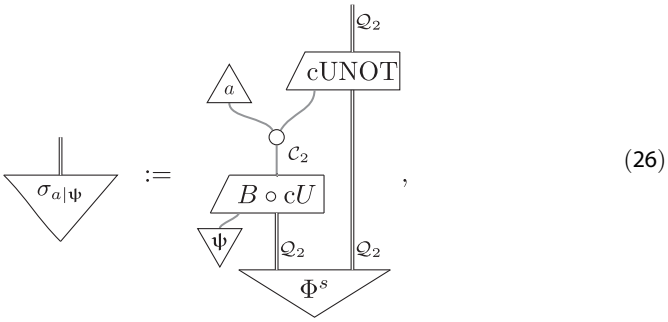
(24)

where

(25)

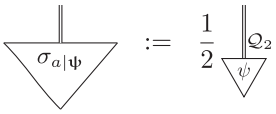
is not the state $|\psi\rangle$, but simply a classical label corresponding to it, used to determine the unitary U_ψ that the transformation cU implements. In addition, $B\circ cU$ is the process that first implements the controlled unitary cU and then the measurement B.

The crucial step here is to notice that each term in the sum in Eq. (24) can be identified with an element of an assemblage $\{\sigma_{a|\boldsymbol{\psi}}\}$ in an instrumental steering scenario (see Definition D.4 in the Supplementary Material) as follows:



(26)

where a denotes Alice's dichotomic outcome, and $\boldsymbol{\psi}$ is the classical variable that denotes her measurement choice. That is, RSP is ultimately an instance of an instrumental steering scenario, and the possible assemblages that Alice can prepare dictates whether RSP is possible for the given cardinality of a. For the particular RSP protocol discussed above,



(27)

It is readily seen that the assemblage $\{\sigma_{a|\boldsymbol{\psi}}\}$ has no quantum realisation: if this was instead the case, this assemblage would provide a quantum RSP protocol that succeeds deterministically with 1 bit of communication, which is fundamentally impossible.

We see therefore how instrumental steering powers RSP, and how the post-quantum steering featured in Witworld makes this theory more efficient than quantum theory at the task of Remote State Preparation.

Let us observe that quantum theory restricted to the reals[63], which has mixed states given by symmetric matrices (a subset of quantum states), also requires a single bit of communication for RSP. A rebit (2 dimensional real quantum system) has mixed states given by the $X-Z$ plane of the Bloch sphere (a disk). The universal NOT is just rotation by π around the Y axis, and is completely positive. Since the singlet state $|\Phi^s\rangle\langle\Phi^s|$ is a real valued density operator (i.e. it is a symmetric matrix) it follows that it is a valid entangled state of two rebits. Hence the protocol outlined above in Witworld can be applied to real quantum theory as well, to give RSP with a single bit of communication.

DISCUSSION

In this work, we explored the scope of post-quantum steering as a stronger-than-quantum resource for information processing. We particularly focused the search on tasks beyond device-independent ones or those that ultimately rely on Bell correlations (such as random access codes[64,65] or device-independent quantum key distribution): we aimed at finding tasks that intrinsically leveraged quantum systems and non-classical steering. We discovered that remote state preparation of qubits systems provides a friendly proof-of-principle of a general phenomenology: steering assemblages in the instrumental scenario serve as a resource for the task, and post-quantum assemblages perform better than quantum ones at it. This is the first time that post-quantum steering – as opposed to post-quantum Bell non-classicality – has been identified as a resource powering information processing which can provably outperform quantum theory.

In order to prove our claims, we defined a generalised probabilistic theory, that we call Witworld, by combining classical, quantum, and Boxworld systems in a simple mathematical way, via the max tensor product. The intuitive formulation of Witworld allowed us to present its powerful post-quantum features in an accessible way: one can readily see how post-quantum Bell nonlocality, post-quantum steering, and post-quantum states emerge within Witworld. The task of remote state preparation can be studied diagrammatically within Witworld, and by doing so we showed how the post-quantum assemblages allowed by the theory makes Witworld perform better at it than quantum theory does.

A feature of Witworld is that, even though it is built in part on quantum systems, it does not contain quantum theory as a subtheory: there are tasks, such as quantum teleportation, that quantum theory can perform whilst Witworld cannot. The reason for this is the choice of composition rule: Witworld composes via the max tensor product, and hence no entangling measurements are allowed in the theory. Nonetheless, Witworld remarkably succeeds at reproducing all the quantum entangled states, quantum steering assemblages, and quantum correlations in Bell scenarios. That is, for the non-classical phenomena usually leveraged in quantum information tasks, Witworld is at least as good as quantum theory at manifesting them.

If we turn our attention to a particular subtheory of Witworld by restricting the system types to classical and quantum only – that is, by removing Boxworld from the theory – we find that this subtheory still features post-quantum properties, such as Bell non-classicality in multipartite scenarios (for example, by utilising the results of Ref. [56]), as well as post-quantum steering and post-quantum states even in bipartite scenarios. Remarkably, the post-quantum advantage for remote state preparation is also featured by this subtheory of Witworld, since the post-quantum advantage provided by it stems from the enlarged set of operations allowed on local quantum systems. We leave it as an open question whether other previously defined GPTs (e.g. Refs. [66,67]) may provide such an advantage for this information processing task.

It is worth noticing that Witworld's simple formulation does not make the theory intrinsically groundbreaking from the perspective of generalised probabilitic theories, however its relevance is not grounded in its appeal as a standalone GPT. Rather, Witworld shows that there exists a compositional theory that could underpin post-quantum effects such as post-quantum steering. This shows that the latter phenomenon in not in principle un-realisable, and hence its study should not be simply dismissed.

Looking ahead, there are a variety of open questions that can be studied, especially about the extent to which post-quantum steering compatible with special relativity can be underpinned by some generalised theory. We know that Witworld may display post-quantum steering but, unlike the case of Bell non-classicality, it is still unknown whether any no-signalling assemblage may have a realisation within Witworld. Any answer to this question would be of interest: if Witworld can realise all no-signalling assemblages, then this theory becomes the first GPT to accommodate steering fully in a common-cause resource theoretic framework[68]; otherwise, understanding the reason behind the gap between no-signalling realisable and Witworld realisable assemblages may lead to an operational principle that could shed light on the characterisation of quantum phenomena.

Finally, the exploration of the information processing power of steering (quantum and beyond) has only just begun. Since Witworld is formulated in an intuitive way leveraging a diagrammatic representation[32,33,69–71], there is plenty of scope for investigating other post-quantum advantages of this theory, and of post-quantum steering, for information processing.

Received: 16 May 2021; Accepted: 3 May 2022;
Published online: 30 June 2022

REFERENCES

1. Einstein, A., Podolsky, B. & Rosen, N. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **47**, 777–780 (1935).
2. Schrödinger, E. Die gegenwärtige situation in der quantenmechanik. *Naturwissenschaften* **23**, 807–812 (1935).
3. Bell, J. S. On the Einstein Podolsky Rosen paradox. *Phys. Phys. Fiz.* **1**, 195–200 (1964).
4. Schrödinger, E. Probability relations between separated systems. *Math. Proc. Camb. Philos. Soc.* **32**, 446–452 (1936).
5. Wiseman, H. M., Jones, S. J. & Doherty, A. C. Steering, entanglement, nonlocality, and the Einstein-Podolsky-Rosen paradox. *Phys. Rev. Lett.* **98**, 140402 (2007).
6. Jones, S. J., Wiseman, H. M. & Doherty, A. C. Entanglement, Einstein-Podolsky-Rosen correlations, bell nonlocality, and steering. *Phys. Rev. A* **76**, 052116 (2007).
7. Hensen, B. et al. Loophole-free bell inequality violation using electron spins separated by 1.3 kilometres. *Nature* **526**, 682–686 (2015).
8. Shalm, L. K. et al. Strong loophole-free test of local realism. *Phys. Rev. Lett.* **115**, 250402 (2015).
9. Giustina, M. et al. Significant-loophole-free test of bell's theorem with entangled photons. *Phys. Rev. Lett.* **115**, 250401 (2015).
10. Bennet, A. J. et al. Arbitrarily loss-tolerant Einstein-Podolsky-Rosen steering allowing a demonstration over 1 km of optical fiber with no detection loophole. *Phys. Rev. X* **2**, 031003 (2012).
11. Wittmann, B. et al. Loophole-free Einstein–Podolsky–Rosen experiment via quantum steering. *N. J. Phys.* **14**, 053030 (2012).
12. Smith, D. et al. Conclusive quantum steering with superconducting transition-edge sensors. *Nat. Commun.* **3**, 1–6 (2012).

13. Brunner, N., Cavalcanti, D., Pironio, S., Scarani, V. & Wehner, S. Bell nonlocality. *Rev. Mod. Phys.* **86**, 419–478 (2014).
14. Cavalcanti, D. & Skrzypczyk, P. Quantum steering: a review with focus on semidefinite programming. *Rep. Prog. Phys.* **80**, 024001 (2016).
15. Uola, R., Costa, A. C. S., Nguyen, H. C. & Gühne, O. Quantum steering. *Rev. Mod. Phys.* **92**, 015001 (2020).
16. Popescu, S. & Rohrlich, D. Quantum nonlocality as an axiom. *Found. Phys.* **24**, 379–385 (1994).
17. Barrett, J., Hardy, L. & Kent, A. No signaling and quantum key distribution. *Phys. Rev. Lett.* **95**, 010503 (2005).
18. Acín, A., Gisin, N. & Masanes, L. From bell's theorem to secure quantum key distribution. *Phys. Rev. Lett.* **97**, 120405 (2006).
19. Brassard, G. et al. Limit on nonlocality in any world in which communication complexity is not trivial. *Phys. Rev. Lett.* **96**, 250401 (2006).
20. Pawłowski, M. et al. Information causality as a physical principle. *Nature* **461**, 1101–1104 (2009).
21. Navascués, M. & Wunderlich, H. A glance beyond the quantum model. *Proc. R. Soc. A* **466**, 881–890 (2010).
22. Fritz, T. et al. Local orthogonality as a multipartite principle for quantum correlations. *Nat. Commun.* **4**, 1–7 (2013).
23. Sainz, A. B. et al. Exploring the local orthogonality principle. *Phys. Rev. A* **89**, 032117 (2014).
24. Acín, A., Fritz, T., Leverrier, A. & Sainz, A. B. A combinatorial approach to nonlocality and contextuality. *Commun. Math. Phys.* **334**, 533–628 (2015).
25. Henson, J. & Sainz, A. B. Macroscopic noncontextuality as a principle for almost-quantum correlations. *Phys. Rev. A* **91**, 042114 (2015).
26. Belén Sainz, A., Brunner, N., Cavalcanti, D., Skrzypczyk, P. & Vértesi, T. Post-quantum steering. *Phys. Rev. Lett.* **115**, 190403 (2015).
27. Sainz, A. B., Hoban, M. J., Skrzypczyk, P. & Aolita, L. Bipartite postquantum steering in generalized scenarios. *Phys. Rev. Lett.* **125**, 050404 (2020).
28. Lo, H.-K. Classical-communication cost in distributed quantum-information processing: a generalization of quantum-communication complexity. *Phys. Rev. A* **62**, 012313 (2000).
29. Leung, D. W. & Shor, P. W. Oblivious remote state preparation. *Phys. Rev. Lett.* **90**, 127905 (2003).
30. Peters, N. A., Barreiro, J. T., Goggin, M. E., Wei, T.-C. & Kwiat, P. G. Remote state preparation: arbitrary remote control of photon polarization. *Phys. Rev. Lett.* **94**, 150502 (2005).
31. Childs, A. M. Secure assisted quantum computation. *Quantum Info Comput.* **5**, 456–466 (2005).
32. Chiribella, G., D'Ariano, G. M. & Perinotti, P. Probabilistic theories with purification. *Phys. Rev. A* **81**, 062348 (2010).
33. Hardy, L. Reformulating and reconstructing quantum theory. Preprint at https://arxiv.org/abs/1104.2066 (2011).
34. Barrett, J. Information processing in generalized probabilistic theories. *Phys. Rev. A* **75**, 032304 (2007).
35. Hardy, L. Quantum theory from five reasonable axioms. Preprint at https://arxiv.org/abs/quant-ph/0101012 (2001).
36. Dakic, B. & Brukner, C. Quantum theory and beyond: Is entanglement special? In *Deep Beauty: Understanding the Quantum World through Mathematical Innovation*, (Cambridge University Press, 2011) https://doi.org/10.1017/CBO9780511976971.011.
37. Chiribella, G., D'Ariano, G. M. & Perinotti, P. Informational derivation of quantum theory. *Phys. Rev. A* **84**, 012311 (2011).
38. Barnum, H., Müller, M. P. & Ududec, C. Higher-order interference and single-system postulates characterizing quantum theory. *N. J. Phys.* **16**, 123029 (2014).
39. Barrett, J., de Beaudrap, N., Hoban, M. J. & Lee, C. M. The computational landscape of general physical theories. *npj Quantum Inf.* **5**, 1–10 (2019).
40. Lee, C. M. & Selby, J. H. Generalised phase kick-back: the structure of computational algorithms from physical principles. *N. J. Phys.* **18**, 033023 (2016).
41. Barnum, H., Lee, C. M. & Selby, J. H. Oracles and query lower bounds in generalised probabilistic theories. *Found. Phys.* **48**, 954–981 (2018).
42. Lee, C. M. & Barrett, J. Computation in generalised probabilisitic theories. *N. J. Phys.* **17**, 083001 (2015).
43. Krumm, M. & Müller, M. P. Quantum computation is the unique reversible circuit model for which bits are balls. *npj Quantum Inf.* **5**, 1–8 (2019).
44. Garner, A. J. P. Interferometric computation beyond quantum theory. *Found. Phys.* **48**, 886–909 (2018).
45. Lee, C. M. & Selby, J. H. Deriving Grover's lower bound from simple physical principles. *N. J. Phys.* **18**, 093047 (2016).
46. Barnum, H. & Wilce, A. Information processing in convex operational theories. *ENTCS* **270**, 3–15 (2011).
47. Lami, L., Palazuelos, C. & Winter, A. Ultimate data hiding in quantum mechanics and beyond. *Commun. Math. Phys.* **361**, 661–708 (2018).
48. Sikora, J. & Selby, J. H. Impossibility of coin flipping in generalized probabilistic theories via discretizations of semi-infinite programs. *Phys. Rev. Res.* **2**, 043128 (2020).
49. Selby, J. H. & Sikora, J. How to make unforgeable money in generalised probabilistic theories. *Quantum* **2**, 103 (2018).
50. Vazirani, U. & Vidick, T. Fully device-independent quantum key distribution. *Phys. Rev. Lett.* **113**, 140501 (2014).
51. Short, A. J. & Barrett, J. Strong nonlocality: a trade-off between states and measurements. *N. J. Phys.* **12**, 033034 (2010).
52. Gross, D., Müller, M., Colbeck, R. & Dahlsten, O. C. O. All reversible dynamics in maximally nonlocal theories are trivial. *Phys. Rev. Lett.* **104**, 080402 (2010).
53. Pusey, M. F. Negativity and steering: a stronger peres conjecture. *Phys. Rev. A* **88**, 032313 (2013).
54. Hoban, M. J. & Sainz, A. B. A channel-based framework for steering, non-locality and beyond. *N. J. Phys.* **20**, 053048 (2018).
55. Belén Sainz, A., Aolita, L., Piani, M., Hoban, M. J. & Skrzypczyk, P. A formalism for steering with local quantum measurements. *N. J. Phys.* **20**, 083040 (2018).
56. Acín, A. et al. Unified framework for correlations in terms of local quantum observables. *Phys. Rev. Lett.* **104**, 140404 (2010).
57. Barnum, H., Fuchs, C. A., Renes, J. M. & Wilce, A. Influence-free states on compound quantum systems. Preprint at https://arxiv.org/abs/quant-ph/0507108 (2005).
58. Cavalcanti, D. et al. Detection of entanglement in asymmetric quantum networks and multipartite quantum steering. *Nat. Commun.* **6**, 1–6 (2015).
59. Gisin, N. Stochastic quantum dynamics and relativity. *Helvetica Phys. Acta* **62**, 363–371 (1989).
60. Hughston, L. P., Jozsa, R. & Wootters, W. K. A complete classification of quantum ensembles having a given density matrix. *Phys. Lett. A* **183**, 14–18 (1993).
61. Colbeck, R. & Renner, R. Free randomness can be amplified. *Nat. Phys.* **8**, 450–453 (2012).
62. Pironio, S. et al. Random numbers certified by bell's theorem. *Nature* **464**, 1021–1024 (2010).
63. Hardy, L. & Wootters, W. K. Limited holism and real-vector-space quantum theory. *Found. Phys.* **42**, 454–473 (2012).
64. Wiesner, S. Conjugate coding. *SIGACT N.* **15**, 78–88 (1983).
65. Ambainis, A., Nayak, A., Ta-Shma, A. & Vazirani, U. Dense quantum coding and a lower bound for 1-way quantum automata, In *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*, STOC '99 (Association for Computing Machinery, New York, NY, USA, 1999) p. 376–383. https://doi.org/10.1145/301250.301347.
66. Barnum, H., Graydon, M. A. & Wilce, A. Composites and categories of Euclidean Jordan Algebras. *Quantum* **4**, 359 (2020).
67. Hefford, J. & Gogioso, S. Hyper-decoherence in density hypercubes. In *Proceedings of QPL*, 340 (EPTCS, 2020) pp. 141–159. https://doi.org/10.4204/EPTCS.340.7.
68. Schmid, D. et al. Postquantum common-cause channels: the resource theory of local operations and shared entanglement. *Quantum* **5**, 419 (2021).
69. Coecke, B. & Kissinger, A. Picturing quantum processes. In *International conference on theory and application of diagrams* (Springer, 2018) pp. 28–31.
70. Gogioso, S. & Scandolo, C. M. Categorical probabilistic theories, EPTCS 266, 2018, pp. 367–385 https://doi.org/10.4204/EPTCS.266.23.
71. Selby, J. H., Scandolo, C. M. & Coecke, B. Reconstructing quantum theory from diagrammatic postulates. *Quantum* **5**, 445 (2021).

ACKNOWLEDGEMENTS

We thank an anonymous referee for their useful feedback on our manuscript. 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). This research was supported by Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Research, Innovation and Science. All of the diagrams within this manuscript were prepared using TikZit.

AUTHOR CONTRIBUTIONS

All authors contributed equally to the writing and editing of the manuscript. T.D.G., A.B.S., J.H.S., J.S. conceived of this project during the PIMan workshop at Chapman University in 2018. The bulk of the results were formally proven by P.J.C. under the supervision of A.B.S. and J.H.S. and in collaboration with T.D.G. and J.S. Diagrams were created by P.J.C. and J.H.S. using TikZiT. Figures were created by P.J.C. and A.B.S.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41534-022-00574-8.

Correspondence and requests for materials should be addressed to Jamie Sikora.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Post-quantum steering is a stronger-than-quantum resource for information processing

Supplemental Material

Paulo J. Cavalcanti,[1] John H. Selby,[1] Jamie Sikora,[2] Thomas D. Galley,[3] and Ana Belén Sainz[1]

[1] *International Centre for Theory of Quantum Technologies, University of Gdańsk, 80-308 Gdańsk, Poland*
[2] *Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA*
[3] *Perimeter Institute for Theoretical Physics, 31 Caroline St. N, Waterloo, Ontario, N2L 2Y5, Canada*

(Dated: June 17, 2022)

Appendix A: Generalised Probabilistic Theories

1. The basics

In this section, we provide a description of generalised probabilistic theories (GPTs) that have the property of being locally tomographic [35]. Roughly speaking, the GPT framework is a general framework to formulate and describe theories (including quantum theory), that allows for the calculation of probabilities of measurement outcomes when system preparations have states associated to them. In this appendix, we aim for a description of GPTs which connects the diagrammatic [32, 33] and the linear algebraic notations, hoping to make it useful to a broader audience. Although a more general kind of GPT could be defined, for the purpose of defining Witworld, restricting the present discussion to locally tomographic GPTs significantly simplifies the task in hand. The interested reader can find a more general definition of GPTs in, for example, Ref. [32].

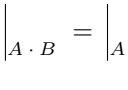
In general, a GPT consists of collections of states that are associated to different system types, a rule for combining state spaces of simple systems into state spaces of composite systems, collections of allowed transformations between these states, and effects – i.e., functions that associate probabilities to each outcome of each measurement for each state in the theory. Here, as mentioned above, we focus on locally tomographic GPTs, which are those where the states of composite systems can be uniquely determined by the information given by local measurements on its parts. Each of these ingredients are defined and compared to quantum theory in what follows.

We start with the states. For each system A of a GPT, there is a vector space V^A associated to it. A convex subset $\Omega^A \subset V^A$, called the state space, defines the allowed states of the system A. This subset has dimension $\mathsf{dim}(\Omega^A) = \mathsf{dim}(V^A) - 1$. We require every state in Ω^A to be normalised in a sense to be defined later in this section when we introduce effects. The convexity property means that, if $\sigma \in \Omega^A$ and $\rho \in \Omega^A$, then $p\sigma + (1-p)\rho \in \Omega^A$ for any $p \in [0,1]$. We require convexity so that the GPT accommodates statistical mixtures of state preparations in a natural way. Diagrammatically, the system A and its associated vector space V^A are represented by a labelled wire:

(A1)

A remark on notation is in order: throughout this section (i.e., Sec. A), we review definitions and properties of a general class of GPTs,which Witworld belongs to, but we do not restrict the presentation to the latter. Hence, the wire type in Eq. (A1) should here be understood as a generic system rather than as a classical system in Witworld (as per Eq. (1)). From Sec. B we will shift the focus back to Witworld, and hence the notation from Eq. (1) will take precedence again. In the case of quantum theory, hence, the wires in Eq. (A1) represent real vector spaces of Hermitian operators on Hilbert spaces. For instance, if A is a qubit system, the wire labelled by A represents the vector space $V^A = \{O \in \mathcal{L}(\mathbb{C}^2, \mathbb{C}^2) : O = O^\dagger\}$, where $\mathcal{L}(\mathbb{C}^2, \mathbb{C}^2)$ is the space of linear operators on $\mathbb{C}^2$. Then, Ω^A is the set of positive operators whose trace is 1, that is, $\Omega^A = \{\rho \in V^A : \rho \geq 0, \mathrm{tr}\,\rho = 1\}$, which is indeed a convex set as required. It is sometimes useful, moreover, to include within the GPT formulation of quantum theory some wires that represent classical variables which store the results of measurements, see, for example, Refs. [69–71]. This is done in section II A of the present work.

As mentioned previously, one can construct composite systems by the combination of simpler systems. We denote by $A \cdot B$ the system composed by a system A and a system B. Its states belong to the set $\Omega^{A\cdot B} \subset V^{A\cdot B}$, which is represented diagrammatically by multiple wires side by side:



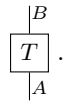
(A2)

In the locally tomographic GPTs that we consider here, such as quantum theory, $V^{A\cdot B} = V^A \otimes V^B$. For convenience,we omit the label when the exact system being discussed is not relevant or it is clear from the context, or we use different kinds of wires to highlight the distinctions, as we do in Sec. II A. If we want to refer diagrammatically to a specific state of A, that is, some element s of Ω^A, we use a box (usually, but not necessarily, a triangular box) with an output wire A connected to its

top:

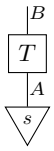
$$s \equiv$$  . (A3)

The transformations in a (tomographically local) GPT are linear functions from the vector space V^A associated to a system of some type A to the vector space V^B associated to some system of type B. Hence, the set of transformations of type $A \to B$, denoted by $\mathcal{T}^{A\to B}$ is a subset of $\mathcal{L}(V^A, V^B)$, the set of linear transformations from V^A to V^B. Diagrammatically, a particular transformation $T \in \mathcal{T}^{A\to B}$ is denoted by a box with an input wire A connected to its bottom and an output wire B connected to its top:

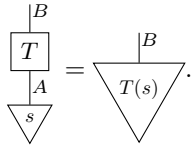
$$T \equiv$$  . (A4)

For quantum theory, the set of transformations $\mathcal{T}^{A\to B}$ is the set of quantum operations, which correspond to completely positive trace-non-increasing maps from V^A to V^B.

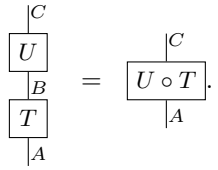
Of course, we may want to represent not just the transformation itself, but its action on a specific state. This is done by connecting the input wire of the transformation with a state of matching type:

$$T(s) \equiv$$  . (A5)

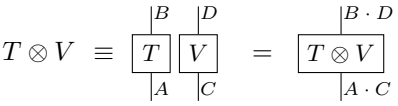
Note that with this, viewing $T(s)$ as a vector can be expressed in diagrams by "sliding" the box representing T until it merges with the box representing s:

$$T \circ s = T(s) \equiv$$  . (A6)

This is a manipulation of diagrams that is used often in this work. The converse operation, where we split a vector into a product where a transformation T is applied on a state s, is also a valid manipulation where we split a diagram with only a state into one where a transformation is connected to a different state. Notice that this mirrors exactly linear algebraic operations where an equation like $s' = T \circ s$ is used for substitutions. Furthermore, boxes representing transformations can also be connected, when the wire types match, to represent the sequential composition of them. Because a sequence of linear transformations T and U, can also be viewed as a single transformation $U \circ T$, the composition of both, the merging of boxes can also be done with transformations that are connected:

$$U \circ T \equiv$$  . (A7)

Naturally, one may need to represent transformations that happen in parallel on the parts of a composite system $A \cdot B$. While in linear algebraic notation this is done with the direct product $\otimes$, so that, for T of type $A \to B$ and V of type $C \to D$, we write $T \otimes V$, in diagrammatic notation we simply put T and V side by side:

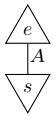
$$T \otimes V \equiv$$  , (A8)

where the order of the wires (from left to right) matters just like the order of the product $T \otimes V$.

The effects of a system A in a GPT are linear functionals over V^A that evaluate to probabilities, i.e., numbers in $[0, 1]$, for every valid state. This means that the set E^A of effects of a system A is a subset of $(V^A)^*$, the dual of V^A, such that $e \in E^A$ implies $e(s) \in [0, 1]$ for every $s \in \Omega^A$. Diagrammatically, the effects are represented as boxes with only inputs, so

$$e \equiv$$  (A9)

represents the element $e \in E^A \subset (V^A)^*$. Just like the linear functions, the action of e on a state s, that is, $e(s)$, is given by connecting the input (bottom) wire of the effect to the output (top) wire of the state:

$$e(s) =$$  . (A10)

Using quantum theory again as an example, its effects are trace inner products with the elements of a positive operator-valued measure (POVM). That is to say that the POVM elements give the Riesz representation of the effects of the theory. So, if M is associated to the effect e_M through the Riesz representation, then $e_M(\rho) = \mathrm{tr}(M\rho)$.

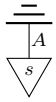
Notice that Diagram (A10), unlike those in the previous examples, contains no loose wires. This means that $e(s)$ is a real number, and, similarly, any diagram in this

formalism without loose wires represents a real number. Moreover, diagrams with only output (top) loose wires are always states, diagrams with only input (bottom) loose wires are always effects, and diagrams with both input and output loose wires are transformations.

Now that we have discussed the effects, we can define what it means for a state to be normalised in a GPT. This is done through a special effect, called the unit effect, which, for a system A, we denote by u^A. We say that a vector $s \in V^A$ is normalised if and only if $u^A(s) = 1$. Therefore, by our definition of the set of states Ω^A, if $s \in \Omega^A$, then $u^A(s) = 1$. This special effect is denoted by a special diagram

(A11)

thus, diagrammatically, state normalisation is captured by the condition

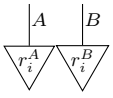


$= 1.$ (A12)

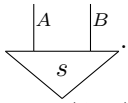
In the example of quantum theory, the unit effect of any system type is the trace operation, or, in other words, the trace inner product with the identity operator.

An important definition to be made, and that appears nicely in diagrammatic notation, is that of separable effects and states. In quantum theory, a separable state is that which can be written as a convex combination of product quantum states. Here, we simply generalise that notion to any GPT state: $s \in \Omega^{A\cdot B}$ is separable if $s = \sum_i p_i r_i^A \otimes r_i^B$, with $p_i \in [0,1]$ and $\sum_i p_i = 1$, $r_i^A \in \Omega^A$, and $r_i^B \in \Omega^B$. In diagrammatic notation, a separable state can be viewed as:

$$s = \sum_i p_i r_i^A \otimes r_i^B \equiv \sum_i p_i$$



$=$



. (A13)

Separable effects are defined similarly, with its diagrammatic representation being like the one above but where the loose wires come from the bottom instead of the top.

2. Further definitions

We can use these fundamental notions to define some concepts that are necessary in this work. These are positive cones, positive vectors, local tomography, the no-restriction hypothesis, the generalised no-restriction hypothesis, the maximal tensor product, trace-preserving transformation, and positive and completely positive transformations. Some of these are present in quantum theory, but here we need definitions that generalise them to arbitrary GPTs.

Definition A.1 (Positive Cone)**.** *A positive cone X_+ generated by a subset X of a vector space V is the set of nonnegative multiples of the elements of X. That is,*

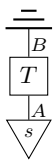
$$X_+ = \{\lambda x : \lambda \geq 0, x \in X\}. \tag{A14}$$

Definition A.2 (Positive Vector)**.** *A vector v of a vector space V^A associated to a system of A of a GPT is said to be positive, denoted $v \geq 0$, if $v \in \Omega_+^A$, the cone generated by the set of states.*

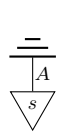
Note that in quantum theory our notion of positive cones recovers the sets of positive operators from the sets of density matrices. This is because, for any quantum system A, the density matrices $\rho \in \Omega^A$ satisfy $\rho \geq 0$ and $\mathrm{tr}(\rho) = 1$, so by multiplying then by positive numbers λ, we are simply dropping the unit trace assumption. Hence, the cone generated is $\Omega_+^A = \{\rho \in V^A : \rho \geq 0\}$, that is, the set of positive operators on the Hilbert space corresponding to A. Conversely, using the unit effects, it is always possible to recover the states from the positive cones by restricting them to normalised vectors.

Definition A.3 (Trace Non-Increasing Operation)**.** *A transformation T in $\mathcal{L}(V^A, V^B)$ is said to be trace non-increasing if $u^B(T(s)) \leq u^A(s)$ for every s in Ω^A.*

Diagrammatically, T is trace non-increasing means, for all $s \in \Omega^A$, that:



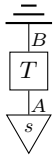
$\leq$



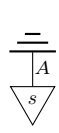
. (A15)

Definition A.4 (Trace-Preserving Operation)**.** *A transformation $T \in \mathcal{L}(V^A, V^B)$ is said to be trace preserving if $\forall s \in \Omega^A, u^B(T(s)) = u^A(s)$.*

Diagrammatically, T is trace preserving means, for all $s \in \Omega^A$, that:



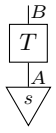
$=$



. (A16)

Definition A.5 (Positive Transformation)**.** *A transformation $T \in \mathcal{L}(V^A, V^B)$ is said to be positive if it takes elements in Ω_+^A to elements in Ω_+^B, that is, $s \in \Omega_+^A \implies T(s) \in \Omega_+^B$.*

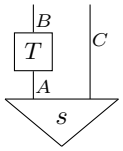
Diagrammatically, for all $s \in \Omega^A$, a positive transformation satisfies:



$\in \quad \Omega_+^B.$ (A17)

Definition A.6 (Completely Positive Transformation)**.** *A transformation $T \in \mathcal{L}(V^A, V^B)$ is said to be completely positive if it is positive and for any system C, the transformation $T \otimes \mathbb{1}_C \in \mathcal{L}(V^{A\cdot C}, V^{B\cdot C})$ is positive, where $\mathbb{1}_C$ is the identity map on V^C.*

Diagrammatically, for all systems C and all bipartite states $s \in \Omega^{AC}$, this means that:



$$\in \quad \Omega_+^{BC}\,. \tag{A18}$$

Any time that a transformation appears in a diagram, it is implied that it is a completely positive transformation for the corresponding GPT because it is an allowed transformation in said theory. The same applies for states: if they appear in a diagram, they must be positive in the corresponding GPT. So, the diagrams drawn in the beginning of this appendix are examples of positive transformations and states. Moreover, note that these notions recover those of positive, completely positive, trace preserving, and trace non-increasing maps when applied to quantum theory, because the cones are generated by the sets of states, and the unit effects are the trace operations.

Definition A.7 (Local Tomography)**.** *A GPT is said to be locally tomographic if any state $\rho^{A_1\cdot\ldots\cdot A_n}$ of a composite system $A_1 \cdot \ldots \cdot A_n$ can be uniquely determined by the information obtained from local effects $\{e^{A_1}\}$,..., $\{e^{A_n}\}$ on its parts A_1,..., A_n. When this holds, the unit vector for the composite system, $u^{A_1\cdot\ldots\cdot A_n}$, is given by $u^{A_1} \otimes \ldots \otimes u^{A_n}$.*

As an example of a GPT satisfying local tomography we have quantum theory. There, any $\rho^{A_1\cdot\ldots\cdot A_n}$ is completely determined by a set of probabilities $(e_{i_1}^{A_1} \otimes \ldots e_{i_n}^{A_n})[\rho^{A_1\cdot\ldots\cdot A_n}]$, where each local effect $e_{i_j}^{A_j}$ denotes the inner product of the quantum state with the corresponding POVM element.

Definition A.8 (Maximal Tensor Product)**.** *The maximal tensor product, $\otimes_{max}$ is a rule for the combination of two systems into one, say, A and B into $A \cdot B$, that defines the positive cone of the composite system as the largest set of vectors in $V^A \otimes V^B$ that is consistent (in the sense of producing sensible probabilities) with all the separable effects of $A \cdot B$. That is*

$$\Omega_+^{A\cdot B} = \Omega_+^A \otimes_{max} \Omega_+^B := \{\rho \in V^A \otimes V^B : (e^A \otimes e^B)[\rho] \geq 0 \quad \forall e^A \in E^A\,,\, e^B \in E^B\}. \tag{A19}$$

The maximal tensor product $\otimes_{max}$ is associative [45], hence one can unambiguously write

$$\Omega_+^{A_1} \otimes_{max} \ldots \otimes_{max} \Omega_+^{A_n}, \tag{A20}$$

which has the explicit form:

$$\left\{\rho \in \bigotimes_{i=1}^n V^{A_i} : \bigotimes_{i=1}^n e^{A_i}[\rho] \geq 0 \quad \forall e^{A_i} \in E^{A_i}\right\}. \tag{A21}$$

As discussed after Def. A.2, this operation fixes the state spaces for the composite systems, because the cone $\Omega_+^{A\cdot B}$ and the unit effect $u^A \otimes u^B$ (which is the unit effect for $u^{A\cdot B}$ in locally tomographic GPTs [32]) can be used to construct the set of states $\Omega^{A\cdot B}$.

Definition A.9 (No-Restriction Hypothesis [32])**.** *A theory is said to satisfy the no-restriction hypothesis (NRH) if any element $e \in (V^A)^*$ that gives $e(\rho) \in [0,1]$ for every $\rho \in \Omega^A$ is an element of E^A. That is, if*

$$E^A = \{e \in (V^A)^* : \forall \rho \in \Omega^A, e(\rho) \in [0,1]\}. \tag{A22}$$

This is to say, the NRH is the statement that, given the set of states, the set of effects is the largest possible that still gives sensible probabilities for every state.

Definition A.10 (Generalised No-Restriction Hypothesis)**.** *A theory is said to satisfy the Generalised No-Restriction Hypothesis (GNRH), if it satisfies the NRH and every completely positive trace non-increasing transformation is an allowed transformation. That is, for any two systems A and B, every transformation $T \in \mathcal{L}(V^A, V^B)$ that takes elements in $\Omega_+^{A\cdot C}$ to elements in $\Omega_+^{B\cdot C}$ for any third system C, and satisfies $u^B(T(\rho)) \leq u^A(\rho)$ for any $\rho \in V^A$, is a valid transformation from A to B.*

This is a convenient assumption to make about a theory because it simplifies its description, since it implies that the specification of the state spaces uniquely fixes both the effects and transformations. Again, we can use quantum theory as an example, as it does satisfy the GNRH.

Our last definition in this section is that of entanglement witnesses in a generic GPT. This definition follows closely to that found in the context of quantum theory.

Definition A.11 (Generic Bipartite Entanglement Witness)**.** *The set $W^{A\cdot B}$ of entanglement witnesses of a bipartite system $A \cdot B$ for a generic locally tomographic GPT is given by*

$$W^{A\cdot B} = \{w \in V^A \otimes V^B : \langle w, s^A \otimes s^B\rangle \geq 0 \quad \forall s^A \in \Omega_+^A, s^B \in \Omega_+^B\}\,. \tag{A23}$$

Note that this will depend on the choice of inner product. In the quantum case the standard choice will be the Hilbert-Schmidt inner product.

That is, an entanglement witness is a vector associated, through the Riesz representation, to a linear functional which evaluates to positive numbers for every product state of the bipartite system. This is a simple generalization of the quantum entanglement witnesses that uses arbitrary GPT states instead of quantum states. The generalization for multipartite entanglement witnesses is straightforward.

Definition A.12 (Generic Multipartite Entanglement Witness)**.** *The set $W^{A_1\cdot\ldots\cdot A_n}$ of the entanglement witnesses of a multipartite system $A_1\cdot\ldots\cdot A_n$ for a generic locally tomographic GPT is given by*

$$W^{A_1\cdot\ldots\cdot A_n} = \{w \in V^{A_1}\otimes\ldots\otimes V^{A_n} : \langle w, s^{A_1}\otimes\ldots\otimes s^{A_n}\rangle \geq 0 \quad \forall s^{A_1}\in\Omega_+^{A_1},\ldots,s^{A_n}\in\Omega_+^{A_n}\}. \tag{A24}$$

Like in the bipartite case, this will depend on the choice of inner product. Again, in the quantum case the standard choice will be the Hilbert-Schmidt inner product.

3. Some useful results

We now prove (or reprove) various results which are useful later on. Firstly we note an important consistency condition for the max tensor product. This is well known in the literature (see, e.g., Ref. [72]) but we reproduce it here for completeness.

Proposition A.13. In a theory where systems compose via the max tensor product, and that satisfies the NRH, one can check that the vectors, $v \in V^A$, that can be steered to from bipartite states in Ω^{AB} correspond to subnormalised states, i.e., live in Ω_+^A and satisfy $u^A(v) \leq 1$.[1]

Proof. Steered states are of the form $v_A = \mathbb{1}_A \otimes e_B(s_{AB})$ for a bipartite state s_{AB} and an effect e_B and the identity transformation $\mathbb{1}_A$. Note that a special case of this is the reduced state which is given by taking $e_B = u^B$. We want that these vectors are local (subnormalised) states. Note that, by definition of the max tensor product we have that for all e_A that $e_A\otimes e_B(s_{AB}) \in [0,1]$ and hence that $e_A(v_A)\in[0,1]$. As in our theory the local effects e_A are defined via the NRH this means that v_A must be in the cone Ω_+^A. Moreover, it is easy to compute that it is (sub)normalised as $u^A(v_A) = u^A\otimes e^B(s_{AB}) \in [0,1]$. □

Another well-known result (see, e.g., Ref. [45]) is the following.

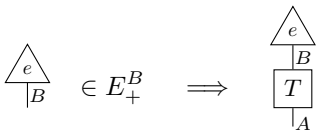
Proposition A.14. In a GPT composed by the max tensor product, every effect on a composite system is a separable effect [45].

Proof. It follows straightforwardly by noticing that the use of $\otimes_{max}$ to combine systems implies that the set of effects of the combined system $A\cdot B$ is just the set of separable effects. □

[1] Note that this is an essential consistency condition for any GPT, but here we see that it is automatically satisfied by GPTs satisfying NRH and composing via the max tensor product and, hence, imposes no further constraints.

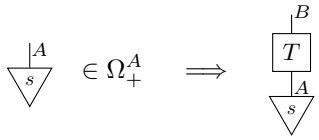
Next we prove a lemma which is useful for proving a key observation.

Lemma 1. In a GPT containing systems A and B, if the no-restriction hypothesis holds and the map $T \in \mathcal{L}(V^A, V^B)$ is positive, then



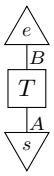
$$\in E_+^B \implies \in E_+^A, \tag{A25}$$

Proof. Recall that positivity of $T \in \mathcal{L}(V^A, V^B)$ means that



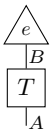
$$\in \Omega_+^A \implies \in \Omega_+^B. \tag{A26}$$

Now, using this positivity and, noting that $e^B \in E_+^B$, we find for all $s^A \in \Omega_+^A$ that



$$\geq 0. \tag{A27}$$

Hence,



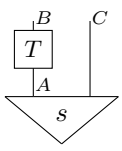
$$\in (\Omega_+^A)^* \overset{\text{NRH}}{=} E_+^A, \tag{A28}$$

which completes the proof. □

The lemma above can be used to prove a useful fact for our work.

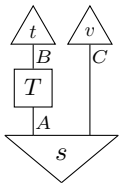
Theorem A.15. *In any GPT that combines systems through $\otimes_{max}$, the max tensor product, and satisfies the no-restriction hypothesis, if a map T is positive, then it is completely positive.*

Proof. For the sake of contradiction, assume that, in a GPT satisfying the NRH where $\otimes_{max}$ is the combination rule, the map $T \in \mathcal{L}(V^A, V^B)$ is positive but not completely positive. That is, that there exists a system C such that $T\otimes\mathbb{1}_C \in \mathcal{L}(V^A\otimes V^C, V^B\otimes V^C)$ is not positive. That means that there must exist some bipartite state $s \in \Omega^{A\cdot B} = \Omega_+^A \otimes_{max} \Omega_+^C$ such that



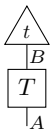
$$\notin \Omega_+^B \otimes_{max} \Omega_+^C. \tag{A29}$$

By the definition of $\otimes_{max}$, this means exists $t \in E^B$ and $v \in E^C$ such that



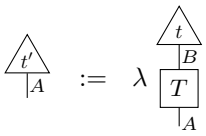
$< 0.$ (A30)

However, as $t \in E^B \subset E^B_+$, we know from lemma 1 that:



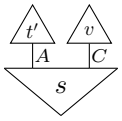
$\in \quad E^A_+$ (A31)

and hence that there exists $\lambda \geq 0$ such that



(A32)

Substituting this into eq. (A30) gives us that:



$< \ 0$ (A33)

but, as $t' \in E^A$ and $v \in E^C$ this means that $s \notin \Omega^A_+ \otimes_{max} \Omega^C_+$ and, hence, we have reached a contradiction. □

Appendix B: Formal definition and features of Witworld

In this Appendix we provide the formal definition of Witworld as a GPT, and the proofs that it does indeed possess the features mentioned in the main text. That task amounts to explicitly saying what are the states, effects and transformations of Witworld, following the formalities of the GPT framework mentioned in App. A. For Witworld, this is simplified because we define it to satisfy the GNRH, so by providing just the state spaces for each system type (including the multipartite ones, which requires the combination rule), we determine the complete GPT.

As was said in the main text, Witworld contains systems which we call atomic systems, and systems which we call composite systems. A system A is atomic if it cannot be considered as being the result of combining a system B with another system C. That is, there are no B and C such that $A = B \cdot C$. This means that any system type can be built from the atomic systems, so to determine all the types in Witworld, we only need to say what the atomic systems are and what the combination rule $\cdot$ is. Regarding the latter, we choose $\cdot$ to be the maximal tensor product $\otimes_{max}$. Regarding the former, we now formally define the atomic systems:

Definition B.1 (Quantum System $\mathcal{Q}_d$, $d \in \mathbb{N}$)**.** *A quantum system of type d has as its vector space $V^{\mathcal{Q}_d}$, as positive cone $\Omega^{\mathcal{Q}_d}_+$, as effects set $E^{\mathcal{Q}_d}$, and as unit effect the function $u^{\mathcal{Q}_d}(\_)$, all defined as follows:*

- $V^{\mathcal{Q}_d} = \{A \in \mathcal{L}(\mathcal{H}_d, \mathcal{H}_d) : A = A^\dagger\}$*: the space of Hermitian operators on a Hilbert space of dimension d.*
- $\Omega^{\mathcal{Q}_d}_+ = \{A \in V^{\mathcal{Q}_d} : A \geq 0\}$*: the set of positive operators on $\mathcal{H}_d$.*
- $u^{\mathcal{Q}_d}(\_) = \mathrm{tr}(\mathbb{1}\_)$*: the trace inner product with the identity operator $\mathbb{1}$ in $\mathcal{L}(\mathcal{H}_d, \mathcal{H}_d)$.*
- $E^{\mathcal{Q}_d} = \{\mathrm{tr}(A\_) : A \in V^{\mathcal{Q}_d}$ *and* $0 \leq A \leq \mathbb{1}\}$*: the set of trace inner products with operators in $V^{\mathcal{Q}_d}$ that are positive and smaller than or equal to the identity, as required by NRH.*

For atomic systems, the quantum type coincides with those of single systems in traditional quantum theory, and indeed local states and effects of atomic systems coincide for quantum types in both theories. However, as we see later on, this no longer holds for either transformations or composite systems – for the latter, this follows from the fact that the combination rule in Witworld, the maximal tensor product, is not the same as in quantum theory, so $\mathcal{Q}_d \cdot \mathcal{Q}_{d'} \neq \mathcal{Q}_{dd'}$.

Definition B.2 (Classical system $\mathcal{C}_v$, $v \in \mathbb{N}$)**.** *A classical system of type v has as its vector space $V^{\mathcal{C}_v}$, as positive cone $\Omega^{\mathcal{C}_v}_+$, as effects set $E^{\mathcal{C}_v}$, and as unit effect the function $u^{\mathcal{C}_v}(\_)$, all defined as follows:*

- $V^{\mathcal{C}_v} = \mathbb{R}^{v-1} \oplus \mathbb{R}^1 \cong \mathbb{R}^v$*: The direct sum of a $(v-1)$-dimensional real vector space with the real numbers. We can work with the isomorphic space $\mathbb{R}^v$ to simplify notation.*
- $\Omega^{\mathcal{C}_v}_+ = \{v \in V^{\mathcal{C}_v} : v = \lambda(q \oplus 1), \lambda \geq 0, q_i \geq 0, \sum_i q_i \leq 1\}$ *: the set of vectors in $\mathbb{R}^v$ that are the null vector or have a positive last component and whose first $v-1$ components divided by the last give the probabilities for $v-1$ outcomes of a measurement with v possible outcomes.*
- $u^{\mathcal{C}_v}(\_) = \langle (0, ..., 0, 1)^T, \_ \rangle$*: the Euclidean inner product with the vector in $\mathbb{R}^v$ whose only nonzero component is the last one, which is 1.*
- $E^{\mathcal{C}_v} = \{\langle e, \_ \rangle : e \in V^{\mathcal{C}_v}, \langle e, s \rangle \in [0,1] \forall s \in \Omega^{\mathcal{C}_v}\}$*: the set of euclidean inner products with vectors in $V^{\mathcal{C}_v}$ that evaluate to probabilities for every vector in $\Omega^{\mathcal{C}_v}$, as required by the NRH.*

Those are the traditional classical systems – probability distributions written as vectors. Such vectors can always be seen as a convex combination of deterministic

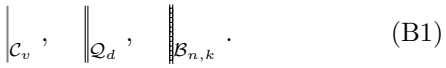
states. Note that, unlike for quantum systems, for classical systems we do have that $\mathcal{C}_d \cdot \mathcal{C}_{d'} = \mathcal{C}_{dd'}$[2]. Writing classical states in this form allows us to further notice that they are just particular cases of the Boxworld type.

Definition B.3 (Boxworld system $\mathcal{B}_{n,k}$, $(n,k) \in \mathbb{N}^2$)**.** *A Boxworld system of type (n,k)[3] has as its vector space $V^{\mathcal{B}_{n,k}}$, as positive cone $\Omega^{\mathcal{B}_{n,k}}$, as effects set $E^{\mathcal{B}_{n,k}}$, and as unit effect the function $u^{\mathcal{B}_{n,k}}(\cdot)$, all defined as follows:*

- $V^{\mathcal{B}_{n,k}} = (\mathbb{R}^n \otimes \mathbb{R}^{k-1}) \oplus \mathbb{R}^1 \cong \mathbb{R}^{n(k-1)+1}$*: The direct sum of the real numbers with the direct product between two real vector spaces of dimensions n and $k-1$. We can work with the isomorphic space $\mathbb{R}^{n(k-1)+1}$ to simplify notation.*
- $\Omega_+^{\mathcal{B}_{n.k}} = \{v \in V^{\mathcal{B}_{n,k}} : v = \lambda\left(\sum_{i=1}^n \boldsymbol{m}_i \otimes \boldsymbol{q}_i \oplus 1\right), \lambda \geq 0, q_{ij} \geq 0, \sum_{j=1}^k q_{ij} \leq 1, m_{ij} = \delta_{ij}\}$*: the vectors in $\mathbb{R}^{n(k-1)+1}$ which are the null vector or that have a positive last component and the first $n(k-1)$ components divided by the last component (if positive) can be viewed as probabilities for the first $k-1$ components of n measurements of k possible outcomes stacked in a list.*
- $u^{\mathcal{B}_{n,k}}(\_) = \langle \overline{u}, \_ \rangle$ *where* $\overline{u} = 0_{\mathbb{R}^n} \otimes 0_{\mathbb{R}^{k+1}} \oplus 1$*: the inner product with the vector in $\mathbb{R}^{n(k-1)+1}$ with only null components except for the last, which is 1.*
- $E^{\mathcal{B}_{n,k}} = \{\langle e, \_ \rangle : e \in V^{\mathcal{B}_{n,k}}, \langle e, s \rangle \in [0,1] \forall s \in \Omega^{\mathcal{B}_{n,k}}\}$*: the set of inner products with vectors in $V^{\mathcal{B}_{n,k}}$ that satisfies the NRH.*

The Boxworld systems can be viewed as classical systems that require many measurements to uniquely determine a state, rather than just 1, and the probability distributions for those measurements are independent of each other. A classical system of type v, then, can be viewed as a Boxworld system of type $(1,v)$. Note, however, that unlike classical systems, the composite of two more general Boxworld systems is no longer an atomic Boxworld system, that is $\mathcal{B}_{n,k} \cdot \mathcal{B}_{n',k'} \neq \mathcal{B}_{n'',k''}$.

When using diagrams, we denote the atomic classical, quantum, and Boxworld systems by different types of wires:



(B1)

When we need to talk about an arbitrary kind of system, the wire we use is the following:

(B2)

As stated previously, from the atomic systems any general system in Witworld can be constructed as an arbitrary composite of the three fundamental system types, $\mathcal{Q}_d, \mathcal{C}_v, \mathcal{B}_{n,k}$. For instance, $\mathcal{Q}_d \cdot \mathcal{Q}_{d'}$ and $\mathcal{Q}_d \cdot \mathcal{C}_v \cdot \mathcal{C}_{v'} \cdot \mathcal{B}_{n,k} \cdot \mathcal{Q}_{d'}$ would both be systems within our theory. More generally, systems correspond to arbitrary strings of elements from the set $\{\mathcal{Q}_d, \mathcal{C}_v, \mathcal{B}_{n,k}\}_{d,v,n,k \in \mathbb{N}}$. The positive cones for these composite systems are obtained through the max tensor product of the cones of the atomic types, and from those we can obtain the set of states by taking the intersection of the cone with the set of normalised vectors in the product vector space. Here, by normalised vector we mean vectors for which the unit effect evaluates to 1. Since Witworld is a locally tomographic GPT, the unit effect for $A_1 \cdot ... \cdot A_n$ is simply $u^{A_1} \otimes ... \otimes u^{A_n}$.

Since Witworld, by definition, satisfies the NRH, once we establish what the states for every type of system are, the effects are also determined. Given any system like $\mathcal{Q}_d \cdot \mathcal{B}_{n,k} \cdot ...$, every linear functional on $V^{\mathcal{Q}_d} \otimes V^{\mathcal{B}_{n,k}} \otimes ...$ that gives probabilities for every vector in $\Omega^{\mathcal{Q}_d \cdot \mathcal{B}_{n,k} \cdot \cdots}$ is a valid effect.

The definition of the transformations in Witworld is similar to that of the effects. Here, we require the theory to satisfy the GNRH, so when we determine the states, the transformations are fixed. In Witworld, any completely positive transformation is allowed. We prove later that this, together with the fact that the combination rule is $\otimes_{max}$, implies that any positive transformation is an allowed transformation for arbitrary systems in Witworld.

Formally, and concisely, Witworld is therefore defined as follows:

Definition B.4 (Witworld)**.** *Witworld is the locally-tomographic GPT that satisfies the generalised no-restriction hypothesis, and whose systems are arbitrary combinations under the max tensor product $\otimes_{max}$ of the atomic system types described in definitions B.1, B.2, and B.3.*

Now that we presented the definition of the theory, we can move on to observing or proving the various features of Witworld which we used in the main text. Note that because Witworld composes via the max tensor product and satisfies the GNRH, that all of the results of App. A 3 hold.

Firstly, Proposition A.14, tells us that in Witworld there are only separable effects. Therefore, for systems that are the combination of atomic quantum systems, there are fewer effects in Witworld than in quantum theory: effects from measurements in an entangled basis are not present in Witworld. For Boxworld and classical systems, such a difference does not exist: that is, Boxworld and classical system types feature separable-only effects both in Witworld and in their respective traditional frameworks.

Theorem A.15, together with the GNRH means that local transformations which in quantum theory are positive but not completely positive maps are indeed valid

[2] Hence strictly the atomic systems should be taken to be prime dimensional classical systems, however, we do not worry about this subtlety here.

[3] Strictly we should demand that $n > 1$ so as not to duplicate the classical systems, however, we do not worry about this subtlety here either.

transformations in Witworld. This is important for constructing many examples of post-quantum assemblages. Again for Boxworld and classical systems this distinction does not exist as for them the notion of positivity and complete positivity already coincide in their respective traditional frameworks.

Next we show that bipartite states for quantum systems within Witworld correspond to quantum entanglement witnesses.

Theorem B.5. *In Witworld, the bipartite system resulting from the combination of atomic quantum systems $\mathcal{Q}_d$ and $\mathcal{Q}_{d'}$ contains every bipartite entanglement witness in its positive cone.*

Proof. Take the set of effects $E^{\mathcal{Q}_d}$. To each effect φ_e in it, there is a vector $e \in V^{\mathcal{Q}_d}$ associated to it by its Riesz representation with the Hilbert-Schmidt inner product. Let us call the set of all $e \in V^{\mathcal{Q}_d}$ associated to some $\varphi_e \in E^{\mathcal{Q}_d}$ by $\tilde{E}^{\mathcal{Q}_d}$. Do the same to define $\tilde{E}^{\mathcal{Q}_{d'}}$. In quantum theory, $\tilde{E}_+ = \Omega_+$ because the effects are inner products with positive operators. Now, note that by changing $E^A \to E^A_+$ and $E^B \to E^B_+$ we do not change the set defined by equation A.11. Hence, we can write it, already using the action of the effect as an inner product, as

$$\Omega^A_+ \otimes_{max} \Omega^B_+ = \{v \in V^A \otimes V^B : \langle v, e^A \otimes e^B \rangle \geq 0 \quad \forall e^A \in \tilde{E}^A_+, e^B \in \tilde{E}^B_+\}. \tag{B3}$$

If $A = \mathcal{Q}_d$ and $B = \mathcal{Q}_{d'}$, then $\tilde{E}^A_+ = \Omega^{\mathcal{Q}_d}_+$ and $\tilde{E}^B_+ = \Omega^{\mathcal{Q}_d}_+$, so the equation above is by definition the set $W^{\mathcal{Q}_d \cdot \mathcal{Q}_{d'}}$ of entanglement witnesses of $\mathcal{Q}_{d'} \cdot \mathcal{Q}_{d'}$. □

This theorem means that if we compare the positive cones in quantum theory with the positive cones in Witworld when combining two atomic quantum systems, we see that the positive cone in Witworld is larger than the positive cone in quantum theory. To see this more explicitly, suppose A and B are two atomic quantum systems, then refer to A and B combined as prescribed by quantum theory by $A \otimes B$, and by $A \cdot B$ when combined as prescribed by Witworld. Then, since definition A.11 is independent of the combination rule and is equivalent to what is used in quantum theory, theorem B.5 implies $\Omega^{A\cdot B}_+ = W^{A\cdot B} = W^{A\otimes B} \supset \Omega^{A\otimes B}_+$, where the last inclusion is given from quantum theory. Nevertheless, since classical and Boxworld systems originally combine through $\otimes_{max}$, in Witworld the combination of atomic systems of said types do not build more states than what we would normally have. Finally, states that are the combination of different types of atomic systems are incomparable to states in quantum, Boxworld or classical systems.

The fact that we can view quantum entanglement witnesses as valid states in Witworld is a key feature which underpins many of our realisations of post-quantum assemblages. This is also true in the multipartite generalisation which we now prove.

Theorem B.6. *In Witworld, the multipartite system resulting from the combination of atomic quantum systems $\mathcal{Q}_{d_1} = A_1, ..., \mathcal{Q}_{d_n} = A_n$ contains every multipartite entanglement witness in its positive cone.*

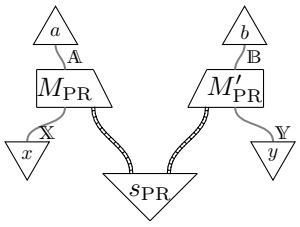
Proof. This is a straightfoward generalization of the bipartite case:

$$\begin{aligned}
&\Omega^{A_1}_+ \otimes_{max} \cdots \otimes_{max} \Omega^{A_n}_+ = \\
&= \left\{ s \in \bigotimes_{i=1}^n V^{A_i} : \bigotimes_{i=1}^n \varphi_e^{A_i}[s] \geq 0, \quad \forall \varphi_e^{A_i} \in E^{A_i} \right\} \\
&= \left\{ s \in \bigotimes_{i=1}^n V^{A_i} : \left\langle \bigotimes_{i=1}^n e^{A_i}, s \right\rangle \geq 0 \quad \forall e^{A_i} \in \tilde{E}^{A_i}_+ \right\} \\
&= \left\{ s \in \bigotimes_{i=1}^n V^{A_i} : \left\langle \bigotimes_{i=1}^n e^{A_i}, s \right\rangle \geq 0 \quad \forall e^{A_i} \in \Omega^{A_i}_+ \right\} \\
&= W^{A_i \otimes \ldots \otimes A_n}
\end{aligned} \tag{B4}$$

where $e^{A_i} \in V^{A_i}$ is associated to $\varphi_e^{A_i} \in (V^{A_i})^*$ by the Riesz representation and $\tilde{E}^{A_i}_+ = \Omega^{A_i}_+$ because the A_i are atomic quantum systems. □

Appendix C: How to realise a PR-box in Boxworld and Witworld

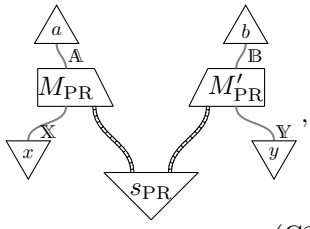
To make explicit that Witworld can, in fact, realise a PR-box, we explicitly write down the elements in the diagram



(C1)

and thereby show that the following holds:

$$p_{\mathrm{PR}}(ab|xy) := \frac{1}{2}\delta_{a\oplus b=xy} =$$

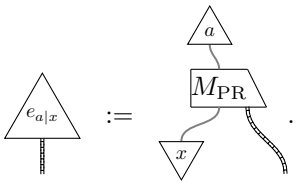


(C2)

where $\oplus$ is addition modulo 2, and $a, b, x, y \in \{0, 1\}$.

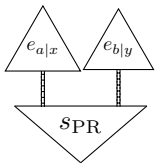
The first step is to define the measurements M_{PR} and M'_{PR}. We characterise these by their associated set of effects. For example, M_{PR} by $\{e^A_{a|x}\}$, for the outcome a when measurement x is performed, diagrammatically

these are defined as:



(C3)

Similarly we can characterise the measurement M'_{PR} by the set of effects $\{e^B_{b|y}\}$ which are defined analogously. We therefore want to verify that there exists measurements and states such that:



$$= \frac{1}{2}\delta_{a\oplus b=xy}, \tag{C4}$$

which, to make this more explicit, can be rewritten symbolically as:

$$e_{a|x} \otimes e_{b|y}(s^{AB}) = \frac{1}{2}\delta_{a\oplus b=xy}. \tag{C5}$$

Following definition B.3, the states of $\mathcal{B}_{2,2}$ are three component real vectors, whose first component is $p(0|0)$, second component is $p(0|1)$ and last component is 1. Therefore, for $e_{a|x}(s) = p(a|x)$ to hold, we need the vectors $\tilde{e}_{a|x}$, associated to $e_{a|x}$ by the Riesz representation, to be given by

$$\begin{aligned} \tilde{e}_{0|0} &= \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \tilde{e}_{1|0} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \\ \tilde{e}_{0|1} &= \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \tilde{e}_{1|1} = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}. \end{aligned} \tag{C6}$$

Also by definition B.3, the Riesz representation of the unit effect u^A is given by

$$\tilde{u}^A = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \tag{C7}$$

so that $\tilde{e}_{0|x} + \tilde{e}_{1|x} = \tilde{u}^A$, which makes $\{e_{a|x}\}$ a valid measurement in $\mathcal{B}_{2,2}$ for each x.

From the vectors above, we can, using the Kronecker product, write $\tilde{e}_{a|x} \otimes \tilde{e}_{b|y}$, which are associated to the product effects of the composite system $\mathcal{B}_{2,2} \cdot \mathcal{B}_{2,2}$. Now, any vector s^{AB} in $V^{\mathcal{B}_{2,2}} \otimes V^{\mathcal{B}_{2,2}} \cong \mathbb{R}^9$ such that $e_{a|x} \otimes e_{b|y}(s^{AB}) = \langle \tilde{e}_{a|x} \otimes \tilde{e}_{b|y}, s^{AB} \rangle \geq 0$ for all $a, b, x, y \in \{0, 1\}$ is in the positive cone $\Omega_+^{\mathcal{B}_{2,2}} \otimes_{max} \Omega_+^{\mathcal{B}_{2,2}}$.

We now show that the following vector describes a normalised state in the positive cone of bipartite states, and, moreover, reproduces the statistics of the PR box as we desire. That is, consider the vector:

$$s_{\mathrm{PR}} = \begin{pmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \\ 0 \\ 1/2 \\ 1/2 \\ 1/2 \\ 1 \end{pmatrix}. \tag{C8}$$

and note that it satisfies

$$\langle \tilde{e}_{a|x} \otimes \tilde{e}_{b|y}, s_{\mathrm{PR}} \rangle = \frac{1}{2}\delta_{a\oplus b,xy}, \tag{C9}$$

which can be verified by direct calculation. Therefore s_{PR} is in the positive cone of $\mathcal{B}_{2,2} \cdot \mathcal{B}_{2,2}$ and moreover reproduces the PR box statistics. Finally it is also normalised as:

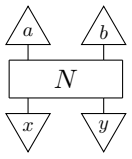
$$u^A \otimes u^B(s_{\mathrm{PR}}) = 1. \tag{C10}$$

Hence, s_{PR} is a valid state of $\mathcal{B}_{2,2} \cdot \mathcal{B}_{2,2}$ which recovers the PR-box under the separable measurements $\{e^A_{a|x} \otimes e^B_{b|y}\}$, which proves that Witworld can indeed realise a PR-box.

Appendix D: Formalities of steering scenarios

Here we introduce the basic concepts in steering, starting from the simple case of a bipartite scenario and building up to more general cases. We present steering scenarios by comparison with Bell scenarios, so that the former can be viewed as a modification of the latter where one or more of the parties does not perform a measurement. After the transition from Bell scenarios to a simple steering scenario, we introduce more general ones and proceed to define some important types of assemblages. In particular, we focus on generalised steering scenarios which display post-quantum features.

Consider the pictorial description[4] of a no-signalling box in a Bell scenario where Alice (Bob) measures $x \in \mathbb{X}$ ($y \in \mathbb{Y}$) and obtains the outcome $a \in \mathbb{A}$ ($b \in \mathbb{B}$):

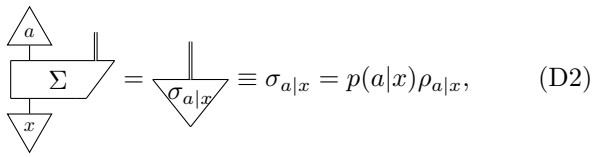


$$= p(ab|xy). \tag{D1}$$

Now, suppose instead that Bob decides not to perform the measurement y or indeed any other measurement,

[4] For now, these diagrams are not, strictly speaking, the same kind of mathematical diagrams from App. A, because drawing such diagrams presupposes that all the parts of them are objects existing in some GPT, and we don't know if there exists a GPT capable of realizing all assemblages to lend us its diagrams.

and, instead, merely keeps his system – by assumption, a quantum one. Then Bob has the subnormalised states $\sigma_{a|x}$ which are given by the following scheme:



$$\equiv \sigma_{a|x} = p(a|x)\rho_{a|x}, \tag{D2}$$

where $\rho_{a|x}$ is the normalised state in possession of Bob when Alice obtains outcome a upon measuring x with probability $p(a|x)$. The complete description of this scenario is specified by the the set $\Sigma_{\mathbb{A}|\mathbb{X}} = \{\sigma_{a|x}\}_{a\in\mathbb{A},x\in\mathbb{X}}$ of subnormalised quantum states, which contains the information about the states that Bob can have by the end of Alice's measurement, each of them conditioned on some measurement x and outcome a on Alice's side, together with the probability of the outcome a happening. This set of subnormalised quantum states, $\Sigma_{\mathbb{A}|\mathbb{X}}$, is known as an assemblage [52]. Note that the assemblage elements $\sigma_{a|x}$ indeed contain the complete information about the scenario because $p(a|x) = \mathrm{tr}(\sigma_{a|x})$ and $\rho_{a|x} = \sigma_{a|x}/\,\mathrm{tr}(\sigma_{a|x})$.

Of course, if signalling is permitted between Alice and Bob, then (within quantum theory) any assemblage can be trivially prepared, so we restrict our discussion to the non-signalling scenarios. The assemblages that can possibly be produced in this case are called non-signalling assemblages. We define them as being those satisfying conditions analogous to those that define non-signalling boxes in Bell scenarios:

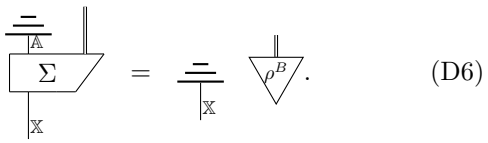
Definition D.1 (No signalling bipartite assemblages)**.** *An assemblage $\Sigma_{\mathbb{A}|\mathbb{X}}$ is no signalling iff*

$$\sigma_{a|x} \in \Omega_+^B, \tag{D3}$$

$$\sum_a \sigma_{a|x} = \sum_a \sigma_{a|x'} = \rho^B \quad \forall x, x', \tag{D4}$$

$$u^B(\rho^B) = 1. \tag{D5}$$

In the channel based picture, these constraints can be captured diagrammatically by the condition that Σ is a 'causal' channel [73]:

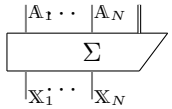


(D6)

It can be seen that this is equivalent to the standard no-signalling condition by noting that whatever input x is chosen for the classical system $\mathbb{X}$ can have no influence over the quantum system, as it is in a fixed normalised state ρ^B. The study of steering is the study of the properties of assemblages – or equivalently, the study of the properties of 'causal' classical-quantum channels. We use this to make a classification of types of assemblages in a meaningful way. Notice that while the set $\Sigma_{\mathbb{A}|\mathbb{X}}$ is a set of (subnormalised) quantum states, the fact that the system that Alice measures is not specified opens up the possibility for post-quantumness in the joint scenario, while keeping the quantum theoretical description valid for the local state of Bob.

Of course, like in the study of Bell non-locality, this scenario can be generalised. The two generalisations that we consider here are: i) adding more parties that are steering Bob; and, ii) allowing for Bob to have a setting variable $y \in \mathbb{Y}$. For the purpose of investigating post-quantumness, this is not only possible, but necessary, as it has been proven [58, 59] that every assemblage in the standard bipartite scenario can be realised in quantum theory. That is, any standard bipartite no signalling assemblage can be constructed by Alice performing a controlled measurement on one half of a bipartite quantum state shared with Bob. We call the assemblages which are beyond the powers of quantum theory in non-signalling scenarios, that is, which cannot be realised in this way, *post-quantum assemblages.*

In multipartite scenarios, the assemblage elements now carry the labels for the outcomes $a_i \in \mathbb{A}_i$ for multiple parties $i \leq N$, and similarly for the measurement choices $x_i \in \mathbb{X}_i$. They form assemblages $\Sigma_{\mathbb{A}_1\ldots\mathbb{A}_N|\mathbb{X}_1\ldots\mathbb{X}_N} = \{\sigma_{a_1\ldots a_N|x_1\ldots x_N}\}$ and are given in diagrammatic notation by



(D7)

Definition D.2 (No-signalling multipartite assemblages)**.** *A multipartite assemblage, $\Sigma_{\mathbb{A}_1\ldots\mathbb{A}_N|\mathbb{X}_1\ldots\mathbb{X}_N}$, is said to be no-signalling if it satisfies the following non-signalling constraints:*

a)

$$\sigma_{a_1\ldots a_N|x_1\ldots N} \in \Omega_+^B, \tag{D8}$$

b)

$$\sum_{a_1\ldots a_N} \sigma_{a_1\ldots a_N|x_1\ldots x_N} = \rho^B, \tag{D9}$$

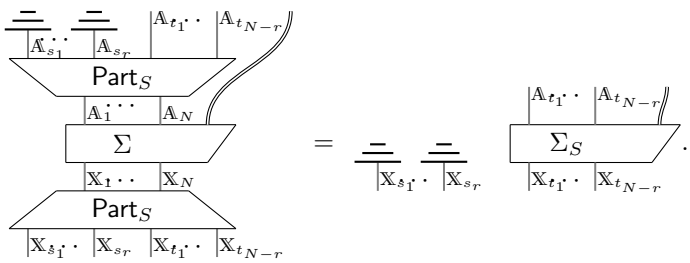
c)

$$u(\rho^B) = 1, \tag{D10}$$

d) Let $S = \{s_1, \ldots, s_r\} \subseteq \{1, ..., N\}$ be an arbitrary set of r parties, with $1 \leq r \leq N$, and denote by $\{t_1, ..., t_{N-r}\} = \{1, ..., N\} \setminus S$. Then, for all such S,

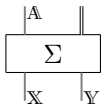
$$\sum_{a_s\,:\,s\in S} \sigma_{a_1\ldots a_N|x_1\ldots x_N} = \sigma_{a_{t_1}\ldots t_{j_{N-r}}|x_{t_1}\ldots x_{t_{N-r}}}. \tag{D11}$$

The no-signalling constraints of Def. D.2 can alternatively be expressed in diagrammatic notation in a simple way. For each arbitrary partitioning of $\{1,..N\} = \{s_1,...,s_r\} \sqcup \{t_1,...,t_{N-r}\}$, with $S = \{s_1,...,s_r\}$ and now $0 \leq r \leq N$, describe this partitioning via a physical splitting of the wires into a left hand group (the s_i) and a right hand group (the t_j) depicted by the process Part_S. Then diagrammatically, Eqs. (D9), (D10), and (D11) read:



(D12)

The other kind of generalization of the bipartite scenario that we consider is to allow Bob to, instead of staying passive, perform a local transformation labelled by $y \in \mathbb{Y}$ to his share of the system. Here, we require that Bob's system is locally a quantum one only after his transformation. In this case, the assemblage is denoted by $\Sigma_{\mathbb{A}|\mathbb{X}\mathbb{Y}} = \{\sigma_{a|xy}\}$, and pictorially is represented by



(D13)

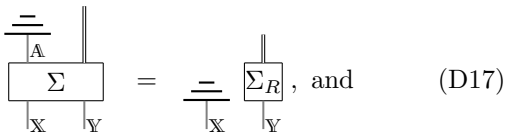
Definition D.3 (No-signalling Bob-with-input assemblages)**.** *A Bob-with-input assemblage $\Sigma_{\mathbb{A}|\mathbb{X}\mathbb{Y}}$ is no-signalling iff the following no-signalling constraints are satisfied:*

$$\sigma_{a|xy} \in \Omega_+^B \qquad \forall a, x, y, \tag{D14}$$

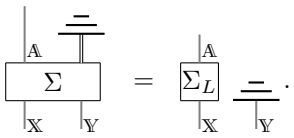
$$\sum_a \sigma_{a|xy} = \sum_a \sigma_{a|x'y} \qquad \forall x, x', y, \tag{D15}$$

$$u^B(\sigma_{a|xy}) = p(a|x) \qquad \forall a, x, y. \tag{D16}$$

These can be pictorially represented by:

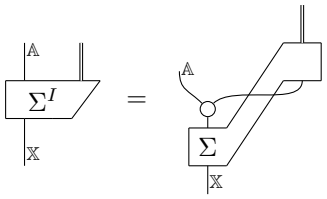


, and (D17)



. (D18)

The last scenario which is important to us is the instrumental steering scenario. This can be seen as a Bob-with-input scenario where Bob's input y is completely determined by Alice's output a. The assemblage $\Sigma^I_{\mathbb{A}|\mathbb{X}} = \{\sigma_{a|x}\}$ in the instrumental scenario can be diagrammatically represented as a wiring of a Bob-with-input assemblage Σ:



, (D19)

where the small circle in the bifurcation represents the copy operation, which is available for classical systems (which are the types of systems carrying measurement inputs and outputs). The class of instrumental assemblages of interest are known as general instrumental assemblages, in contrast to the other cases in which they were no-signalling assemblages. The reason for this is that in this scenario there is explicit signalling from Alice to Bob so the term no-signalling would be inappropriate.

Definition D.4 (General instrumental assemblage)**.** *An instrumental assemblage $\Sigma^I_{\mathbb{A}|\mathbb{X}}$ is said to be a general instrumental assemblage iff it is a wiring (as per Eq. (D19)) of a no-signalling Bob-with-input assemblage (Def. D.3).*

As was said previously, not only the scenarios are important but also some types of assemblages in each scenario should be defined. By doing so, we become able to talk precisely about what post-quantumness means for steering. These types are defined as follows.

Definition D.5 (Local Hidden State (LHS) (N+1)-Partite Assemblage)**.** *An assemblage $\Sigma_{\mathbb{A}_1...\mathbb{A}_N|\mathbb{X}_1...\mathbb{X}_N}$ in the (N+1)-partite steering scenario has a local hidden state model iff it can be prepared by the parties A_i performing local measurements on a shared classical random variable λ, whilst B prepares a quantum state conditioned on this random variable. That is,*

$$\sigma_{a_1...a_n|x_1...x_N} = \sum_\lambda p(\lambda) p_\lambda(a_1|x_1)...p_\lambda(a_N|x_N)\rho^B_\lambda \tag{D20}$$

for some probability distribution $p(\lambda)$ over λ, and quantum states ρ^B_λ.

Definition D.6 (Quantum (N+1)-Partite Assemblage)**.** *An assemblage $\Sigma_{\mathbb{A}_1...\mathbb{A}_N|\mathbb{X}_1...\mathbb{X}_N}$ in the (N+1)-partite steering scenario has a quantum realization iff it can be prepared by the parties A_i performing local quantum measurements on a shared quantum system. That is,*

$$\sigma_{a_1...N|x_1...x_N} = \mathrm{tr}_{A_1...A_N}\left(M^{x_1}_{a_1} \otimes ... \otimes M^{x_N}_{a_N} \otimes \mathbb{1}_B \cdot \rho^{A_1...A_N B}\right), \tag{D21}$$

for some local POVMs $\{M_{a_i}^{x_i}\}$ *and joint quantum state* $\rho^{A_1\ldots A_N B}$.

Definition D.7 (Local Hidden State (LHS) Bob-with-input Assemblage)**.** *An assemblage* $\Sigma_{\mathbb{A}|\mathbb{X}\mathbb{Y}}$ *in the Bob-with-input steering scenario has a local hidden state model if and only if it can be prepared by the parties performing local operations on a shared classical system. That is,*

$$\rho_{a|xy} = \sum_{\lambda} p(\lambda) p_\lambda(a|x)\, \rho^B_{\lambda,y}\,, \tag{D22}$$

where $\rho^B_{\lambda,y}$ *are local quantum states.*

Definition D.8 (Quantum Bob-with-input Assemblage)**.** *An assemblage* $\Sigma_{\mathbb{A}|\mathbb{X}\mathbb{Y}}$ *in the Bob-with-input steering scenario has a quantum realization if and only if it can be prepared by the parties performing local operations on a shared quantum system. That is,*

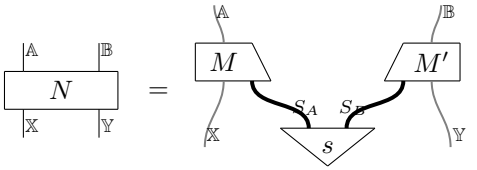
$$\sigma_{a|xy} = T_y\left(\mathrm{tr}_A\left[M_a^x \otimes \mathbb{1}_B \cdot \rho^{AB}\right]\right) \tag{D23}$$

for some local POVM $\{M_a^x\}$*, joint quantum state* ρ^{AB}*, and quantum operations* $\{T_y\}$.

The LHS and quantum assemblages in the instrumental scenario are defined just like in the Bob-with-input scenario, but with the constraint that $y = a$.

With the steering scenarios and assemblages defined in a general way, we can proceed to describe how these appear within the GPT framework. We follow a similar path, starting from Bell nonlocality scenarios and legitimately use the GPT diagrams for each case.

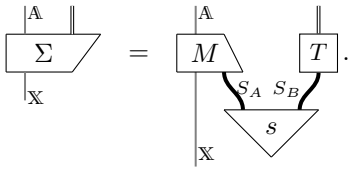
In a GPT, a Bell scenario where the no-signalling condition is satisfied is produced when Alice (Bob) makes local measurements M_A (M'_B) with input x (y) on a shared state s. The set of no-signalling boxes that can be realised in such a way are equivalent to the set of 'causal' classical channels, N, that can be realised by:



(D24)

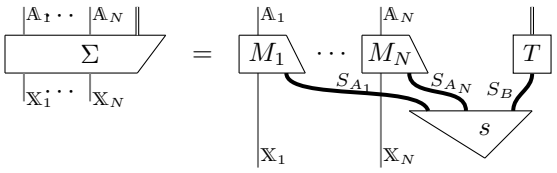
Again, we now let Bob be passive and perform no measurement, under the assumption that his system is a quantum one, and the resulting diagram is an assemblage element in the bipartite steering scenario:

Definition D.9 (GPT realisable assemblages)**.** *i) A bipartite assemblage* $\Sigma_{\mathbb{A}|\mathbb{X}}$ *is GPT realisable for a given GPT iff the channel associated to it can be written as:*



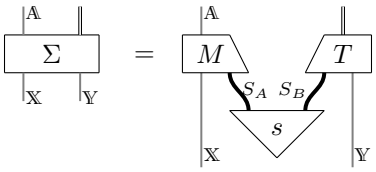
(D25)

Note that the transformation T *can be viewed as the process by which Bob characterises his system, which could be a post-quantum system, as a quantum system. This could be incorporated into the state* s *and we could view Bob as being given a quantum system to start with, this picture, however, is useful for later generalisations.*
ii) A multipartite assemblage $\Sigma_{\mathbb{A}_1\ldots\mathbb{A}_N|\mathbb{X}_1\ldots\mathbb{X}_N}$ *is GPT realisable if and only if its associated causal channel can be written as:*



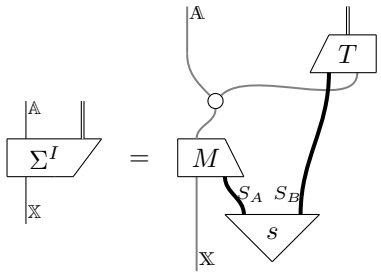
(D26)

iii) A Bob-with-input assemblage $\Sigma_{\mathbb{A}|\mathbb{X}\mathbb{Y}}$ *is GPT realisable iff its associated channel can be written as:*



(D27)

iv) An Instrumental assemblage $\Sigma^I_{\mathbb{A}|\mathbb{X}}$ *is GPT realisable if and only if it is a wiring of a GPT realisable assemblage in the Bob-with-input scenario:*

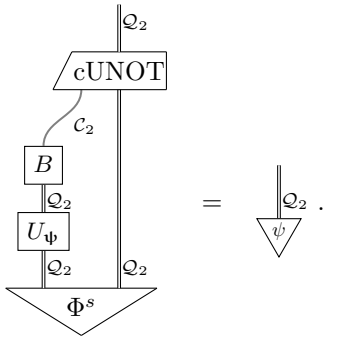


(D28)

With the definition of GPT realisable assemblages in place, we can revisit the LHS and quantum assemblages, and see how these amount to restrictions on the shared state s and the GPT to which it belongs. That is, if we consider the assemblages that are realisable in the GPT of quantum theory, then, within this GPT, the GPT realisable assemblages are exactly the Quantum assemblages. If we moreover restrict to the state s being a separable state, then we recover the LHS assemblages.

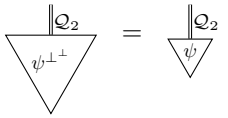
Appendix E: Details on the Remote State Preparation protocol

In this section, we prove that Eq. (23) is true. Namely, that for all ψ we have:



(E1)

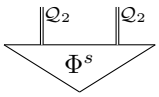
To see this we note some basic results regarding the various components of the diagram on the left hand side. First note that, as we are considering a qubit system $\mathcal{Q}_2$, then for each pure state ψ there is a unique orthogonal state, which we denote as $\psi^\perp$. This uniqueness, in particular, means that:



(E2)

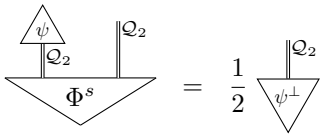
Now, turning to basic properties of the diagrammatic elements we have:

i. the singlet state



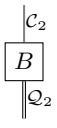
(E3)

satisfies



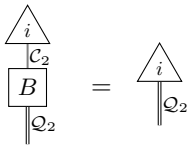
$\forall\psi$; (E4)

ii. the measurement



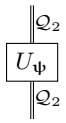
(E5)

is the computational basis measurement satisfying



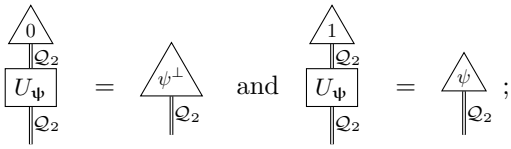
$\forall\, i \in \{0,1\}$; (E6)

iii. the family of unitaries



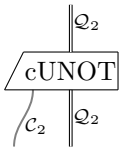
(E7)

satisfies (for all $\boldsymbol{\psi}$)



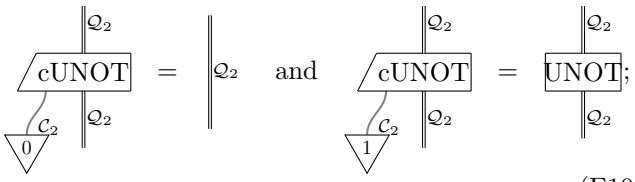
; (E8)

iv. the controlled transformation



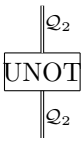
(E9)

satisfies



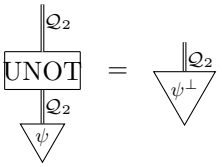
(E10)

v. the universal not gate



(E11)

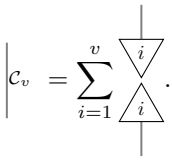
satisfies



$\forall\psi$. (E12)

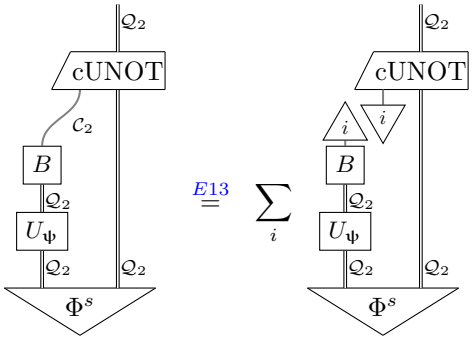
Finally, note that we can decompose the classical system as a sum of projectors. Consider the basis for classical

system $\mathcal{C}_v$ labelled by $i \in \{1, ..., v\}$, then

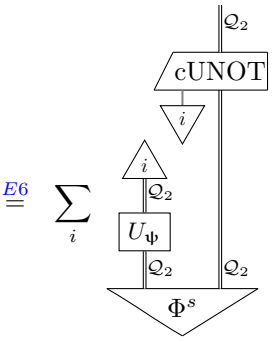


(E13)

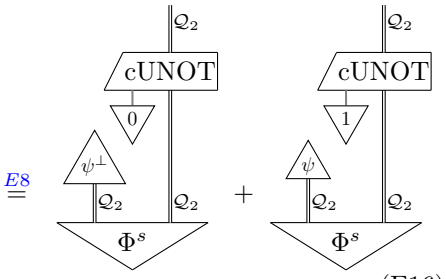
From these, together with linearity of GPTs, we can conclude that the protocol does indeed work as we want:



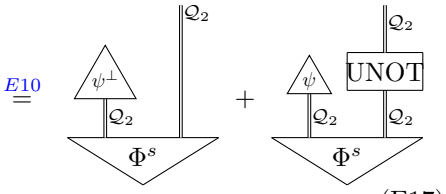
(E14)



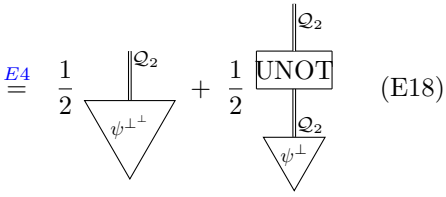
(E15)



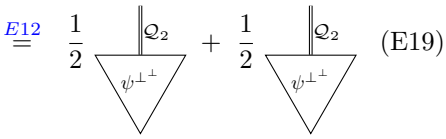
(E16)



(E17)



(E18)



(E19)

(E20)

which is the state $|\psi\rangle\langle\psi|$ chosen by Alice but prepared at Bob's lab, as required.

REFERENCES

[72] Peter Janotta and Raymond Lal, "Generalized probabilistic theories without the no-restriction hypothesis," Phys. Rev. A 87, 052131 (2013).
[73] David Beckman, Daniel Gottesman, M. A. Nielsen, and John Preskill, "Causal and localizable quantum operations," Phys. Rev. A 64, 052309 (2001).
[74] Ciarán M Lee, and John H Selby, "Deriving Grovers lower bound from simple physical principles," New J. Phys. 18, 093047 (2016).

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

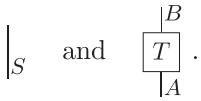
2. Generalised probabilistic theories: the basics

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.

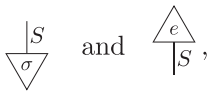
From this notion of how the experimental processes connect, or compose, a convenient diagrammatic notation can be defined so as to capture the entire structure of the GPTs. We can represent any process by a box, and encode the type of system on which it happens as a labelled input wire at the bottom 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

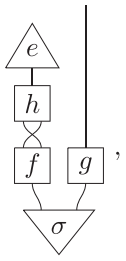


(2)

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)

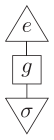
These diagrammatic pieces can be connected when the input/output wire types match. This represents the sequential composition of processes. When processes are instead drawn side by side, we are representing their parallel composition. By connecting boxes, therefore, we can then construct more complex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

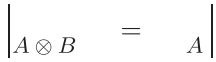
When a diagram has no loose wires, they are interpreted as numbers, which in the case of GPTs are the probabilities generated by the theory. For instance,



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

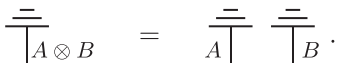
Of course, we might need to describe systems that are composed by simpler parts—multipartite systems—so we can emphasize that some system is composite by drawing the wires of its parts side by side



(6)

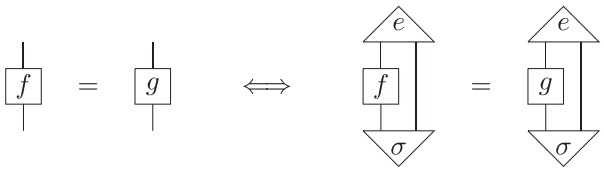
[3] The uniqueness of this discarding effect means that we are dealing with so-called causal GPTs [92].

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:



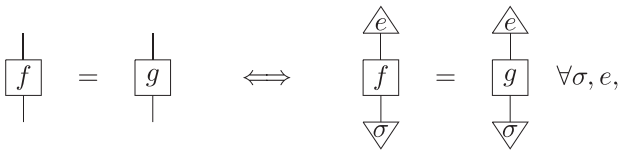
(7)

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



$\forall \sigma, e.$ (8)

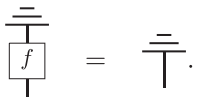
In fact, in this paper we will work with a special class of GPTs which satisfy the principle of *tomographic locality* [9, 90]. This means that:



$\forall \sigma, e,$ (9)

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

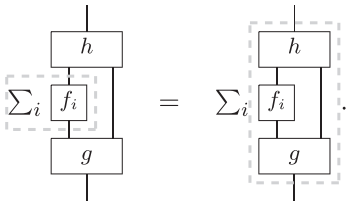
An important type of processes is that of those that are *discard-preserving*, which means they satisfy the following:



(10)

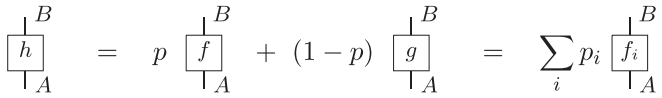
In the case of quantum theory, these correspond to the trace-preserving maps. Physically-realisable discard-preserving processes in a GPT are known as *channels*. Discard-preservation also defines a notion of causality for processes [92, 93]: a process is said to be causal if it is discarding preserving. 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)

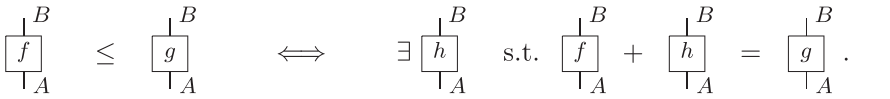
This definition implies that we can only sum processes with the same input/output types. From this, since a probability p can be a number (diagram without loose wires) of the theory, we can write



(12)

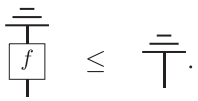
to describe convex mixtures of processes.

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



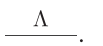
(14)

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:

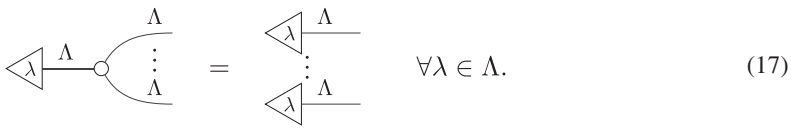


(15)

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

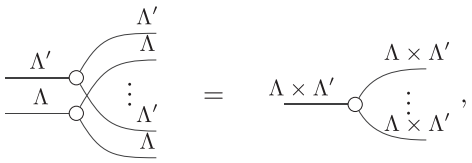
(16)

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:



$$, \tag{18}$$

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

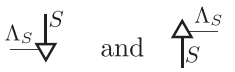
In contrast to the physical (sub)stochastic maps, we will draw mathematically well defined but (potentially) unphysical processes as black boxes such as:

$$, \tag{19}$$

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.

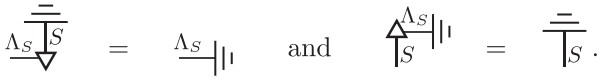
In contrast to the approach of reference [90], rather than labelling horizontal systems by black and white dots, we instead label the processes as being either black or white. This is equivalent but more convenient for us as, on the one hand, we can interpret the color as representing whether or not 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:



$$\tag{20}$$

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:



$$. \tag{21}$$

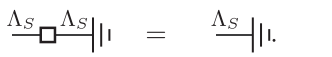
Note that here we follow the convention of reference [91] rather than reference [90], as the former demands that the fiducial effects form a measurement whilst the latter does not. This

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:

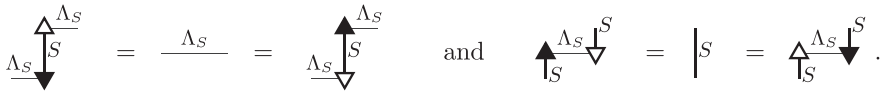


(23)

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:

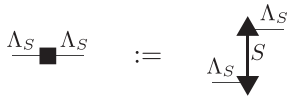
(24)

such that:



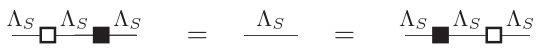
(25)

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:



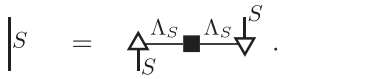
(27)

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:

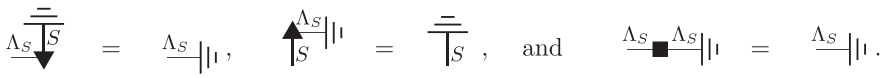
(28)

which, in particular, means that:



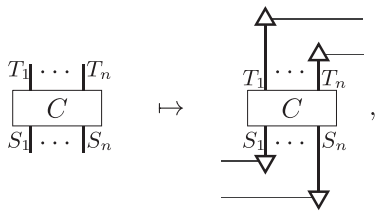
(29)

Moreover, it is also easy to show that:



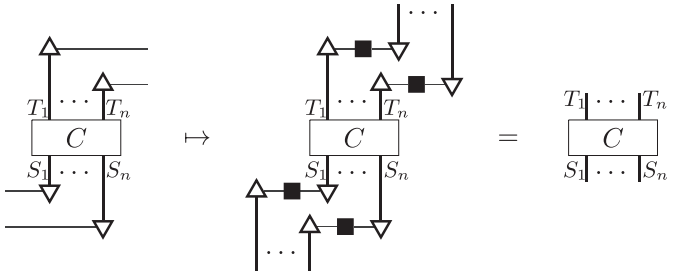
(30)

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



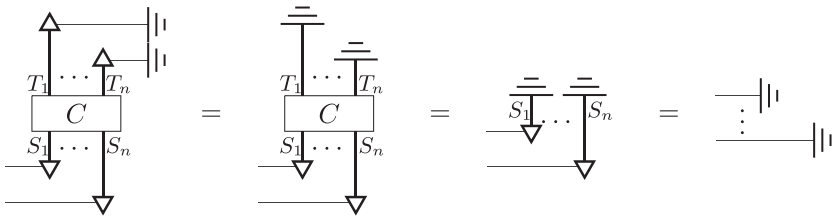
(31)

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:



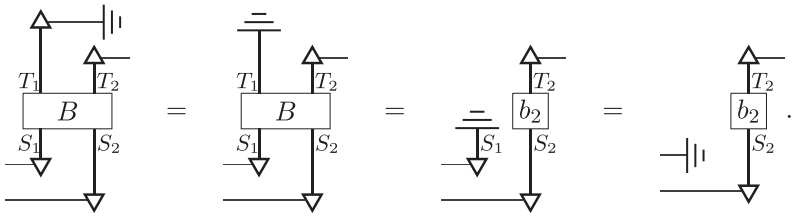
(33)

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)

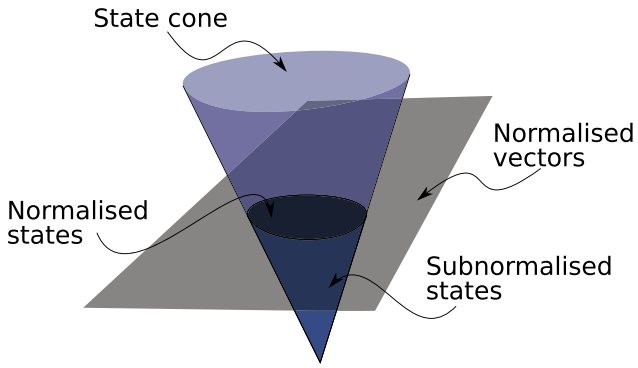
This straightforwardly generalises to multipartite GPT channels, and also to the case where only some of the no-signalling conditions hold. That is, the non-signalling structure of the channel and of the associated stochastic map are the same.

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:

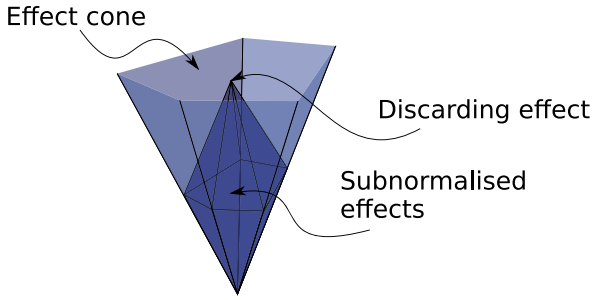


(36)

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:



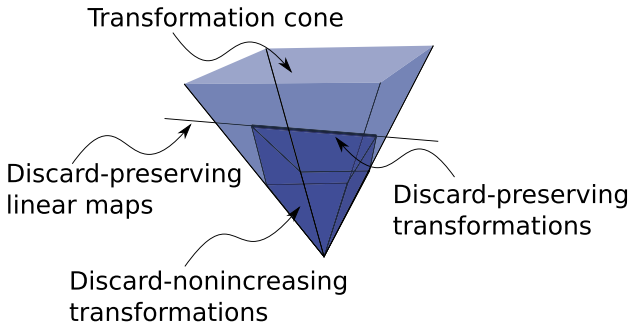
(37)

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

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

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

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

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:

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

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

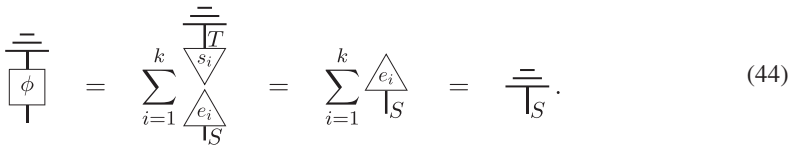
Of particular interest is the set of physical transformations referred to as *measure-and-prepare*, which we denote as MP:

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

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

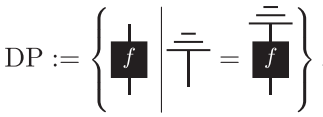
Since these are physically possible in any GPT, they are a subset[5] of the valid transformations, that is, they live inside the convex cone $\mathcal{T}_S^T$ and, in fact, MP $\subseteq \mathcal{K}$.

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.*

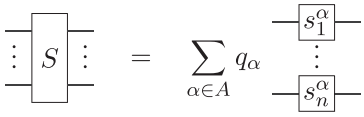
We provide a proof of theorem 4.1 preceded by a background on convex geometry in appendix A.

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

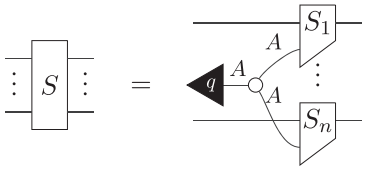
[5] In quantum theory, for example, these are a proper subset of all quantum channels and are known as entanglement-breaking channels.

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)

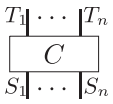
where the white dot is the copy operation, the quasiprobability distribution q is a black triangle because it is not physically realisable as it can have negative coefficients, and the S_i are stochastic maps controlled by the variable A.

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.*

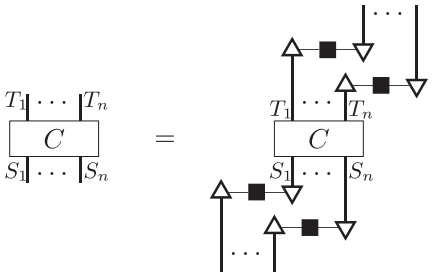
Proof. The if direction is trivial, hence we will focus on the only if direction here.

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



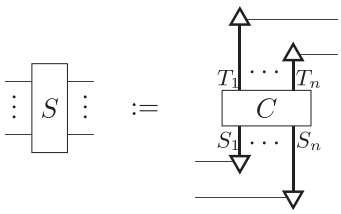
(49)

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



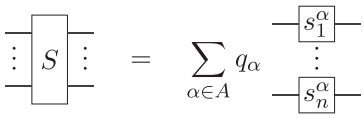
(50)

We can then observe that:



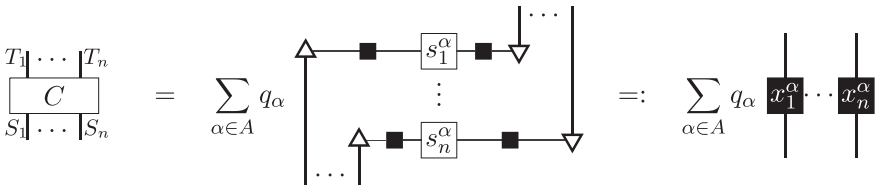
(51)

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



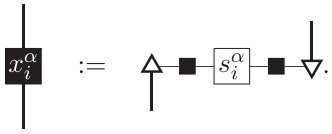
(52)

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:



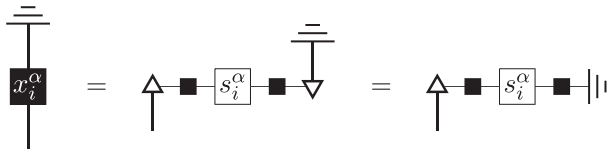
(53)

where

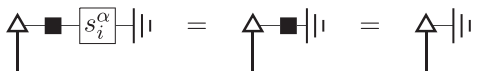


(54)

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



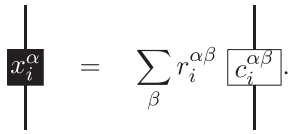
(55)



(56)

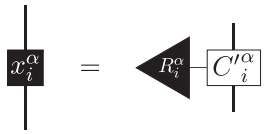
(57)

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



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

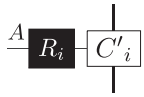
We can write this instead as:



$$= \tag{59}$$

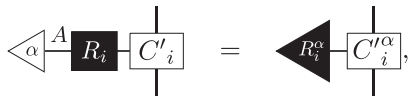
where C'^α_i is a classically controlled channel and R_i^α is a quasidistribution.

Now, let us define:



(60)

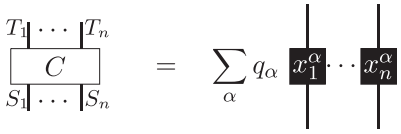
such that



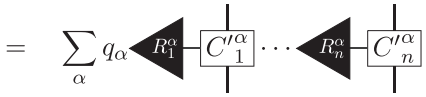
(61)

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

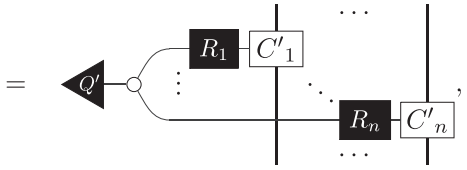
Putting this together with equation (52) we find that:



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

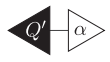


(63)



(64)

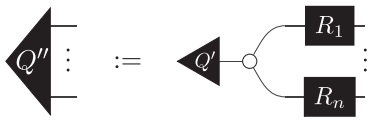
where Q' is the quasidistribution defined by the q_α, i.e.:



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

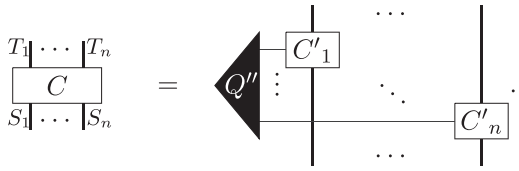
for all α.

Next, let us define Q'' by



(66)

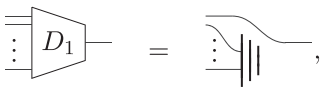
such that we can now combine this with equation (64) to write our channel as



(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.

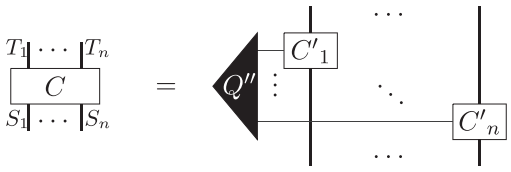
Now, define 'all but system i' marginalisation maps, D_i as:



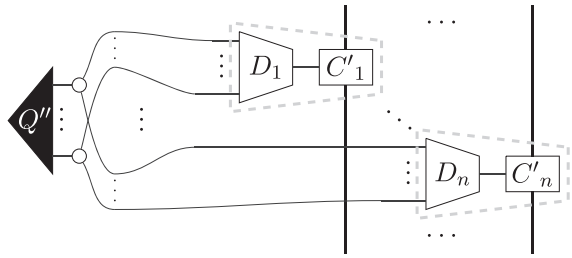
(68)

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

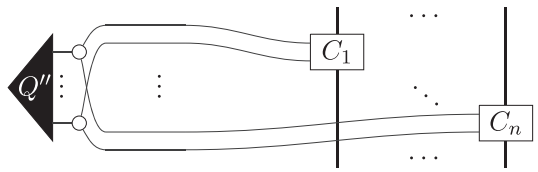
We can then write:



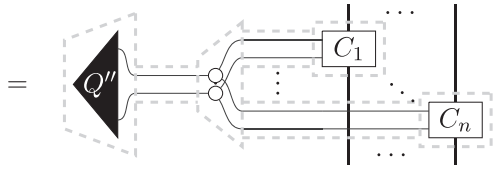
(69)



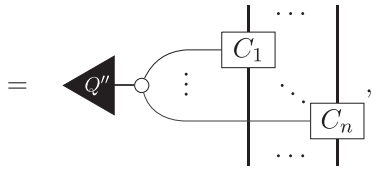
(70)



(71)



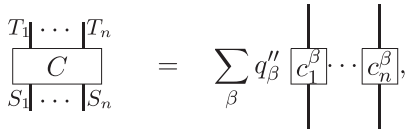
(72)



(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.

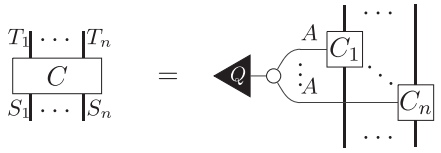
By decomposing the quasidistribution Q'' we can equivalently write this as:



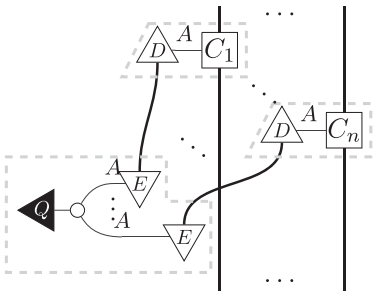
(74)

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. □

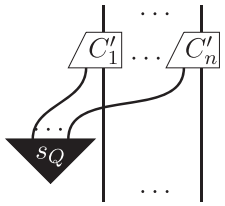
If we have a GPT, such as quantum theory, in which one can always reversibly encode classical data into a GPT system, then we can rewrite this as:



(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.

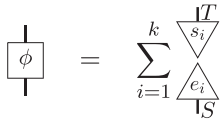
Conceptually, a convex set contains all the line segments between all pairs of points in the set. An affine set contains all the lines that extend beyond the endpoints of the line segments. This brings us to the definition of the *core* of a set. Thinking of line segments, $x \in S$ is in the core of 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. $\text{MP} = \text{DP} \cap \mathcal{K}$.

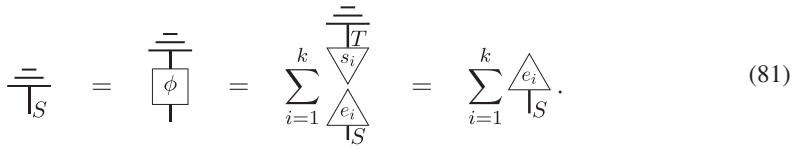
Proof. Since $\text{MP} \subseteq \text{DP} \cap \mathcal{K}$, all that remains to show is the opposite containment. Let $\phi \in \text{DP} \cap \mathcal{K}$ be a fixed, arbitrary vector. Since $\phi \in \mathcal{K}$, we can write it as



(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 $x \in \text{core}(K)$.

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

$$x - ats_i \in K \tag{85}$$

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

We now have

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

which is in K for all $t \in [0, t_v]$ where $t_v \coloneqq \min_i\{t_i\}$ (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 $\mu \in \text{int}(\mathcal{T}^T)$ be a normalised state. Then*

$$\begin{array}{c} \overset{T}{\underset{\mu}{\triangledown}} \\ \overset{=}{\underset{S}{\top}} \end{array} \in \text{DP} \cap \text{core}(\mathcal{K}), \tag{90}$$

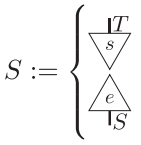
where, recall, $\overset{=}{\top} \in \text{int}(E_S)$ is the discarding effect.

Proof. Clearly

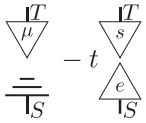

$\in \mathrm{DP},$ (91)

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

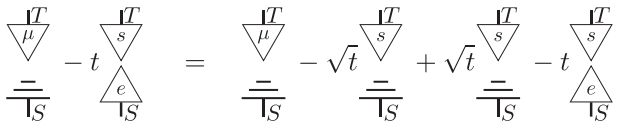
Define

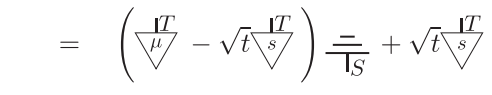

(92)

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

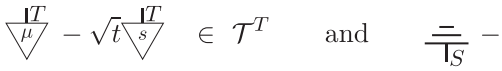

(93)

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

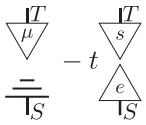

(94)


(95)

Note that


(96)

for all sufficiently small $t > 0$ as μ and $\doteq$ are interior in their respective cones. Therefore,

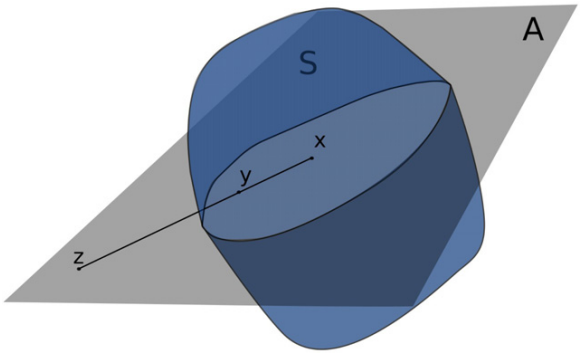

(97)

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

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

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 $e_i + \bar{e}_i = \bar{\bar{\top}}$. We also define $M_0 = \{\bar{\bar{\top}}, 0\}$ which is a binary outcome measurement such that the $\bar{\bar{\top}}$ 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

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

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

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

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)

I. INTRODUCTION

There has been a great deal of recent interest in the study of resource theories [3] in which the free operations are either Local Operations and Shared Randomness (LOSR) [4–9], for the purposes of studying nonlocality and entanglement, or Local Operations and Shared Entanglement (LOSE) [1], for the purposes of studying post-quantum nonlocality. In particular, it has been shown that these can be studied in a type-independent manner [1, 10, 11] such that resources of various types (entangled states, nonlocal boxes, steerable assemblages, etc.) can be treated in a uniform and unified way. These resource theories are motivated by the idea that the best way to understand Bell's theorem is from the perspective of causal models [6, 12], and that the lesson to be learnt from Bell's theorem is that we need an intrinsically quantum notion of causality and of common causes [6, 13].

Defining a resource theory requires a specification of both a *free* and an *enveloping* theory [3]. The free theory specifies the things that can be done effectively without cost, whilst the enveloping theory specifies the things that can be done irrespective of cost. Whilst in the study of LOSE and LOSR it is clear how the free theory should be defined, it is not clear how the *enveloping theory* should be defined [1]. There are two options for this, each of which has pros and cons. On the one hand, we have the choice which is typically made, which is to use the enveloping theory which describes *non-signalling* resources. The benefit of this choice is that it is mathematically simple to characterise, since in the cases of interest so far the set of such resources can often be expressed in a computationally-easy way (polytope, or semidefinite programme) [14, 15]. Its downside, however, is that this enveloping theory is not so well motivated from a causal perspective – it makes sense to say that resources should be non-signalling, but why should *all* non-signalling resources be considered? On the other hand, we can take the enveloping theory to describe arbitrary *common cause* resources, typically described using the framework of generalised probabilistic theories (GPTs) subsuming classical and quantum common causes as special cases. The benefit of this approach is that it is conceptually well motivated, from the causal perspective [6]. Its downside, however, is that providing a clean mathematical characterisation of this enveloping theory is an open problem. The characterisation and the relationship between these two options was cleanly articulated as an open question in Ref. [1, Open Question 1].

In this paper we resolve the tension between these two choices, by showing that these two options actually coincide. This means that we get the benefits of both approaches with none of the downsides. It is well established that every common-cause realisable resource is non-signalling, so here we just focus on the converse direction. In particular, we show that there exists a GPT in which all non-signalling resources of a target locally tomographic GPT, such as quantum theory, can be realised in a common-cause setting. On the one hand, we can view this result as, for the first time, providing a clear characterisation of the set of GPT-realisable resources. On the other hand, we can also view it as providing a principled justification, backed by the causal perspective, for choosing the set of non-signalling resources as the enveloping theory in resource theories of common-cause processes. We moreover show that this result holds not only in the bipartite case, which has so far dominated the literature, but also in the general multipartite scenario, thereby setting the stage for explorations of multipartite generalisations of LOSR and LOSE resource theories. A corollary of this result answers one of the open questions posed in Ref. [2], namely it shows that indeed any non-signalling assemblage can be given a GPT-common-cause explanation.

The scheme by which we build the GPT where all non-signalling resources can be realised in a common-cause setting differs from the standard approach to GPT construction in the literature. Usually, GPTs are constructed by making reference to the geometry of their states, effects, and transformations spaces, requiring, for example, that they are convex subsets of linear spaces (see, e.g., Ref. [16]). Here, instead of putting emphasis on the geometry, we focus our attention on compositionality, that is we take a process-theoretic [17–20] approach to constructing one GPT from another. By focusing on the compositional properties of the theory, our method also has the potential to be applied to other problems.

To be more formal, let us define a *common-cause completion* of a given GPT $\mathbf{G}$ as an supertheory of $\mathbf{G}$ which can realise all of the non-signalling resources $\mathbf{G}$ in a common-cause scenario. If some theory is the common-cause completion of itself, then we call it *common-cause complete*, in contrast to quantum and classical theory which have non-signalling resources which cannot be realised in common cause scenarios, being therefore common-cause incomplete. In this paper, we define a common-cause completion map, $\mathcal{C}$, which takes an arbitrary tomographically-local GPT $\mathbf{G}$ as an input, and gives a common-cause completion of it, $\mathcal{C}[\mathbf{G}]$, as output. Specifically, this means that $\mathbf{G}$ is a full subtheory of $\mathcal{C}[\mathbf{G}]$ and that every non-signalling resource in $\mathbf{G}$ can be realised with only common-cause resources in $\mathcal{C}[\mathbf{G}]$. Proving the existence of such a common-cause completion map demonstrates the main claims of this paper: i) all non-signalling resources in $\mathbf{G}$ are common-cause realisable in the GPT $\mathcal{C}[\mathbf{G}]$, and so the non-signalling resources in $\mathbf{G}$ coincide with its GPT-common-cause realisable processes; and ii) every non-signalling assemblage in $\mathbf{G}$ is realisable in an EPR scenario in $\mathcal{C}[\mathbf{G}]$.

II. GENERALISED PROBABILISTIC THEORIES (GPTS)

In this section, we provide a concise overview of Generalized Probabilistic Theories (GPTs) [16, 21], emphasizing their compositional attributes. We provide a brief introduction here, and refer the interested reader to, for example, Refs. [22, 23] for more details. Specifically, we are following the formalism of Refs. [19, 20].

Conceptually, a GPT is a theory about experiments that assigns probabilities to observation events, equipped with a compositional structure that mirrors the possibility we have to perform actions sequentially or in parallel. Formally speaking, the compositional aspects of the theory are captured by the fact that a GPT is a (strict) symmetric monoidal category (SMC) (see App. A). The probabilistic aspects are captured by the fact that we have a classical (stochastic) interface with the full theory in order to represent outcomes and control variables, formally, this means that we have the SMC **Stoch** (Sec. II B) as a full subtheory. This leads to a convex structure (Sec. II C 2) on the sets of processes with a given input and output, and allows us to define a notion of tomography (Sec. II C 3). Finally, we capture the requirement that the theory interact well with relativistic causal structure, by demanding the existence of unique discarding maps (Sec. II C 4).

In the rest of this section, we will introduce the diagrammatic notation used throughout this work, and discuss the defining features of a GPT that we mentioned above.

A. Diagrammatic notation

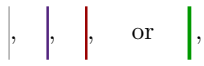
An interesting feature of SMCs is that they have a diagrammatic representation with which we can perform every calculation that we could using their axiomatic definition [24–26]. In the context of GPTs, we can represent their processes as boxes with input and output wires, and encode the composition of these processes by how they are wired together.

In the diagrammatic notation, each wire is named to represent a system type, and we follow the convention where those connected to the bottom of the boxes represent the input types of the process, while those at the top are the outputs. Note that this means that, in our convention, "time" in the diagrams flows from the bottom up. In this way we can represent a process $f : A \to B$, that takes a system of type A to a system of type B, as follows:

$$f : A \to B \quad \doteq \quad \boxed{f}^{\,B}_{\,A}\,, \tag{1}$$

where we are using $\doteq$ to indicate the translation from one notation into another.

We often omit wire labels for simplicity, and/or use colors to encode certain information about the system type. For instance, in this paper we will use



(2)

where, for example, the first of these represents a classical system of unspecified dimension, and the meaning of the others will be explained in section IV. To represent composite types such as $A \otimes B$, we just put their wires side by side, as in

$$A \otimes B \quad \doteq \quad {}_{A}\;{}_{B}\,. \tag{3}$$

Using this notation for composite systems, a process with composite input or output wires is depicted as having multiple input/output wires, e.g.,

$$f : A \otimes B \to C \otimes D \otimes E \quad \doteq \quad \boxed{f}^{\,C\;D\;E}_{\,A\;\;B}\,. \tag{4}$$

One system type that every GPT must contain is the trivial system, which corresponds to having no system at all. We refer to it in text as I. Since the trivial system is the unit for parallel composition (i.e., the monoidal unit of the

symmetric monoidal category), we have $A \otimes I = A = I \otimes A$, diagrammatically, I is represented by empty space:

$$I \quad \doteq \quad . \tag{5}$$

States and effects can be seen as preparation and observation procedures, respectively, which are processes that start and end in the trivial system, i.e., they must not have input or output wires respectively. For example, if s is a state and e an effect, then we denote them as

$$s : I \to A \quad \doteq \quad \text{and} \quad e : A \to I \quad \doteq \quad . \tag{6}$$

There can also be processes with both input and output as the trivial system, $p : I \to I$, which are represented by diagrams without open wires. The compositional properties of the SMC imply that diagrams of this kind can be composed together with a multiplicative structure, and hence can be called numbers. For instance, we could have

$$1 : I \to I \quad \doteq \quad . \tag{7}$$

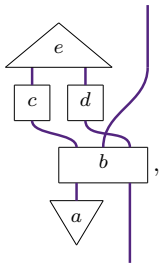
Finally, we represent the parallel composition $\otimes$ of processes by drawing their boxes side by side, and their sequential composition $\circ$ by connecting the input and output wires of matching types. That is, for $f : A \to B$ and $g : C \to D$

$$f \otimes g : A \otimes C \to B \otimes D \quad \doteq \quad , \tag{8}$$

and for $f : A \to B$ and $g : B \to C$

$$g \circ f : A \to C \quad \doteq \quad . \tag{9}$$

One example of a more complex diagram is



, (10)

where we omit the labels of the wires, but it should be understood that connections are allowed only when types match.

This notation and the rules for composing diagrams are common to all (strict) symmetric monoidal categories. Now, it remains to discuss features that are shared only by those who can be considered as GPTs. Since one of the ingredients of a GPT is that they contain **Stoch** as a full subtheory, we start from the definition of that theory.

B. Example: classical stochastic maps

As we mentioned, any GPT must have **Stoch** as a full subtheory. The simplest possible GPT, then is the one that contains nothing else (if the other properties are satisfied, of course, which is the case).

In order to define **Stoch**, all we have to do is to define what concrete mathematical objects correspond to its system types, states, effects, transformations, and composition rules (parallel and sequential composition). We organized this information in the following table:

Element	Definition	Example
System types	Real vector spaces	$\mathbb{R}^2$
States	Probability column vectors	$\begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$
Effects	Row vectors whose all entries are equal to 1	$\begin{pmatrix} 1 & 1 \end{pmatrix}$
Transformations	Stochastic matrices	$\begin{pmatrix} 1/2 & 1/3 \\ 1/2 & 2/3 \end{pmatrix}$
Sequential Composition	Matrix multiplication	$\begin{pmatrix} 1/2 & 1/3 \\ 1/2 & 2/3 \end{pmatrix} \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$
Parallel Composition	Kronecker product (or tensor product)	$\begin{pmatrix} 1/2 \\ 2/3 \end{pmatrix} \otimes \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$

Because we require that GPTs have this theory as a full subtheory, it will act as an interface to provide the GPT with the probabilistic interpretation that we need. For example, in this framework we describe a measurement as a process from a general system to a system in **Stoch**.

C. Defining properties of causal GPTs

Not every SMC can be considered as being a hypothetical theory of physics. In this section, we characterise those that can. In particular, what we are looking for with this characterization is to use **Stoch** as an interface to the theory that enables us to make statistical predictions in a manner coherent with its compositional structure, and where we can characterize the objects by the statistics that they can generate.

The additional features that an SMC has to satisfy in order to be a causal GPT are:

1. The SMC contains **Stoch as a full subtheory**.
2. There is a **convex structure** compatible with the one from **Stoch**.
3. There is a notion of **tomography**.
4. There is a **unique effect** associated to each system type.

In this manuscript we further focus on GPTs that satisfy the following additional property:

5. The theory is **locally tomographic**.

We now discuss each of those points in turn.

1. **Stoch** *is a full subtheory*

This means that all of the systems from **Stoch** and all of the processes from **Stoch** are also in the GPT, and, moreover, that when we compose these systems and processes in the GPT this matches the composition in **Stoch** [19]. Moreover, if we have a process in the GPT which only has inputs and outputs coming from **Stoch**, then this must be a process coming from **Stoch**.

The importance of that, is that inside a GPT, we can take the maps that go from a classical system (i.e. a system interpreted as a system of **Stoch** to another one as a stochastic process. Then, these processes, with all their internal probabilities, provide a probabilistic interpretation to the diagrams. Note that if it were not a *full* subtheory, then there would necessarily be situations in which the theory failed to make sensible probabilistic predictions, for example, giving negative probabilities for measurement outcomes.

For example, suppose we have a state of some general system in the GPT, A, then a (destructive) measurement for A would be a process with A as an input and some system X in **Stoch** as an output, when we compose these we are left with a process which must be a state in **Stoch**, namely, a probability distribution. It is precisely these probability distributions which encode the probabilistic predictions of the GPT.

We denote the systems coming from the subtheory **Stoch** as:

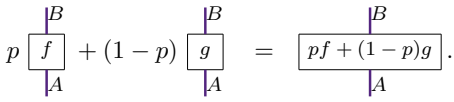
(11)

where we use a thin gray wire to distinguish the systems in the subtheory from generic systems in the GPT.

Note that in many other approaches to GPTs, the probabilities are encoded as scalars in the theory. In the approach we take here this is not the case, as, in particular, we find here that there is a unique scalar, the number 1. Instead, we obtain probability distributions over measurement outcomes via the states of the subtheory, **Stoch**. For example, this is what we obtain when we compose a state of a generic system in sequence with a measurement on that system.

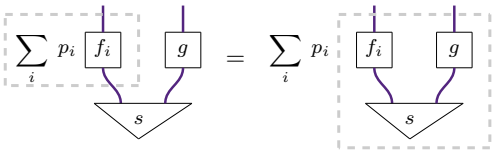
2. *Convex structure*

In order to naturally express statistical mixtures in the GPTs, we require them to be closed under convex mixtures of processes of matching input and output types. We require further that this composition is consistent with the convex composition from **Stoch** [19]. To start illustrating that, note that if we have $f : A \to B$ and $g : A \to B$, there must exist some $pf + (1-p)g : A \to B$ in the theory where we denote this as



(12)

Note that these combinations are allowed only when the input/output systems are the same for each of the combined processes. Moreover, these must distribute over diagrams, that is, they must satisfy, for example:



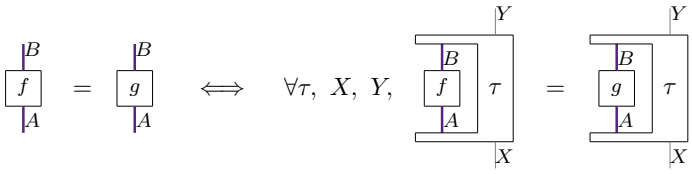
(13)

Finally, these convex combinations must match up with the standard notion of convex-combinations when specialised to the subtheory **Stoch**. This ensures that we can consistently view these convex combinations as describing our classical uncertainty about which process is happening.

3. *Tomography*

The next requirement that a GPT must satisfy is to have a notion of tomography [27]. What that means is that we should be able to characterize its elements – i.e., the states, effects and transformations – by the statistics that they are capable of generating. In this way, an experimentalist would be able to characterize the theoretical objects describing their experiment by connecting the statistics to the probabilities that the theory predicts.

To have a notion of tomography of processes, we need to always be able to establish equalities between them by looking at the statistics that they can generate. In a GPT, this means the following: we require that if it is the case that whenever we swap the process $f : A \to B$ by the process $g : A \to B$ in any diagram that represents a stochastic map, that map is kept unchanged, then it must be that $f = g$. That is,



(14)

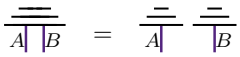
Here we are using τ to represent an arbitrary diagram that, after inserting f in some specific spot thereof, has only classical inputs and outputs left, and so is a process in **Stoch**. Note that this includes the case where any of the input/output wires of τ are the trivial system, because the trivial system is a classical (that is, **Stoch**) system. This condition can be phrased in the following way: two processes f and g from A to B are equal (left hand side of Eq. (14)) if and only if they are *operationally equivalent* (right hand side of Eq. (14)).

4. *Causality*

In this work we are interested in GPTs that are causal [28, 29]. By that, we mean that for each system type A, there is a unique effect that we can think of as discarding, or simply ignoring, a given system. This property is called causality because it can be used to impose compatibility of the GPT with a relativistic causal structure [30]. When the theory satisfies causality, we use a special diagram to denote the unique (for each system A) discarding effect:

(15)

Note that the uniqueness of the discarding effects is given for each fixed system type. In particular, this means that for composite systems the discarding is obtained by parallel composition of the discarding of the subsystems:

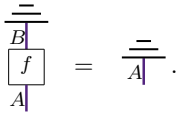


(16)

The discarding effects will be used in the next section to define non-signalling channels for a general GPT, just like the trace is in quantum theory.

The fact that there is a unique effect immediately means that all of the processes are discard-preserving [29]:

Definition II.1 (Deterministic, or Discard-Preserving, Process)**.** *A process $f : A \to B$ is deterministic if it is discard preserving, that is,*

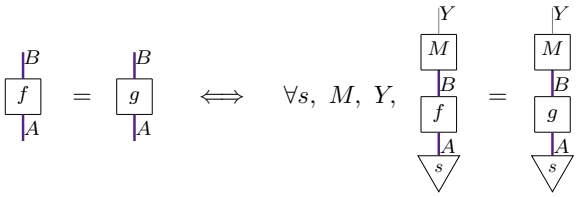


(17)

In quantum theory, since discarding is the trace operation, this corresponds to the trace-preserving property. That is, the formalism that we are using here is the analogue of working with only CPTP maps rather than working with CPTNI maps. Typically, CPTNI maps are used to describe the potential outcomes of some measurement, we can instead equally well work only with CPTP maps, by instead considering all possible outcomes at once, and keeping track of which outcome occurred by means of an auxiliary classical system.

5. *Local tomography*

In this work, we are interested in GPTs that satisfy a stricter notion of tomography. We require that the tomography of the processes can be done by evaluating the probabilities produced by local effects, that is, we require our GPT to satisfy local tomography [21]. This is expressed diagrammatically by the following:



(18)

where s is an arbitrary state of A, Y is an arbitrary classical system, and M is an arbitrary measurement of B. Note that this is taking a particular, less general, shape for τ in the definition of tomography.

Remark II.2. *A very convenient fact about locally-tomographic GPTs is that they are all subtheories of* $\mathbb{R}$**Linear** *[31–33] (Example 1 in Appendix A), in the sense that all of the processes of the former are in the latter (or more rigorously, there is an injective map between their processes and system types), and they compose according to* $\mathbb{R}$**Linear** *compositional rules. This will come in handy, as in our construction we will use the fact that our GPT is one of* $\mathbb{R}$**Linear***'s subtheories to write its processes in a mathematically concrete way. In particular, both classical and quantum theory satisfy local tomography, and therefore are also subtheories of* $\mathbb{R}$**Linear***.*

Now that we are done discussing the structure of the generalised probabilistic theories, we can proceed and focus on the properties of the processes that we are interested in investigating inside those theories. Namely, we can talk about the non-signalling channels.

III. CHANNELS IN GENERALISED PROBABILISTIC THEORIES

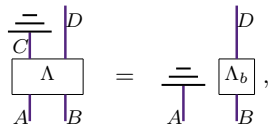
In this section we discuss, in the context of generalised probabilistic theories, the two classes of channels of interest for this paper: the non-signalling channels, and the common-cause channels (which form a subset of the non-signalling channels, as we will see).

A. Non-signalling channels

A practical starting point to understand what non-signalling channels in GPTs are is to remind ourselves of what they are in quantum or classical theory.

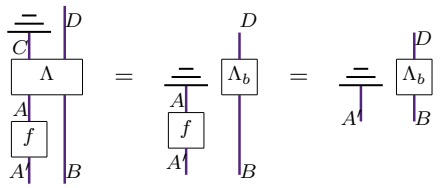
Quantum channels are formally completely-positive trace-preserving maps on density matrices, and specify ways in which quantum systems can be transformed. The properties of quantum channels are widely studied in the literature [34], and of particular interest are the quantum channels that satisfy a form of the no-signalling principle [35], introduced first by Beckman, Gottesman, Nielsen, and Preskill [36] in bipartite setups. These non-signalling quantum channels are sometimes referred to as "causal channels" [37], and do not permit superluminal quantum (nor classical) communication between two parties – i.e., two wings of the experiment. Non-signalling channels were discussed in the context of multipartite setups by Schumacher and Westmoreland [37].

In general theories – not necessarily quantum or classical – one can also define the concept of a channel as a transformation in the theory that is discard-preserving (Def. II.1), that is, one that preserves, on any state, the result of the application of the discarding process. In this context, we can talk about the property of a channel being non-signalling. In this section, we present a convenient definition of non-signalling channels in the diagrammatic language that we presented in Sec. II. Specifically, we want to diagrammatically represent the idea that no information can flow between the parties. Consider, for example, a bipartite process $\Lambda : A \otimes B \to C \otimes D$. If by discarding system C the resulting process $A \otimes B \to D$ is such that changing system A does not produce any changes in system D, then Λ cannot signal from the AC wing of the experiment to the BD wing of the experiment. In other words, we say that $\Lambda : A \otimes B \to C \otimes D$ is non-signalling from AC to BD if and only if



(19)

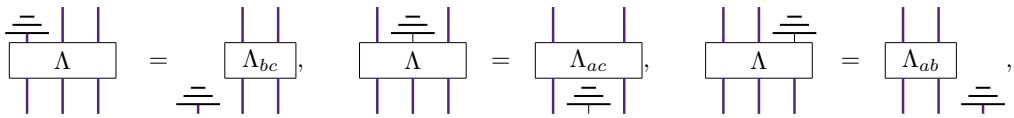
where $\Lambda_b : B \to D$ is a valid channel within the theory [29]. Note, in particular, that this implies that the application of any deterministic process (Def. II.1) in the AC wing does not change the marginal channel Λ_b:



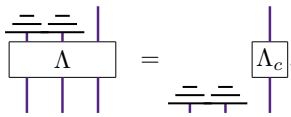
(20)

hence, no information can flow from the AC wing to the BD wing of the experiment. A channel is then said to be non-signalling when it satisfies that property in both directions between the wings of the experiment.

So far we have presented the case of bipartite non-signalling channels, but the notion of a multipartite non-signalling channel has also been defined in the literature [37]. Here we present a convenient diagrammatic definition of multipartite non-signalling channels. In order to define the multipartite generalisation of this condition we need a convenient way to represent discarding an arbitrary subset of the outputs. To see why, suppose that Λ is a tripartite channel. If we want to guarantee that no information can flow from any of the subsystems to any other, we need to have that



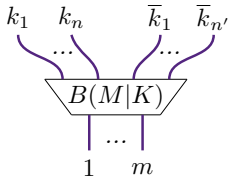
(21)



(22)

and so on. It is easy to see that this can become quite complex quickly as we increase the number of parties. In order to capture this in a succinct diagrammatic form, we need a notation which allows us to describe discarding an arbitrary subset of the outputs (or inputs), for this purpose we first introduce a bipartitioning processes as follows:

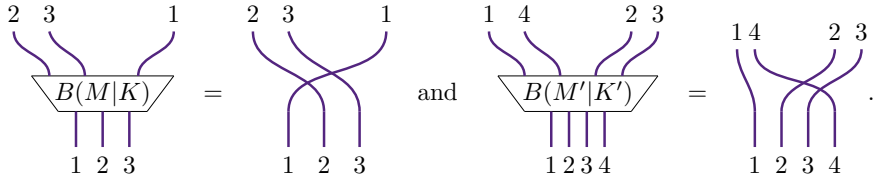
Definition III.1 (Bipartitioning processes $B(K)$)**.** *Given a set $M = \{1, ..., m\}$ take a labelled subset $K = \{k_1, ..., k_n\} \subset M$ and its complement $\overline{K} = \{\overline{k}_1, ..., \overline{k}_{n'}\} = M \setminus K$, where $n + n' = m$. Then, the bipartitioning process $B(K)$ is the permutation which takes $(1, ..., m)$ to $(k_1, ..., k_n, \overline{k}_1, ..., \overline{k}_{n'})$. Diagrammatically, we represent this by*

$B(M|K) \doteq$ 

. (23)

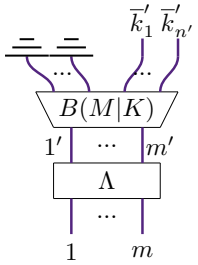
where we are using numbers, instead of system type names, to refer to the wires for the sake of clarity. ●

For example, if we take $M = \{1, 2, 3\}$ and $K = \{2, 3\}$, or $M' = \{1, 2, 3, 4\}$ and $K' = \{1, 4\}$ then we have, respectively,



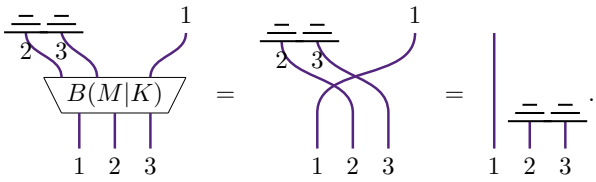
. (24)

We can then use this bipartitioning operation to concisely notate discarding some subset$K$ of the outputs M of a channel Λ, i.e.,



, (25)

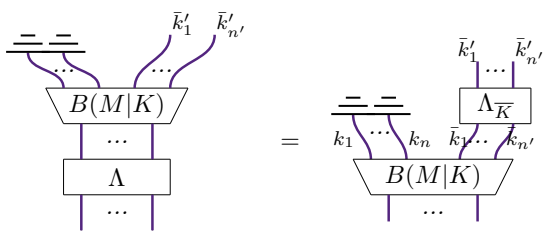
which in quantum theory would represent the partial trace $\mathrm{tr}_K(\Lambda)$, up to a permutation of the surviving systems. For example, in the tripartite case we can represent discarding the second and third outputs by



. (26)

We can now present the definition of multipartite non-signalling channels in a succinct diagrammatic form:

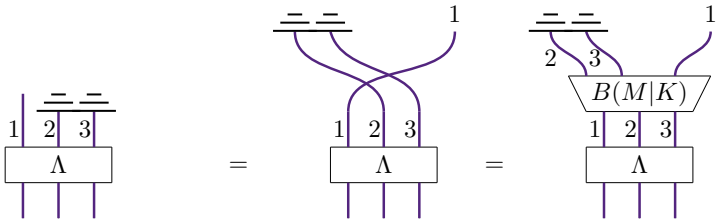
Definition III.2 (Non-signalling channel)**.** *An m-partite channel $\Lambda : \otimes_{i=1}^{m} i \to \otimes_{i'=1}^{m} i'$ is non-signalling iff for all labelled subsets $K \subset \{1, ..., m\}$, there exists a channel $\Lambda_{\overline{K}} : \otimes_{i=1}^{n'} \overline{k}_i \to \otimes_{i=1}^{n'} \overline{k}'_i$, with $\overline{K} = \{1, \ldots, m\} \setminus K$, such that*



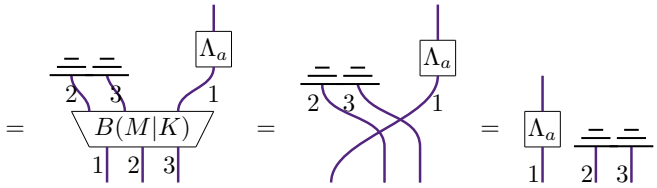
. (27)

●

To illustrate this, one of the conditions that this definition would impose on the tripartite case ($M = \{1, 2, 3\}$) would be for $K = \{2, 3\}$, which would give



(28)



(29)

that is, we can see explicitly how our condition gives us no signalling from $2 \otimes 3$ to 1. It is straightforward to similarly verify that the other conditions in the tripartite case are recovered by varying over the subsets $K \subseteq M$.

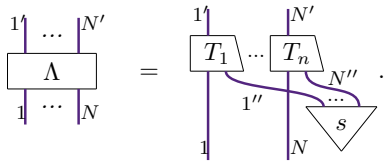
Notice that this definition of a non-signalling channel treats each pair of input/output systems (i, i') as a different wing of the experiment. Therefore, when specifying the experimental scenario and the channel Λ the systems should be represented via 'one wire per wing'. As an example, consider the case where one wing of the experiment consists of two qubits forming a 4-dimensional quantum system as an input: then this must be represented by one 4-dimensional system – rather than by two wires representing two qubits – when Definition III.2 is applied, since signalling is allowed between the wing's internal two qubits.

B. Common cause channels

To formally state the question tackled in this paper, we first need to specify the notion of a common-cause channel that we use in this manuscript. Broadly speaking, the common-cause channels are a subset of the non-signalling channels. Namely, we say that a channel is common-cause if, in the GPT of interest, it can be constructed by the parties via the application of some local operations to a shared multipartite state. A good example of such a channel is the one obtained in a Bell experiment, where, for example, Alice and Bob each make measurements on their shares of a Bell state. One can view the result of the Bell experiment as being a bipartite classical channel which is realised by local operations on a shared quantum state, i.e., a quantum common-cause.

Based on this example, we can define the notion of a common-cause decomposition within a given GPT $\mathbf{G}$.

Definition III.3 (Common-cause decomposition)**.** *Let Λ be a channel in a given GPT $\mathbf{G}$. Λ admits of a common-cause decomposition if there are N systems $\{1'', \ldots, N''\}$ from $\mathbf{G}$, a state s in the state space of the multipartite system $1'', \ldots, N''$ and a collection $\{T_i\}_{i=1\ldots N}$ of transformations in $\mathbf{G}$, such that*



(30)

●

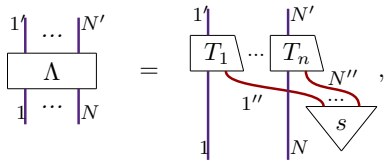
One can compare this formal diagrammatic definition to the conceptual definition to see that indeed the idea of construction by local operations (the transformations T_i) on a shared common cause (the state s) is indeed captured by this diagram.

Now, the idea of common-cause decomposition within a GPT might not be enough if one is considering the possible existence of some hypothetical cause that might not be modeled by the GPT under consideration. In particular, this is precisely the kind of situation that is considered in the resource theories of Refs. [1, 6]. In such cases, the more appropriate question is not whether Λ can be realised with a common cause in $\mathbf{G}$, but whether or not there exists a theory $\mathbf{G}'$ in which it can be realised with a common cause. Going back to our example of the Bell experiment, if we

violate a Bell inequality, then we know that the resulting channel cannot be realised via common case within **Stoch**, but it can be realised via a quantum common cause, that is, within the quantum GPT, **Quant**.

For that purpose, we define the notion of *GPT-common-cause realisable*, by asking whether the common-cause decomposition of Λ exists in *any* GPT.

Definition III.4 (GPT-Common-cause realisable channel)**.** *Let Λ be a channel in a given GPT $\mathbf{G}$. Λ is GPT-common-cause realisable if there exists a GPT $\mathbf{G}'$ which contains $\mathbf{G}$ as a full subtheory, N systems $\{1'', \dots, N''\}$ from $\mathbf{G}'$, a state s in the state space of the multipartite system $1'', \dots, N''$ in $\mathbf{G}'$, and a collection $\{T_i\}_{i=1\dots N}$ of transformations in $\mathbf{G}'$, such that*



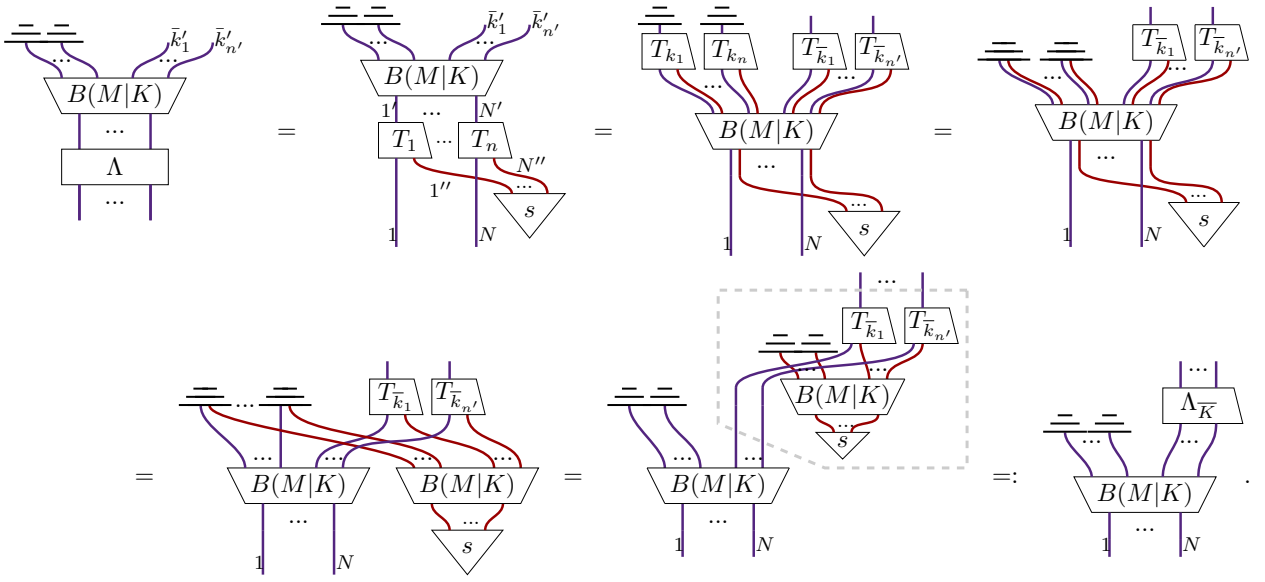
(31)

where we changed the colors of the i'' wires to stress the fact that they can be present only in the hypothetical GPT $\mathbf{G}'$, whilst the wires i and i' are required to live in the original subtheory $\mathbf{G}$. ●

Common-cause realisable channels are well known to be non-signalling, here we present this result using the diagrammatic notation that we have set up so far.

Proposition III.5. *Any GPT-common-cause realisable channel is non-signalling.*

Proof. Consider a fixed but arbitrary channel Λ in a GPT $\mathbf{G}$. Let $\mathbf{G}'$ be the GPT that provides the common-cause realisation of Λ. First, notice that, because in Eq. 27 the T_i channels are discard-preserving, if we take Λ to be decomposed as in Eq. 31, we get



(32)

where $\Lambda_{\overline{K}}$, the channel defined by combining the elements within the dashed box, must be a valid channel from $\mathbf{G}$ because all of its inputs and outputs are from $\mathbf{G}$, and $\mathbf{G}$ is assumed to be a full subtheory of $\mathbf{G}'$. □

The main aim of the paper is hence to explore the converse direction to Proposition III.5, namely, whether non-signalling channels can in general be common-cause realisable. The first observation to make is the well-known fact that the non-signalling classical channel known as Popescu-Rohrlich (PR) box[1] [35] does not have a common-cause realisation within classical theory [38], but it does have one such realisation within the GPT known as Boxworld [16].

[1] The PR box can be thought of as a non-signalling classical channel that takes classical systems to classical systems.

In this sense, hence, we say that the classical GPT is *common-cause incomplete*. Moreover, we further view Boxworld as adding extra common causes to classical theory, and so can be thought of as a *common-cause completion* of classical theory. This discussion motivates the following definition:

Definition III.6 (Common-cause complete GPT)**.** *A GPT is said to be common-cause complete if a common-cause decomposition can be found for each of its non-signalling channels within the theory. That is, given a non-signalling channel Λ in the GPT, we can decompose it as in Definition III.4 taking $\mathbf{G}' = \mathbf{G}$.* ●

The previous observation shows that there are some GPTs – such as classical and quantum theory – which are not common-cause complete. However, classical theory does have a common-cause completion. The question we therefore ask is whether or not this is generic? That is:

Given some GPT $\mathbf{G}$, can we find a common-cause completion $\mathbf{G}'$ such that all of the non-signalling channels of $\mathbf{G}$ have a GPT common-cause realisation in $\mathbf{G}'$ (Def. III.4)?

Formally, we defined the common-cause completion as follows:

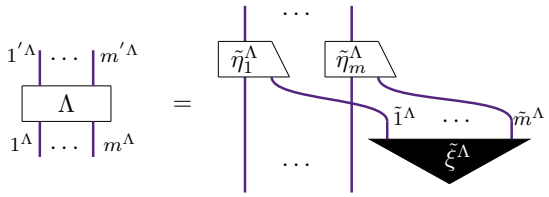
Definition III.7 (Common-cause completion)**.** *A GPT $\mathbf{G}'$ is a common-cause completion of a GPT $\mathbf{G}$ if $\mathbf{G}$ is a subtheory of $\mathbf{G}'$, and $\mathbf{G}'$ contains a common-cause decomposition (as per Definition III.4) of all of the non-signalling channels of $\mathbf{G}$. Note that this definition does not require $\mathbf{G}'$ to be common-cause complete itself.* ●

In the following section we show that any tomographically-local GPT does indeed have a common-cause completion.

IV. COMMON-CAUSE COMPLETION

In this section, we provide a construction $\mathcal{C}$ which takes an arbitrary locally-tomographic causal GPT $\mathbf{G}$ into a common-cause completion thereof, $\mathcal{C}[\mathbf{G}]$. The starting point of our construction relies on the following Lemma, proven in Ref. [39].

Lemma IV.1. *(Affine common-cause decomposition of non-signalling channels [39]) In a locally-tomographic GPT $\mathbf{G}$, any m-partite non-signalling channel, Λ, can be written as:*



(33)

where $\tilde{\eta}_i^{\Lambda}$ are discard-preserving processes in $\mathbf{G}$, and $\tilde{\xi}^{\Lambda}$ is an affine combination of states from $\mathbf{G}$ (e.g., when $\mathbf{G}$ is quantum theory, $\tilde{\xi}^{\Lambda}$ is a unit trace Hermitian operator). Note that we have drawn $\tilde{\xi}^{\Lambda}$ as a black box to indicate that, whilst it is a mathematically valid object, it is not necessarily a physical process within the GPT $\mathbf{G}$[2]*.*

Proof. Theorem 5.1 of Ref. [39]. □

This lemma, at first glance, provides a common-cause realisation of any non-signalling channel. However, these affine combinations of states $\tilde{\xi}^{\Lambda}$ are not (in general) going to be valid states in the GPT. One route to a solution could therefore be to define a common-cause-completion by enlarging the state-space so that it now includes these non-physical states. The problem with this approach, however, is that it does not necessarily yield a well-defined GPT, since this procedure will often lead to negative probabilities for measurement outcomes when we start composing these states in ways other than the diagram described in Eq. (33).

In order to prevent negative numbers from arising, then, one can by fiat forbid certain 'undesired' compositions. That is, one needs to equip the produced theory with *restrictions* on how the processes may be composed – type-matching conditions would no longer be a sufficient compositional criterion. Such a theory is, in the language of Ref. [40], called a "non-free" GPT as one is not free to compose processes solely based on their system types. Whilst mathematically consistent, we find it difficult to justify such restrictions on physical grounds, and hence we will not pursue its study further in this paper. In what follows, we instead provide a construction of a valid common-cause completion map, which, given a causal tomographically local GPT will always build a valid GPT, where composition precisely follows the GPT rules as per section II.

[2] That is, when we view the tomographically local theory as a subtheory of $\mathbb{R}$**Linear** we can then take arbitrary affine (or more generally linear) combinations and have a well defined process in $\mathbb{R}$**Linear** but this then could be outside of the subtheory $\mathbf{G}$.

A. Constructing the $\mathcal{C}$ map

Here we define a common-cause completion map, $\mathcal{C}$ which takes an arbitrary tomographically-local GPT, $\mathbf{G}$, as an input and then constructs a common-cause completion of it, $\mathcal{C}[\mathbf{G}]$, which is its output. The basic idea of this construction is to include all the non-physical states $\tilde{\xi}^{\Lambda}$ and $\tilde{\eta}_i^{\Lambda}$ from Lemma IV.1, but now with the caveat that the output systems of each $\tilde{\xi}^{\Lambda}$ (and consequently the inputs to the $\tilde{\eta}_i^{\Lambda}$) are taken to be new system types which are added to the theory. It will then be the type matching constraints (which are part of the basic definition of a GPT) which will prevent negative probabilities from arising when freely composing processes. It is not immediately clear, however, whether having done so we satisfy all of the other conditions of a GPT, and indeed this turns out not to be the case. Therefore, some extra steps are needed in the construction, in particular, to ensure that the theory is convex and tomographic.

In more detail, the steps followed in the construction, along with what they aim to achieve and how we denote them, are the following:

1. Take the non-signalling channels in $\mathbf{G}$ and decompose them as per Lemma IV.1. Take each output system of each $\tilde{\xi}^{\Lambda}$ and promote it to a new primitive system type. Collect all these new system types and, together with the system types from $\mathbf{G}$, define a new set of systems types including them all. Moreover, include as processes within the theory all of the processes from $\mathbf{G}$ together with all processes which are required such that these new systems can realise the common-cause channels as per Lemma IV.1.

   Aim: **To ensure that the common-cause decompositions** for non-signalling channels of $\mathbf{G}$ exist in $\mathcal{C}[\mathbf{G}]$.

   Notation: $\mathbf{G} \mapsto \mathbf{G} \sqcup \boldsymbol{\eta}$

2. Take the closure of those systems and processes under composition, and of the processes under convex combinations.

   Aim: **To ensure the compositionality and convexity rules are obeyed**.

   Notation: $\mathbf{G} \sqcup \boldsymbol{\eta} \mapsto \mathsf{Conv}[\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}]$

3. Quotient the theory $\mathsf{Conv}[\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}]$ via operational equivalence.

   Aim: **To ensure the theory satisfies tomography**.

   Notation: $\mathsf{Conv}[\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}] \mapsto \mathsf{Conv}[\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}]/\sim$

It is this theory that we will define as our common-cause completion, i.e. $\mathcal{C}[\mathbf{G}] := \mathsf{Conv}[\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}]/\sim$.

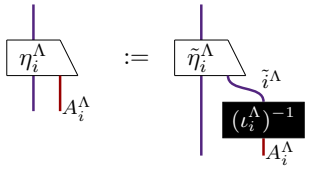
As we progress through the steps, we will show that they do indeed achieve the stated aim. In the end, we will therefore see that the outcome $\mathcal{C}[\mathbf{G}]$ of this construction is a valid causal GPT (in particular, that there are no extra restrictions on composing systems and processes) and that it is a common-cause completion of $\mathbf{G}$.

In this section we will be dealing with many system types from different GPTs (due to the nature of the problem of extending a theory), and therefore we shall use colors to differentiate the wires corresponding to different theories' system types. The convention we follow is given by the following table:

System Type	Wire Type
System from the classical subtheory, **Stoch**	
Generic system from the target GPT, $\mathbf{G}$	
Extra system to be added to $\mathbf{G}$	
Generic system in the new GPT	

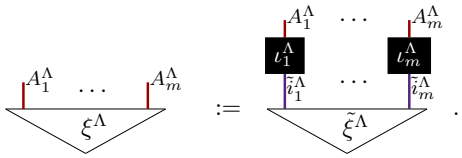
Step 1 - Add generating system types and processes

Starting from **G**, for each Λ in **G** decomposed as in Eq. (33), let us define a vector space A_i^Λ which is isomorphic to $\tilde{i}^\Lambda$ with isomorphism $\iota_i^\Lambda : \tilde{i}^\Lambda \to A_i^\Lambda$. Then, we define the following linear maps:



(34)

and

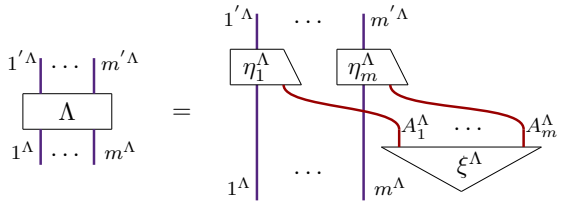


(35)

Note that the isomorphisms ι_i^Λ and their inverses are *not* taken to be physically realisable processes within the theory that we are constructing, hence, we denote them, as above, with black boxes. We will, however, take the above composites of them with the $\tilde{\eta}_i^\Lambda$ and $\tilde{\xi}^\Lambda$, to give η_i^Λ and ξ^Λ, to be valid processes in the theory we are defining, hence why the left-hand-side of Eqs. (34) and (35) are white-coloured boxes.

We therefore obtain the following straightforward corollary of Lemma IV.1:

Corollary IV.2. Any m-partite non-signalling channel, N, can be written as:



(36)

Proof. This immediately follows from the definition of the η_i^Λ and the ξ^Λ (Eqs. (34) and (35)) together with the fact that the ι_i^Λ are isomorphisms.

We include these extra systems A_i^Λ and processes ξ^Λ, η_i^Λ within the GPT we are building, thereby extending **G** and enabling the realisation of arbitrary non-signalling channels from **G** within the common-cause scenario.

Step 2 - Take closure under compositions and convex combinations

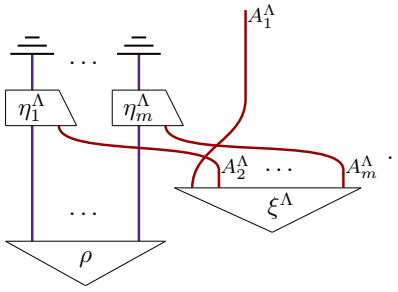
For the second step, let us denote by $|\mathbf{G}|$ the collection of systems of **G**, and (with slight abuse of notation) by **G** the collection of its processes. In order to define the closure properties that we want, we will note that we can view all of the processes that we have defined as living within the process theory of real linear maps, $\mathbb{R}$**Linear**. To see this, recall that **G** is, by assumption, tomographically local, and hence is a subtheory of $\mathbb{R}$**Linear**, and that the new systems and processes that we have added are all, by definition, real linear maps.

We therefore define another subtheory of $\mathbb{R}$**Linear** which is, by construction, closed under composition as follows:

Definition IV.3. *We denote by* $\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}$ *the subtheory of* $\mathbb{R}$**Linear** *whose objects (system types) are the closure of* $|\mathbf{G}| \sqcup \{A_i^\Lambda\}_{\Lambda,i}$ *under* $\otimes$*, and whose morphisms (processes) are the closure of* $\mathbf{G} \sqcup \{\eta_i^\Lambda, \xi^\Lambda\}_{\Lambda,i}$ *under* $\circ$ *and* $\otimes$ *as the operations in* $\mathbb{R}$**Linear**. ●

Note that, even though we did not explicitly mention the states of the A_i^Λ systems, these are implicitly defined by the above closure to obtain $\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}$. For example, by varying over ρ in the following diagram, we can obtain many

states of A_1^Λ:



(37)

In the same way, effects and other general processes on the new system types A_i^Λ can also be defined. The fact that we only have an implicit definition of the state and effect space is in stark contrast to traditional ways of constructing GPTs, in which the convex geometry of the state and effect spaces is typically the first thing to be defined and then the compositional structure is built on top of this. Here we invert this, first starting with the compositional structure and then defining the geometry of the states and effects which this provides.

Next we will check whether $\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}$ leads to sensible probabilistic predictions, namely, whether it contains **Stoch** as a full subtheory. To answer this we note that **Stoch** is a full subtheory of **G** and show that **G** is a full subtheory of $\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}$, hence, by transitivity, that **Stoch** is a full subtheory of $\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}$.

Specifically, what we need to show is that any process with all inputs and outputs in $|\mathbf{G}|$, such as

(38)

yields a valid process from **G**. Note that this is not guaranteed apriori, due to the fact that the new systems A_i^Λ appear in the interior of the diagram. However, in our case it turns out that this is true as is proven in the following lemma.

Lemma IV.4. *Any process in $\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}$ with only input and output system types in $|\mathbf{G}|$ is a valid process in* **G**.

Proof. The proof can be found in Appendix B 1. □

Next we show that $\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}$ is compatible with relativistic causal structure, in the sense that there is a unique effect for each system [29, 30].

Lemma IV.5. *There is a unique effect for each system in $\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}$.*

Proof. The proof can be found in Appendix B 2. □

A GPT must also be closed under convex combinations so as to model probabilistic mixtures of processes, and so far we have not proven that this is the case for $\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}$. Indeed, it is conceivable that this property has been lost when adding in the new systems and processes and arbitrary diagrams thereof. Hence, we take the convex closure of $\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}$, via the convex combinations of linear maps provided by the supertheory $\mathbb{R}$**Linear**.

Definition IV.6. *We denote by $\mathsf{Conv}[\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}]$ the convex closure of $\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}$ under convex combinations of processes taken as linear combinations of linear maps from* $\mathbb{R}$**Linear**. ●

Notice that the properties of 'has **Stoch** as a full subtheory' and 'is causal' that we proved for $\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}$ are properties which must hold in any GPT, hence we next show they also hold for $\mathsf{Conv}[\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}]$:

Lemma IV.7. *i) Any process in $\mathsf{Conv}[\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}]$ with only input and output system types in $|\mathbf{G}|$ is a valid process in* **G**. *ii) There is a unique discarding effect for each system in $\mathsf{Conv}[\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}]$.*

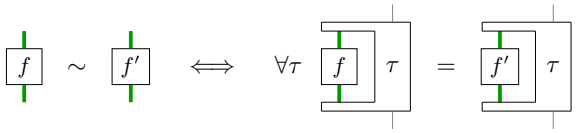
Proof. The proof can be found in Appendix B 3. □

Step 3 - Quotient the theory

There is one final property which must be satisfied in order to have a GPT on our hands, that is, tomography. That means that we need to be able to establish the equality between two processes when the probabilities that they can produce are the same. At this point, however, we do not know that $\mathsf{Conv}[\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}]$ satisfies this property. Hence, we need a way to "merge" any two differently-labelled but operationally-equivalent processes (defined shortly) into a single one.

To enforce this, we simply take the quotient $\mathsf{Conv}[\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}]$ under operational equivalence. That amounts to defining processes to be equivalence classes and also the operations of sequential, parallel, and convex compositions thereof. For this, let us first formally specify what we mean by "operational equivalence".

Definition IV.8. *Processes f and f' (with the same input systems and the same output systems) are operationally equivalent if they give the same statistical predictions when composed with any circuit fragment τ such that the resulting process has only classical inputs and outputs:*



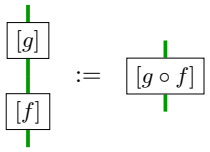
(39)

●

Note that we are using green wires to denote arbitrary systems which may be **G**-type, the new systems A_i^{Λ}, or even systems of the quotiented theory, because operational equivalence is a concept defined independently of the theory. In any case, we will apply this here only to $\mathsf{Conv}[\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}]$ in order to construct the quotiented theory. We denote the equivalence classes defined by this by square brackets, hence we can write that $f \sim f' \iff [f] = [f']$, and moreover think of some $f' \in [f]$ as providing a representative for the equivalence class of operations that f' belongs to.

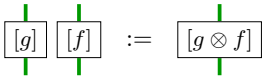
In order to build a theory in which processes are labelled by equivalence classes of processes, we must first define a notion of composition for the equivalence classes.

Definition IV.9. *The equivalence classes of processes compose sequentially as*



(40)

and compose in parallel as

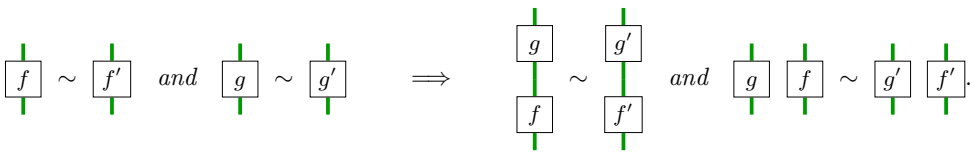


(41)

●

For these to be valid operations between equivalence classes, they must not depend on the choices of representatives:

Lemma IV.10. *Composition as defined in Def. IV.9 is independent of the choices of representatives. That is,*

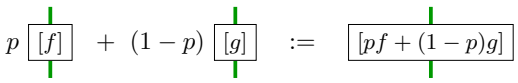


(42)

Proof. The proof can be found in Appendix B 4. □

In a similar way we can define convex combinations of equivalence classes as follows:

Definition IV.11. *Convex mixtures of equivalence classes of processes are given by the following:*

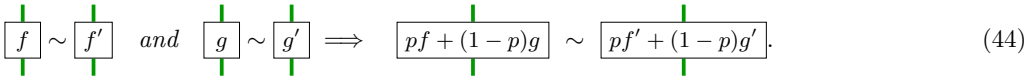


(43)

●

It is easy to see that the relevant properties of convex combinations, for example distributivity over $\circ$ and $\otimes$, are immediately inherited from the analogous property in the prequotiented theory. Again, for consistency, we prove the following:

Lemma IV.12. *Convex mixtures as defined in Def. IV.11 are independent of the choice of representative. That is:*



(44)

Proof. The proof can be found in Appendix B 5. □

These operations allow us to define the quotiented theory as follows:

Definition IV.13. *We denote the theory whose processes are operational equivalence classes of the processes in* $\mathsf{Conv}[\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}]$*, with composition and convex mixtures given by Defs. IV.9 and IV.11, by* $\mathsf{Conv}[\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}]/\sim$ ●

Note that, as $\mathbf{G}$ is a GPT, and hence satisfies tomography, for a valid process f in $\mathbf{G}$, we have $[f] = \{f\}$, that is, each equivalence class of processes in $\mathbf{G}$ contains a single element. It is then clear that Lemma IV.4 also holds for our quotiented theory. Moreover, it is also clear that Lemma IV.5 continues to hold even in our quotiented theory, as quotienting could only identify effects for a particular system with one another, and as we only have a unique effect for a given system in the first place we have a unique effect after quotienting.

The theory $\mathsf{Conv}[\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}]/\sim$ therefore satisfies all of the desired properties to be considered a causal GPT.

While the GPT that we constructed is $\mathsf{Conv}[\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}]/\sim$, it is clear that it is much easier to perform calculations within $\mathsf{Conv}[\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}]$ as it is simply a subtheory of $\mathbb{R}\mathbf{Linear}$. Luckily one can always perform calculations in $\mathsf{Conv}[\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}]/\sim$ by picking suitable representative elements for the equivalence classes, doing a computation within $\mathbb{R}\mathbf{Linear}$, and then requotienting to determine the resultant equivalence class.

Definition IV.14 (Common-cause completion map). *The map* $\mathcal{C}$ *given by* $\mathcal{C}[\mathbf{G}] \equiv \mathsf{Conv}[\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}]/\sim$ *is a common-cause completion map on the set of causal locally-tomographic GPTs.* ●

This is because $\mathcal{C}[\mathbf{G}]$ is a valid GPT which contains $\mathbf{G}$ as a full subtheory and where every $\Lambda \in \mathbf{G}$ has a common-cause realisation in $\mathcal{C}[\mathbf{G}]$.

V. RESULTS AND DISCUSSION

The construction we have presented for a common-cause completion map is useful as it allows us to understand possible causal explanations of physical phenomena. To elaborate on this, let us first introduce our main theorem and a useful corollary.

Theorem V.1. Given a locally-tomographic causal GPT $\mathbf{G}$, its set of multipartite non-signalling channels (Def. III.2) is the same as its set of multipartite common-cause realisable (Def. III.4) channels. Notice these common-causes might not be state-preparations allowed in $\mathbf{G}$

Proof. Consider the GPT $\mathcal{C}[\mathbf{G}]$. By Prop. III.5, the common-cause realisable channels in $\mathbf{G}$ are non-signalling. In the other direction, by construction, $\mathcal{C}[\mathbf{G}]$ can provide a common-cause realisation of any non-signalling channel of $\mathbf{G}$. □

Noting that **Quant** is a locally-tomographic causal GPT we immediately obtain the following:

Corollary V.2. There exists a causal GPT that provides a common-cause realisation of every non-signalling quantum channel. Such a GPT is given by $\mathcal{C}[\mathbf{Quant}]$.

This corollary is important for two reasons. Firstly, it answers in the negative 'Open Question 1' posed in Ref. [1]: *Do there exist bipartite non-signalling quantum channels which cannot be realized by GPT common causes?*.

Secondly, recall the phenomenon of Einstein-Podolski-Rosen (EPR) inference [41] (a.k.a. steering) where a party (say Alice) learns about the state preparation of a physical system (held by a distant party, hereon called Bob) by performing measurements on her share of the bipartite physical system [42, 43]. Here the object of study is the collection of *subnormalised conditional states* that Bob's subsystem may be prepared in, usually called an *assemblage* [44]. Similarly to the case of non-signalling correlations in Bell experiments, one may mathematically define general assemblages as those which comply with the no-signalling principle. Given the particular causal structure that underpins these EPR experiments, then, a crucial foundational question is whether these general assemblages could be realised within some (beyond quantum) GPT as a common-cause process. This question can be readily answered in the affirmative by Cor. V.2, given that assemblages in EPR scenarios can be formalised in terms of non-signalling quantum-classical channels [45]. This sets the foundation stone to be able to study the non-classicality of EPR assemblages based on the properties of the common-cause process within the GPT that may realise them. In particular, this observation answers in the affirmative the question posed in Ref. [2]: there exists a causal GPT $\mathbf{Q}'$ that provides a common-cause realisation of every general assemblage.

More generally, our result provides the fundamental justification of the possibility to assess and quantify the non-classicality of arbitrary non-signalling processes by means of the non-classicality of the common-cause required to realise them. This has previously been argued at length for the case of correlations in Bell scenarios [6], where the existence of common-cause realisations of non-signalling boxes had already been provided by the GPT known as Boxworld [16]. In this light, hence, our work enables the possibility of extending this causal reasoning to scenarios beyond Bell experiments, which involve other local systems types rather than strictly classical ones.

Looking forward, there are many open questions pertaining to the common-cause completion construction that we defined:

- Is $\mathcal{C}[\mathbf{G}]$ common-cause complete? Intuitively it seems that this should be the case, but conceivably there may be non-signalling channels between the new systems which are not realisable in common cause scenarios within $\mathcal{C}[\mathbf{G}]$. Note that $\mathcal{C}[\mathcal{C}[\mathbf{G}]]$ may not be well defined because we do not yet know whether or not:
  - $\mathcal{C}[\mathbf{G}]$ is tomographically local, or
  - whether or not there is a way to extend the common-cause completion to tomographically-nonlocal GPTs, or to more general kinds of process theories.

Whilst being of technical nature, we expect the answers to these questions to also help us deepen our understanding on the possible non-signalling processes that can be motivated, understood, and studied from the causal perspective.

[1] D. Schmid, H. Du, M. Mudassar, G. Coulter-de Wit, D. Rosset, and M. J. Hoban, Quantum **5**, 419 (2021).
[2] P. J. Cavalcanti, J. H. Selby, J. Sikora, T. D. Galley, and A. B. Sainz, npj Quantum Information **8**, 1 (2022).
[3] B. Coecke, T. Fritz, and R. W. Spekkens, Information and Computation **250**, 59 (2016).
[4] J. I. de Vicente, J. Phys. A **47**, 424017 (2014).
[5] J. Geller and M. Piani, J. Phys. A **47**, 424030 (2014).
[6] E. Wolfe, D. Schmid, A. B. Sainz, R. Kunjwal, and R. W. Spekkens, Quantum **4**, 280 (2020).
[7] D. Schmid, T. C. Fraser, R. Kunjwal, A. B. Sainz, E. Wolfe, and R. W. Spekkens, arXiv preprint arXiv:2004.09194 https://doi.org/10.48550/arXiv.2004.09194 (2020).
[8] B. Zjawin, D. Schmid, M. J. Hoban, and A. B. Sainz, Quantum **7**, 926 (2023).
[9] B. Zjawin, D. Schmid, M. J. Hoban, and A. B. Sainz, arXiv preprint arXiv:2209.10177 https://doi.org/10.48550/arXiv.2209.10177 (2022).
[10] D. Schmid, D. Rosset, and F. Buscemi, Quantum **4**, 262 (2020).
[11] D. Rosset, D. Schmid, and F. Buscemi, Phys. Rev. Lett. **125**, 210402 (2020).
[12] C. J. Wood and R. W. Spekkens, New Journal of Physics **17**, 033002 (2015).
[13] D. Schmid, J. H. Selby, and R. W. Spekkens, arXiv preprint arXiv:2009.03297 10.48550/arXiv.2009.03297 (2020).
[14] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Rev. Mod. Phys. **86**, 419 (2014).
[15] A. B. Sainz, N. Brunner, D. Cavalcanti, P. Skrzypczyk, and T. Vértesi, Phys. Rev. Lett. **115**, 190403 (2015).
[16] J. Barrett, Phys. Rev. A **75**, 032304 (2007).
[17] B. Coecke, Deep beauty: understanding the quantum world through mathematical innovation , 129 (2011).
[18] B. Coecke and A. Kissinger, *Picturing Quantum Processes: A First Course in Quantum Theory and Diagrammatic Reasoning* (Cambridge University Press, 2017).

[19] S. Gogioso and C. M. Scandolo, arXiv preprint arXiv:1701.08075 https://doi.org/10.48550/arXiv.1701.08075 (2017).
[20] J. H. Selby, C. M. Scandolo, and B. Coecke, Quantum **5**, 445 (2021).
[21] L. Hardy, arXiv preprint quant-ph/0101012 https://doi.org/10.48550/arXiv.quant-ph/0101012 (2001).
[22] M. Müller, SciPost Physics Lecture Notes , 028 (2021).
[23] M. Plávala, arXiv preprint arXiv:2103.07469 10.48550/arXiv.2103.07469 (2021).
[24] A. Joyal and R. Street, Advances in mathematics **88**, 55 (1991).
[25] P. Selinger, in *New structures for physics* (Springer, 2010) pp. 289–355.
[26] E. Patterson, D. I. Spivak, and D. Vagner, Electronic Proceedings in Theoretical Computer Science **333**, 49 (2021).
[27] L. Hardy and W. K. Wootters, Foundations of Physics **42**, 454 (2012).
[28] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. A **81**, 062348 (2010).
[29] B. Coecke, Electronic Proceedings in Theoretical Computer Science **172**, 27 (2014).
[30] A. Kissinger, M. Hoban, and B. Coecke, arXiv preprint arXiv:1708.04118 https://doi.org/10.48550/arXiv.1708.04118 (2017).
[31] L. Hardy, arXiv preprint arXiv:1104.2066 https://doi.org/10.48550/arXiv.1104.2066 (2011).
[32] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Quantum theory: informational foundations and foils , 171 (2016).
[33] D. Schmid, J. H. Selby, M. F. Pusey, and R. W. Spekkens, arXiv preprint arXiv:2005.07161 https://doi.org/10.48550/arXiv.2005.07161 (2020).
[34] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, 2010).
[35] S. Popescu and D. Rohrlich, Found. Phys. **24**, 379 (1994).
[36] D. Beckman, D. Gottesman, M. A. Nielsen, and J. Preskill, Phys. Rev. A **64**, 052309 (2001).
[37] B. Schumacher and M. D. Westmoreland, Quantum Information Processing **4**, 13 (2005).
[38] J. S. Bell, Phys. Phys. Fiz. **1**, 195 (1964).
[39] P. J. Cavalcanti, J. H. Selby, J. Sikora, and A. B. Sainz, Journal of Physics A: Mathematical and Theoretical 10.1088/1751-8121/ac8ea4 (2022).
[40] J. Barrett, N. de Beaudrap, M. J. Hoban, and C. M. Lee, npj Quantum Inf. **5**, 1 (2019).
[41] V. P. Rossi, M. J. Hoban, and A. B. Sainz, Journal of Physics A: Mathematical and Theoretical **55**, 264002 (2022).
[42] E. Schrödinger, Mathematical Proceedings of the Cambridge Philosophical Society **32**, 446–452 (1936).
[43] H. M. Wiseman, S. J. Jones, and A. C. Doherty, Phys. Rev. Lett. **98**, 140402 (2007).
[44] M. F. Pusey, Phys. Rev. A **88**, 032313 (2013).
[45] M. J. Hoban and A. B. Sainz, New J. Phys. **20**, 053048 (2018).

Appendix A: (strict) Symmetric Monoidal Categories

Since we define a generalised probabilistic theory (GPT) in terms of a strict symmetric monoidal category (SMC), we devote this appendix to define the latter. We follow that with a brief commentary on interpreting that structure in terms of processes, which is key for understanding how to see that GPTs are SMCs, and end with the most important example of SMC for this paper.

A (strict) symmetric monoidal category consists of (i) a collection of objects $A, B...$, (ii) for each pair of objects A, B, a collection of morphisms $f : A \to B$, and (iii) two operations, $\circ$ and $\otimes$ under which the category is closed. The first operation, $\circ$, maps certain pairs of morphisms to morphisms. In particular, it combines $f : A \to B$ and $g : B \to C$ into $g \circ f : A \to C$, and can be performed only when the domain of g matches the codomain of f (in this example, the matching is given by the object B). Furthermore, $\circ$ is associative, so it is similar to function composition. (iv) An identity morphism $1_A : A \to A$ that is a unit for $\circ$, is moreover associated to each object A. The second operation, $\otimes$, combines arbitrary pairs of objects, taking A and B to $A \otimes B$ as well as arbitrary pairs of morphisms, taking $f : A \to B$ and $g : C \to D$ into $f \otimes g : A \otimes C \to B \otimes D$. Furthermore, $\otimes$, is associative and has a unit object which we denote I, so it is a monoid operation on the collection of objects, being therefore responsible for the monoidal structure of the category. Finally, the two operations satisfy a consistency condition, namely that $(g \circ f) \otimes (g' \circ f') = (g \otimes g') \circ (f \otimes f')$.

An interesting property of the symmetric monoidal categories is that they feature a diagrammatic calculus, which provides an intuitive and expressive way to write and perform mathematical calculations. For a description of that, we refer the reader to the section II A of the main text.

The bare structure of the SMC has a nice interpretation in terms of processes [17]. We take the objects to represent system types, and call the monoidal unit, denoted I, the trivial system. The morphisms $f : A \to B$ are interpreted as processes that take a system of type A into a system of type B. The processes that start (but do not end) in the unit object (the trivial system), i.e., those like $s : I \to A$, are called states, the ones that end (but do not start) in I, like $e : A \to I$, are called effects, and the ones who neither start nor end in I, such as $f : A \to B$, are called transformations. This is intuitive because $s : I \to A$ can be viewed as some preparation procedure of a system of type A, and $e : A \to I$ as a destructive operation. Next, processes that start and end in I, such as $p : I \to I$, are called *numbers*, or scalars. Now, processes can happen sequentially or in parallel, and this is captured by the SMC

– we interpret $g \circ f$ as the sequential composition of the processes f and g, where f is followed by g (which acts on the output of f), and $f \otimes g$ as the composite process given by f and g occurring in parallel. This interpretation of $\circ$ and $\otimes$ motivates the consistency condition that they had to satisfy, since that is the natural relationship between processes happening in parallel and in sequence.

We now illustrate this abstract definition of an SMC by means of the key example for this paper:

Example 1 ($\mathbb{R}$**Linear**)**.** The SMC $\mathbb{R}$**Linear** takes objects (system types) to be real vector spaces, and, morphisms (processes) to be linear maps between the vector spaces. The $\circ$ operation is the composition of linear maps, and $\otimes$ is the tensor product. The identity morphisms are given by the identity linear maps, and the monoidal unit is given by the one dimensional vector space $\mathbb{R}$.

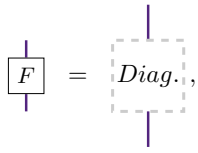
Appendix B: Proofs

1. Proof of Lemma IV.4

Lemma IV.4. Any process in $\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}$ with only input and output system types in $|\mathbf{G}|$ is a valid process in $\mathbf{G}$.

Proof. First note that, by definition, any process in $\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}$ can always be written as a diagram involving only our generating processes, that is, processes in $\mathbf{G}$, and the processes in $\{\eta_i^\Lambda, \xi^\Lambda\}_{\Lambda,i}$.

Now consider an arbitrary process F in $\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}$ with input and output system types in $|\mathbf{G}|$. This process can be written in terms of the above-mentioned generating processes:

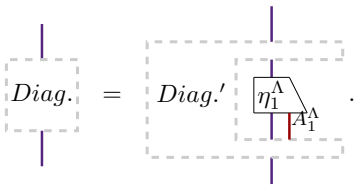


(B1)

where we do not specify the internal structure of the dashed box as the actual compositional structure of F has no generic specification but we assume it is a diagram consisting of generating processes.

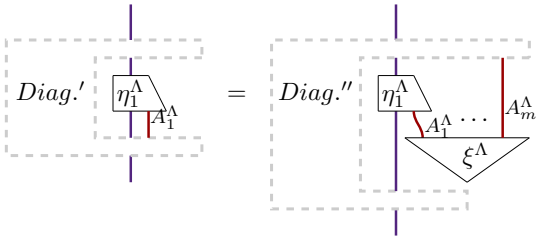
We will now show that this box associated to the process F can always be rewritten into a diagram which only involves processes in $\mathbf{G}$. This follows from the fact the we can rewrite any diagram using only generating processes.

Suppose, for example, that the diagram involves the process η_1^Λ, that is:



(B2)

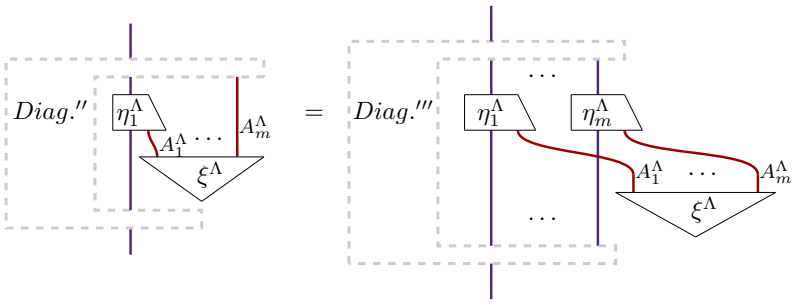
Since A_1^Λ is not an input to the process F (as we are assuming that the inputs and outputs system types are in $|\mathbf{G}|$), there must be a process in the diagram $Diag'$ for which this system, A_1^Λ, is an output. There is only one generating process which has A_1^Λ as an output, namely, ξ^Λ. Hence, we can write diagram $Diag'$ as:



(B3)

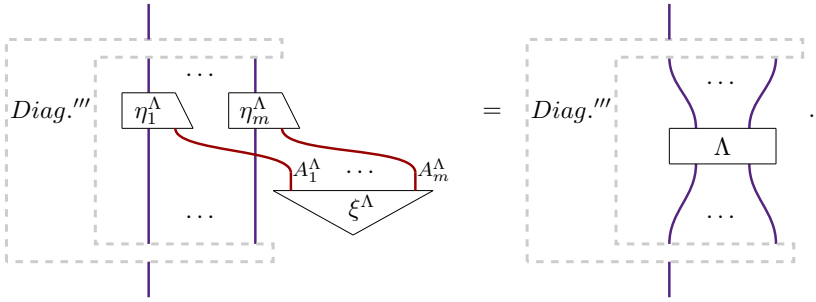
We also know that none of the A_i^Λ are outputs of the process, hence, they must be the input of some process within the diagram $Diag''$. For each of these there is a single generating process which has A_i^Λ as an input, namely, η_i^Λ. This

means we can rewrite the diagram $Diag''$ as:



(B4)

The explicitly drawn part of the diagram, however, is now nothing but the non-signalling channel Λ, which is a process that lives in $\mathbf{G}$:



(B5)

Hence, we have shown that we can redraw the diagram associated to F so as not to use the generating process η_1^Λ. This argument clearly also applies to any other η_i^Λ that may appear in the specification of F, and a very minor modification of it applies to any ξ^Λ.

This means that any process in $\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}$ whose input and output system have types in $|\mathbf{G}|$ can always be written in a way that only involves generating processes from $\mathbf{G}$ and does not involve any generating processes from $\{\eta_i^\Lambda, \xi^\Lambda\}$. As $\mathbf{G}$ is closed under composition, we have therefore shown that any process with input and output system types in $|\mathbf{G}|$ is necessarily a valid process in $\mathbf{G}$. □

Notice that, in particular, Lemma IV.4 implies that the theory $\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}$ that we have defined will make sensible probabilistic predictions, since the classical systems are valid systems in $\mathbf{G}$ and any processes with only classical inputs and outputs is necessarily a stochastic map.

2. Proof of Lemma IV.5

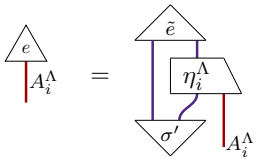
Lemma IV.5. There is a unique discarding effect for each system in $\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}$.

Proof. Here we show that every generating type has a unique discarding effect, as the generalisation to composite types is straightforward.

For each generating system i from the GPT $\mathbf{G}$, Lem. IV.4 implies that the discarding effect for i is itself a valid process $\mathbf{G}$. Since $\mathbf{G}$ is causal, this means that the discarding effect for i in $\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}$ is unique.

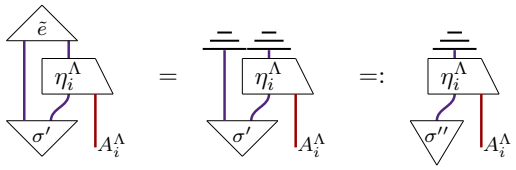
Now we need to show that the claim also holds for system types beyond those present in the GPT $\mathbf{G}$, i.e., the systems $\{A_i^\Lambda\}$.

Since all processes of $\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}$ can decomposed in terms of generating processes, we can write a generic effect for A_i^Λ as



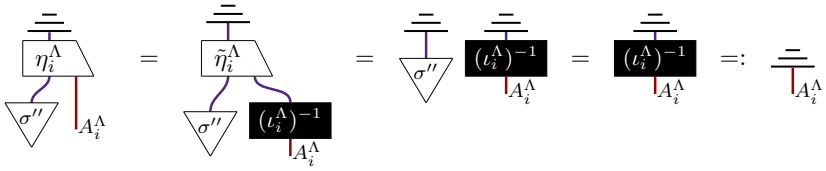
(B6)

where, thanks to Lemma B 1, we know that $\tilde{e}$ and σ' are necessarily in $\mathbf{G}$. Then, as there is a unique effect for each system in $\mathbf{G}$ we have that



(B7)

Now, using the definition of η_i^Λ we have that:



(B8)

Hence the extra systems in the enlarged theory still satisfy the property of having an unique effect to each system type. □

3. Proof of Lemma IV.7

Lemma IV.7. i) Any process in $\mathsf{Conv}[\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}]$ with only input and output system types in $|\mathbf{G}|$ is a valid process in $\mathbf{G}$. ii) There is a unique discarding effect for each system in $\mathsf{Conv}[\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}]$.

Proof.

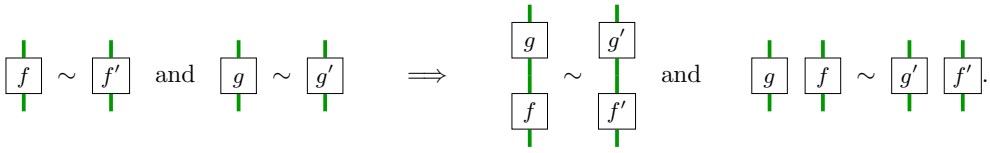
i) From Lem. IV.4 we know that any process in $\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}$ with only input and output system types in $|\mathbf{G}|$ is a valid process in $\mathbf{G}$. Since processes in $\mathbf{G}$ are closed under convex combinations, this implies that any convex combination of processes in $\overline{\mathbf{G} \sqcup \boldsymbol{\eta}}$ with only input and output system types in $|\mathbf{G}|$ is a valid process in $\mathbf{G}$, which proves the claim.

ii) That there is a unique discarding effect for each system immediately follows from Lem. IV.5 together with the fact that since there is a unique discarding effect for each generating system type it is impossible to obtain other discarding effects by composition and convex combinations.

□

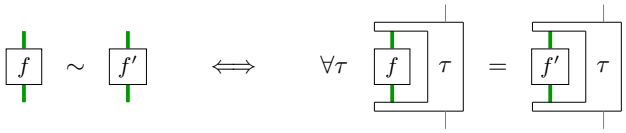
4. Proof of Lemma IV.10

Lemma IV.10. Composition as defined in Def. IV.9 is independent of the choices of representatives. That is,



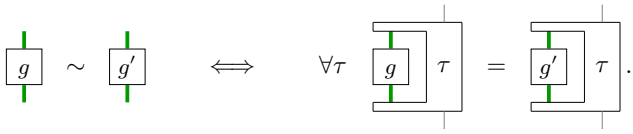
(B9)

Proof. We start by rewriting the assumptions of the theorem using the definition of equivalence:



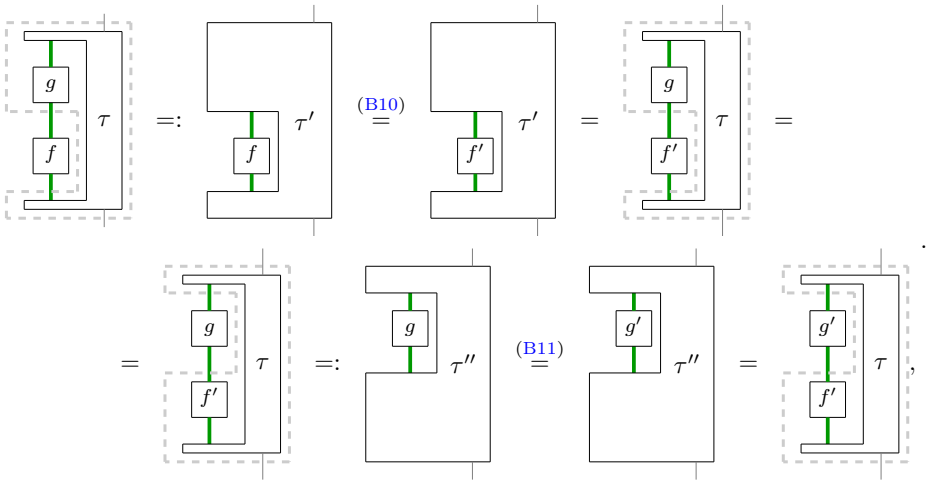
(B10)

and



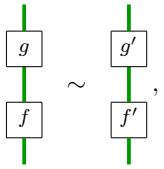
(B11)

Hence, for an arbitrary τ, the following holds:



(B12)

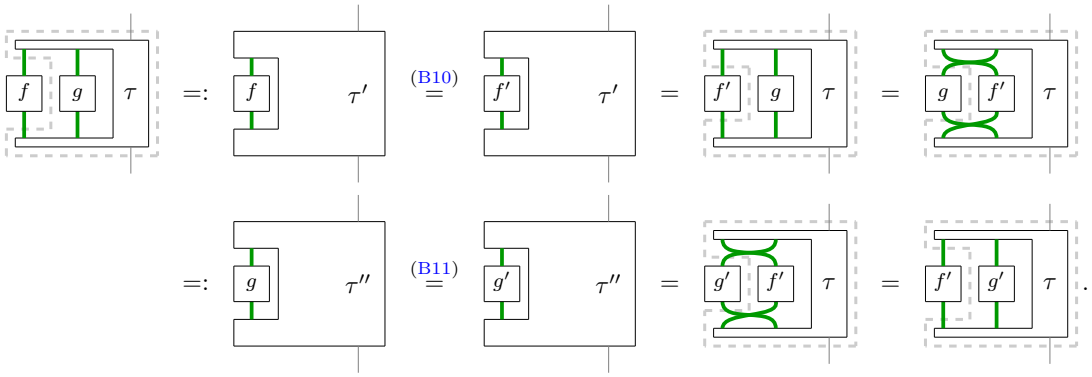
This implies that



(B13)

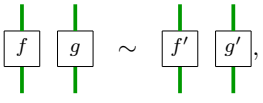
hence proving the first part of the lemma.

The proof of the second part of the lemma follows in a similar way: for an arbitrary τ,



(B14)

Hence,

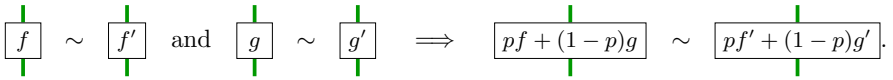


(B15)

which completes the proof. □

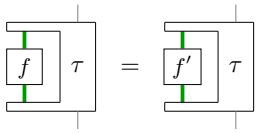
5. Proof of Lemma IV.12

Lemma IV.12. Convex mixtures as defined in Def. IV.11 are independent of the choice of representative. That is:



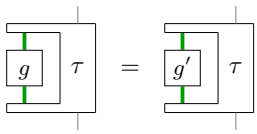
(B16)

Proof. The assumptions of the Lemma can be equivalently stated as



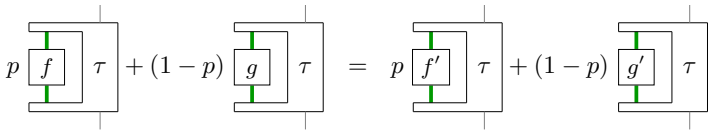
(B17)

and



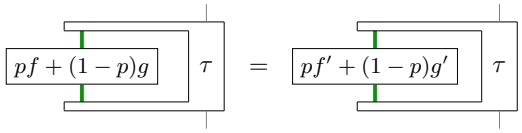
(B18)

In particular, this means that $\forall\tau$



(B19)

Linearity of τ implies



(B20)

for all τ, which proves the claim. □